université
PARIS-SACLAY

Analyse Visuelle de Réseaux Sociaux Historiques: Traçabilité, Exploration et Analyse

Visual Analytics for Historical Social Networks: Traceability, Exploration, and Analysis

Thèse de doctorat de l'université Paris-Saclay et de Telecom Paris

École doctorale n°580 : Sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat: Informatique
Graduate School : Informatique et Sciences du Numérique
Référent : Faculté des sciences d'Orsay

Thèse préparée au Laboratoire interdisciplinaire des sciences du numérique (Université Paris-Saclay, CNRS, Inria), et à Telecom Paris, sous la direction de Jean-Daniel FEKETE, Directeur de recherche et la co-direction de Christophe PRIEUR, Professeur des universités.

Thèse soutenue à Paris-Saclay, le xx décembre 2022, par

Alexis PISTER

Composition du jury

Ulrik Brandes Professeur, ETH Zürich	Rapporteur & Examinateur
Guy Melançon Professeur, Univerité de Bordeaux	Rapporteur & Examinateur
Wendy Mackay Directrice de recherche, Univ. Paris-Saclay, CNRS, Inria, LISN	Examinatrice
Uta Hinrichs Professeur, University of Edinburgh	Examinatrice
Laurent Beauguitte Chargé de recherche, CNRS	Examinateur
Jean-Daniel Fekete Directeur de recherche, Univ. Paris-Saclay, CNRS, Inria, LISN	Directeur de thèse
Chritophe Prieur Professeur, Université Gustave Eiffel	Directeur de thèse

THESE DE DOCTORAT

NNT : 2020UPASA001

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Analyse Visuelle de Réseaux Sociaux Historiques: Traçabilité, Exploration et Analyse

Mots clés: analyse visuelle, analyse de réseau sociaux, visualisation de réseaux sociaux, histoire sociale, réseaux historiques

Résumé: Cette thèse vise à identifier comment l'analyse visuelle peut supporter les historiens dans leur processus d'analyse de réseaux sociaux, de la collecte de documents historiques jusqu'à la formulation de conclusions socio-historiques. L'analyse de réseaux sociaux historiques est une méthode permettant d'étudier les relations sociales au sein de groupes d'acteurs (familles, institutions, entreprises, etc.) pour comprendre leurs structures sous-jacentes tout en décrivant des comportements spécifiques. Les chercheurs en histoire sociale reconstruisent les relations du passé à partir du contenu de documents historiques, tel que des actes de mariage, formulaires de migration, ou des recensements. Utilisant des méthodes analytiques et de visualisation, les historiens peuvent décrire la structure de ces groupes et expliquer des comportements individuels à partir de motifs locaux. Cependant, l'inspection, l'encodage et la modélisation des sources pour obtenir un réseau finalisé provoquent souvent des erreurs, distorsions et des problèmes de traçabilité. Pour ces raisons, ainsi que des problèmes d'utilisabilité, les historiens ne sont pas toujours en position de faire des conclusions approfondies sur leur réseau à partir des systèmes de visualisation actuels. Je vise dans cette thèse à identifier comment l'analyse visuelle (la combinaison d'algorithmes statistiques intégrés à des interfaces graphiques à l'aide d'interaction) peut supporter les historiens dans leur processus, de la collecte des données jusqu'à l'analyse finale. Vers ce but, je formalise le processus d'une analyse de réseau historique en partant de collaborations avec des historiens, de l'acquisition des sources jusqu'à l'analyse visuelle, et pointe que les outils supportant ce processus devraient satisfaire des principes de traçabilité, simplicité et de réalité documentaire pour faciliter les va-et-vient entre les différentes étapes, avoir des outils faciles à utiliser, et à ne pas distordre le contenu des sources. Particulièrement, je propose de modéliser les sources historiques en réseaux sociaux bipartis multivariés dynamiques avec rôles pour satisfaire ces propriétés. Ce modèle représente concrètement les documents historiques, permettant aux utilisateurs d'encoder, corriger et analyser leurs données avec le même modèle et les mêmes outils. Je propose deux interfaces d'analyse visuelle pour manipuler, explorer et analyser ce type de données, avec un appui sur les principes de traçabilité, simplicité, et réalité documentaire. Je présente d'abord ComBiNet, qui permet une exploration visuelle à partir de la topologie, dynamique, localisation et att ributs du réseau à l'aide de vues coordonnées, un système de requêtes visuelles, et de comparaisons. En trouvant des motifs facilement et en les comparant, les historiens peuvent trouver des erreurs dans leurs annotations tout en répondant à des questions historiques. Le second système, PK-Clustering, constitue une proposition concrète pour améliorer l'utilisabilité et l'efficacité des mécanismes de clustering dans les systèmes de visualisation de réseaux sociaux. L'interface permet de créer des regroupements pertinent à partir de la connaissance à priori, le consensus algorithmique et l'exploration du réseau dans un cadre d'initiative mixte. Les deux systèmes ont été conçu à partir des besoins et de retours continus d'historiens, et visent à augmenter la traçabilité, simplicité, et la vision réelle des sources dans l'analyse de réseaux historiques. Je conclus sur des discussions sur la fusion des deux systèmes et plus globalement sur la convergence vers une meilleure intégration des outils d'analyse visuelle sur le processus global des historiens. De tels systèmes avec une attention les propriétés de traçabilité, simplicité, et réalité documentaire peuvent limiter l'introduction de biais et abaisser les exigences pour l'utilisation de méthodes quantitatives, qui a toujours été une discussion controversée en Histoire.

Title: Visual Analytics for Historical Social Networks: Traceability, Exploration, and Analysis
Keywords: visual analytics, social network analysis, social network visualization, social history, historical networks

Abstract: This thesis aims at identifying how Visual Analytics can support historians in their social network analysis process, from the collection of historical documents to the formulation of high-level socio-historical conclusions. Historical Social Network Analysis is a method to study social relationships between groups of actors (families, institutions, companies, etc.) to understand their underlying structure while characterizing specific behaviors. Social historians are able to reconstruct relationships of the past using historical documents' content, such as marriage acts, migration forms, birth certificates, and censuses. Through visualization and analytical methods, they can describe the global structure of studied groups and explain individual behaviors through local network patterns. However, the inspection, encoding, correction, and modeling process of the historical documents leading to a finalized network is intricate and often results in inconsistencies, errors, distortions, simplifications, and traceability issues. For these reasons, social historians are not always able to make thorough historical conclusions with current analytical and visualization tools. I aim in this thesis to identify how visual analytics—the integration of data mining capabilities into visual interfaces with interaction—can support social historians in their process, from the collection of their data to the answer to high-level historical questions. Towards this goal, I formalize the workflow of historical network analysis in collaboration with social historians, from the acquisition of their sources to their final visual analysis, and point out that visual analytics tools supporting this process should satisfy traceability, simplicity, and document reality principles to ease bask and forth between the different steps, provide tools easy to manipulate, and not distort the content of sources with modifications and simplifications. Particularly, I propose to model historical sources into bipartite multivariate dynamic social networks with roles to satisfy those properties. This modeling allows a concrete representation of historical documents, hence letting users encode, correct, and analyze their data with the same abstraction and tools. Leveraging this data model, I propose two interactive visual interfaces to manipulate, explore, and analyze this type of data with a focus on usability for social historians. First, I present ComBiNet, which allows an interactive exploration leveraging the structure, time, localization, and attributes of the data model with the help of coordinated views, a visual query system, and comparison mechanisms. Finding specific patterns easily and comparing them, social historians are able to find inconsistencies in their annotations and answer their high-level questions. The second system, PK-Clustering, is a concrete proposition to increase the usability and effectiveness of clustering mechanisms in social network visual analytics systems. It consists in a mixed-initiative clustering interface that let social scientists create meaningful clusters with the help of their prior knowledge, algorithmic consensus, and interactive exploration of the network. Both systems have been designed with continuous feedback from social historians, and aim to increase the traceability, simplicity, and document reality of visual analytics supported historical social network research. I conclude with discussions on the potential merging of both systems and more globally on research directions towards better integration of visual analytics systems on the whole workflow of social historians. Such systems with a focus on those properties—traceability, simplicity, and document reality—can limit the introduction of bias while lowering the requirements for the use of quantitative methods for historians and social scientists which has always been a controversial discussion among practitioners.

Contents

List of Figures

List of Tables

1 Introduction

The goal of this thesis is to characterize and produce visual analytics tools that can support social historians conducting research on their sources—particularly when using network methods—with a focus on exploration, analysis, traceability, and usability. Historical Social Network Analysis (HSNA) is a method—sometimes referred as a paradigm [205]—followed by social historians to study sociological phenomena through the observation of relationships of actors of the past, modeled into a network. The usage of networks as an abstraction to represent and study social relationships—such as friendships, kinship, or business ties—grew in popularity in the last 40 years [71, 187] and constitute a powerful metaphor, especially in our time when many of our digital connections and interactions use an explicit network structure[1]. This approach has first been formalized in sociology under the term Social Network Analysis (SNA) [71] and is now widely used in anthropology, geography, and history [104]. Historians leverage historical documents—which are at the core of their profession [111]—to extract relationships between actors of interest that they model with networks constructed from nodes and links that respectively represent actors (often persons) and relationships (like kinship). Using social network visualization techniques and leveraging network measures and computations, they can then test hypotheses they have and gain insight on the structural aspect of the relational phenomena they are studying [104, 203]. This approach has been followed successfully to study various subjects such as kinship [87], entrepreneurship [167], maritime routes [112], political power [145], political oppositions [144], and persecution [127]. Yet, history is considered by many as a literary and qualitative science, and many critics emerged from the history community concerning quantitative and network methods [85, 101, 116, 119], pointing to problems such as the leading to trivial conclusions, anachronisms, simplifications, and mismatches between network and historical concepts. Moreover, quantitative and network analysis are complex processes, and demand many efforts in data collection, encoding, modification, and processing before being able to make efficient observations. This thesis considers the whole workflow of social historians to better support it with visual analytics.

Social historians have to take many annotation (sometimes called encoding) and modeling decisions, concerning *what* to model from their sources into a network, and *how* to model it [42, 54], i.e., should the information of interest be represented as a node, a link, an attribute, or not reflected in the network at all, and what format should be used. Practically, they usually use ad hoc processing and analysis scripts to transform historical documents to analyzable networks, which is time-consuming, sometimes to end up with trivial or hard to interpret results [5]. Still, HSNA led to many highly regarded studies with thorough conclusions, such as the study of families of power in Florence by Padgett and Ansell where they explained the rise of the Medici family through its central position in the economical, political, and trading networks of powerful families [145] or Gribaudi and Blum work on the social and professional shift during the 19th century in France [84].

[1] This analogy goes to the point that the term “Social Network” can refer both to the sociological metaphor for social relationships and to the social media platforms such as Facebook.

The usage of visualization to graphically display networks is common in SNA[2] as it allows to unfold the structure of networks to the eyes, thus letting social scientists confirm hypotheses they had when collecting and exploring their data as well as gaining new insight through the discovery of interesting patterns and trends [43]. Images of networks also constitute an efficient mean of communication, especially in scientific productions [70]. Many visualization techniques and softwares have thus been developed since the birth of SNA, but most popular tools are usually not designed for historians specifically, meaning that they do not regard on the provenance and process leading to the network, and focus on analysis aspects only. Moreover, they usually enforce simple network models without proposing exploration mechanisms, beyond allowing to look at the network structure and computed measures. In result, many HSNA studies show a plot of their network and describe it qualitatively, often by identifying the central actors—sometimes with the help of centrality—but do not go beyond that [117]. *In this thesis, I therefore investigate how visualization can support social historians in their work, first during the pre-analysis process and secondly during the analysis step, with the right level of expressiveness, usability, and traceability.*

1.1 Social History and Historical Social Network Analysis

Social History has continuously evolved since its beginning in the 1930s, especially with the rise of quantitative and network methods based on the development of computer science during the end of the 20th century. If these computer-supported methods are now widely used in history [104, 150], they attracted many criticism from the start—some are which still relevant.

We can trace back the birth of Social History with the formation of the "Annales School" in the 1930s, where historians gained interest in socio-economic questions and started to rely heavily on the exhaustive extraction and analysis of historical documents coming from archives [20, 156]. Beforehand, History was mainly political and event-centered, as the majority of work consisted in narrating and characterizing specific events—such as wars and diplomatic alliances—while eliciting their causes and consequences, and describing the lives of historic figures, such as sovereigns [156]. Social History shifted the focus by aiming to link together sociological, economical, and political issues and by placing individuals at the center of these questions [190]. Later on in the 1960s, with the development of computer science, historians started to use quantitative methods to analyze data extracted from historical documents and make conclusions grounded in statistical results, in various subjects such as demographics [93] and economics [81]. Around the same time, the use and study of networks started to become popular in various disciplines to study real-world relational phenomena based on mathematical computations and measures, especially in sociology and anthropology [32]. A network is an abstraction based on graph theory concepts which can be used to model phenomena based on relationships (called links) between entities (called nodes).

[2] Historians and sociologists following network analyses typically use similar techniques and tools for analyzing their data. The difference between SNA and HSNA hence come from the provenance and process leading to the construction of the network. I therefore use the SNA acronym for practices common in both fields and the HSNA acronym for history specificities.

Sociologists appropriated this concept to model social relationships between agents of interest, allowing them to study the sociological structure of groups of interest—such as families, institutions, and companies—and concepts like friendship, oppression, and diffusion using real world observation and mathematical computations. This SNA approach allows analysts to ground results in formal network measures and metrics based on real observations instead of relying on traditional social categories such as age, job, and gender [71]. This shift in the object of study from traditional social classes and aggregates to the observation of relationships of individuals remind the microhistory movement [77] which theorized that following the life of single individuals and small groups enable the making of higher level conclusions about the social structures they live in. Social historians followed this tradition and started to appropriate network concepts to study relational aspects of the past and formalized it under the term Historical Network Research or Historical Social Network Analysis [203]. However, historians do not have the possibility to run surveys or directly observe interactions of the past and are thus constrained by the information contained in historical documents they find in archives. These documents can be anything mentioning social relationships between actors of interest, such as marriage acts, birth certificates, census, migration acts, business transactions, journals. After selecting a corpus of documents, they typically read and inspect in depth several documents while taking notes to have a deeper insight on the content of the sources, which allow them to start eliciting hypotheses. Following this exploration phase, they manually annotate each document and encode the desired information—the mention of persons and their social relationship in the case of a network analysis. This is a long and tedious process that can result in small to large networks that they analyze using network measures to make conclusions on the structure of social groups or social behaviour of individual of interests. Figure 1.1 shows for example an original business document of the 17th century from Nantes (France). The historian have to inspect these documents in depth, extract useful information, and cross-reference the sources to do her quantitative analysis afterwards. The investigation and reading of the historical documents is therefore an exploratory process, where historians start to generate sociological hypotheses from the continuous extraction of insight and revelations of this process, similarly to grounded theory [79]. Once they finalised a network, they can test their hypotheses using qualitative or quantitative methods—based on statistical and network measures. Lemercier and Zalc write "Although history is not an exact science, counting, comparing, classifying, and modeling are nevertheless useful methods for measuring our degree of doubt or certainty, making our hypotheses explicit, and evaluating the influence of a phenomenon." [117] Social historians, therefore, have hypotheses about their subject of study, that they can back up or refute with the help of quantitative and network results, in a way similar to the competing hypotheses workflow of Intelligence Analysis [51]. By pointing to evidence supporting or refuting hypotheses, they can give insight into the level of the plausibility of different claims.

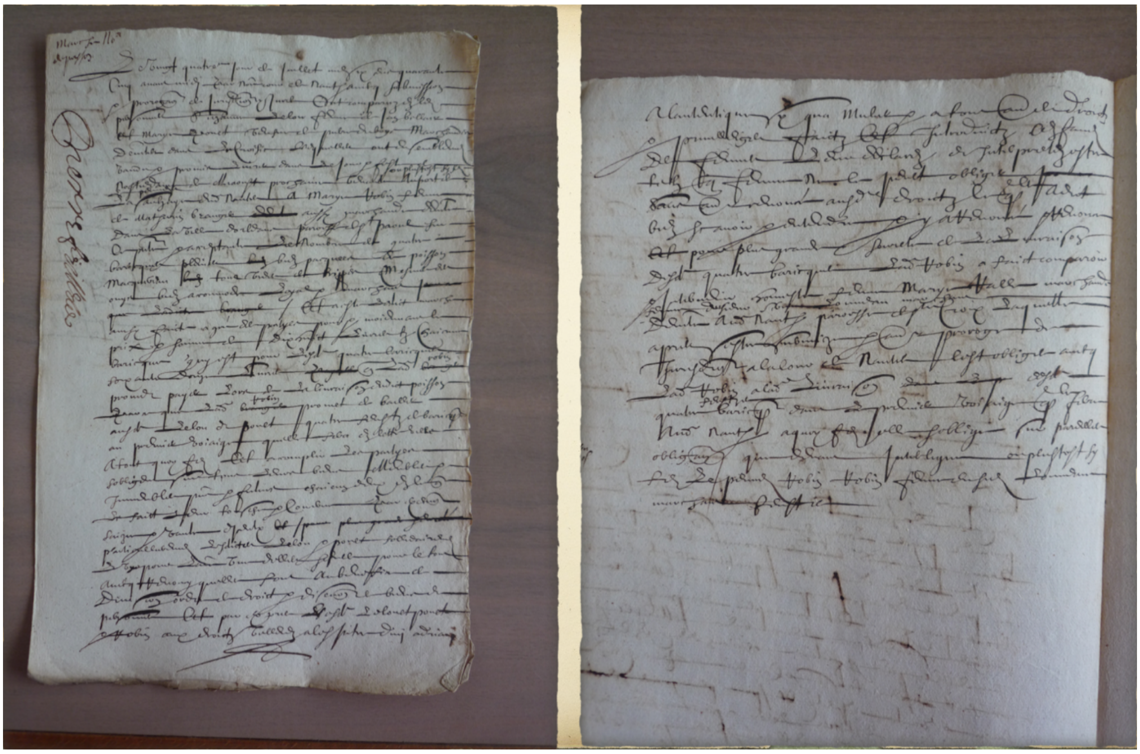

Figure 1.1 – Business contract originated from Nantes (France) during the 17th century. See [55] for more detail of the historian process to analyze her sources.

1.2 Visualization and Visual Analytics

Visualization has been said to be a central part in the development of SNA [70, 209]—as it the case for many scientific fields[3]. Social scientists now widely use visual and analytical tools to unfold their network structure, allowing them to confirm or deny hypotheses, or follow exploration analysis.

Visualization is the process of displaying data visually to leverage the human visual system and enhance cognition to gain insight into data [35]. Using visual abstractions (such as size, color, and position) to display abstract data allows us to rapidly see structure and patterns otherwise hidden in raw text and numbers. As data keeps growing in size with time due to the increase of hardware and storage capabilities, visualization is a powerful tool to gain insight into the underlying structure of various complex datasets.

Visualization has traditionally been used for confirmatory and communication purposes, particularly in empirical sciences [179]. By showing data visually, analysts are able to confirm or refute hypotheses and communicate their findings in scientific productions.

However, visualization can also be used for exploration, which can help to understand the underlying structure of data and generate new hypotheses. Tukey defined this process as

[3]the historian Alfred Crosby went as far as claiming that visualization is one of the two factors—with measurement—which led to the development of modern science [46].

Exploratory Data Analysis in the 1960s [195], as a procedure to gain insight into the structure of the data by identifying outliers, trends, and patterns with the usage of visualization and statistical measures. Social network visualization is used for communication of findings in the field, but is also often following this exploration process as showing the network visually allows social scientists to reveal the structure of their data. As freeman writes "Images of social networks have provided investigators with new insights about network structures and have helped them to communicate those insights to others" [70]. Social scientists very often represent their data using node-link diagrams, that we find in every production of reference in the field [114, 187, 202].

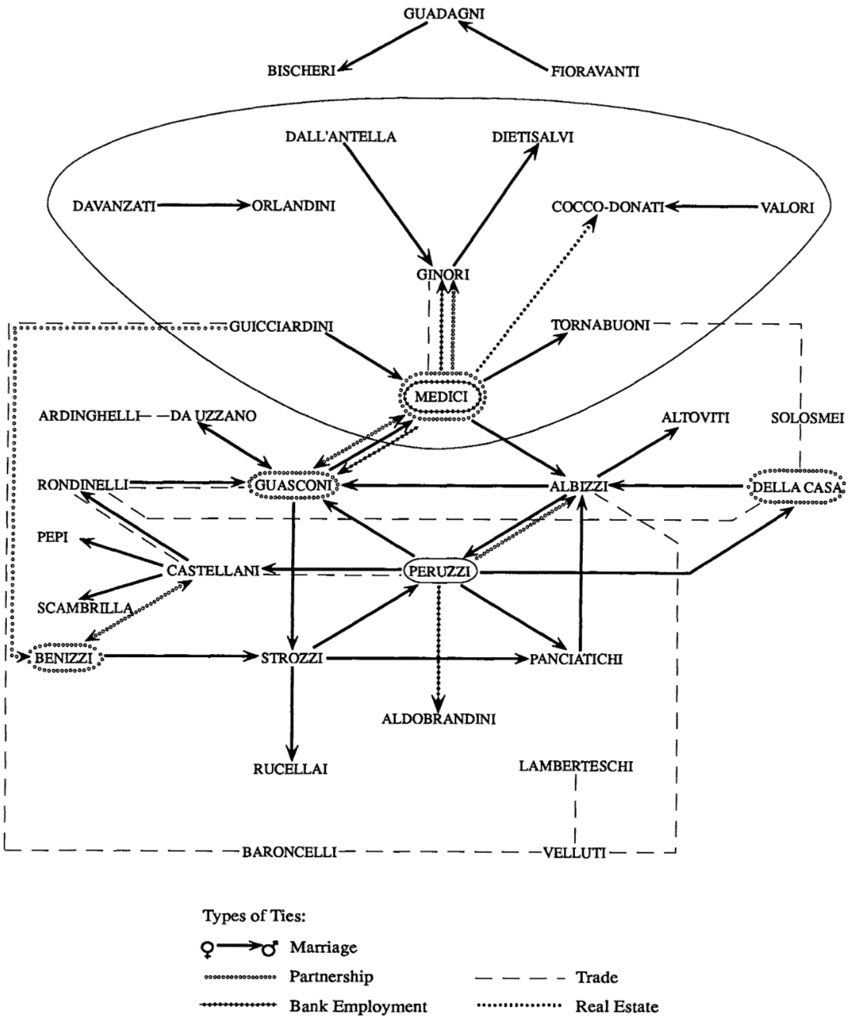

Figure 1.2 – Marriage, partnership, trading, banking, and real estate networks of the powerful families of Florence from [145]. We can see the central position in the network of the Medici Family.

Figure 1.2 shows a node-link representation of the network constructed by Padgett and Ansell in their work on the Medici. At that time, diagrams were often drawn by hand, practice which have now been replaced by automatic layout algorithms. Most visual software for SNA such as Gephi [13], Pajek [146], NodeXl [181], or Ucinet [22] are based on this representation, and allow an exploration of the data with the help of basic interaction mechanisms and the computation of network measures. The detection of patterns and trends can also be facilitated with automatic methods coming from data mining and machine learning fields, directly implemented in the visual analysis loop. This coupling of visual exploration and automatic data mining algorithms has been coined as Visual Analytics (VA) and is defined as the process of using interactive visualizations, transformations, and models of the data in an interactive analysis workflow to create knowledge [102].

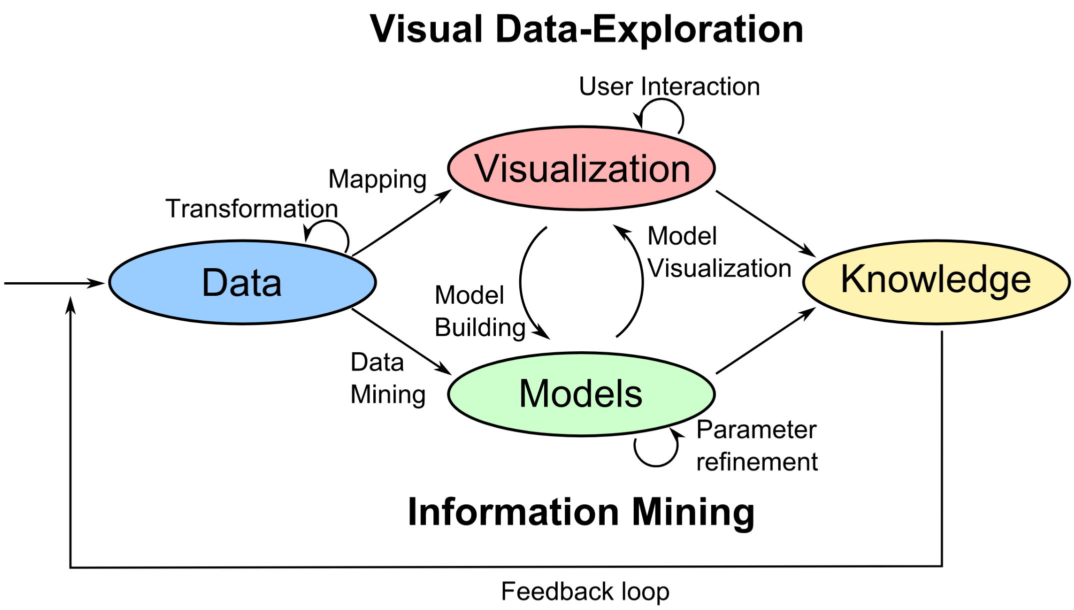

Figure 1.3 – Abstraction of the VA process. It is characterized by continuous interactions between the data, visualizations, models, and knowledge. Image from [102].

Figure 1.3 illustrate the schematic process of VA: the coupling of visualization and data mining models operated by the user through interaction lead to the generation of knowledge "extracted" from the data. If most widely used visual interface for HSNA do not yet provide complex interactions or high data mining capabilities, more recent tools are oriented towards VA, as the combination of automatic knowledge extraction with interaction and exploration can be a powerful support for social scientists to gain insight on the structure of their network, especially that the data they study keep growing in size and complexity [100].

1.3 Visual Analytics Supported Historical Network Research

Most visual tools for SNA are designed for the analysis of already curated networks, without taking into account the context in which those networks have been produced, where they come from, and the workflow that led to their creation. Moreover, many practitioners have trouble using current computer-supported tools, due to misconception in their encoding and modeling process or usability problems [5]. VA should therefore support social historians in the entirety of their process, with a focus on usability and simplicity.

Currently, social historians spend a long time in their data acquisition, processing, encoding, and modeling steps which lead them to the construction of a network [55, 118]. They typically visualize and analyze their network at the end of this process, first to verify hypotheses they formulated during the inspection of their sources, then to gain a better view of the structure of the network, allowing them to potentially generate new hypotheses [116]. However, research showed that all the steps preceding the analysis can introduce errors and misconceptions, especially since social scientists are often not trained in computer science and data science [5, 117]. Social scientists usually visualize their network using SNA tools like Gephi, Pajek, and NodeXl which encompass basic interactions, node-link visualization, SNA measure computations, and clustering algorithms. Once they visualize their data, they typically notice errors and inconsistencies in the data, such as duplication of the same entities, merging of different entities, or geolocation errors [5, 52].

Practitioners also have to decide on a network model [42] (see §2.3.4 for more details) when encoding their documents, which sometimes do not match the final analysis goals. Simple models typically oversimplify the relationships contained in the sources [116] and too complicated models are hard to manipulate [143]. They, therefore, have to go back and forth between the visualization software and the encoding process which can be tedious, especially since it can be complicated to trace back the entities of the data model back to the original documents for correction. VA tools that encompass the whole process of social historians should therefore be beneficial for the flow of their work and could help detect and correct errors or analysis plans way before the visualization of a finalised network. Proposing how to design such interfaces with proof-of-concepts is one of the goal of this thesis.

Furthermore, several historians highlighted the fact that many social history studies leveraging network methods simply use networks in a metaphorical sense, in what Rollinger calls "soft SNA" or "informal network research" [163]. Such studies typically show one—or a couple—node-link diagram(s) which they describe with qualitative terms [117] to refer to the global structure of the network (dense, parse, connected, etc.), the place of actors (central, distant), or interesting patterns (cliques, bridges, communities). In case of dense networks, such descriptions become obsolete, as diagrams start to look like what have been called a "spaghetti monster" [39, 117] i.e., an unreadable image due to the high level of cluttering. Figure 1.4 shows for example a medieval social network of peasants proximity relationships between 1250 and 1350, extracted from agrarian contracts. The graphic do not convey much information, especially that the links represent a constructed notion of proximity without indicating the types of relationships the individuals were mentioned in the contracts.

The lack of use of network analytical methods—which are numerous in modern SNA softwares[4]—have been in part explained by "math anxiety" [148]: it takes long effort to learn the mathematical concepts behind network measures and algorithms, and their relationships

[4] See for example the long technical manuals of Pajek [146] and Ucinet [99]

Figure 1.4 – Node-link diagram of a medieval social network of peasants, produced with a force-directed layout, commonly used in SNA softwares. Image from [26].

to sociological concepts [163], especially for practitioners without formal computer science and mathematical training. My claim is that current HSNA tools do not support social scientists enough in their analysis due to 1) the lack of interaction, direct manipulation, and exploration mechanisms in current interfaces and 2) the lack of network measures and algorithm interpretations and explainability. For example, clustering algorithms are often included in such systems, letting social scientists partition networks into groups, but many algorithms exist in the literature, potentially giving diverse results. Scientists often run several algorithms until finding a satisfying enough partition, which can bias the result of an analysis [153]. Usability and traceability of the results are therefore primordial in VA interfaces aimed at supporting social historians in their analysis.

VA could therefore help social historians using network methods for their research, first by supporting their entire workflow to help them explore, encode, correct, and model their data with simple tools and without introducing oversimplifications, but also to provide guidance and exploration mechanisms during the purely analytical step. For this, such interfaces should therefore 1) be simple enough to manipulate, 2) model the original documents and annotations without distortions, and 3) let historians trace back their network entities to the original sources and analytical results in explainable frameworks. In other terms, they should satisfy *simplicity*, *document reality*, and *traceability* principles. We discuss and explain them more in depth in chapter 3.

1.4 Contributions and Research Statement

The goal of this thesis is to characterize how VA can support social historians in their HSNA process and present proofs of concepts of tools supporting it. Most social network visualization tools are agnostic to the process of social historians leading to a polished network, even though it has an high impact of the network model and structure. Using visualization only at the end of the process often reveals potential errors, inconsistencies, or mismatches between the network model and analysis goals [5]. Moreover, due to lack of usability and interaction mechanisms, social historians often simply visualize statically their network and partially describe their structure, leading to conclusion which would have been easier to reach with simpler methods [61]. VA could therefore 1) assist social historians in their overall workflow, starting at the documents' acquisition to the final analysis step, with the help of data mining and interaction mechanisms in the data acquisition, encoding, modeling steps, and 2) provide exploration and analysis mechanisms to answer complex historical questions, beyond simply plotting the network with a node-link diagram.

The goal of this thesis is hence to give answers to the high-level question "How can VA support social historians in their entire HSNA process?". To answer this question, I first characterize the HSNA process from start to finish from discussions and collaborations with social historians, with the goal of identifying pitfalls that regularly arise and characterizing social historians' needs. From this, I give answers and directions—illustrated by proof-of-concepts—to three questions concerning the modeling aspect of HSNA and how VA and automatic tools can support social historians in different parts of their process, while satisfying *traceability*, *reality*, and *simplicity* properties:

Q1: How to model historical documents into analyzable networks with the right balance between expressiveness and simplicity?

Q2: What representations and interactions would allow social historians answer complex historical questions—with a focus on usability?

Q3: How to design VA tools and interactions that leverage algorithmic power but keep historians in control of their analyses and biases?

In chapter 3, I start by describing the HSNA workflow and identify recurring pitfalls we encountered in our collaborations with historians and answer **Q1** by proposing a network model for modeling historical documents. In the following chapter 4, I give answers to **Q2** by providing a VA interface to explore bipartite multivariate dynamic networks, with queries and comparison interactions with the aim of letting historians find errors easily, transform their network data, answer their questions, and generate interesting hypotheses. Finally, in chapter 5, I propose PK-Clustering, a mixed-initiative clustering technique for social scientists based on their prior knowledge, algorithmic consensus, and traceability of results, as a concrete example of a system providing answers to **Q3**.

6 Conclusion

Contents

6.1 Summary

In this thesis, I gave answers and directions to the high-level question of how VA can help historians follow network analysis, in their entire process, from the collection of documents to the final analysis and visualization of constructed networks. Indeed, social historians currently use visual and analysis tools to get insights from curated networks, but the process leading to those is tedious and error-prone, which can result in simplification, distortions, errors, and inconsistencies [5, 116]. Moreover, current tools typically lead to qualitative description of the network data [163], mostly due to usability and interpretability issues. VA could therefore support historians in 1) their data preparation process, and 2) the final analysis of curated networks. I identified three principles such VA interfaces should follow: *traceability*, *simplicity*, and *document reality*, to respectively ease the back and forth between the different process steps while assuring reproducibility of analyses, have expressive representations and tools which are simple enough to manipulate for social scientists, and ground results in the concrete reality of the documents, hence without introducing bias and distortion. More precisely, I tried to answer three questions in respect of those properties: **Q1**: How to model historical documents into analyzable networks with the right balance between expressiveness and simplicity, **Q2** What representations and interactions would allow social historians answer complex historical questions—with a focus on usability, and **Q3**: How to design VA tools and interactions that leverage algorithmic power but keep historians in control of their analyses and biases. In chapter 3, I answered **Q1** by proposing to model historical documents into bipartite multivariate dynamic networks to have a right balance between expressiveness and *simplicity*, while satisfying *traceability* and *document reality* properties. More globally, I formalized the HSNA process from collaborations with social historians, into five steps: textual sources acquisition, digitization, annotation, network creation, and network visualization/analysis, and identified recurring pitfalls for each step, such as wrongly chosen network models or named entity recognition errors. The identification of pitfalls and continuous discussions with practitioners led to the definition of the *traceability*, *simplicity*, and *document reality* properties. Leveraging the proposed network model, I proposed the ComBiNet system in chapter 4 as a proof-of-concept answer for **Q2**. By proposing easy-to-use exploration, visual querying, and comparison interactions on a model encoding all different dimensions of the content of historical documents (roles, social structure, location, time, other attributes), social historians were able to 1) reflect on their annotation process and potentially detect errors, while 2) answering complex historical questions on the specificity and difference of groups and individuals of interest. Finally, I proposed PK-Clustering in chapter 5, a new method for clustering based on the prior knowledge of social scientists, consensus of automatic algorithms, and exploration interactions. This system gives a concrete answer to **Q3** by providing the right balance between user control and data mining automatic capabilities, while keeping *traceability*, *simplicity*, and *document reality* properties, through detailed report of interactions and results, simple interactions mechanisms, and the use of bipartite multivariate dynamic networks as a data model. These two systems demonstrate that VA tools can support social historians in their overall workflow, and increase the traceability and control of the process while leveraging complex representations and algorithmic power. While ComBiNet and PK-Clustering have several limitations that I discuss in §6.2, I believe they lead the way towards a better integration of VA tools to support social historians in their overall workflow, with the right levels of simplicity and usability. I discuss the perspective of new VA tools for HSNA, social history, and more globally the future of Digital Humanities in §6.3.

6.2 Discussion

I discuss in this section different limitations of my work:

Network modeling. I proposed with my collaborators to model historical documents using bipartite multivariate dynamic networks, as it allows to satisfy *traceability*, *simplicity*, and *document reality* properties. However, this type of modeling have some limitation in 1) the types of sources it can model, 2) how persons are represented in the network, and 3) how uncertainty is managed. We elaborated this model from collaborations with several social historians who study semi-structured documents, such as marriage acts, birth certificates, business contracts, construction documents, censuses, and migration forms. These types of documents have a repetitive structure and mention people in a restricted number of relationships (spouses and witnesses for marriages, parents and child for birth certificate, etc.) that can be encoded as roles in a consistent manner. However, other types of textual documents can be leveraged by historians, which can be less structured or without any predefined structure at all. One example is correspondence letters, which is a type of document often studied in history [**?**, 58]. The content of letters is more verbose and vary from one to another, making the process of defining a set of relationships to encode more difficult. bipartite multivariate dynamic networks would therefore not necessarily be an efficient model to encode this type of data, and other network models may be a better fit (such as directed networks). Moreover, in the proposed model, if documents are concretely represented as one type of nodes, person nodes constitute a merging of several mentions of the same person through several sources. Historians therefore have to follow a named-entity-recognition and disambiguation process to give identifiers to the different person mentions in the documents, and merge the information from several sources into one node. Person nodes are therefore not the concrete representation of the mentions inside the sources, but constructed concepts, resulting of the inspection and cross-referencing of the documents by historians. From one of my discussion with a historian, she told that in this model "document node can be considered as emic and person nodes ethic concepts" [90]. Historians hence have to make decisions, especially when there is ambiguity in the identity of persons, and when potential contradicting information is written on the same individuals (concerning age, origin, profession, etc.). This process raise a problem which is widespread in quantitative social history, but also most empirical science, which is the handling of *uncertainty*. Practitioners typically dismiss the uncertainty inherent to most textual data when constructing networks and encoding specific entities, thus removing it in the making of final conclusions. This is particularly true in history, where many mentions are ambiguous and not always precise [54]. Almost none work has been done on the handling of uncertainty directly in network models [**?**, **?**], even though it would allow to ground results in a more rational and real vision of the textual data (at the cost of an increasing complexity).

Temporality. The time is a key information for historians, as they want to contextualize the phenomena they study in a period, relative to other events. This is why we encode time in our suggested model of bipartite multivariate dynamic networks through the time mentioned in historical documents, so historians can explore and analyse this dimension of their data. However, dynamic graphs are complex to visualize and analyse. In ComBiNet, if users can explore the dynamic aspect of the data through time distribution, overlays, and dynamic filters, it currently does not propose a layout unfolding the time structure. It may therefore be harder to detect time-related patterns compared to topological and geolocated ones (even if possible with interactions). PK-Clustering lets visualize the data through a static or dynamic layout. However, the current prototype considers only static clustering, which can be seen as a simplification of the real world groups which can often evolve with time [164]. Indeed, persons often can often meet new people, change affiliations, or move places, provoking groups to merge, split, and disappear. PK-Clustering is already a complex process for static clustering, but could be extended to the building of dynamic groups with the use of time-dependant prior knowledge and dynamic graph clustering algorithms.

HSNA and Social History. HSNA is now a widely used method in quantitative history to study relational structures and phenomena of the past [104, 150, 203], and my reflexions and tools described in this thesis aim at improving the workflow of historians following this type of method. Yet, historians usually have heterogeneous and various documents when they are researching an area and era of interest, and usually apply different methods at the same time to make their historical conclusions [145, 150]. The core of their work consists in extracting knowledge from rigorous inspection and cross-referencing of historical documents. If providing VA tools for their HSNA analysis from start to finish is useful to them, other types of analysis methods should also be implemented in their work environments to allow them a larger set of options to make socio-historical conclusions. This includes methods like text analysis, correlations, and statistical testing [117]. History is also often considered a qualitative process, meaning that historians often make conclusions and hypothesis based on the reading of other sources and the qualitative analysis of their documents. VA tools which aim to encompass the whole historical workflow should be able to support this type of analysis, for example by managing textual annotation management on the digital documents, similar to Jigsaw's feature for intelligence analysis [183]. Some quantitative methods can also let users express some of their qualitative knowledge to influence the results. For example, bayesian statistics and semi-supervised machine learning methods are based on the expression of prior knowledge which will influence the computational results. Similarly, With PK-Clustering, historians can express their prior knowledge and use it as a start to find meaningful clusters, by seeing how the diversity of algorithms match their qualitative knowledge of the data. VA tools for social history should therefore let users follow both qualitative and quantitative inspection of their documents, from data collection to final analysis, with combinations of several methods and prior knowledge expression.

Globality of HSNA workflow. The key point of this thesis is to show that VA tools should support the overall HSNA workflow of historians. VA can be used to help them from data collection to their final analysis in the same environment, to ease back and forth between the steps, allowing easier exploration of different analysis goals, and better traceability/reproducibility for the overall analysis. By modeling historical documents into bipartite multivariate dynamic networks (see chapter 3), we represent the documents and their content as a network, allowing a traceability between the network entities and the original documents. If historians find errors in the network, they can rapidly trace it back from which document the errors come from, and correct it either directly in the visual interface, or in their annotation software using the unique identifier of the document. This modeling choice is a first step towards a better integration of the different steps into the same VA loop. Moreover, with ComBiNet, social scientists can apply filters to study specific visions of the network and follow multiple analysis paths on different dimensions of the data. ComBiNet therefore allow a better integration of the annotation/encoding, modeling, and analysis/visualization steps, using the same interface. However, it does not allow complex network transformations (such as creating simple unipartite networks) nor adding new annotations in the documents texts. Historians still need to use ad-hoc methods for data collection and encoding, and may want to make more complex network transformations for specific analysis goals.

6.3 Perspectives

I list in this section how this work could be extended, and interesting research directions for social history VA applications.

Uncertainty models. As discussed in §6.2, historical sources are filled with ambiguity and imprecise information. Practitioners have to disambiguation homonyms mentions to know if they refer to different persons or not. They also have to deal with potential surnames, or plain errors in the writing of names. Similar problems arise for other entities mentions such as locations, since many places in the world have the same name or have changed name with time. With non-contemporary documents, location mentions can also have inconsistent resolution and refer to places with non defined borders, such as "county of XX", or "kingdom of XX", which do not exist anymore. Moreover, historians can find contradictory information in several sources, for example concerning persons and events. When encoding their sources, practitioners hence have to make decision on all this ambiguous information, by cross-referencing the documents, and using their common sense and intuition, to decide what seems the most probable choices. However, we could think about encoding schemes and data models which encapsulate the uncertainty inherent to historical data, to ground analysis results in a less biased and more real vision of the sources. I think this is a promising research direction, which have not stirred a lot of interest until now.

Dynamic Layouts and Clustering. As discussed in §6.2, temporality is one of the key dimension of historical networks modeled as bipartite multivariate dynamic networks. Several layouts have been proposed to show the time aspect of dynamic graphs and bipartite dynamic graph, such as PAOHVIS [196], which is the layout PK-Clustering is based on. However, this type of layout do not allow to see the attributes of persons and documents, nor the geolocation of entities. Moreover, most proposed layouts do not scale beyond approximately one hundred nodes. One way to solve this scalability problem is to aggregate the network, for example using dynamic clustering. Several dynamic clustering algorithms have been developed for dynamic networks, but they often struggle to take into account complex dynamic of clusters, such as merging, splitting, or community drifting. Furthermore, no layout currently exist to visually display dynamic communities specifically for dynamic hypergraphs. More work can thus still be done in this space, to visualize the dynamic aspect of bipartite multivariate dynamic networks, and visualizing dynamic communities. I started to develop a prototype to visualize the dynamic aspect of bipartite multivariate dynamic networks, with a focus on the document. Figure 6.1 shows the prototype. Each column corresponds to a timeslot, and each square to a document. Persons mentioned in documents are displayed in the document nodes as smaller rectangles, colored by their roles. For one person, all its mentions are linked through arcs inside one timestep, and splines through different timeslots. The splines are optimized to be of minimum size between each timestep. This representation should allow to reveal the community structure of this type of data, by placing documents sharing many persons nearby. It also allows to rapidly see properties associated with documents, for example by encoding the document nodes with color, as shown in Figure 6.1 (bottom).

This layout prototype could serve as a base to develop a process similar to PK-Clustering, but for creating meaningful dynamic groups, based on consensus of dynamic clustering algorithms, prior knowledge of social historians, and exploration capabilities.

Machine Learning, automation, and agency. Machine learning went through rapid progress in the last 10 years, mainly due to the increase of data storage, computing power, and the rise of deep learning architectures. It has been applied to various tasks such as automatic driving, fraud detection, computer vision, and medicine diagnostics. In the context of SNA, machine learning methods have been used to automatically extract knowledge through tasks such as node classification, clustering, and link prediction [128]. More broadly, it has also been used for historical documents digitization [151]. If machine learning can give state-of-the-art accuracy on many of those tasks, it also poses issues on the explainability and reliability of the results in real world applications. It can be particularly frustrating it the context of social history, as historians needs to be able to understand and explain structure of their networks, as discussed in chapter 5. Several methods and approaches now focus on trying to explain the outputs of these black-box algorithms to the end-user [?]. Similarly, research is done on how to design interactive systems which leverage machine learning algorithms to guide and advise users, which are at the center of the decision-making process. This concept of utilizing

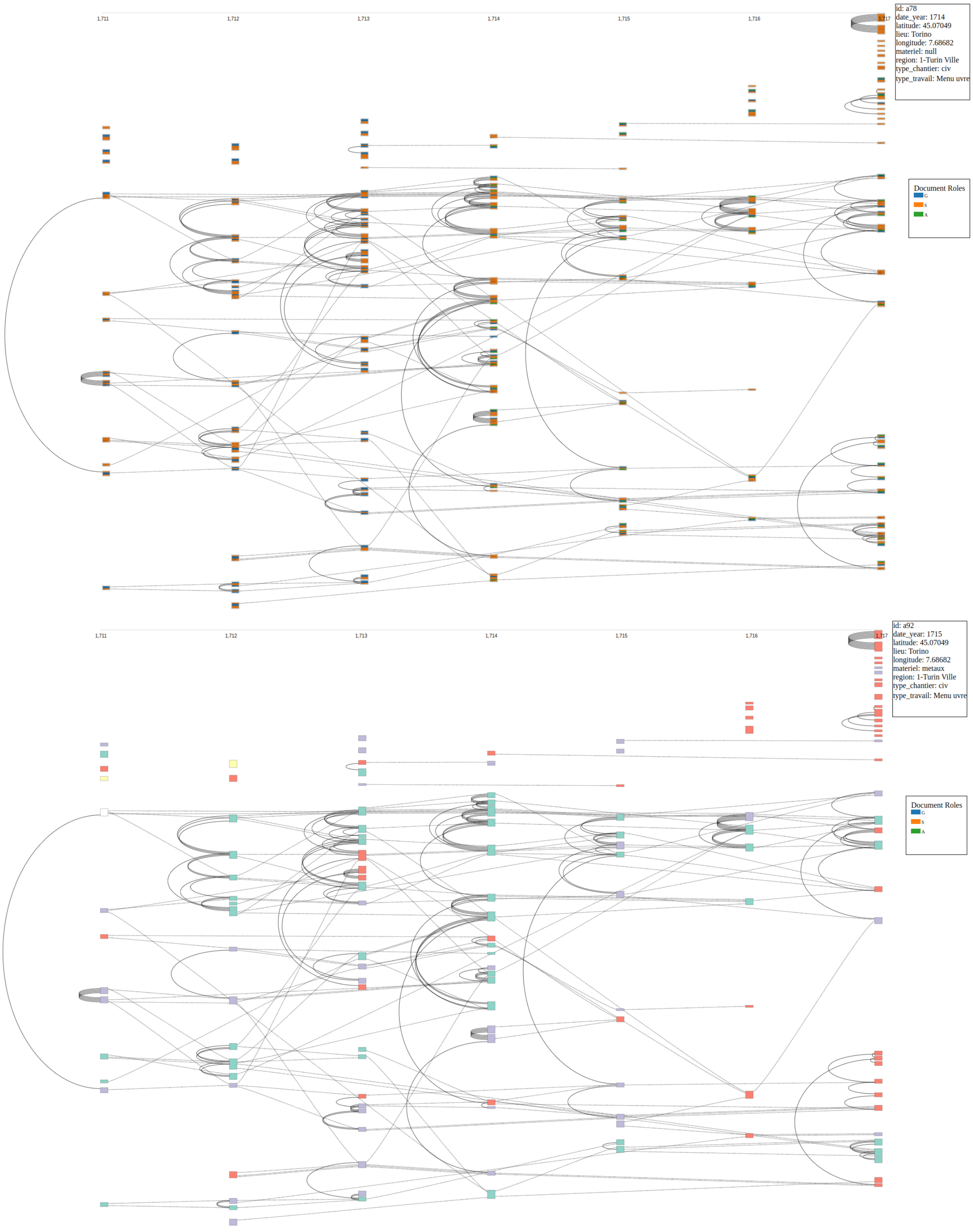

Figure 6.1 – Prototype of a document-centered dynamic layout for bipartite multivariate dynamic networks, visualized with data of construction contracts in Piedmont (see collaboration #1 in chapter 3). The layout can show the mention of persons encoded with their roles (top) or the documents and their properties (bottom). Here, the *region* attribute is selected, hence coloring construction contract depending on their location.

artificial intelligence power to support human decision-making through interactive systems have been coined as "agency" [92] and "human-centered articificial intelligence" [?]. PK-clustering is based on this core idea that machine learning should support users while not removing their decision-making process, by providing automatic suggestions through clustering results, while letting social historians decide. ComBiNet could also be extended with agency, for example to suggest social scientists recurring subgraphs in the data, that could be interesting to them. Over-represented subgraphs could be a query start that the users would refine through the easy-to-use visual query system. This idea of human-centered artificial intelligence could be applied not only to the analysis part, but also the data preparation workflow, for example in document transcription, named entity recognition, disambiguation, all tasks that machine learning is efficient at.

A common workflow interface. Currently, most social scientists have to use a lot of different pieces of software, files, and ad-hoc processes to follow quantitative analyses. I provided two VA interfaces to help historians analyse their data and ease back and forth between the different steps of their analysis. Both interfaces use the same data format to lower the time-cost to switch between them. However, historians still have to collect and annotate/encode their data manually with ad-hoc methods, and may have to convert their data to various formats when using several visual analysis tools. All these operations usually break the traceability and reproducibility of their analyses, and make their process tedious, especially that it often requires the writing of conversion scripts, which they do not necessary have the programming skills to do. I therefore argue there is a need for visual interfaces which integrate the whole workflow of social scientists, from the data collection to the formulation of high-level conclusions. If all the processes they do is integrated in the same visual environment, it would ease the flow of the analysis, increase the traceability and replicability of the actions and results, and allow them to take several explorations paths more easily. ComBiNet and PK-Clustering could for example be integrated in the same environment, with added possibilities of managing documents in the same place, apply annotations/encoding, and see in real time the creation of networks and transformations from the annotation process instead of having to do many back and forth. This constitutes an interesting research direction as it would allow social historians to collect, annotate, apply transform, analyse, and visualize their historical documents in the same environment, with easy-to-use interactions and artificial intelligence support.

6.4 Conclusion

To conclude, the goal of this thesis was to provide answers and directions towards the question of how VA, and more globally computer science, can help and support historians in the analyses they want to make. Towards this goal, I first formalized the current HSNA process from collaborations and discussions, defined three properties tools supporting this process should satisfy (*traceability*, *simplicity*, and *document reality)*, and proposed two interfaces showcasing visualization and interactions mechanisms to support social historian in their workflow, leveraging historical documents modeled as bipartite multivariate dynamic networks. ComBiNet allows to explore this data model, reflect on annotations, reveal specific facets of the data, and to globally highlight and compare specific groups and behaviors to either detect erroneous patterns or answer socio-historical questions. PK-Clustering aims at integrating better the clustering task in the social historians workflow, by providing a mixed-initiative approach for clustering on bipartite multivariate dynamic networks leveraging the prior knowledge of practitioners, the consensus of automatic algorithms, and exploration capabilities. Both systems have been validated with real use cases, and aim at providing simple-to-use yet expressive tools, which let historians at the center of the analysis loop. Indeed, ths use of quantitative methods in history and more globally the humanities have led to many expectations in the last 50 years, but also many disappointments due to usability and interpretability issues. More recently, with the rise of popularity of machine learning, many propositions are made towards automatic inspection and extraction of historical data. Yet, many criticism emerged towards those methods, as it can easily lead to disembodied work without a deep understanding of historical content and phenomena. Moreover, historians actually regard highly the inspection process of the sources. As one historian told me, "What many people promoting artificial intelligence to automatically read and inspect historical sources do not understand, is that this part is actually the most fun aspect of the historical work, and why many of us do it." Computer-supported tools for digital humanities should therefore support practitioners with the help of interactive visualization and quantitative-supported suggestions, instead of only providing automatic uninterpretable results. Propositions of this work aim in this direction, which I think is where lies the future of digital humanities as a field, i.e., methods and systems that social scientists can use easily with low friction and entry-barriers, which provide data-supported suggestions to help the decision-making process of scientists through visualization and interaction.

Bibliography

[1] Interchange: The Promise of Digital History. *Journal of American History*, 95(2):452–491, September 2008. doi:10.2307/25095630. 33

[2] Moataz Abdelaal, Nathan D. Schiele, Katrin Angerbauer, Kuno Kurzhals, Michael Sedlmair, and Daniel Weiskopf. Comparative Evaluation of Bipartite, Node-Link, and Matrix-Based Network Representations, August 2022. arXiv:2208.04458. 42

[3] Ruth Ahnert, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. The Network Turn: Changing Perspectives in the Humanities. *Elements in Publishing and Book Culture*, December 2020. doi:10.1017/9781108866804. 33

[4] Michael C. Alexander and James A. Danowski. Analysis of an ancient network: Personal communication and the study of social structure in a past society. *Social Networks*, 12(4):313–335, December 1990. doi:10.1016/0378-8733(90)90013-Y. 9, 39, 41

[5] Mashael Alkadi, Vanessa Serrano, James Scott-Brown, Catherine Plaisant, Jean-Daniel Fekete, Uta Hinrichs, and Benjamin Bach. Understanding barriers to network exploration with visualization: A report from the trenches. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 2022. 15, 21, 23, 26, 40, 43, 47, 48, 50, 54, 56, 62, 121

[6] Keith Andrews, Martin Wohlfahrt, and Gerhard Wurzinger. Visual Graph Comparison. In *2009 13th International Conference Information Visualisation*, pages 62–67, July 2009. doi:10.1109/IV.2009.108. 69

[7] F. J. Anscombe. Graphs in Statistical Analysis. *The American Statistician*, 27(1):17–21, February 1973. doi:10.1080/00031305.1973.10478966. 9, 28

[8] Thomas J. Archdeacon. *Correlation and Regression Analysis: A Historian's Guide*. Univ of Wisconsin Press, 1994. 50

[9] Mariona Coll Ardanuy, Federico Nanni, Kaspar Beelen, Kasra Hosseini, Ruth Ahnert, Jon Lawrence, Katherine McDonough, Giorgia Tolfo, Daniel CS Wilson, and Barbara McGillivray. Living Machines: A study of atypical animacy, November 2020. arXiv:2005.11140, doi:10.48550/arXiv.2005.11140. 33

[10] David Auber, Daniel Archambault, Romain Bourqui, Maylis Delest, Jonathan Dubois, Antoine Lambert, Patrick Mary, Morgan Mathiaut, Guy Melançon, Bruno Pinaud, Benjamin Renoust, and Jason Vallet. TULIP 5. In Reda Alhajj and Jon Rokne, editors, *Encyclopedia of Social Network Analysis and Mining*, pages 1–28. Springer, August 2017. doi:10.1007/978-1-4614-7163-9_315-1. 9, 30

[11] Trevor J Barnes. Big data, little history. *Dialogues in Human Geography*, 3(3):297–302, November 2013. doi:10.1177/2043820613514323. 49

[12] Allen H. Barton. Survey Research and Macro-Methodology. *American Behavioral Scientist*, 12(2):1–9, November 1968. doi:10.1177/000276426801200201. 35

[13] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. In Eytan Adar, Matthew Hurst, Tim Finin, Natalie S. Glance, Nicolas Nicolov, and Belle L. Tseng, editors, *ICWSM*. The AAAI Press, 2009. 20, 43, 67, 75, 98

[14] Sugato Basu, Ian Davidson, and Kiri Wagstaff. *Constrained Clustering: Advances in Algorithms, Theory, and Applications*. Chapman & Hall/CRC, first edition, 2008. 101

[15] Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall PTR, USA, 1st edition, 1998. 42

[16] Leilani Battle and Jeffrey Heer. Characterizing Exploratory Visual Analysis: A Literature Review and Evaluation of Analytic Provenance in Tableau. *Computer Graphics Forum*, 38(3):145–159, 2019. doi:10.1111/cgf.13678. 70

[17] Michael Baur, Marc Benkert, Ulrik Brandes, Sabine Cornelsen, Marco Gaertler, Boris Köpf, Jürgen Lerner, and Dorothea Wagner. Visone Software for Visual Social Network Analysis. In Petra Mutzel, Michael Jünger, and Sebastian Leipert, editors, *Graph Drawing*, Lecture Notes in Computer Science, pages 463–464, Berlin, Heidelberg, 2002. Springer. doi:10.1007/3-540-45848-4_47. 43

[18] Jacques Bertin. *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Paris: Gauthier-Villars, 1967. 9, 26, 27

[19] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist, and J.d. Fekete. GraphDice: A System for Exploring Multivariate Social Networks. *Computer Graphics Forum*, 29(3):863–872, 2010. doi:10.1111/j.1467-8659.2009.01687.x. 75

[20] Marc Bloch. *Apologie Pour l'histoire*. A. Colin, 1949. 16

[21] Christian Böhm and Claudia Plant. HISSCLU: A hierarchical density-based method for semi-supervised clustering. In *Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '08, pages 440–451, New York, NY, USA, 2008. ACM. doi:10.1145/1353343.1353398. 101

[22] S.P. Borgatti, M. G. Everett, and L. C. Freeman. UCINET 6 for Windows: Software for Social Network Analysis. Harvard, MA, Analytic Technologies, 2002. 20

[23] Stephen Borgatti. Social Network Analysis, Two-Mode Concepts in. *Computational Complexity: Theory, Techniques, and Applications*, January 2009. doi:10.1007/978-0-387-30440-3_491. 40, 59

[24] Christian Bors, John Wenskovitch, Michelle Dowling, Simon Attfield, Leilani Battle, Alex Endert, Olga Kulyk, and Robert S. Laramee. A Provenance Task Abstraction Framework. *IEEE Computer Graphics and Applications*, 39(6):46–60, November 2019. doi:10.1109/MCG.2019.2945720. 70

[25] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011. doi:10.1109/TVCG.2011.185. 75, 85

[26] Romain Boulet, Bertrand Jouve, Fabrice Rossi, and Nathalie Villa. Batch kernel SOM and related Laplacian methods for social network analysis. *Neurocomputing*, 71(7):1257–1273, March 2008. doi:10.1016/j.neucom.2007.12.026. 9, 22, 40

[27] Pierre Bourdieu. Sur les rapports entre la sociologie et l'histoire en Allemagne et en France. *Actes de la Recherche en Sciences Sociales*, 106(1):108–122, 1995. doi:10.3406/arss.1995.3141. 31

[28] Paul Bradshaw. Data journalism. In *The Online Journalism Handbook*. Routledge, second edition, 2017. 29

[29] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):172–188, February 2008. doi:10.1109/TKDE.2007.190689. 54

[30] Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital_Humanities*. MIT Press, February 2016. 33

[31] Peter Burke. *History and Social Theory*. Polity, 2005. 31

[32] Mitchell J. C. The Concept and Use of Social Networks. *Social Networks in Urban Situations*, 1969. 16, 35

[33] Steven P. Callahan, Juliana Freire, Emanuele Santos, Carlos E. Scheidegger, Cláudio T. Silva, and Huy T. Vo. VisTrails: Visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data - SIGMOD '06*, page 745, Chicago, IL, USA, 2006. ACM Press. doi:10.1145/1142473.1142574. 55, 70

[34] Charles-Olivier Carbonell. *L'Historiographie*. FeniXX, January 1981. 31

[35] Stuart-K. Card, Jock-D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers In, San Francisco, Calif, February 1999. 18, 26

[36] Raphaël Charbey and Christophe Prieur. Stars, holes, or paths across your Facebook friends: A graphlet-based characterization of many networks. *Network Science*, 7(4):476–497, December 2019. doi:10.1017/nws.2019.20. 37, 38, 68

[37] Duen Horng Chau, Christos Faloutsos, Hanghang Tong, Jason I. Hong, Brian Gallagher, and Tina Eliassi-Rad. GRAPHITE: A Visual Query System for Large Graphs. In *2008 IEEE International Conference on Data Mining Workshops*, pages 963–966, December 2008. doi:10.1109/ICDMW.2008.99. 69

[38] J. S. Coleman. Introduction to mathematical sociology. *Introduction to mathematical sociology.*, 1964. 36

[39] Anna Collar, Fiona Coward, Tom Brughmans, and Barbara J. Mills. Networks in Archaeology: Phenomena, Abstraction, Representation. *J Archaeol Method Theory*, 22(1):1–32, March 2015. doi:10.1007/s10816-014-9235-6. 21

[40] TEI Consortium. TEI P5: Guidelines for electronic text encoding and interchange, February 2021. doi:10.5281/zenodo.4609855. 53

[41] Ryan Cordell and David Smith. Viral texts: Mapping networks of reprinting in 19th-Century newspapers and magazines, 2017. 33

[42] Pascal Cristofoli. Aux sources des grands réseaux d'interactions. *Reseaux*, 152(6):21–58, 2008. 15, 21, 39, 48, 50, 52, 54, 58, 67

[43] Pascal Cristofoli. Principes et usages des dessins de réseaux en SHS. *La visualisation des données en histoire*, page 35, 2015. 16, 42, 75

[44] Pascal Cristofoli and Nicoletta Rolla. Temporalités à l'œuvre dans les chantiers du bâtiment. *Temporalités. Revue de sciences sociales et humaines*, (27), June 2018. doi:10.4000/temporalites.4456. 10, 51, 57, 70, 74

[45] Tarik Crnovrsanin, Chris W. Muelder, Robert Faris, Diane Felmlee, and Kwan-Liu Ma. Visualization techniques for categorical analysis of social networks with multiple edge sets. *Social Networks*, 37:56–64, 2014. doi:10.1016/j.socnet.2013.12.002. 57

[46] Alfred W. Crosby. *The Measure of Reality*. Cambridge University Press, Cambridge, reprint édition edition, March 1998. 18

[47] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. 69, 75

[48] Erick Cuenca, Arnaud Sallaberry, Dino Ienco, and Pascal Poncelet. VERTIGo: A Visual Platform for Querying and Exploring Large Multilayer Networks. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2021. doi:10.1109/TVCG.2021.3067820. 69, 96

[49] Zach Cutler, Kiran Gadhave, and Alexander Lex. Trrack: A Library for Provenance-Tracking in Web-Based Visualizations. In *2020 IEEE Visualization Conference (VIS)*, pages 116–120, October 2020. doi:10.1109/VIS47514.2020.00030. 82, 85

[50] Allison Davis, Burleigh Bradford Gardner, and Mary R. Gardner. *Deep South: A Social Anthropological Study of Caste and Class*. Univ of South Carolina Press, 2009. 59

[51] Mandeep K. Dhami, Ian K. Belton, and David R. Mandel. The "analysis of competing hypotheses" in intelligence analysis. *Applied Cognitive Psychology*, 33(6):1080–1090, 2019. doi:10.1002/acp.3550. 17

[52] Jana Diesner, Craig Evans, and Jinseok Kim. Impact of Entity Disambiguation Errors on Social Network Properties. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1):81–90, 2015. 21, 45, 48, 52, 53

[53] Dana Diminescu. The migration of ethnic germans from romania to west germany: Insights from the archives of the former communist regime. In *CERS, Public Lecture, UCLA*, Los Angeles, United States, March 2020. 51, 72

[54] Nicole Dufournaud. La recherche empirique en histoire à l'ère numérique. *Gazette des archives*, 240(4):397–407, 2015. doi:10.3406/gazar.2015.5321. 15, 123

[55] Nicole Dufournaud. Comment rendre visible le rôle économique des femmes sous l'Ancien Régime ? Étude méthodologique sur les marchandes à Nantes aux XVIe et XVIIe siècles. In Bernard Michon and Nicole Dufournaud, editors, *Femmes et Négoce Dans Les Ports Européens (Fin Du Moyen Age - XIXe Siècle)*, pages 65–84. Peter Lang, 2018. 9, 18, 21, 51, 56

[56] Nicole Dufournaud and Jean-Daniel Fekete. Comparaison d'outils pour la visualisation de sources historiques codées en XML/TEI. *Document numérique*, 9(2):37–56, April 2006. doi:10.3166/dn.9.2.37-56. 53

[57] Cody Dunne, Nathalie Henry Riche, Bongshin Lee, Ronald Metoyer, and George Robertson. GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1663–1672, New York, NY, USA, May 2012. Association for Computing Machinery. doi:10.1145/2207676.2208293. 70

[58] Dan Edelstein, Paula Findlen, Giovanna Ceserani, Caroline Winterer, and Nicole Coleman. Historical Research in a Digital Age: Reflections from the Mapping the Republic of Letters Project. *The American Historical Review*, 122(2):400–424, April 2017. doi:10.1093/ahr/122.2.400. 9, 33, 34, 47, 122

[59] P. Erdös and A. Rényi. On the evolution of random graphs. In *On the Evolution of Random Graphs*, pages 38–82. Princeton University Press, October 2011. doi:10.1515/9781400841356.38. 35

[60] Emily Erikson and Peter Bearman. Malfeasance and the Foundations for Global Trade: The Structure of English Trade in the East Indies, 1601–1833. *American Journal of Sociology*, 112(1):195–230, July 2006. doi:10.1086/502694. 57

[61] Michael Eve. Deux traditions d'analyse des reseaux sociaux. *Réseaux*, 115(5):183–212, 2002. 23, 37, 38

[62] Wenfei Fan. Graph pattern matching revised for social network analysis. In *Proceedings of the 15th International Conference on Database Theory*, ICDT '12, pages 8–21, New York, NY, USA, March 2012. Association for Computing Machinery. doi:10.1145/2274576.2274578. 69

[63] Lucien Febvre. VERS UNE AUTRE HISTOIRE. *Revue de Métaphysique et de Morale*, 54(3/4):225–247, 1949. 31

[64] Jean-Daniel Fekete, Danyel Fisher, Arnab Nandi, and Michael Sedlmair. *Progressive Data Analysis and Visualization*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, April 2019. doi:10.4230/DagRep.8.10.1. 81

[65] Roderick Floud. *An Introduction to Quantitative Methods for Historians*. Routledge, London, September 2013. doi:10.4324/9781315019512. 50

[66] Robert Fogel. *Railroads and American Economic Growth: Essays in Econometric History*. 1964. 32

[67] Robert William Fogel. The Limits of Quantitative Methods in History. *The American Historical Review*, 80(2):329–350, 1975. doi:10.2307/1850498. 49

[68] Robert William Fogel and Stanley L Engerman. *Time on the Cross: Evidence and Methods, a Supplement*, volume 2. Little, Brown, 1974. 32

[69] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, February 2010. doi:10.1016/j.physrep.2009.11.002. 37

[70] L. Freeman. Visualizing Social Networks. *J. Soc. Struct.*, 2000. 16, 18, 19, 41

[71] L.C. Freeman. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004. 15, 17, 25, 34, 35, 36, 37, 54

[72] Manuel Freire, Catherine Plaisant, Ben Shneiderman, and Jen Golbeck. ManyNets: An interface for multiple network analysis and visualization. In *CHI '10*, CHI '10, pages 213–222, New York, NY, USA, 2010. ACM. doi:10.1145/1753326.1753358. 69

[73] Michael Friendly. Visions and Re-Visions of Charles Joseph Minard. *Journal of Educational and Behavioral Statistics*, 27(1):31–51, March 2002. doi:10.3102/10769986027001031. 26

[74] Michael Friendly. A Brief History of Data Visualization. In Chun-houh Chen, Wolfgang Härdle, and Antony Unwin, editors, *Handbook of Data Visualization*, Springer Handbooks Comp.Statistics, pages 15–56. Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-33037-0_2. 28

[75] GEDCOM: The genealogy data standard. 41

[76] Mohammad Ghoniem, J.-D. Fekete, and Philippe Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, pages 17–24. Ieee, 2004. 42

[77] Carlo Ginzburg and Carlo Poni. La micro-histoire. *Le Débat*, 17(10):133, 1981. doi:10.3917/deba.017.0133. 17, 38, 50

[78] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, June 2002. doi:10.1073/pnas.122653799. 37

[79] Barney G. Glaser and Anselm L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction, New Brunswick, 5. paperback print edition, 2010. 17

[80] Michael Gleicher. Considerations for visualizing comparison. *IEEE Trans. Vis. Comput. Graphics*, 24(1):413–423, 2018. doi:10.1109/TVCG.2017.2744199. 69

[81] Claudia Goldin. Cliometrics and the Nobel. *Journal of Economic Perspectives*, 9(2):191–208, June 1995. doi:10.1257/jep.9.2.191. 16

[82] Martin Grandjean. Social network analysis and visualization: Moreno's Sociograms revisited, 2015. 9, 36

[83] Martin Grandjean. Analisi e visualizzazioni delle reti in storia. L'esempio della cooperazione intellettuale della Società delle Nazioni. *ME*, (2/2017), 2017. doi:10.14647/87204. 66

[84] Maurizio Gribaudi and Alain Blum. Des catégories aux liens individuels : l'analyse statistique de l'espace social. *Annales*, 45(6):1365–1402, 1990. doi:10.3406/ahess.1990.278914. 15

[85] Jo Guldi and David Armitage. *The History Manifesto*. Cambridge University Press, October 2014. 15

[86] Klaus Hamberger, Cyril Grange, Michael Houseman, and Christian Momon. Scanning for patterns of relationship: Analyzing kinship and marriage networks with Puck 2.0. *The History of the Family*, 19(4):564–596, October 2014. doi:10.1080/1081602X.2014.892436. 39, 41, 57, 59

[87] Klaus Hamberger, Michael Houseman, and R. White, Douglas. Kinship network analysis. In John Scott & Peter J. Carrington, editor, *The Sage Handbook of Social Network Analysis*, pages 533–549. Sage Publications, 2011. 15, 41

[88] Mountaz Hascoët and Pierre Dragicevic. Interactive graph matching and visual comparison of graphs and clustered graphs. In Genny Tortora, Stefano Levialdi, and Maurizio Tucci, editors, *AVI '12*, pages 522–529. ACM, 2012. doi:10.1145/2254556.2254654. 69

[89] Loren Haskins and Kirk Jeffrey. *Understanding Quantitative History*. Wipf and Stock Publishers, March 2011. 30

[90] Thomas N. Headland, Kenneth L. Pike, and Marvin Harris, editors. *Emics and Etics: The Insider/Outsider Debate*. Emics and Etics: The Insider/Outsider Debate. Sage Publications, Inc, Thousand Oaks, CA, US, 1990. 56, 123

[91] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 32–39, October 2005. doi:10.1109/INFVIS.2005.1532126. 76

[92] Jeffrey Heer. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116(6):1844–1850, 2019. 127

[93] Louis Henry and Michel Fleury. Des registres paroissiaux a l'histoire de la population: Manuel de dépouillement et d'exploitation de l'état civil ancien. *Population (French Edition)*, 11(1):142–144, 1956. doi:10.2307/1525715. 16

[94] 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, November 2007. doi:10.1109/TVCG.2007.70582. 9, 42, 43

[95] Martin Hilbert and Priscila López. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025):60–65, April 2011. doi:10.1126/science.1200970. 28

[96] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, and Sabrina Kirrane et al. Knowledge graphs. *ACM Comput. Surv.*, 54(4), July 2021. doi:10.1145/3447772. 57

[97] Infovis SC policies FAQ. 115

[98] Piers J Ingram, Michael PH Stumpf, and Jaroslav Stark. Network motifs: Structure does not determine function. *BMC Genomics*, 7:108, May 2006. doi:10.1186/1471-2164-7-108. 68

[99] J. David Johnson. UCINET: A software tool for network analysis. *Communication Education*, 36(1):92–94, January 1987. doi:10.1080/03634528709378647. 21, 43

[100] Frédéric Kaplan. The Venice Time Machine. In *Proceedings of the 2015 ACM Symposium on Document Engineering*, DocEng '15, page 73, New York, NY, USA, September 2015. Association for Computing Machinery. doi:10.1145/2682571.2797071. 20, 33

[101] Karine Karila-Cohen, Claire Lemercier, Isabelle Rosé, and Claire Zalc. Nouvelles cuisines de l'histoire quantitative. *Annales. Histoire, Sciences Sociales*, 73(4):773–783, December 2018. doi:10.1017/ahss.2019.90. 15, 49, 50, 55

[102] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual Analytics: Definition, Process, and Challenges. In Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, Lecture Notes in Computer Science, pages 154–175. Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-70956-5_7. 9, 20, 29

[103] Daniel A Keim. Visual Analytics. page 6. 43

[104] Florian Kerschbaumer, Linda Keyserlingk, Martin Stark, and Marten Düring. *The Power of Networks. Prospects of Historical Network Research*. January 2015. 15, 16, 25, 40, 123

[105] Steffen Klamt, Utz-Uwe Haus, and Fabian Theis. Hypergraphs and cellular networks. *PLoS computational biology*, 5(5):e1000385, 2009. 94

[106] Jon Kleinberg. An Impossibility Theorem for Clustering. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002. 99

[107] Elena V. Konstantinova and Vladimir A. Skorobogatov. Application of hypergraph theory in chemistry. *Discrete Mathematics*, 235(1-3):365–383, May 2001. doi:10.1016/S0012-365X(00)00290-9. 94

[108] C. Kosak, J. Marks, and S. Shieber. Automating the layout of network diagrams with specified visual organization. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(3):440–454, March 1994. doi:10.1109/21.278993. 9, 42

[109] Ernest Labrousse. *La Crise de l'économie Française à La Fin de l'Ancien Régime et Au Début de La Révolution*, volume 1. Presses Universitaires de France-PUF, 1990. 31

[110] David S. Landes and Charles Tilly. *History as Social Science. The Behavioral and Social Sciences Survey*. Prentice Hall, Inc, 1971. 32

[111] Charles-Victor Langlois and Charles Seignobos. *Introduction aux études historiques*. ENS Éditions, February 2014. 15, 30

[112] Katherine A. Larson. Thomas F. Tartaron, Maritime Networks in the Mycenaean World. New York: Cambridge University Press, 2013. *Comparative Studies in Society and History*, 56(4):1064–1065, October 2014. doi:10.1017/S0010417514000516. 15

[113] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, January 2008. doi:10.1016/j.socnet.2007.04.006. 59, 76

[114] Emmanuel Lazega. *Réseaux sociaux et structures relationnelles*. Presses universitaires de France, Paris, 1998. 19, 38

[115] Claire Lemercier. Analyse de réseaux et histoire. *Revue dhistoire moderne contemporaine*, 522(2):88–112, 2005. 52

[116] Claire Lemercier. 12. Formal network methods in history: Why and how? In Georg Fertig, editor, *Social Networks, Political Institutions, and Rural Societies*, volume 11, pages 281–310. Brepols Publishers, Turnhout, January 2015. doi:10.1484/M.RURHE-EB.4.00198. 15, 21, 25, 37, 39, 40, 41, 44, 47, 55, 66, 121

[117] Claire Lemercier and Claire Zalc. *Quantitative Methods in the Humanities: An Introduction*. University of Virginia Press, March 2019. 16, 17, 21, 32, 33, 41, 48, 49, 50, 55, 67, 123

[118] Claire Lemercier and Claire Zalc. Back to the Sources: Practicing and Teaching Quantitative History in the 2020s. *Capitalism*, 2(2):473–508, 2021. doi:10.1353/cap.2021.0010. 21, 32, 47, 48, 49, 50, 55

[119] Bernard Lepetit. L'histoire quantitative : deux ou trois choses que je sais d'elle. *Histoire & Mesure*, 4(3):191–199, 1989. doi:10.3406/hism.1989.1355. 15, 49

[120] Carola Lipp. Kinship Networks, Local Government, and Elections in a Town in Southwest Germany, 1800-1850. *Journal of Family History*, 30(4):347–365, October 2005. doi:10.1177/0363199005278726. 39

[121] Carola Lipp and Lothar Krempel. Petitions and the Social Context of Political Mobilization in the Revolution of 1848/49: A Microhistorical Actor-Centred Network Analysis. *Int Rev of Soc His*, 46(S9):151–169, December 2001. doi:10.1017/S0020859001000281. 59

[122] Stephen Makonin, Daniel McVeigh, Wolfgang Stuerzlinger, Khoa Tran, and Fred Popowich. Mixed-Initiative for Big Data: The Intersection of Human + Visual Analytics + Prediction. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 1427–1436, January 2016. doi:10.1109/HICSS.2016.181. 96

[123] Gribaudi Maurizio. *Espaces, Temporalités, Stratifications :. Exercices Méthodologiques Sur Les Réseaux Sociaux*. Editions de l'Ecole des Hautes Etudes en Sciences Sociales, Paris, January 2000. 37

[124] Philip Mayer. Migrancy and the Study of Africans in Towns. *American Anthropologist*, 64(3):576–592, 1962. 38

[125] Fintan McGee, Benjamin Renoust, Daniel Archambault, Mohammad Ghoniem, Andreas Kerren, and Bruno Pinaud et al. *Visual Analysis of Multilayer Networks*. Synthesis Lectures on Visualization. Morgan & Claypool Publishers, 2021. 57

[126] Michael J. McGuffin. Simple algorithms for network visualization: A tutorial. *Tsinghua Science and Technology*, 17(4):383–398, August 2012. doi:10.1109/TST.2012.6297585. 42

[127] Pierre Mercklé and Claire Zalc. Peut-on modéliser la persécution ?: Apports et limites des approches quantifiées sur le terrain de la Shoah. *Annales. Histoire, Sciences Sociales*, 73(4):923–957, December 2018. doi:10.1017/ahss.2019.95. 15

[128] R. Michalski, P. Kazienko, and D. Krol. Predicting Social Network Measures Using Machine Learning Approach. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1056–1059, Istanbul, August 2012. IEEE. doi:10.1109/ASONAM.2012.183. 125

[129] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, October 2002. doi:10.1126/science.298.5594.824. 37, 68

[130] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Lulu.com, 2019. 100

[131] Megan Monroe, Krist Wongsuphasawat, Catherine Plaisant, Ben Shneiderman, Jeff Millstein, and Sigfried Gold. Exploring point and interval event patterns: Display methods and interactive visual query. *University of Maryland Technical Report*, 2012. 94

[132] J. L. Moreno. *Who Shall Survive?: A New Approach to the Problem of Human Interrelations*. Who Shall Survive?: A New Approach to the Problem of Human Interrelations. Nervous and Mental Disease Publishing Co, Washington, DC, US, 1934. doi:10.1037/10648-000. 9, 36, 41

[133] J. L. Moreno. Foundations of Sociometry: An Introduction. *Sociometry*, 4(1):15, February 1941. doi:10.2307/2785363. 35

[134] Zacarias Moutoukias. Buenos Aires, port between two oceans: Mobilities, networks, stratifications (2nd half of the 18th century). *E-SPANIA-REVUE ELECTRONIQUE D ETUDES HISPANIQUES MEDIEVALES*, 25, 2016. 51, 72

[135] Zacharias Moutoukias. Réseaux personnels et autorité coloniale : Les négociants de Buenos Aires au XVIIIe siècle. *Annales. Histoire, Sciences Sociales*, 47(4-5):889–915, October 1992. doi:10.3406/ahess.1992.279084. 39

[136] Andrej Mrvar and Vladimir Batagelj. Analysis and visualization of large networks with program package Pajek. *Complex Adaptive Systems Modeling*, 4(1), April 2016. doi:10.1186/s40294-016-0017-8. 43, 75

[137] Natural earth. 75

[138] Neo4j graph data platform. 67, 69, 85, 94

[139] Mark Newman. *Networks*. Oxford university press, 2018. 36

[140] Rolla Nicoletta. Mobilité et conflits. Travailler sur les chantiers de construction piémontais dans la première moitié du XVIIIe siècle. In Andrea Caracausi and Marco Schnyder, editors, *Travail et Mobilité En Europe (XVIe-XIXe Siècles)*, Coll. Histoire et Civilisations. Presses universitaires du Septentrion, Villeneuve d'Ascq, 2018. 51

[141] Carolina Nobre, Marc Streit, and Alexander Lex. Juniper: A Tree+Table Approach to Multivariate Graph Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):544–554, January 2019. doi:10.1109/TVCG.2018.2865149. 42

[142] Gérard Noiriel. Naissance du métier d'historien. *Genèses. Sciences sociales et histoire*, 1(1):58–85, 1990. doi:10.3406/genes.1990.1014. 30

[143] Juri Opitz, Leo Born, and Vivi Nastase. Induction of a Large-Scale Knowledge Graph from the Regesta Imperii. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 159–168, Santa Fe, New Mexico, August 2018. Association for Computational Linguistics. 21, 67

[144] Maryjane Osa. *Solidarity And Contention: Networks Of Polish Opposition*. Univ Of Minnesota Press, Minneapolis, first edition edition, July 2003. 15, 54

[145] John F. Padgett and Christopher K. Ansell. Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, 98(6):1259–1319, May 1993. doi:10.1086/230190. 9, 15, 19, 39, 40, 123

[146] Pajek — Analysis and visualization of very large networks. 20, 21, 98

[147] Terence J. Parr and Russell W. Quong. ANTLR: A predicated-LL (k) parser generator. *Software: Practice and Experience*, 25(7):789–810, 1995. 85

[148] Pamela Paxton. Dollars and Sense: Convincing Students That They Can Learn and Want to Learn Statistics. *Teach Sociol*, 34(1):65–70, January 2006. doi:10.1177/0092055X0603400106. 21, 56

[149] Vanessa Peña-Araya, Tong Xue, Emmanuel Pietriga, Laurent Amsaleg, and Anastasia Bezerianos. HyperStorylines: Interactively untangling dynamic hypergraphs. *Information Visualization*, 21(1):38–62, January 2022. doi:10.1177/14738716211045007. 61

[150] Cindarella Sarah Maria Petz. *On Combining Network Research and Computational Methods on Historical Research Questions and Its Implications for the Digital Humanities*. PhD thesis, Technische Universität München, 2022. 16, 40, 50, 123

[151] James P. Philips and Nasseh Tabrizi. Historical Document Processing: Historical Document Processing: A Survey of Techniques, Tools, and Trends, September 2020. arXiv:2002.06300, doi:10.48550/arXiv.2002.06300. 125

[152] Robert Pienta, Fred Hohman, Alex Endert, Acar Tamersoy, Kevin Roundy, Chris Gates, Shamkant Navathe, and Duen Horng Chau. VIGOR: Interactive Visual Exploration of Graph Query Results. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):215–225, January 2018. doi:10.1109/TVCG.2017.2744898. 69

[153] Alexis Pister, Paolo Buono, Jean-Daniel Fekete, Catherine Plaisant, and Paola Valdivia. Integrating Prior Knowledge in Mixed-Initiative Social Network Clustering. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1775–1785, February 2021. doi:10.1109/TVCG.2020.3030347. 22, 43

[154] Alexis Pister, Nicole Dufournaud, Pascal Cristofoli, Christophe Prieur, and Jean-Daniel Fekete. From Historical Documents To Social Network Visualization: Potential Pitfalls and Network Modeling. In *7th Workshop on Visualization for the Digital Humanities (VIS4DH)*, 2022. 48

[155] Alexis Pister, Christophe Prieur, and Jean-Daniel Fekete. Visual Queries on Bipartite Multivariate Dynamic Social Networks. The Eurographics Association, 2022. doi:10.2312/evp.20221115. 66

[156] Antoine Prost. *Douze Leçons sur l'histoire*. Média Diffusion, April 2014. 16, 25, 30, 31

[157] Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, January 2007. doi:10.1093/bioinformatics/btl301. 37

[158] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. 69, 98

[159] Eric Ragan, Endert Alex, Jibonananda Sanyal, and Jian Chen. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), January 2016. doi:10.1109/TVCG.2015.2467551. 70

[160] Ramana Rao and Stuart K. Card. The table lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 318–322, New York, NY, USA, 1994. Association for Computing Machinery. doi:10.1145/191666.191776. 119

[161] Donghao Ren, Bongshin Lee, and Matthew Brehmer. Charticulator: Interactive Construction of Bespoke Chart Layouts. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):789–799, January 2019. doi:10.1109/TVCG.2018.2865158. 70

[162] Pedro Ribeiro and Fernando Silva. Discovering Colored Network Motifs. In Pierluigi Contucci, Ronaldo Menezes, Andrea Omicini, and Julia Poncela-Casasnovas, editors, *Complex Networks V*, Studies in Computational Intelligence, pages 107–118, Cham, 2014. Springer International Publishing. doi:10.1007/978-3-319-05401-8_11. 68

[163] Christian Rollinger. Prolegomena. Problems and Perspectives of Historical Network Research and Ancient History. *Journal of Historical Network Research*, 4:1–35, May 2020. doi:10.25517/jhnr.v4i0.72. 21, 22, 25, 56, 121

[164] Giulio Rossetti and Rémy Cazabet. Community discovery in dynamic networks: A survey. *ACM computing surveys (CSUR)*, 51(2):1–37, 2018. 40, 123

[165] Giulio Rossetti, Letizia Milli, and Rémy Cazabet. CDLIB: A python library to extract, compare and evaluate communities from complex networks. *Appl Netw Sci*, 4(1):52, July 2019. doi:10.1007/s41109-019-0165-9. 101

[166] Fabrice Rossi, Nathalie Vialaneix, and Florent Hautefeuille. Exploration of a large database of French notarial acts with social network methods. *Digital Medievalist*, 9:2013, July 2014. 66

[167] Juan A. Rubio-Mondejar and Josean Garrues-Irurzun. Women entrepreneurs and family networks in Andalusia (Spain) during the second industrial revolution. *Business History*, pages 1–22, May 2022. doi:10.1080/00076791.2022.2068524. 15

[168] C.J. Rueda and Catedral de Buenos Aires. *Matrimonios de La Catedral de Buenos Aires, 1747-1823*. Number v. 2 in Fuentes Históricas y Genealógicas Argentinas. Fuentes Históricas y Genealógicas Argentinas, 1989. 51

[169] Anni Sairio. Methodological and practical aspects of historical network analysis: A case study of the Bluestocking letters. In Arja Nurmi, Minna Nevala, and Minna Palander-Collin, editors, *Pragmatics & Beyond New Series*, volume 183, pages 107–135. John Benjamins Publishing Company, Amsterdam, 2009. doi:10.1075/pbns.183.08sai. 57

[170] Bahador Saket, Paolo Simonetto, and Stephen Kobourov. Group-level graph visualization taxonomy. In N. Elmqvist, M. Hlawitschka, and J. Kennedy, editors, *EuroVis - Short Papers*. The Eurographics Association, 2014. doi:10.2312/eurovisshort.20141162. 102

[171] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. Vega-lite: A grammar of interactive graphics. *IEEE Trans. Vis. Comput. Graphics*, 23(1):341–350, 2016. 29, 85

[172] Shrutika S. Sawant and Manoharan Prabukumar. A review on graph-based semi-supervised learning methods for hyperspectral image classification. *The Egyptian Journal of Remote Sensing and Space Science*, 2018. doi:10.1016/j.ejrs.2018.11.001. 101

[173] John Scott. Social Network Analysis. *Sociology*, 22(1):109–127, February 1988. doi:10.1177/0038038588022001007. 25, 35, 36, 43, 54

[174] Vanessa Serrano Molinero, Benjamin Bach, Catherine Plaisant, Nicole Dufournaud, and Jean-Daniel Fekete. Understanding the use of the vistorian: Complementing logs with context mini-questionnaires. In *Visualization for the Digital Humanities Workshop*, Phoenix, United States, October 2017. 9, 42, 43, 44, 53

[175] Rachel Shadoan and Chris Weaver. Visual Analysis of Higher-Order Conjunctive Relationships in Multidimensional Data Using a Hypergraph Query System. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2070–2079, December 2013. doi:10.1109/TVCG.2013.220. 69

[176] Termeh Shafie, David Schoch, Jimmy Mans, Corinne Hofman, and Ulrik Brandes. Hypergraph Representations: A Study of Carib Attacks on Colonial Forces, 1509-1700. *Journal of Historical Network Research*, pages 52–70 Pages, October 2017. doi:10.25517/JHNR.V1I1.6. 59, 66

[177] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, September 1996. doi:10.1109/VL.1996.545307. 28, 29

[178] Ben Shneiderman. Dynamic queries for visual information seeking. *IEEE Softw.*, 11(6):70–77, November 1994. doi:10.1109/52.329404. 78

[179] Ben Shneiderman. Inventing Discovery Tools: Combining Information Visualization with Data Mining. *Information Visualization*, 1(1):5–12, March 2002. doi:10.1057/palgrave.ivs.9500006. 18

[180] Georg Simmel. *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. Duncker & Humblot, Berlin, 7. aufl edition, 2013. 37

[181] Marc A. Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave. Analyzing (social media) networks with NodeXL. In John M. Carroll, editor, *Proceedings of the Fourth International Conference on Communities and Technologies, C&T 2009, University Park, PA, USA, June 25-27, 2009*, pages 255–264. ACM, 2009. doi:10.1145/1556460.1556497. 20, 43, 67, 75

[182] John Snow. On the Mode of Communication of Cholera. *Edinb Med J*, 1(7):668–670, January 1856. 26

[183] John T. Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Inf. Vis.*, 7(2):118–132, 2008. doi:10.1057/palgrave.ivs.9500180. 45, 60, 61, 123

[184] Chris Stolte, Diane Tang, and Pat Hanrahan. Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 8(1):14, 2002. 61

[185] Lawrence Stone. The Revival of Narrative: Reflections on a New Old History. *Past & Present*, (85):3–24, 1979. 32

[186] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617, 2002. 102

[187] Shazia Tabassum, Fabiola S. F. Pereira, Sofia Fernandes, and João Gama. Social network analysis: An overview. *WIREs Data Mining and Knowledge Discovery*, 8(5):e1256, 2018. doi:10.1002/widm.1256. 15, 19, 36

[188] Melissa Terras. Quantifying digital humanities. *UCL Centre for Digital Humanities*, 2011. 33

[189] J.J. Thomas and K.A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, January 2006. doi:10.1109/MCG.2006.5. 29

[190] Charles Tilly. Retrieving european lives. 1984. 16, 25, 29, 48, 66

[191] Charles Tilly. Observations of Social Processes and Their Formal Representations. *Sociological Theory*, 22(4):595–602, 2004. doi:10.1111/j.0735-2751.2004.00235.x. 25, 49

[192] Natkamon Tovanich, Alexis Pister, Gaelle Richer, Paola Valdivia, Christophe Prieur, Jean-Daniel Fekete, and Petra Isenberg. VAST 2020 Contest Challenge: GraphMatchMaker: Visual Analytics for Graph Comparison and Matching. *IEEE Computer Graphics and Applications*, pages 1–1, 2021. doi:10.1109/MCG.2021.3091955. 68, 69

[193] Francesca Trivellato. Is There a Future for Italian Microhistory in the Age of Global History? *California Italian Studies*, 2(1), 2011. doi:10.5070/C321009025. 32, 50

[194] John W. Tukey. The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33(1):1–67, 1962. 26

[195] John W. Tukey. *Exploratory Data Analysis*. Pearson, Reading, Mass, 1er édition edition, January 1977. 19, 29

[196] Paola Valdivia, Paolo Buono, Catherine Plaisant, Nicole Dufournaud, and Jean-Daniel Fekete. Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. *IEEE Trans. Visual. Comput. Graphics*, 27(1):1–13, January 2021. doi:10.1109/TVCG.2019.2933196. 42, 61, 96, 125

[197] Guido van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995. 69

[198] Ingeborg van Vugt. Using multi-layered networks to disclose books in the republic of letters. *Journal of Historical Network Research*, 1(1):25–51, October 2017. 57

[199] Corinna Vehlow, Fabian Beck, and Daniel Weiskopf. The state of the art in visualizing group structures in graphs. In R. Borgo, F. Ganovelli, and I. Viola, editors, *Eurographics Conference on Visualization (EuroVis) - STARs*. The Eurographics Association, 2015. doi:10.2312/eurovisstar.20151110. 102

[200] VisMaster: Visual analytics — Mastering the information age. 116

[201] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, et al. Constrained k-means clustering with background knowledge. In *Icml*, volume 1, pages 577–584, 2001. 101

[202] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, November 1994. 19, 37, 43

[203] Charles Wetherell. Historical Social Network Analysis. *Int Rev of Soc His*, 43(S6):125–144, December 1998. doi:10.1017/S0020859000115123. 15, 17, 25, 38, 39, 48, 54, 57, 66, 123

[204] Robert Whaples. Where Is There Consensus Among American Economic Historians? The Results of a Survey on Forty Propositions. *The Journal of Economic History*, 55(1):139–154, March 1995. doi:10.1017/S0022050700040602. 32

[205] Douglas White, Douglas R. White, and Ulla Johansen. *Network Analysis and Ethnographic Problems: Process Models of a Turkish Nomad Clan*. Lexington Books, 2005. 15

[206] Hadley Wickham and Maintainer Hadley Wickham. The ggplot package. *Google Scholar*, 2007. 29

[207] Leland Wilkinson. *The Grammar Of Graphics*. Springer, New York, 1993. 28

[208] Ian Winchester. The Linkage of Historical Records by Man and Computer: Techniques and Problems. *Journal of Interdisciplinary History*, 1(1):107, 1970. doi:10.2307/202411. 90

[209] Alvin W. Wolfe. The rise of network thinking in anthropology. *Social Networks*, 1(1):53–64, January 1978. doi:10.1016/0378-8733(78)90012-6. 18

[210] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum*, 39(3):757–783, June 2020. doi:10.1111/cgf.14035. 55, 70

[211] Franciszek Zakrzewski. The 1932 population register, May 2020. 50

[212] Michelle X. Zhou. "Big picture": Mixed-initiative visual analytics of big data. In *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*, VINCI '13, page 120, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2493102.2499786. 101