THESE DE DOCTORAT

NNT : 2022UPASG081

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

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

Thèse de doctorat de l'université Paris-Saclay

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 l'Institut interdisciplinaire de l'innovation (Télécom Paris, Institut Polytechnique de 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 15 décembre 2022, par

Alexis PISTER

Composition du jury

Membres du jury avec voix délibérative

Guy Melançon Professeur, Univerité de Bordeaux	Président & Rapporteur
Ulrik Brandes Professeur, ETH Zürich	Rapporteur & Examinateur
Laurent Beauguitte Chargé de recherche, CNRS Paris	Examinateur
Uta Hinrichs Professeur, Université d'Édimbourg	Examinatrice
Wendy Mackay Directrice de recherche, Univ. Paris-Saclay, CNRS, Inria, LISN	Examinatrice

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 théoriquement et concrètement comment l'analyse visuelle peut aider les historiens dans leur processus d'analyse de réseaux sociaux. L'analyse de réseaux sociaux est une méthode utilisée en histoire sociale qui vise à étudier les relations sociales au sein de groupes d'acteurs (familles, institutions, entreprises, etc.) en reconstruisant les relations du passé à partir de documents historiques, tels que des actes de mariages, des actes de naissances, ou des recensements. L'utilisation de méthodes visuelles et analytiques leur permet d'explorer la structure sociale formant ces groupes ainsi que de relier des mesures structurelles à des hypothèses sociologiques et à des comportements individuels. Cependant, l'encodage et la modélisation des sources menant à un réseau finalisé donnent souvent lieu à des erreurs, des distorsions et des problèmes de traçabilité, et les systèmes de visualisation actuels présentent souvent des défauts d'utilisabilité et d'interprétabilité. En conséquence, les historiens ne sont pas toujours en mesure d'aboutir à des conclusions approfondies à partir de ces systèmes : beaucoup d'études se limitent à une description qualitative d'images de réseaux, surlignant la présence de motifs d'intérêts (cliques, îlots, ponts, etc.). Le but de cette thèse est donc de proposer des outils d'analyse visuelle adaptés aux historiens afin de leur permettre une meilleure intégration de leur processus global et des capacités d'analyse guidées. En collaboration avec des historiens, je formalise le processus d'une analyse de réseau historique, de l'acquisition des sources jusqu'à l'analyse finale, en posant comme critère que les outils utilisés dans ce processus devraient satisfaire des principes de traçabilité, de simplicité et de réalité documentaire (i.e., que les données présentées doivent être conformes aux sources) pour faciliter les va-et-vient entre les différentes étapes et la prise en main par l'utilisateur, et ne pas distordre le contenu des sources. Pour satisfaire ces propriétés, je propose de modéliser les sources historiques en réseaux sociaux bipartis multivariés dynamiques avec rôles. Ce modèle intègre explicitement les documents historiques sous forme de nœuds, ce qui permet aux utilisateurs d'encoder, de corriger et d'analyser leurs données avec les mêmes outils. Je propose ensuite deux interfaces d'analyse visuelle permettant, avec une bonne utilisabilité et interprétabilité, de manipuler, d'explorer et d'analyser ce modèle de données. Le premier système ComBiNet offre une exploration visuelle de l'ensemble des dimensions du réseau à l'aide de vues coordonnées et d'un système de requêtes visuelles permettant d'isoler des individus ou des groupes et de comparer leurs structures topologiques et leurs propriétés. L'outil permet également de détecter les motifs inhabituels et ainsi de déceler les éventuelles erreurs dans les annotations. Le second système, PK-Clustering, est une proposition d'amélioration de l'utilisabilité et de l'efficacité des mécanismes de clustering dans les systèmes de visualisation de réseaux sociaux. L'interface permet de créer des regroupements pertinents à partir des connaissances a priori de l'utilisateur, du consensus algorithmique et de l'exploration du réseau dans un cadre d'initiative mixte. Les deux systèmes ont été conçus à partir des besoins et retours continus d'historiens, et visent à augmenter la traçabilité, la simplicité, et la réalité documentaire des sources dans le processus d'analyse de réseaux historiques. Je conclus sur la necessité d'une meilleure intégration des systèmes d'analyse visuelle dans le processus de recherche des historiens. Cette intégration nécessite des outils plaçant les utilisateurs au centre du processus avec un accent sur la flexibilité et l'utilisabilité, limitant ainsi l'introduction de biais et les barrières d'utilisation des méthodes quantitatives, qui subsitent 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 theoretically and concretely how visual analytics can support historians in their social network analysis process. Historical social network analysis is a method to study social relationships between groups of actors (families, institutions, companies, etc.) through a reconstruction of relationships of the past from historical documents, such as marriage acts, migration forms, birth certificates, and censuses. The use of visualization and analytical methods lets social historians explore and describe the social structure shaping those groups while explaining sociological phenomena and individual behaviors through computed network measures. However, the inspection and encoding of the sources leading to a finalized network is intricate and often results in inconsistencies, errors, distortions, and traceability problems, and current visualization tools typically have usability and interpretability issues. For these reasons, social historians are not always able to make thorough historical conclusions: many studies consist of qualitative descriptions of network drawings highlighting the presence of motifs such as cliques, components, bridges, etc. The goal of this thesis is therefore to propose visual analytics tools integrated into the global social historians' workflow, with guided and easy-to-use analysis capabilities. From collaborations with historians, I formalize the workflow of historical network analysis starting at the acquisition of sources to the final visual analysis. By highlighting recurring pitfalls, I point out that 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. To satisfy those properties, I propose to model historical sources into bipartite multivariate dynamic social networks with roles as they provide a good tradeoff of simplicity and expressiveness while modeling explicitly the documents, hence letting users encode, correct, and analyze their data with the same abstraction and tools. I then propose two interactive visual interfaces to manipulate, explore, and analyze this data model, with a focus on usability and interpretability. The first system ComBiNet allows an interactive exploration leveraging the structure, time, localization, and attributes of the data model with the help of coordinated views and a visual query system allowing users to isolate interesting groups and individuals, and compare their position, structures, and properties. It also lets them highlight erroneous and inconsistent annotations directly in the interface. The second system, PK-Clustering, is a concrete proposition to enhance 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 tools, and more globally on research directions towards better integration of visual analytics systems on the whole workflow of social historians. 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

2 Related Work

Contents

Social historians rely on textual historical documents to study social groups through their structures and socio-economic characteristics in societies of the past [40, 234]. They read and analyze documents they can find from a period and subject of interest, and make their conclusions through deep inspection and cross-referencing of the information they found. Several methods have been developed in history to extract and analyze the information contained in the documents in a methodical way [235], based on qualitative or quantitative methods—among which HSNA is now widely popular [200]. HSNA is a method consisting in modeling the relational information mentioned in the documents—such as family, business, or friendship ties—in a network, to be able to characterize and explain social behaviors through the description of the network's structure [133, 248]. This approach is directly inspired by SNA, which is a well-known method that sociologists theorized to understand and describe real-world social relationships modeled as networks [86, 212]. Historians appropriated this method, by extracting relationships from historical documents. The specifics of HSNA in contrast to its sociology counterpart is, therefore, the modeling of the network from the historical documents—which are at the core of the historical work [193]—and the integration of the temporal dimension which is often disregarded in traditional SNA but central in history. Once they successfully constructed a network—which is a long and tedious process—they typically use network measures and visualization techniques to confirm or generate new hypotheses [151]. Visualization let them unfold the structure of their data, revealing potentially interesting social patterns between actors and groups of actors. Analytics and visualization systems for SNA typically allow the exploration of such data with the help of interaction, network measures, and data mining capabilities such as clustering directly implemented in the interfaces. Yet, most HSNA studies only give a qualitative description of their network—which Rollinger calls "soft" or "informal" network research [200]—probably due to usability and formalism issues [6]. The coupling of visualization and data mining through interaction to support the generation of knowledge has been described as VA and can therefore provide support to social historians for their network construction, but also to go beyond simple qualitative description of their data. In this chapter, I first present a general overview of the field of visualization in §2.1 to share its utility and potential for social history. Then, I present the social history discipline with its use of quantitative methods in §2.2, before describing in depth how network analysis has been applied in the field in §2.3. Finally, I present in §2.4 how visualization and VA have been used in the context of HSNA, along with the most popular systems currently used by social scientists and their limitations.

2.1 Visualization

Visualization is often defined as "the use of computer-supported, interactive, visual representations of data to amplify cognition" [44]. Graphically displaying data allows us to leverage our visual system to gain a better acquisition of knowledge, leading to better decision-making, communication, and potential discoveries. The field of visualization can be split into three subdomains: *Scientific visualization* focuses on visualizing continuous physically based data such as weather, astrophysics, and anatomical data, sometimes produced with simulations whereas *Information Visualization* is centered around the visualization of discrete abstract data points, often multidimensional. *Visual Analytics* emerged later from Information Visualization by mixing data mining and more complex analysis process with traditional information visualization displays. I focus in this thesis on the two latter branches of visualization, as social scientists can use both information visualization and VA systems to gain insight into the structure of the networks they are studying.

2.1.1 Information Visualization

Information Visualization focuses on displaying abstract data to amplify cognition and gain insight into real-world phenomena [44]. History is filled with classical examples of visual data displays which helped understand better specific events, such as Minard's map of Napoleon's march in Russia [88], or Snow's dot map of cholera cases in London which showed the proximity between street pumps and cholera infections [224]. If several examples of information visualization can be found thorough history, it mainly developed as a scientific field in the 1960s with Tukey's work on data analysis and visualization [238] and Bertin's publication of *Semiology of graphics* [25].

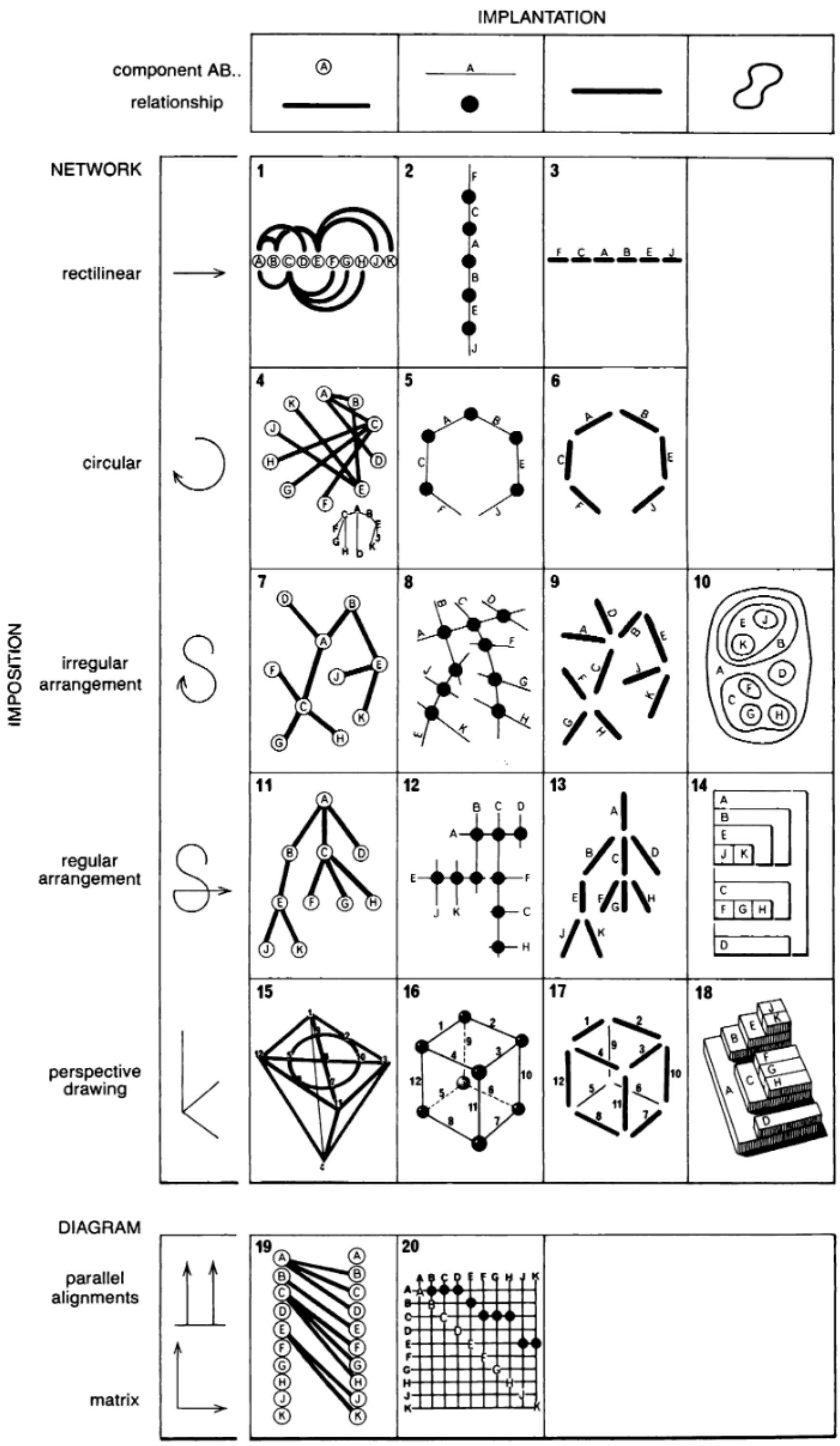

Figure 2.1 – Categorization of visual variables which can be used to represent network data, resulting in many different network representations. Image from [25].

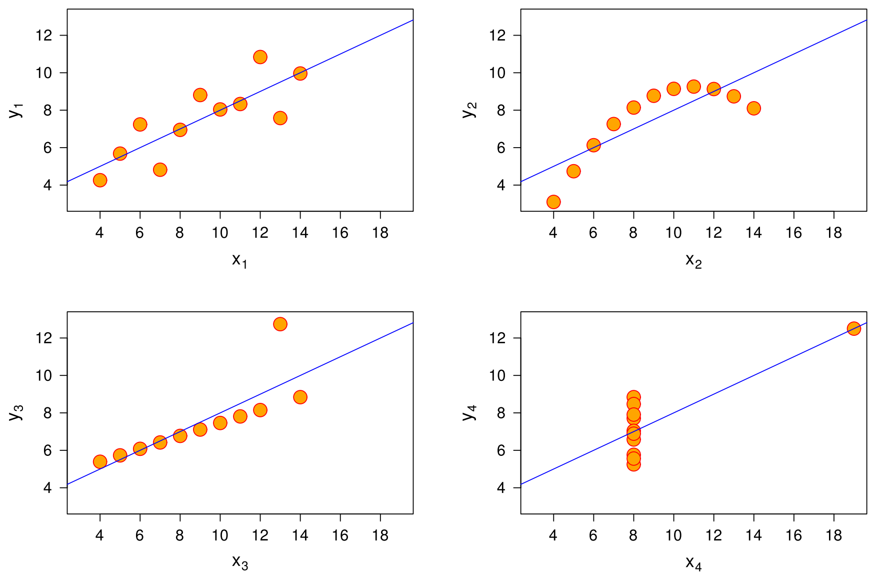

Figure 2.2 – Anscombe quartet. The four datasets have the same descriptive statistics (average, variance, correlation coefficient) but very different structures. Image from [9].

In this foundational work, Bertin described and organized the different visual elements usable in graphical information displays, and linked them to data features and relations types. An illustration of this work of categorization for network data is illustrated in Figure 2.1. Michael Friendly writes "To some, this appeared to do for graphics what Mendeleev had done for the organization of the chemical elements" [89]. The development of computer science and the rise of hardware capabilities at the same time created a big need for data visualization. The amount of data stored increased exponentially [114] and descriptive statistics were not enough to understand the underlying structure of the amount and diversity of produced data. Visualization, leveraging the human visual system, enabled to rapidly see the hidden structure of a dataset and detect interesting and unexpected patterns very often unseen with classical statistical methods. One classical illustration of this is Anscombe's quartet [9] which consists of four datasets of 11 points in $\mathbb{R}^2$ with the same statistical measures (mean, variance, correlation coefficient, etc.) but with very different structures, that are immediately revealed when plotting the data. The four datasets are illustrated in Figure 2.2.

A large number of visualization techniques emerged to make sense of the diversity of data produced, such as multidimensional, temporal, spatial, or network data [218]. Instead of using taxonomies classifying graphics into categories such as histograms, pie charts, and stream graphs, some theorized how to describe graphics in a more systematic and structural way. In 1993, Wilkinson extended Bertin's work and developed the *Grammar of Graphics* [252] as a way to describe the deep structure unifying every possible graphic, thus allowing to characterize and create graphics using common terms and rules. In this framework, a graphic can be defined as a function of six components: *data* (a set of data points and attributes from a dataset), *transformations* (statistical operations which modify the original data, e.g., mean and rank transformations), *scales* (e.g., linear and log scales), *coordinate systems* (e.g., cartesian and polar coordinate systems), *elements* (graphical marks such as rectangular or circular marks, and their aesthetics, e.g., color, and size), and *guides* (additional information such as axes and legend). Many well-known visualization toolkits are now based on this framework, such as vega [209] and ggplot [251], as it enables a greater expressiveness and reusability for graphic creation. Visualization allows to gain insight into the structure of a given dataset and has traditionally been used for confirmation and communication purposes [220], for example, to verify hypotheses on empirical sciences, and later on to communicate findings, first to scientific peers, and nowadays to broader audiences for example through the means of data journalism [35].

2.1.2 Visual Analytics

VA consists of the coupling of visualization and data mining techniques to better support users in their knowledge generation process through continuous interaction with the data and statistical models [232]. It draws inspiration from exploratory visualization, interaction, and data mining. The process of exploratory visualization to gain new insights on the general structure of the data and potentially generate new hypotheses has been characterized by Tukey in 1960 as *exploratory data analysis* [239]. It consists in trying to characterize the structure of a dataset with the help of continuous visualization and statistical measurements of different dimensions of the data. Visual exploration is enhanced by direct manipulation interfaces through interaction and usually follows the information-seeking mantra formalized by Shneiderman: "Overview first, zoom and filter, then details-on-demand" [218]. It allows users to first have a visual overview of the data and get an idea of its overall structure, to then change the point of focus to highlight interesting patterns with the help of filtering, querying, sorting, and zooming mechanisms. As the average size of datasets keeps growing, exploratory tools are often needed to make sense of large datasets and generate pertinent hypotheses.

More recent visual exploration interfaces also incorporate automatic analytical tools along with graphical displays, letting users apply data mining algorithms directly in the exploratory loop. This coupling of visualization and analytical methods such as data mining has been defined as VA and is still a very active research field. Keim et al. define it as "a combination of automatic and visual analysis methods with a tight coupling through human interaction in order to gain knowledge from data" [131].

VA consists of the generation of knowledge using visualizations and statistical models of the data, that the user can explore using interaction. Such systems have been developed in various empirical domains, such as biology, astronomy, engineering, and social sciences, to explore various data types: multidimensional, temporal, geolocated, or relational (i.e., modeled into a network). Figure 2.3 shows the TULIP system, an example of a VA system developed for the analysis of network data. I discuss the uses of VA specifically for SNA in §2.4.2.

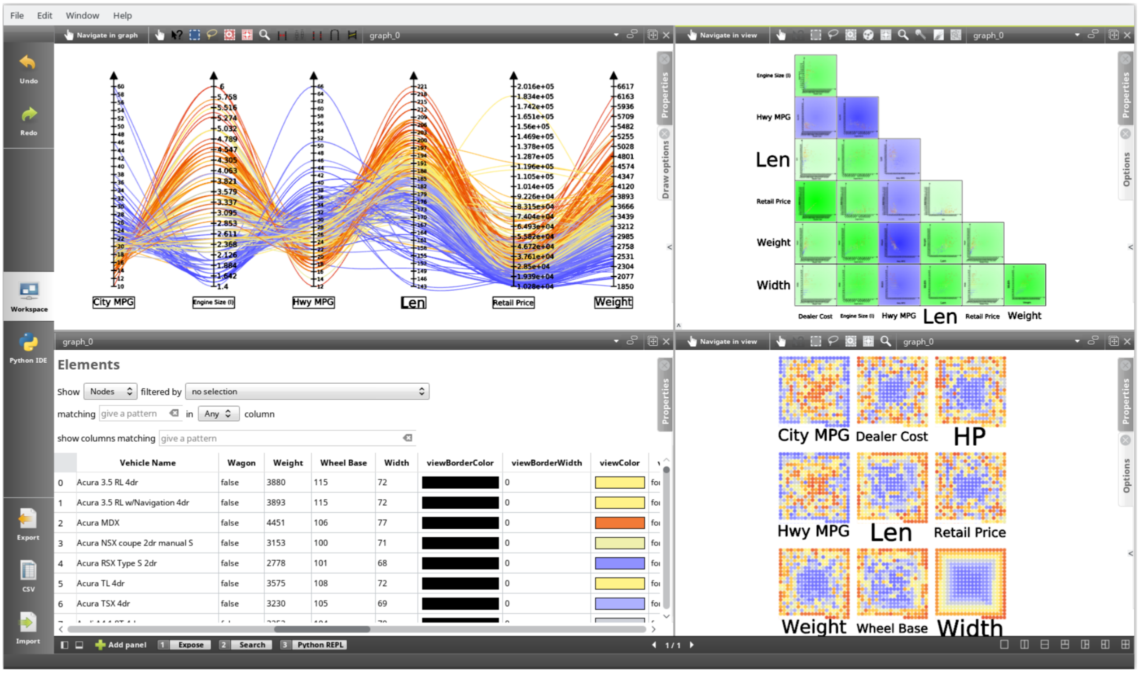

Figure 2.3 – TULIP software is designed for application-independent network visual analytics [12]. The view shows a dataset among multiple interactive coordinated views. Users can also apply data mining algorithms on the data to extract interesting patterns.

2.2 Quantitative Social History

Social History is a branch of history that aims at studying socio-economic aspects of past societies, with a focus on groups instead of specific individuals only. Charles Tilly writes that its goal is to "(i) documenting large structural changes, (2) reconstructing the experiences of ordinary people in the course of those changes, and (3) connecting the two" [234]. If the purpose of social history remained the same across time, methods and formalisms have evolved since its beginning in the 1930s. Specifically, the rise of computer science led to the development of quantitative history methods in the 1960s—now often referred to as *digital humanities*—which brought new ways of grounding results in formalisms and quantitative models, instead of solely relying on qualitative inspection of historical documents [108]. I discuss in this section the evolution of social history from the context of its beginning to the use of more recent quantitative approaches.

2.2.1 History, Social History, and Methodology

The concept of history is hard to define as its practice and codes highly evolved through time. Prost writes "history is what historians do. The discipline called history is not an eternal essence, a Platonic idea. It is a reality that is itself historical, i.e. situated in time and space, carried out by men who call themselves historians and are recognized as such, received as history by various publics [193]." Retrospectively, history of a given time can thus be characterized by the different historical work produced at that time. Nevertheless, History can be characterized as the collection and study of historical documents in the aim of retrieving facts about past societies people living in them. As Langlois and Seignobos write, "The search for and the collection of documents is thus a part, logically the first and most important part, of the historian's craft" [145]. History emerged as a field with its own rules, conventions, and journals in the 1880s from faculties of letters, to counterbalance previous history works which were judged as too "literary" [180]. At that time and until now, two facets characterize the field, which are sometimes overlapping: one is political whereas the other one is methodological. The former aspect of history serves to create a shared story for countries and a sense of unity among citizens. Antoine Prost says for example that "it's through history than France thinks itself" [193]. The latter aspect of history constitutes a methodology to describe the past through methodical inspection of historical sources, with the aim of inferring dated facts about the past and trying to minimize possible bias. Historical documents are thus at the core of the work of historians and having to cite historical documents and previous peers' work for new claims is primordial to be considered rigorous History work. However, methodological and epistemological facets (how historians should read and analyze their sources, how to cite them, what to report/not report, and what is the status of proof) of History have not been studied and discussed for a long time, until the end of the 1980s [193]. Some historians were interested in historiography [43], but none were going to philosophical and epistemological debates on the History discipline. For Lucien Fèbvre, philosophizing was even constituting a "capital crime" [77].

Retrospectively, we can still observe shifts in the objects of study of historians through time, and their relation to sources. History was at first mainly event-centered and was focusing on characterizing central figures of the past like rulers and artists or shedding light on central events like wars or political crises. This narrative approach to history has been criticized for its open interpretation of historical documents, which can introduce bias from the authors [34]. In the 1930s, March Bloch and Lucien Fèbvre detached from traditional history by creating the "Annales School" (École des Annales) which aimed at placing humans as a component of a broader sociological, political, and economic system with influences on each other [40]. They strongly advised exhaustively searching from archives, to ground historical results in documents, texts, and numbers. This new way of studying past events and societies became successful in a profession in crisis, by bringing a new lens of study on various societal subjects more grounded in sources and with better intelligibility. This school of thought can be seen as one of the biggest milestones for social history, which focuses on the socio-economical aspects of societies and their changes through time, rather than an event-centric view of History. For example, in his thesis, Ernest Labrousse—a well-known figure in social history—tries to describe and explain the economic crisis of France at the end of the "Ancien Régime"[1] through the evolution of the economic power of different social groups such as farmers, workers, property owners etc., instead of solely describing memorable facts about the period [143]. Social history continued to evolve since the 1930s, introducing new methods and concepts, but always with the goal

[1]The "Ancien Régime" is a historical period of France which starts from the beginning of the reign of the Bourbon house at 1589 until the Revolution in 1789.

to describe periods and historical facts through a sociological lens and with a strong focus on sources and traceability.

2.2.2 Quantitative History

With the development of statistical methods and computer science, quantitative approaches to history emerged in the 1960s with the goal of analyzing numeric data directly extracted from historical documents. Economists led this first wave of quantification by studying past events using economical concepts and data. This approach called "new economic history" or "cliometrics" was popularized by Fogel's study on the economic impact of the development of railroads in America [80] and Fogel and Engerman's controversial work on the economy of slavery [82]. In the latter study, they extracted numbers of a sample of 5000 bills of slave sales from New Orleans to support the controversial claims that slavery was economically viable and that slaves had a decent material life, which brought up heated debate among the scientific community and the broad audience [249]. Despite the controversy, these kinds of approaches rapidly started to be used in other related domains such as demography, social history, and political history, sometimes rebranded as "new social history" and "new political history" [153]. Using computer-supported methods, historians were able to store data extracted from historical documents and make conclusions based on computational methods such as regression and statistical testing. Many saw the future of social sciences in computer programming, as Le Roy-Ladurie who wrote in 1968 "The historian of tomorrow will be a programmer, or he will not exist" [152].

However, quantitative methods started to be criticized in the 1980s with a wave of disillusionment, for several reasons. Stone was the first to raise his voice in 1979, after participating himself in several of those ambitious projects: "It is just those projects that have been the most lavishly funded, the most ambitious in the assembly of vast quantities of data by armies of paid researchers, the most scientifically processed by the very latest in computer technology, the most mathematically sophisticated in presentation, which have so far turned out to be the most disappointing" [227]. First, many researchers of this first wave dispensed themselves with source criticism, leading to simplification, anachronisms—such as using modern analytical categories and indices like the GDP—and taking the numeric data from historical documents as objective. These problems could be in part explained by the fact that the work process was highly divided, meaning that the people analyzing the data did not necessarily inspect and read the original historical documents in depth. Indeed, "new history" projects often relied on a high division of labor among researchers, assistants, and students who operated with punch card operators [144], since extracting the data from raw documents and uploading it to computers—which were shared among whole departments—was very time-consuming at that time. Secondly, the popularity of these methods made practitioners forget about the many biases inherent to statistics, such as the sampling bias, or the fact that historical data is essentially incomplete data. This resulted in the computation of long data series and aggregates which were sometimes nonsensical given the gaps in the sources [152]. Finally, many historians raised their voices against the study of long-term trends instead of focusing on specific events and individuals. They challenged aggregation procedures and their assumptions, trying to go back to a more complex history by pointing out that phenomena have to be studied and understood through several scales [237]. Indeed, computing correlations and aggregates at a national level greatly simplify complex phenomena and misses specific group and individual related behaviors. Still, if their adoption remains slow and sometimes criticized among historians, quantitative methods provide tools to store, explore, and analyze historical documents systematically if used appropriately (i.e. not trying to bias the analysis, and not losing the trace of the original sources), especially that those methods highly evolved since the 1960s.

2.2.3 Digital Humanities

Digital Humanities is sometimes described as the second wave of computational social sciences [152]. The term has gained popularity since the 2010s and refers to "research and teaching taking place at the intersection of digital technologies and humanities. Digital Humanities aims to produce and use applications and models that make possible new kinds of teaching and research, both in the humanities and in computer science (and its allied technologies). Digital Humanities also studies the impact of these techniques on cultural heritage, memory institutions, libraries, archives and digital culture" [231]. If the first wave of computational social sciences focused a lot on statistical methods such as regression models, correlation testing, and descriptive measures (mean, median, and variance) to make conclusions, digital humanities focuses also on the use of digital tools for exploration, teaching, and communication of humanities concepts and data, leveraging design, infographics, and interactive systems [39]. In the context of historical research, the term *digital history* has been coined as "an approach to examining and representing the past that works with the new communication technologies of the computer, the Internet network, and software systems. On one level, digital history is an open arena of scholarly production and communication, encompassing the development of new course materials and scholarly data collections. On another, it is a methodological approach framed by the hypertextual power of these technologies to make, define, query, and annotate associations in the human record of the past. To do digital history, then, is to create a framework, an ontology, through the technology for people to experience, read, and follow an argument about a historical problem" [1]. Research that labels itself as digital history pivots around the curation and digitization of historical archives, the identification of historical concepts through computational and exploration methods, and also their communication to the general audience through digital technologies. Many Digital History projects are thus multidisciplinary by essence and involve several teams of researchers, such as the *Mapping the Republic of Letters* project which consisted of digitizing, storing, and exploring letters of scholars across the world in the 17th and 18th centuries, in a common hub and using shared visualization tools [72]. It resulted in the elaboration of curated datasets and visualizations concerning the correspondence of various scholars such as Voltaire, Benjamin Franklin (see Figure 2.4), and John Locke, accessible in the same place by researchers and the general audience. With modern technologies and infrastructures, it also becomes possible to study large historical databases—often labeled under the term "big data"—as with the *Venice Time Machine* project [128] which aims at digitizing and analyzing thousands of documents from the archives of Venice to understand the political, geographical, and sociological dynamics of the cities across generations and centuries. Yet, some social scientists raised concern about this type of project, fearing that it could rapidly bring the same type of issues encountered during the first wave of quantification, especially

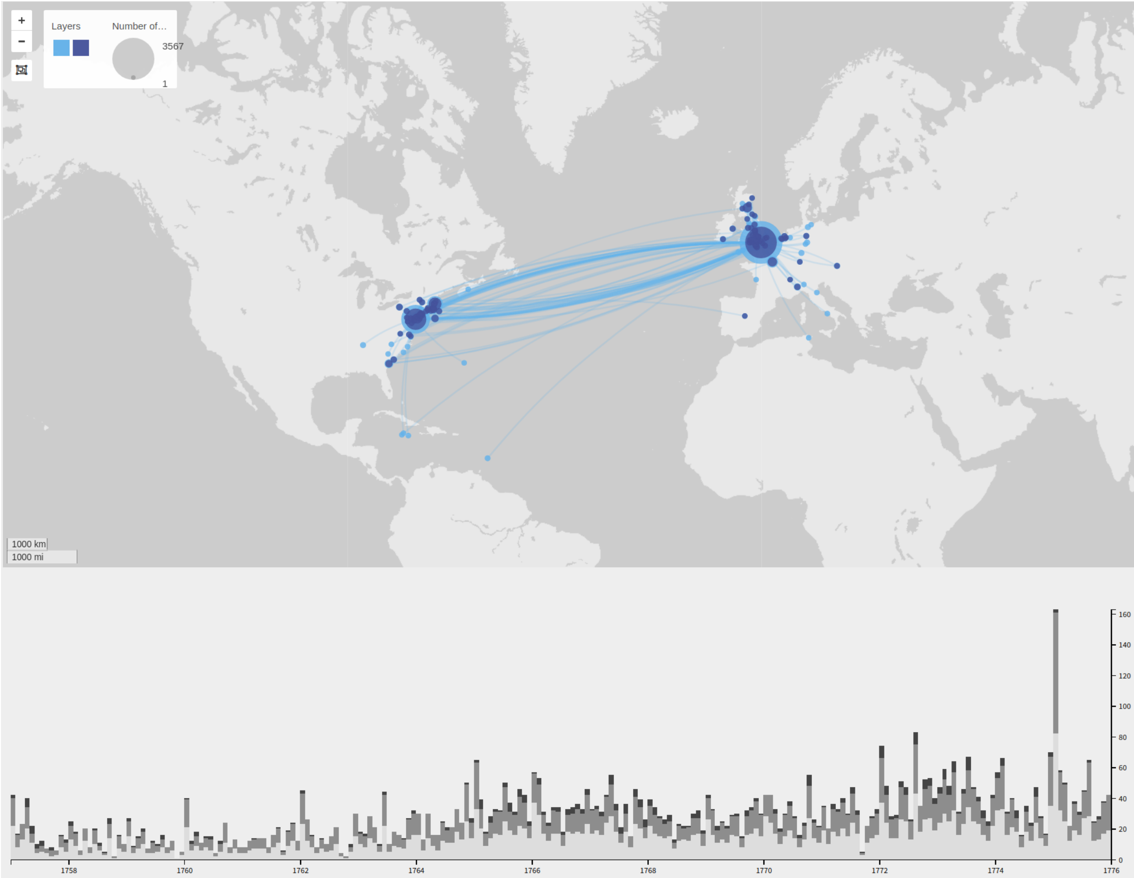

Figure 2.4 – Correspondence letters of Benjamin Franklin and his close relationships, visualized with a map and a histogram, accessible online on the republic of letter website [72].

for big projects involving many actors and highly ambitious goals [?, 152]. Guildi and Aritage went as far as criticizing the decrease of interest of historians working in archives [101], since many historical dataset are now directly available online through web technologies and archives portal.

Many projects which claim themselves as digital history also leverage new methods compared to the 1960s and 1970s, such as the use of network methods and concepts [4]. Examples are the *Viral Texts* [52] and *Living with Machines* [11] projects which respectively study nineteenth-century newspapers and the industrial revolution by translating their sources into analyzable networks. I discuss more in detail the related work of network analysis for historical research in §2.3.

2.3 Historical Social Network Analysis

Historians started to use network analysis to study relational structures and phenomena of past societies in the 1980s, using similar methods developed by sociologists under the label of SNA. SNA can be defined as an "*approach grounded in the intuitive notion that the patterning of social ties in which actors are embedded has important consequences for those actors. Network analysts, then, seek to uncover various kinds of patterns. And they try to determine the conditions under which those patterns arise and to discover their consequences*" [86]. the use of networks emerged in response to traditional sociology methods using pre-defined taxonomies and social categories to understand and explain sociological behaviors and phenomena, which could introduce bias [212]. By modeling real observed social relationships and interactions with networks and by using mathematical and statistical methods to study those, sociologists have been able to explain sociological phenomena and describe sociological interactions through their direct observation modeled as networks. SNA is now a well-praised methodology in sociology and has been extended to historical research to study relational concepts such as kinship, friendships, and business, through the characterization of groups of past societies. Social historians leverage their documents to extract relationships between entities—often persons—that they model into networks. Leveraging network measures and visualization, they can make conclusions through structural observations of such networks.

2.3.1 Sociometry to Social Network Analysis

One of Sociology's main goals is to study social relationships between individuals and find recurrent patterns and structures allowing us to generalize on how social relations operate, and what are the social specificity of specific groups and individuals [212]. Traditional methods try to answer those questions using classical social classifications such as age, social status, profession, and gender, typically collected from surveys and interviews. Criticism pointed out that this type of division is often partially biased and comes from predefined categories which are not always grounded in reality [86] and that using random sampling of individuals with such methods remove them from their sociological context. The sociologist Allen Barton wrote in this regard "For the last thirty years, empirical social research has been dominated by the sample survey. But as usually practiced, using random sampling of individuals, the survey is a sociological meatgrinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it" [17]. Sociometry is considered one of the bases of SNA and had the goal of redefining social categories through the lens of real social interactions and ties between persons, which sociologists wanted to observe in real conditions. It is in the 1930s that Moreno started to develop this new method by trying to depict real social interactions as a way to understand how groups and organizations were socially structured [173]. He developed sociograms to visually show friendships between people with the help of circles representing persons and lines modeling friendships. Figure 2.5 shows one of Moreno's original sociograms to depict friendships in a class of first grades (left).

Sociometry tremendously helped disseminate the metaphor of networks to model and understand social structures and phenomena, but was not using any formalization of network concepts and computation yet. It was not until the 1950s that mathematicians such as Harary

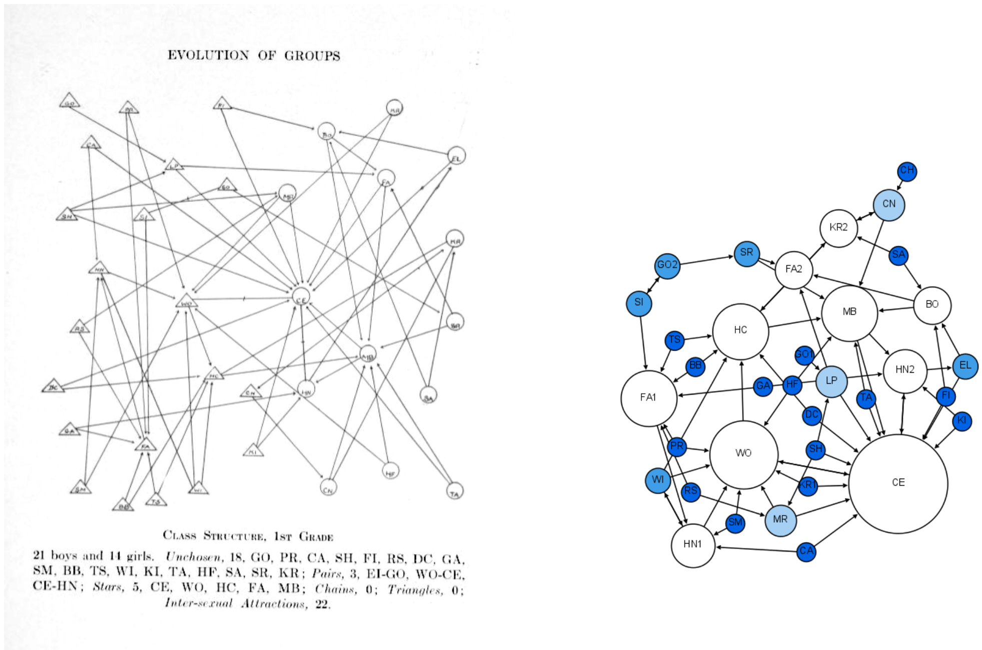

Figure 2.5 – Moreno's original sociogram of a class of first grades from [172] (left). The diagram shows 21 boys (triangles) and 14 girls (circles). The same sociogram using modern practices generated from Gephi from [96] (right). The color encodes the number of incoming connections.

and Roberts started to develop graph theory as a field[2] and apply it to sociology [105]. Then, many sociologists and anthropologists started to formalized SNA using graphs[3] and computations [41, 86], while mathematicians were studying in parallel the mathematical properties of graphs, such as Erdős on his work on random graph models [73]. Sociologists already had structural theories of social phenomena, and they rapidly saw the potential of graphs to model social relationships between actors of interest. Typically, a graph is noted

$$G = (V, E) \tag{2.1}$$

with V a set of vertices representing the actors of interest—typically persons—and $E \subseteq V^2$ a set of edges modeling social relationships. This simple model which does not take into account the diversity and extent of social relationships still allows the characterization of the sociological structure of groups and institutions—which is the primary focus of SNA [86, 212]. More complex network models have been proposed with time to better take into account concrete properties of social relationships. I discuss those more in depth in §2.3.4. Graph theory brought a panoply

[2] The first book on graph theory was published in 1936 by König [136].

[3] Graphs and networks refer to the same thing but are often used in different contexts. The term graph is preferred in a mathematical and abstract setting, while the term network is mostly used when modeling real-world phenomena. We talk about nodes and links for networks and vertices and edges for graphs.

of concepts and methods to characterize the structure of networks, that sociologists such as Coleman started to codify to use in a sociology setting [48]. The use of network measures let sociologists explain social phenomena through the formal description of real observations of relationships modeled as networks. I describe commonly used methods and measures in the following subsection.

2.3.2 Methods and Measures

Many measures and algorithms have been proposed in the network science and SNA literatures to characterize the structure of simple networks as defined in Equation 2.1 and relate it to social behaviors and phenomena [212, 229]. Network measures are either global or local, which allows one to either make high-level conclusions on the general structure of social relationships constituting the network or individual behaviors. Widely used global measures include for example the density and the diameter, which give insight into the sparsity of the network and how distant on average are two random pairs of nodes. Conversely, local measures give information on the structural position of a node compared to the rest of the network. Centrality—probably the most used local measure—allows to formally compute a measure of how important or central are individuals in the network [177]. As the definition of what is an important node can vary, several types of centrality have been proposed such as the degree, betweenness, and closeness centrality, which respectively measure the number of connections, how nodes connect to different groups, and how close are the nodes compared to the rest of the network.

More generally, sociologists aim at identifying recurring patterns of sociability between actors and linking them to other behaviors, measures, or qualitative knowledge. These patterns can for example be small unconnected components, cliques, or bow-tie structures [247]. Groups of nodes similarly located (central or distant) and having similar shapes are sometimes referred to as "structurally equivalent" [151]. Instead of observing complex shapes, network scientists have also been interested in studying relationships at the lowest possible scale, i.e., observing relations between sets of 2 and 3 nodes at once, also called dyads and triads [247]. This reflects Simmel's formal sociology, where he already referred to dyads and triads as the primal form of sociability [222]. More recently, graphlet analysis extended this concept to every pattern of N-entities [194].

Graphlets are defined as small connected *induced*, *non-isomorphic* subgraphs composing any network [169]. In an *induced* subgraph, two vertices linked in the original graph remain linked in the subgraph. For instance, if the original graph is a triangle we can only induce the simple edge or triangle subgraph (graphlet). The path of length 2 has all vertices of the original graph but misses an edge and is, therefore, not a possible graphlet.

Figure 2.6 shows all graphlets of size 2 to 5, for undirected networks. Graphlets counting shows that graphlets are not found in a uniform distribution in social networks [46], thus revealing that social networks do not have the same structure that random networks. Precisely, entities in real-world networks tend to agglomerate into groups (also called *clusters*) where entities in the same groups interact more between them than with entities from other groups [92]. From a sociological perspective, it means that people tend to interact and socialize in groups and interact more rarely with other people from outside groups. These groups are

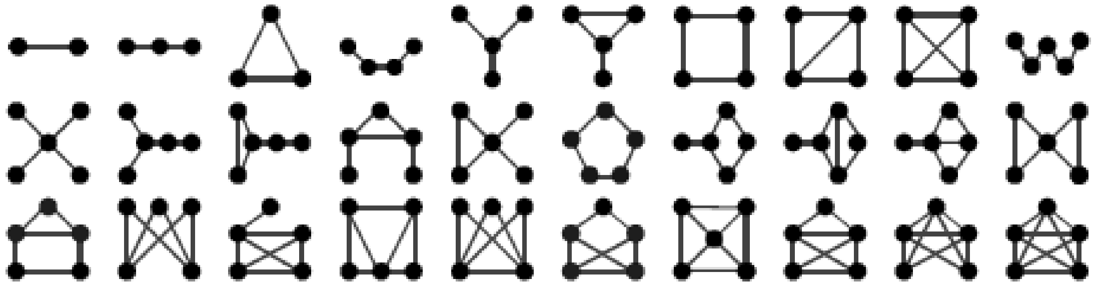

Figure 2.6 – All possible graphlets of size 2 to 5 for undirected graphs

often referred to as *communities*, and many algorithms have been proposed to find these automatically [84].

However, network concepts, measures, and algorithms have not been used only to study groups, organizations, and societies, but also to focus on separate specific individuals. Indeed, two distinct methodologies emerged through the history of SNA: the structuralists and the school of Manchester [75, 86, 161].

Structuralists are interested in observing the relational structures and patterns forming a network, to make parallels between them and the social behaviors of actors in real life [148]. Accordingly, sociologists in this school usually study organizations and specific groups—such as institutions, companies, and families—and want to explain their functioning through the description of the internal shapes and structures of the networks. Thus, they try to construct networks that exhaustively model all the interactions between the actors constituting the groups, as missing links would misrepresent the reality of interactions.

In contrast, the school of Manchester constituted by anthropologists focuses on studying specific individuals and all their interactions in the different facets of their lives and through time. They typically want to explain certain behaviors and social characteristics of individuals by their relationships and interactions in all their complexity and highlight the influence of different social aspects between them in one's life. One famous example is Mayer's study on austral African rural migrants going to cities [162] where he showed that the integration of urban mores and customs was directly correlated to the persons' relationships networks in the city. Xhosa[4] people still interacting with rural people of their village in the city were less changing their customs. This school of thought typically relies on the concept of ego and multiplex networks [75]. Ego networks are networks modeling all the direct relations of one central node—in this case, a person—including the relations existing between the persons of this small network. They typically try to model the different types of relationships of a person, like their family, work, and friendship ties and study them through time. By studying the ego network structure of someone, sociologists of this school try to leverage explanations on other social aspects of the persons like their social status, job, and gender. It is also common to

[4] Xhosa people are an ethnic group living in South Africa who talk the Xhosa language.

compare several ego networks to make correlations between the social relationships of individuals and other interesting social categories [46].

These two methodologies of SNA are often not exclusives and current studies are typically inspired by those two traditions. This is especially true in history where even if historians may want to describe exhaustively a group or institution of the past, they are almost always interested in specific individuals they study in depth.

2.3.3 Historical Social Network Analysis

History started to use concepts and methods from SNA in the 1980s [248] in response to quantitative history, and to develop historical approaches—like *Microstoria* [91]—that focus on the study of individuals and small groups through the lens of their interactions and relationships directly extracted from historical documents. Beforehand, historians were already describing and studying relational structures such as families and organizations with qualitative methods and with classical taxonomies, without necessarily studying the relational aspect of these structures. Network research allowed them to model those relational entities more thoroughly using networks, hence allowing them to make new observations that it was not possible to make without taking into account the relational aspect of these entities [53]. Since then, HSNA—a term coined by C. Wetherell in 1998 [248]—has been applied by historians to study multiple types of relationships, like kinship [103], political mobilization [156], and administrative/economic patronage [175]. If these approaches fall under some of the same criticism as quantitative history [151] like leading to trivial conclusions, it still led to classical work and interesting discoveries, such as the study of the rise of the Medici family in Florence in the 15th century by Padgett & Ansell [184], or Alexander & Danowski study on Cicero's personal communications [5]. In the latter work, they modeled the communication of Cicero into a network using 280 letters written by him between 68 B.C. and 42 B.C. It allowed them to study the relationships between knights and senators—which is a subject of interest in Roman history—and concluded that knight-knight interactions were very rare compared to senator-senator and senator-knight interactions. Cicero communication network is illustrated in Figure 2.7.

Several historians are using and continuously reflecting on HSNA methods [53, 151] which can be very effective to study relational historical phenomena [133]. However, contrary to sociologists and anthropologists who base their networks on direct observations of the real world, historians first have to go through a deep inspection, encoding, and modeling of their sources.

2.3.4 Network Modeling

Constructing a network from historical documents, which can vary tremendously in their formats and structures, is not a trivial task [6]. The most straightforward and well-known approach consists in using simple graphs such as in Equation 2.1, where the nodes refer to the persons mentioned in the documents and links refer to one type of social relationship or a notion of *proximity* constructed from appearance in common documents [33, 151].

This enables to have simple networks to visualize and analyze, but it does not always reflect the sociological complexity of information contained in the documents. HSNA network models have evolved over time to better take into account concrete properties of social networks, such

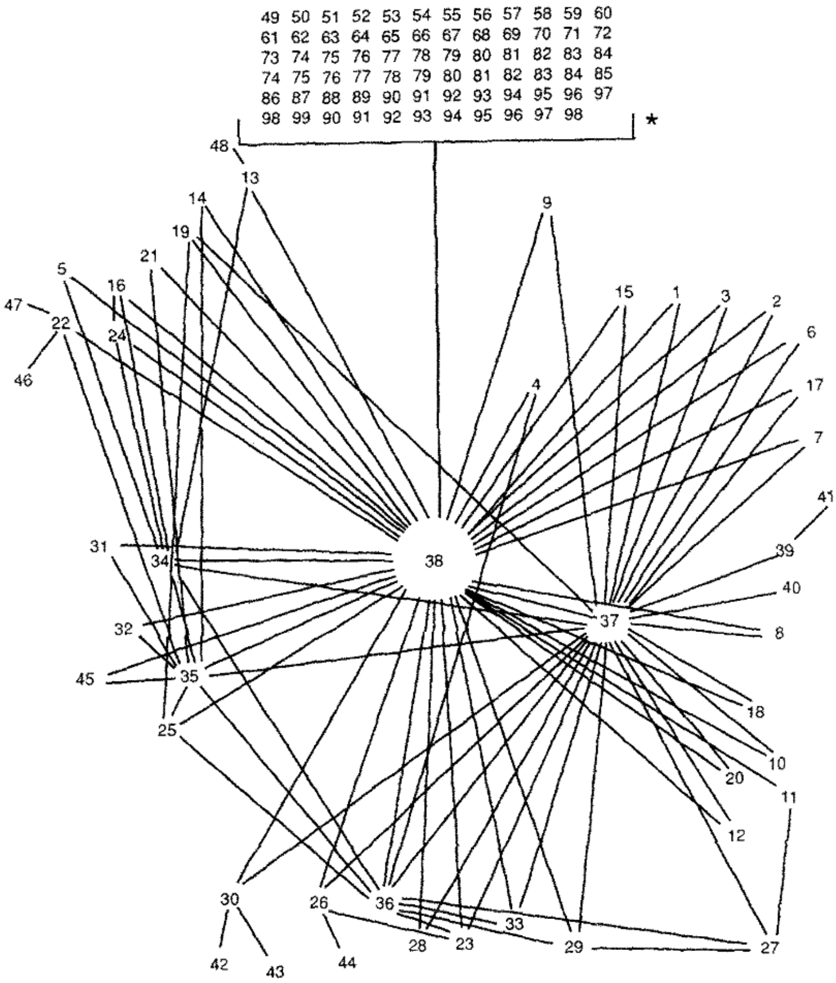

Figure 2.7 – Cicero's personal communication network represented with a node-link diagram. Image from [5]

as the importance of actors or relations with weighted networks, multiple relationships with multiplex networks, and dynamics of relationships with dynamic networks.

Weighted networks model the importance of relations, with a weight w attributed to each edge $e = (u, v, w)$, with $u, v \in V$, $e \in E$, and $w > 0$. Multiplex networks allow the modeling of multiple kinds of relationships between actors, such as spouses and witnesses relations for a historical network constructed from marriage acts. In that case, each edge $e = (u, v, d)$ of the graph

$$G = (V, E, D) \tag{2.2}$$

have a type $d \in D$ which characterize the relation. In the example of marriages, $D = \{spouse, witness\}$. Most relations extracted from historical documents also often contain time information, which can be modeled into dynamic networks. Many dynamic network models have been proposed [201], depending if the time is encoded in the nodes, the links, or both, and if it is modeled as discrete or interval values. As it is often hard to infer the end of social relationships from the trace of historical documents, I only consider in this thesis models which give a timestamp to either nodes or edges, such that a dynamic graph

$$G = (V, E, T) \tag{2.3}$$

have vertices consisting of tuples (u, t) and/or edges of the form (u, v, t), with $t \in T$, where T corresponds to a set of discrete timestamps, such as calendar dates.

Bipartite networks have been proposed to model relations between two types of entities, such as organizations and employees where the relations link employees to organizations but not employees to employees or organizations to organizations [30]. Formally, each node of the graph

$$G = (V, E, B) \tag{2.4}$$

have a type $b \in B$, with $card(B) = 2$. For each edge $e = (u, v) \in E$, the types b_u and b_v of u and v are not equal: $b_u \neq b_v$. Many social situations or documents can be modeled in these terms (affiliation lists or co-authoring). Multivariate networks, i.e., graphs, where vertices and edges can be assigned multiple "properties" or "attributes", are less used in SNA. These attributes are often considered secondary, the emphasis of SNA being on the topology, its features, measures, and evolution.

Historians, demographers, sociologists, and anthropologists have also been designing specific data models for their social networks, based on genealogy or more generally kinship [104]. For genealogy, the standard GEDCOM [106] format models a genealogical graph as a bipartite graph with two types of vertices: individuals and families. This format also integrates an "event" object but it is diversely adapted in genealogical tools. The Puck software [103] has extended its original genealogical graph with the concept of "relational nodes" to adapt the data model to more family structures and to integrate other social relationships for anthropology and historical studies.

When creating a network, sociologists and anthropologists can use direct observations of the real world, which is not the case for historians who only have access to biased and partial sources. Indeed, the documents historians inspect are often produced by the political and economical elite of the time, and include the subjective view of the authors, especially for literary sources (letters, journals, books, etc.). Historians, therefore, need to take a critical view of the sources by acknowledging the position of the authors of the documents compared to the rest of the society and include it in the analysis [152]. Furthermore, the partiality of the sources often does not allow to have access to all possible relationship types of individuals. For example, if many formal relations can be extracted from official documents such as marriage acts and censuses, informal relations such as friendships can exist without leaving any written trace [151]. Even for official relationships such as parents and witnesses, there are high chances for missing documents, which do not allow to make too general and finite claims, such as "X is always the case" or "XX is never the case" [5]. Social historians, therefore, have to take into account the partiality and ambiguity of their sources in their analysis, in order to avoid including the bias inherent to their data in their high-level historical conclusions.

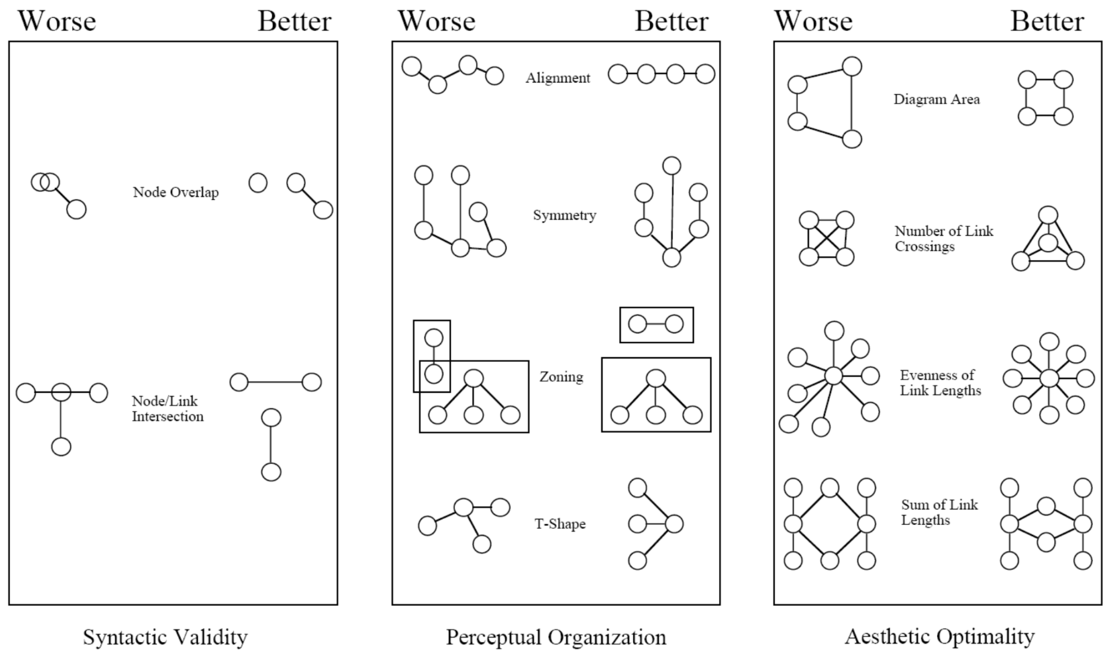

Figure 2.8 – Different criteria are proposed to enhance node-link diagram readability. Image from [139]

2.4 Social Network Visualization

Practitioners of SNA and HSNA have always visually depicted their network data for validation, exploration, and communication, mostly using node-link diagrams. With the use of more complex network models and the increase in average network size and density, new visualization techniques have been proposed to represent the diversity of studied networks. Moreover, more and more social scientists are following exploratory approaches using visualization tools with a focus on interaction, allowing them to unfold the structure of their data and generte new interesting hypotheses.

2.4.1 Graph Drawing

Sociologists rapidly saw the potential of graphically showing relationships between individuals, to better comprehend the underlying social structure and communicate their findings [85]. Moreno elaborated sociograms to visually show friendships among schoolchildren with circles and lines to respectively show children and friendships ties [172]. This type of representation—commonly called node-link diagram—is the most widely used in social sciences, as it is rapidly understandable and effective for small to medium-sized networks which are predominant in the field. Finding an optimal placement for the nodes is however not that simple as several metrics can be optimized depending on the desired drawing, such as the number of edge crossings, the variance of edge length, orthogonality of edges, etc [54, 139]. Figure 2.8 shows some of these metrics, synthesized by Kosara et al. [139]. In Figure 2.5 we can see the difference in readability between the original manual layout (left) and an automatic one (right). Automatic

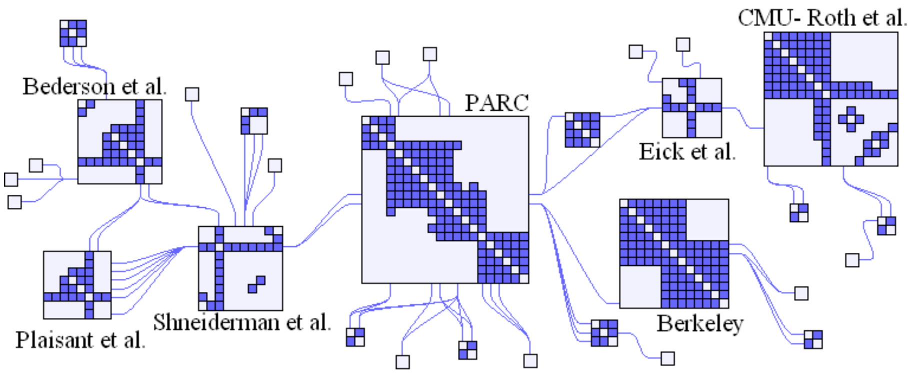

Figure 2.9 – NodeTrix system showing a scientific collaboration social network with clusters. Each cluster is represented as a matrix, Image from [113].

layouts which aim at optimizing readability metrics give clearer diagrams. The number of edge crossings is often considered the most important measure, but finding a drawing with the optimal number of crossings is an NP-Hard problem, meaning that heuristics are needed for most real-world use cases. A large number of algorithms have been designed such as force-directed ones [20], modeling the nodes as particles that repulse each other and are attracted together when connected with a link using a string analogy. Other visual techniques have been proposed to represent networks such as matrices, circular layouts, and arcs, but are less used in social sciences [164]. Still, Matrices have been shown to be more effective than node-link diagrams for several tasks such as finding cluster-related patterns, especially for medium to large networks [2, 90].

As social scientists are using more complex network models such as bipartite or temporal networks, more sophisticated representations are needed. The visualization community developed new representations to visualize other network types such as dynamic hypergraphs with PAOHVis [240], clustered graphs with NodeTrix [113] (illustrated in Figure 2.9), geolocated social networks with the Vistorian [214], and multivariate networks with Juniper [179]. However, these new network representations take time to be adopted by social scientists who rarely use them.

2.4.2 Social Network Visual Analytics

Social scientists use visualization and analytical tools to gain insight on the structure of their finalized network data. The most widely used tools are Gephi [18], Pajek [176], Ucinet [126], and NodeXl [223], which provide node-link diagrams, implementations of network measures, algorithms, and clustering capabilities. Other SNA visualization tools have been proposed in the past such as Visone [22]. However, those tools often have usability issues as they do not include interaction and direct manipulation mechanisms, making the analysis more tedious for social historians. In contrast, the Vistorian [214] let social historians visualize their network with multiple coordinated views (node-link, matrice, arc-diagram, and map), filters and direct manipulation, but do not integrate analytical options. Figure 2.10 shows the Vistorian interface used to explore a historical social network. I propose a qualitative classification of all

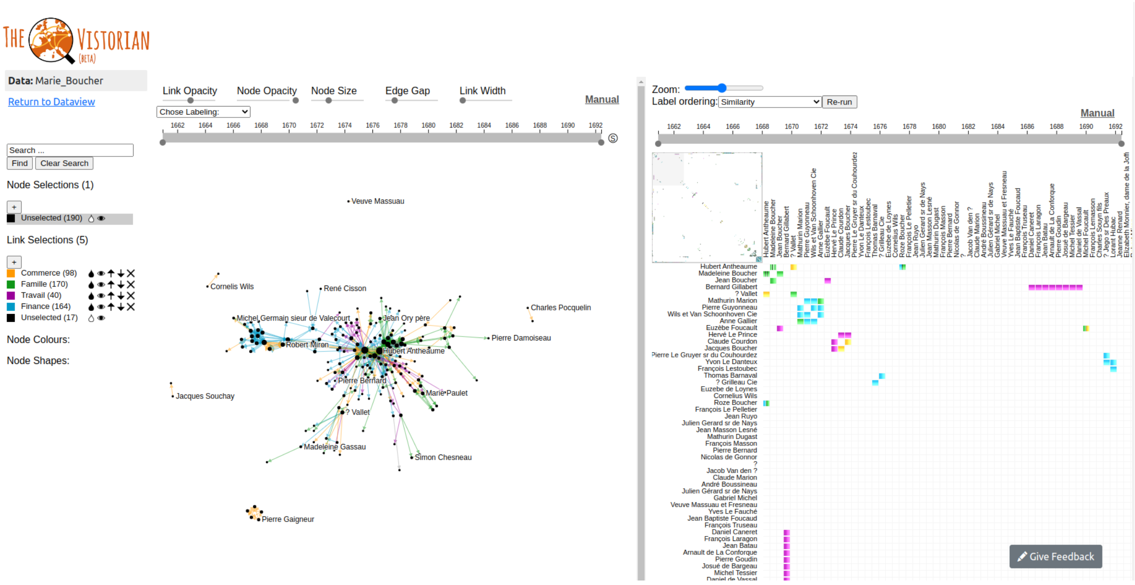

Figure 2.10 – Vistorian interface [214] used to explore a historical social network of business trades in the 17th century, with a coordinated node-link diagram and a matrice view.

those tools in 2.1 which rank them according their visualization, SNA measures and modes, clustering, filtering, and interaction capabilities. It illustrates that the most used tools are analytical-oriented (Pajek, Ucinet, Gephi, NodeXl) without proposing many visualization and direct manipulation options, while the Vistorian is an interactive visualization tool without any analytical method integration. None of those systems therefore fully correspond to the VA label [131].

	Visualizations	SNA Measures and Models	Clustering	Filtering	Interaction and Direct Manipulation
Pajek [176]	■□□	■■■	■■□	■□□	□□□
Ucinet [126]	■□□	■■□	■□□	□□□	□□□
Gephi [18]	■■□	■□□	■□□	■□□	■□□
NodeXl [223]	■□□	■□□	■■□	■□□	□□□
Vistorian [13]	■■■	□□□	□□□	■■■	■■■

Table 2.1 – Comparison table of most widely used visualization and analytical tool for SNA. Visualizations: number of different visualization techniques, and layouts. SNA Measures and Models: number of proposed SNA measures and algorithms. Clustering: Number of proposed clustering algorithms. Filtering: Possibilities of filtering according to various criteria. Interaction/Direct Manipulation: Number of possible interaction mechanisms directly applicable to the visualizations.

If analytical methods such as the computation of network measures, triad computation, and clustering provide a good framework to describe the structure of a network and link it to sociological explanations [212, 247], many social scientists, such as historians, are not trained in computer science and mathematical methods, and, therefore, have trouble to 1) use those methods without guidance and without high usability [6, 200], and 2) interpret their results. This is particularly the case for black-box algorithms such as for clustering tasks: they typically end up trying several algorithms until they stumble upon a satisfactory enough solution [190].

Moreover, preparing and importing the data into visual and analytical software is complicated, as the annotation and network modeling processes have not been globally formalized and every historian use different methods, formats and models. Many users do not succeed in importing their data into those systems without concrete help and guidance [6, 214] due to mismatches with data models, formats, or data inconsistencies (null values, white spaces, etc.). If they succeed in visualizing their data, it often shows them these inconsistencies or errors such as duplications of entities or wrong attribute values. In other cases, they realize the network does not allow them to answer their sociological questions [151]. It leads to continuous back and forth between their analysis process inside the analysis tool they are using, and their annotation/modeling process, to correct errors or modify annotations. Interestingly, the network model choice plays a crucial role, as a simple network model representing only the persons (as is often the case) makes it harder to trace back to the original documents containing the annotations from the network entities. Yet the majority of SNA systems enforce simple network models, making this retroactive process harder.

Some interfaces not primarily designed for social scientists incorporate data models encapsulating document representations, such as Jigsaw [225] which is a VA system using textual documents as a data model, originally developed for intelligence analysis. It allows an analysis of the documents and their mentions of entities (persons, locations, institutions, etc.) through multiple coordinated views. Using such a model allows us to rapidly see errors and inconsistencies in the document annotations that the user can directly correct, while still following complex analyzes.

Finally, more work is still to be done on social network VA tools, to provide more guidance and power to social scientists while doing their analysis, and to help them to do easier back and forth between the annotation, network modeling, correcting, and analysis steps, as errors and inconsistencies can cause high variations in the network structure and hence the analysis results [64].

Bibliography

[1] Interchange: The Promise of Digital History. *Journal of American History*, 95(2):452–491, September 2008. doi:10.2307/25095630. 19

[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. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–11, 2022. doi:10.1109/TVCG.2022.3209427. 29

[3] Eytan Adar and Christopher Re. Managing Uncertainty in Social Networks. *IEEE Data Eng. Bull.*, 30:15–22, 2007. 111

[4] 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. 20

[5] 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. xiii, 25, 26, 27

[6] 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. doi:10.1109/TVCG.2022.3209487. 2, 5, 6, 9, 12, 25, 31, 33, 34, 36, 40, 42, 49, 109

[7] Taher Alzahrani and K. J. Horadam. Community Detection in Bipartite Networks: Algorithms and Case studies. In Jinhu Lü, Xinghuo Yu, Guanrong Chen, and Wenwu Yu, editors, *Complex Systems and Networks: Dynamics, Controls and Applications*, Understanding Complex Systems, pages 25–50. Springer, Berlin, Heidelberg, 2016. doi:10.1007/978-3-662-47824-0_2. 87

[8] 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. 55

[9] F. J. Anscombe. Graphs in Statistical Analysis. *The American Statistician*, 27(1):17–21, February 1973. doi:10.1080/00031305.1973.10478966. xiii, 14

[10] Thomas J. Archdeacon. *Correlation and Regression Analysis: A Historian's Guide*. Univ of Wisconsin Press, 1994. 36

[11] 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. 20

[12] 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. xiii, 5, 16

[13] Benjamin Bach, Nathalie Henry Riche, Roland Fernandez, Emmanoulis Giannisakis, Bongshin Lee, and Jean-Daniel Fekete. NetworkCube: Bringing dynamic network visualizations to domain scientists. Posters of the Conference on Information Visualization (InfoVis), October 2015. 30

[14] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. Visualizing dynamic networks with matrix cubes. In Matt Jones, Philippe A. Palanque, Albrecht Schmidt, and Tovi Grossman, editors, *CHI Conference on Human Factors in Computing Systems, CHI'14, Toronto, ON, Canada - April 26 - May 01, 2014*, pages 877–886. ACM, 2014. doi:10.1145/2556288.2557010. 113

[15] Juhee Bae, Tove Helldin, Maria Riveiro, Sławomir Nowaczyk, Mohamed-Rafik Bouguelia, and Göran Falkman. Interactive Clustering: A Comprehensive Review. *ACM Comput. Surv.*, 53(1):1:1–1:39, February 2020. doi:10.1145/3340960. 88

[16] Trevor J Barnes. Big data, little history. *Dialogues in Human Geography*, 3(3):297–302, November 2013. doi:10.1177/2043820613514323. 35

[17] Allen H. Barton. Survey Research and Macro-Methodology. *American Behavioral Scientist*, 12(2):1–9, November 1968. doi:10.1177/000276426801200201. 21

[18] 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. 5, 29, 30, 53, 60, 84

[19] Sugato Basu, Ian Davidson, and Kiri Wagstaff. *Constrained Clustering: Advances in Algorithms, Theory, and Applications*. CRC Press, August 2008. 87

[20] 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. 29

[21] 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. 56

[22] 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. 29

[23] Laurent Beauguitte. Théorie des graphes et analyse de réseau en géographie : Histoire d'un lien faible (1950-1963). *PasserelleSHS*, 2022. 1

[24] Amine M. Bensaid, Lawrence O. Hall, James C. Bezdek, and Laurence P. Clarke. Partially supervised clustering for image segmentation. *Pattern Recognition*, 29(5):859–871, May 1996. doi:10.1016/0031-3203(95)00120-4. 87

[25] Jacques Bertin. *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Paris: Gauthier-Villars, 1967. xiii, 12, 13

[26] Anastasia Bezerianos, Fanny Chevalier, Pierre Dragicevic, Niklas Elmqvist, and Jean-Daniel 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. 62

[27] Marc Bloch and Jacques Le Goff. *Apologie pour l'histoire ou Métier d'historien*. Armand Colin, 1997. 2

[28] 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. 87

[29] Georges-Pierre Bonneau, Hans-Christian Hege, Chris R. Johnson, Manuel M. Oliveira, Kristin Potter, Penny Rheingans, and Thomas Schultz. Overview and State-of-the-Art of Uncertainty Visualization. In Charles D. Hansen, Min Chen, Christopher R. Johnson, Arie E. Kaufman, and Hans Hagen, editors, *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, Mathematics and Visualization, pages 3–27. Springer, London, 2014. doi:10.1007/978-1-4471-6497-5_1. 89

[30] 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. 27, 46

[31] 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. 56

[32] 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. 60, 71

[33] 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. xiii, 8, 25

[34] 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. 17

[35] Paul Bradshaw. Data journalism. In *The Online Journalism Handbook*. Routledge, second edition, 2017. 15

[36] Ulrik Brandes. *Network Analysis: Methodological Foundations*. Springer Science & Business Media, February 2005. 5

[37] 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. 40, 91

[38] Ulrik Brandes and Bobo Nick. Asymmetric Relations in Longitudinal Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2283–2290, December 2011. doi:10.1109/TVCG.2011.169. 113

[39] Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital_Humanities*. MIT Press, February 2016. 19

[40] Peter Burke. *History and Social Theory*. Polity, 2005. 11, 17

[41] Mitchell J. C. The Concept and Use of Social Networks. *Social Networks in Urban Situations*, 1969. 3, 22

[42] 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. 41, 56

[43] Charles-Olivier Carbonell. *L'Historiographie*. FeniXX, January 1981. 17

[44] 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. 4, 12

[45] Marco Cavallo and Çağatay Demiralp. Clustrophile 2: Guided Visual Clustering Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):267–276, January 2019. doi:10.1109/TVCG.2018.2864477. 88

[46] 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. 23, 25, 54

[47] 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. 55

[48] James Samuel Coleman. *Introduction to Mathematical Sociology*. Free Press of Glencoe / Collier-Macmillan, first edition edition, January 1964. 23

[49] 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. 7

[50] TEI Consortium. TEI P5: Guidelines for electronic text encoding and interchange, February 2021. doi:10.5281/zenodo.4609855. 39

[51] Kristin Cook, Nick Cramer, David Israel, Michael Wolverton, Joe Bruce, Russ Burtner, and Alex Endert. Mixed-initiative visual analytics using task-driven recommendations. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 9–16, October 2015. doi:10.1109/VAST.2015.7347625. 88

[52] Ryan Cordell and David Smith. Viral texts: Mapping networks of reprinting in 19th-Century newspapers and magazines, 2017. 20

[53] Pascal Cristofoli. Aux sources des grands réseaux d'interactions. *Reseaux*, 152(6):21–58, 2008. 1, 7, 25, 34, 36, 38, 40, 44, 45, 53

[54] Pascal Cristofoli. Principes et usages des dessins de réseaux en SHS. *La visualisation des données en histoire*, page 35, 2015. 2, 28, 60

[55] 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. xiv, 37, 43, 57, 61

[56] 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. 43

[57] Alfred W. Crosby. *The Measure of Reality*. Cambridge University Press, Cambridge, reprint édition edition, March 1998. 4

[58] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. 55, 60

[59] 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. 55, 82

[60] 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. 69, 71

[61] 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. 45

[62] 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. 4

[63] Inderjit S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 269–274, New York, NY, USA, August 2001. Association for Computing Machinery. doi:10.1145/502512.502550. 92

[64] 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. 6, 31, 34, 38, 39

[65] 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. 37, 58

[66] 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. 1, 111

[67] 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. xiii, 4, 6, 37, 42

[68] 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. 39

[69] Nicole Dufournaud, Bernard Michon, Benjamin Bach, and Pascal Cristofoli. L'analyse des réseaux, une aide à penser : Réflexions sur les stratégies économique et sociale de

Marie Boucher, marchande à Nantes au XVII e siècle. In *Réseaux de Femmes, Femmes En Réseaux (XVIe-XXIe Siècles)*, pages 109–137. 2017. 100

[70] 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. 56

[71] Natural Earth. Free vector and raster map data. *Available: naturalearthdata. com*, 2015. 61

[72] 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. xiii, 19, 20, 33, 111

[73] 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. 22

[74] 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. 43

[75] Michael Eve. Deux traditions d'analyse des reseaux sociaux. *Réseaux*, 115(5):183–212, 2002. 9, 24

[76] 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. 55

[77] Lucien Febvre. Vers une autre histoire. *Revue de Métaphysique et de Morale*, 54(3/4):225–247, 1949. 17

[78] 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. 67, 107

[79] Roderick Floud. *An Introduction to Quantitative Methods for Historians*. Routledge, London, September 2013. doi:10.4324/9781315019512. 36

[80] Professor Robert William Fogel. *Railroads and American Economic Growth: Essays in Econometric History*. The Johns Hopkins University Press, Baltimore, March 1970. 18

[81] Robert William Fogel. The Limits of Quantitative Methods in History. *The American Historical Review*, 80(2):329–350, 1975. doi:10.2307/1850498. 35

[82] Robert William Fogel and Stanley L Engerman. *Time on the Cross: Evidence and Methods, a Supplement*, volume 2. Little, Brown, 1974. 18

[83] Stefania Forlini, Uta Hinrichs, and John Brosz. Mining the Material Archive: Balancing Sensate Experience and Sense-Making in Digitized Print Collections. *Open Library of Humanities*, 4(2):35, November 2018. doi:10.16995/olh.282. 41

[84] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, February 2010. doi:10.1016/j.physrep.2009.11.002. 24, 87

[85] Linton Freeman. Visualizing Social Networks. *J. Soc. Struct.*, 2000. 2, 4, 5, 28

[86] Linton Freeman. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004. 1, 3, 11, 21, 22, 24, 40

[87] Manuel Freire, Catherine Plaisant, Ben Shneiderman, and Jen Golbeck. ManyNets: An interface for multiple network analysis and visualization. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, page 213, Atlanta, Georgia, USA, 2010. ACM Press. doi:10.1145/1753326.1753358. 56

[88] 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. 12

[89] 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. 14

[90] Mohammad Ghoniem, Jean-Daniel 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, Austin, TX, USA, 2004. IEEE. doi:10.1109/INFVIS.2004.1. 29

[91] Carlo Ginzburg and Carlo Poni. La micro-histoire. *Le Débat*, 17(10):133, 1981. doi:10.3917/deba.017.0133. 3, 25, 36

[92] 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. 23, 87

[93] 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. 3

[94] Michael Gleicher. Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, 2018. doi:10.1109/TVCG.2017.2744199. 55

[95] Claudia Goldin. Cliometrics and the Nobel. *Journal of Economic Perspectives*, 9(2):191–208, June 1995. doi:10.1257/jep.9.2.191. 3

[96] Martin Grandjean. Social network analysis and visualization: Moreno's Sociograms revisited, 2015. xiii, 22

[97] 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. 52

[98] Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hanspeter Pfister, and Marc Streit. Domino: Extracting, Comparing, and Manipulating Subsets Across Multiple Tabular Datasets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2023–2032, December 2014. doi:10.1109/TVCG.2014.2346260. 88

[99] 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. 2

[100] Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral. Module identification in bipartite and directed networks. *Phys. Rev. E*, 76(3):036102, September 2007. doi:10.1103/PhysRevE.76.036102. 92

[101] Jo Guldi and David Armitage. *The History Manifesto*. Cambridge University Press, October 2014. 1, 20

[102] Aric Hagberg and Drew Conway. NetworkX: Network analysis with python. *URL: https://networkx. github. io*, 2020. 55, 84

[103] 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. 25, 27, 43, 46

[104] 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. 1, 27

[105] Frank Harary and Robert Z Norman. *Graph Theory as a Mathematical Model in Social Science*. Number 2. University of Michigan, Institute for Social Research Ann Arbor, 1953. 22

[106] J. Tuomas Harviainen and Bo-Christer Björk. Genealogy, GEDCOM, and popularity implications. *INF*, 37(3), October 2018. doi:10.23978/inf.76066. 27

[107] 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. 56

[108] Loren Haskins and Kirk Jeffrey. *Understanding Quantitative History*. Wipf and Stock Publishers, March 2011. 16

[109] 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. 42, 111

[110] Jeffrey Heer. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116(6):1844–1850, 2019. 114

[111] Jeffrey Heer and danah boyd. Vizster: Visualizing Online Social Networks. In John T. Stasko and Matthew O. Ward, editors, *IEEE Symposium on Information Visualization (InfoVis 2005), 23-25 October 2005, Minneapolis, MN, USA*, pages 32–39. IEEE Computer Society, 2005. doi:10.1109/INFVIS.2005.1532126. 62

[112] 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. 3

[113] 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. xiii, 29

[114] 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. 14

[115] 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. 44

[116] Paul W. Holland and Samuel Leinhardt. Local Structure in Social Networks. *Sociological Methodology*, 7:1, 1976. doi:10.2307/270703. 55

[117] Paul Holleis and Franz J. Brandenburg. QUOGGLES: Query On Graphs – A Graphical Largely Extensible System. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, and János Pach, editors, *Graph Drawing*, volume 3383, pages 465–470. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. doi:10.1007/978-3-540-31843-9_48. 55

[118] Danny Holten. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, September 2006. doi:10.1109/TVCG.2006.147. 88

[119] Andreas Holzinger. From Machine Learning to Explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pages 55–66, August 2018. doi:10.1109/DISA.2018.8490530. 114

[120] Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 159–166, New York, NY, USA, May 1999. Association for Computing Machinery. doi:10.1145/302979.303030. 85, 88

[121] Infovis SC policies FAQ. 102

[122] 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. 55

[123] Petra Isenberg, Florian Heimerl, Steffen Koch, Tobias Isenberg, Panpan Xu, Charles D. Stolper, Michael Sedlmair, Jian Chen, Torsten Möller, and John Stasko. Vispubdata.org: A Metadata Collection About IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2199–2206, September 2017. doi:10.1109/TVCG.2016.2615308. 103

[124] Mikkel R. Jakobsen and Kasper Hornbæk. Evaluating a fisheye view of source code. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 377–386, New York, NY, USA, April 2006. Association for Computing Machinery. doi:10.1145/1124772.1124830. 107

[125] Radu Jianu, Adrian Rusu, Yifan Hu, and Douglas Taggart. How to Display Group Information on Node-Link Diagrams: An Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 20(11):1530–1541, November 2014. doi:10.1109/TVCG.2014.2315995. 88

[126] J. David Johnson. UCINET: A software tool for network analysis. *Communication Education*, 36(1):92–94, January 1987. doi:10.1080/03634528709378647. 5, 29, 30

[127] J. David Johnson. UCINET: A software tool for network analysis. *Communication Education*, 36(1):92–94, January 1987. doi:10.1080/03634528709378647. 8

[128] 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. 5, 19

[129] 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. 1, 35, 36, 41

[130] D.A. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):59–78, January 2000. doi:10.1109/2945.841121. 107

[131] 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. xiii, 5, 7, 15, 30

[132] Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis, and Florian Mansmann. Mastering the information age: Solving problems with visual analytics. 2010. 103

[133] Florian Kerschbaumer, Linda Keyserlingk, Martin Stark, and Marten Düring. *The Power of Networks. Prospects of Historical Network Research*. Routledge, 2020. 1, 2, 11, 25, 112

[134] Steffen Klamt, Utz-Uwe Haus, and Fabian Theis. Hypergraphs and cellular networks. *PLoS computational biology*, 5(5):e1000385, 2009. doi:10.1371/journal.pcbi.1000385. 81

[135] Jon Kleinberg. An Impossibility Theorem for Clustering. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002. 85

[136] Dénes König. Theory of Finite and Infinite Graphs. In Dénes König, editor, *Theory of Finite and Infinite Graphs*, pages 45–421. Birkhäuser, Boston, MA, 1990. doi:10.1007/978-1-4684-8971-2_2. 22

[137] Dmitry A. Konovalov, Bruce Litow, and Nigel Bajema. Partition-distance via the assignment problem. *Bioinformatics*, 21(10):2463–2468, May 2005. doi:10.1093/bioinformatics/bti373. 92

[138] 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. 81

[139] 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. xiii, 28

[140] R. Kosara, F. Bendix, and H. Hauser. Parallel Sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):558–568, July 2006. doi:10.1109/TVCG.2006.76. 88

[141] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955. doi:10.1002/nav.3800020109. 93

[142] Alexander Kumpf, Bianca Tost, Marlene Baumgart, Michael Riemer, Rüdiger Westermann, and Marc Rautenhaus. Visualizing Confidence in Cluster-Based Ensemble Weather Forecast Analyses. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):109–119, January 2018. doi:10.1109/TVCG.2017.2745178. 89

[143] 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. 17

[144] David S. Landes and Charles Tilly. *History as Social Science. The Behavioral and Social Sciences Survey*. Prentice Hall, Inc, 1971. 18

[145] Charles-Victor Langlois and Charles Seignobos. *Introduction aux études historiques*. ENS Éditions, February 2014. 1, 17

[146] 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. 1

[147] 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. 46, 62

[148] Emmanuel Lazega. *Réseaux sociaux et structures relationnelles*. Presses universitaires de France, Paris, 1998. 5, 24

[149] Levi Lelis and Jörg Sander. Semi-supervised Density-Based Clustering. In *2009 Ninth IEEE International Conference on Data Mining*, pages 842–847, December 2009. doi:10.1109/ICDM.2009.143. 87

[150] Claire Lemercier. Analyse de réseaux et histoire. *Revue d'histoire moderne contemporaine*, 522(2):88–112, 2005. 38

[151] 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. 1, 6, 7, 12, 23, 25, 27, 31, 33, 41, 52, 109

[152] Claire Lemercier and Claire Zalc. *Quantitative Methods in the Humanities: An Introduction*. University of Virginia Press, March 2019. 2, 3, 6, 7, 18, 19, 20, 27, 34, 35, 36, 41, 53, 112

[153] 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. 6, 18, 33, 34, 35, 36, 41

[154] 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. 1, 35

[155] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1983–1992, December 2014. doi:10.1109/TVCG.2014.2346248. 96

[156] 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. 25

[157] 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. 45

[158] Sehi L'Yi, Bongkyung Ko, DongHwa Shin, Young-Joon Cho, Jaeyong Lee, Bohyoung Kim, and Jinwook Seo. XCluSim: A visual analytics tool for interactively comparing multiple clustering results of bioinformatics data. *BMC Bioinformatics*, 16(11):S5, August 2015. doi:10.1186/1471-2105-16-S11-S5. 88

[159] Alan M. MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, Robert Murray, Mark Gahegan, and Elisabeth Hetzler. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3):139–160, January 2005. doi:10.1559/1523040054738936. 89

[160] 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. 82, 88

[161] Gribaudi Maurizio. *Espaces, Temporalités, Stratifications :. Exercices Méthodologiques Sur Les Réseaux Sociaux*. Editions de l'Ecole des Hautes Etudes en Sciences Sociales, Paris, January 2000. 24

[162] Philip Mayer. Migrancy and the Study of Africans in Towns. *American Anthropologist*, 64(3):576–592, 1962. 24

[163] 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. 43

[164] 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. 29

[165] 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. 1

[166] 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. 114

[167] Justin J Miller. Graph Database Applications and Concepts with Neo4j. page 7, 2013. 54, 55, 71, 80

[168] Barbara J. Mills. Social Network Analysis in Archaeology. *Annu. Rev. Anthropol.*, 46(1):379–397, October 2017. doi:10.1146/annurev-anthro-102116-041423. 1

[169] 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. 23, 54

[170] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Independently published, 2020. 86

[171] 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. 80

[172] 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. xiii, 22, 28

[173] J. L. Moreno. Foundations of Sociometry: An Introduction. *Sociometry*, 4(1):15, February 1941. doi:10.2307/2785363. 21

[174] Zacarias Moutoukias. Buenos Aires, port between two oceans: Mobilities, networks, stratifications (2nd half of the 18th century). *e-Spania. Revue interdisciplinaire d'études hispaniques médiévales et modernes*, 25, 2016. 37, 58

[175] 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. 25

[176] 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. 5, 8, 29, 30, 60, 84

[177] Mark Newman. *Networks*. Oxford university press, 2018. 23

[178] 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. 37

[179] 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. 29

[180] Gérard Noiriel. Naissance du métier d'historien. *Genèses. Sciences sociales et histoire*, 1(1):58–85, 1990. doi:10.3406/genes.1990.1014. 17

[181] Jeremi K. Ochab, Jan Škvrňák, and Michael Škvrňák. Detecting Ottokar II's 1248–1249 uprising and its instigators in co-witnessing networks. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 0(0):1–20, May 2022. doi:10.1080/01615440.2022.2065397. 45

[182] 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. 7, 53

[183] Maryjane Osa. *Solidarity And Contention: Networks Of Polish Opposition*. Univ Of Minnesota Press, Minneapolis, July 2003. 1, 40

[184] 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. xiii, 1, 2, 6, 25, 112

[185] Terence J. Parr and Russell W. Quong. ANTLR: A predicated-LL (k) parser generator. *Software: Practice and Experience*, 25(7):789–810, 1995. 72

[186] 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. 8, 42

[187] 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. 47

[188] 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. 2, 36, 112

[189] 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. 114

[190] 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. 9, 31

[191] 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. 34

[192] 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. 52

[193] Antoine Prost. *Douze Leçons sur l'histoire*. Média Diffusion, April 2014. 2, 11, 16, 17

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

[195] 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. 56

[196] 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. 107

[197] 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. 56

[198] 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. 55

[199] Christian Rollinger. Ciceros supplicatio und aristokratische Konkurrenz im Senat der Späten Republik. *Klio*, 99(1):192–225, August 2017. doi:10.1515/klio-2017-0007. 111

[200] 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. 7, 8, 11, 12, 31, 42, 109

[201] Giulio Rossetti and Rémy Cazabet. Community discovery in dynamic networks: A survey. *ACM Comput. Surv.*, 51(2), February 2018. doi:10.1145/3172867. 26, 87, 112

[202] 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. 87

[203] 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. doi:doi:10.16995/dm.52. 52

[204] Céline Rozenblat and Guy Melançon, editors. *Methods for Multilevel Analysis and Visualisation of Geographical Networks*. Number volume 11 in Methodos Series. Springer, Dordrecht, 2013. 1

[205] 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. 1

[206] 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. 37

[207] 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. 43

[208] 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. 88

[209] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2016. doi:10.1109/TVCG.2016.2599030. 15, 71

[210] 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. 88

[211] Christoph Schulz, Arlind Nocaj, Jochen Goertler, Oliver Deussen, Ulrik Brandes, and Daniel Weiskopf. Probabilistic Graph Layout for Uncertain Network Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):531–540, January 2017. doi:10.1109/TVCG.2016.2598919. 111

[212] John Scott. Social Network Analysis. *Sociology*, 22(1):109–127, February 1988. doi:10.1177/0038038588022001007. 11, 21, 22, 23, 31, 40

[213] Michael Sedlmair, Christoph Heinzl, Stefan Bruckner, Harald Piringer, and Torsten Möller. Visual Parameter Space Analysis: A Conceptual Framework. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2161–2170, December 2014. doi:10.1109/TVCG.2014.2346321. 92

[214] 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. xiii, 29, 30, 31, 39

[215] 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. 55

[216] 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. 45, 52

[217] Changxing Shang, Shengzhong Feng, Zhongying Zhao, and Jianping Fan. Efficiently detecting overlapping communities using seeding and semi-supervised learning. *Int. J. Mach. Learn. & Cyber.*, 8(2):455–468, April 2017. doi:10.1007/s13042-015-0338-5. 87

[218] 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. 14, 15

[219] Ben Shneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11(6):70–77, November 1994. doi:10.1109/52.329404. 65

[220] 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. 5, 15, 85

[221] Ben Shneiderman. *Human-Centered AI*. Oxford University Press, January 2022. 114

[222] Georg Simmel. *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. Duncker & Humblot, Berlin, 7. aufl edition, 2013. 23

[223] 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. 5, 29, 30, 53, 60

[224] John Snow. On the Mode of Communication of Cholera. *Edinb Med J*, 1(7):668–670, January 1856. 12

[225] 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. 31, 47, 48, 112

[226] 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. doi:10.1109/2945.981851. 47

[227] Lawrence Stone. The Revival of Narrative: Reflections on a New Old History. *Past & Present*, (85):3–24, 1979. 18

[228] 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. doi:10.1162/153244303321897735. 89

[229] 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. 1, 5, 23

[230] R Core Team. R language definition. *Vienna, Austria: R foundation for statistical computing*, 2000. 55, 84

[231] Melissa Terras. Quantifying digital humanities. *URL: https://www.ucl.ac.uk/infostudies/melissa-terras/DigitalHumanitiesInfographic.pdf*, 2011. 19

[232] J. Joshua Thomas and Kris Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, January 2006. doi:10.1109/MCG.2006.5. 15

[233] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. The bohemian bookshelf: Supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1461–1470, Austin Texas USA, May 2012. ACM. doi:10.1145/2207676.2208607. 41

[234] Charles Tilly. *Retrieving European Lives*. Reliving the Past. Olivier Zunz, 1984. 3, 11, 16, 34, 52

[235] 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. 1, 11, 35

[236] 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. 54, 56

[237] 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. 19, 36

[238] John W. Tukey. The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33(1):1–67, 1962. 12

[239] John W. Tukey. *Exploratory Data Analysis*. Pearson, Reading, Mass, January 1977. 5, 15

[240] Paola Valdivia, Paolo Buono, Catherine Plaisant, Nicole Dufournaud, and Jean-Daniel Fekete. Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(1):1–13, January 2021. doi:10.1109/TVCG.2019.2933196. 29, 47, 82, 89, 96, 113

[241] Guido van Rossum. Python Reference Manual. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995. 55

[242] 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. 43

[243] 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. 88

[244] 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. 87

[245] Emily Wall, Subhajit Das, Ravish Chawla, Bharath Kalidindi, Eli T. Brown, and Alex Endert. Podium: Ranking Data Using Mixed-Initiative Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):288–297, January 2018. doi:10.1109/TVCG.2017.2745078. 88

[246] Wenjia Wang. Some fundamental issues in ensemble methods. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2243–2250, June 2008. doi:10.1109/IJCNN.2008.4634108. 88

[247] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, November 1994. 5, 23, 31

[248] Charles Wetherell. Historical Social Network Analysis. *Int Rev of Soc His*, 43(S6):125–144, December 1998. doi:10.1017/S0020859000115123. 1, 3, 11, 25, 34, 40, 43, 52, 112

[249] 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. 18

[250] Douglas R. White and Ulla Johansen. *Network Analysis and Ethnographic Problems: Process Models of a Turkish Nomad Clan*. Lexington Books, 2005. 1

[251] Hadley Wickham. An introduction to ggplot: An implementation of the grammar of graphics in R. *Statistics*, pages 1–8, 2006. 15

[252] Leland Wilkinson, D. Wills, D. Rope, A. Norton, and R. Dubbs. *The Grammar Of Graphics*. Springer-Verlag New York Inc., New York, 2nd ed. 2005 édition edition, 2005. 14

[253] 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. 76

[254] 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. 4

[255] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. The Sandbox for analysis: Concepts and methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 801–810, New York, NY, USA, April 2006. Association for Computing Machinery. doi:10.1145/1124772.1124890. 88

[256] 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. 41, 56

[257] Jaewon Yang, Julian McAuley, and Jure Leskovec. Community Detection in Networks with Node Attributes. In *2013 IEEE 13th International Conference on Data Mining*, pages 1151–1156, December 2013. doi:10.1109/ICDM.2013.167. 87

[258] Franciszek Zakrzewski. The 1932 population register, May 2020. 36

[259] 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. 88

Résumé

Cette thèse vise à identifier théoriquement et concrètement comment l'analyse visuelle peut aider les historiens dans leur processus d'analyse de réseaux sociaux. L'analyse de réseaux sociaux est une méthode utilisée en histoire sociale qui vise à étudier les relations sociales au sein de groupes d'acteurs (familles, institutions, entreprises, etc.) en reconstruisant les relations du passé à partir de documents historiques, tels que des actes de mariages, des actes de naissances, ou des recensements. L'utilisation de méthodes visuelles et analytiques leur permet d'explorer la structure sociale formant ces groupes ainsi que de relier des mesures structurelles à des hypothèses sociologiques et des comportements individuels. Cependant, l'encodage et la modélisation des sources menant à un réseau finalisé donnent souvent lieu à des erreurs, des distorsions et des problèmes de traçabilité, et les systèmes de visualisation actuels présentent souvent des défauts d'utilisabilité et d'interprétabilité. En conséquence, les historiens ne sont pas toujours en mesure d'aboutir à des conclusions approfondies à partir de ces systèmes : beaucoup d'études se limitent à une description qualitative d'images de réseaux, surlignant la présence de motifs d'intérêts (cliques, îlots, ponts, etc.). Le but de cette thèse est donc de proposer des outils d'analyse visuelle adaptés aux historiens afin de leur permettre une meilleure intégration de leur processus global et des capacités d'analyse guidées. En collaboration avec des historiens, je formalise le processus d'une analyse de réseau historique, de l'acquisition des sources jusqu'à l'analyse finale, en posant comme critère que les outils utilisés dans ce processus devraient satisfaire des principes de traçabilité, de simplicité et de réalité documentaire (i.e., que les données présentées doivent être conformes aux sources) pour faciliter les va-et-vient entre les différentes étapes du processus et la prise en main par l'utilisateur, ainsi que pour ne pas distordre ou simplifier le contenu des sources. Pour satisfaire ces propriétés, je propose de modéliser les sources historiques en réseaux sociaux bipartis multivariés dynamiques avec rôles. Ce modèle intègre explicitement les documents historiques sous forme de nœuds dans le réseau, ce qui permet aux utilisateurs d'encoder, de corriger et d'analyser leurs données avec les mêmes outils. Je propose ensuite deux interfaces d'analyse visuelle permettant, avec une bonne utilisabilité et interprétabilité, de manipuler, d'explorer et d'analyser ce modèle de données. Le premier système ComBiNet offre une exploration visuelle de l'ensemble des dimensions du réseau à l'aide de vues coordonnées et d'un système de requêtes visuelles permettant d'isoler des individus ou des groupes et de comparer leurs structures topologiques et leurs propriétés à l'aide de visualisations et d'intéractions adaptées. L'outil permet également de détecter les motifs inhabituels et ainsi de déceler les éventuelles erreurs dans les annotations. Le second système, PK-Clustering, est une proposition d'amélioration de l'utilisabilité et de l'efficacité des mécanismes de clustering dans les systèmes de visualisation de réseaux sociaux. L'interface permet de créer des regroupements pertinents à partir des connaissances a priori de l'utilisateur, du consensus algorithmique et de l'exploration du réseau dans un cadre d'initiative mixte. Les deux systèmes ont été conçus à partir des besoins et retours continus d'historiens, et visent à augmenter la traçabilité, la simplicité, et la réalité documentaire des sources dans le processus

d'analyse de réseaux historiques, tout en donnant un niveau de contrôle optimal aux utilisateurs sur le processus. Je conclus sur la necessité d'une meilleure intégration des systèmes d'analyse visuelle dans le processus de recherche des historiens. Cette intégration nécessite des outils plaçant les utilisateurs au centre du processus avec un accent sur la flexibilité et l'utilisabilité, limitant ainsi l'introduction de biais et les barrières d'utilisation des méthodes quantitatives, qui subsitent en histoire.