ComBiNet: Visual Query and Comparison of Bipartite Multivariate Dynamic Social Networks

Alexis Pister, Christophe Prieur, Jean-Daniel Fekete

▶ To cite this version:

ComBiNet: Visual Query and Comparison of Bipartite Multivariate Dynamic Social Networks

A. Pister[1,2] ID C. Prieur[3,2] ID and J.-D. Fekete[1] ID

[1]Université Paris-Saclay, CNRS, Inria, LISN, [2]I3, Telecom Paris, Institut Polytechnique de Paris
[3]Université Gustave Eiffel

Figure 1: *The ComBiNet system used to compare two subgroups of a social network of contracts from [CR18], extracted with dynamic visual queries. (A) and (B) show the two visual queries created by the user in the query panel using an interactive node-link diagram editor (V6), dynamic query widgets (V7), and the equivalent Cypher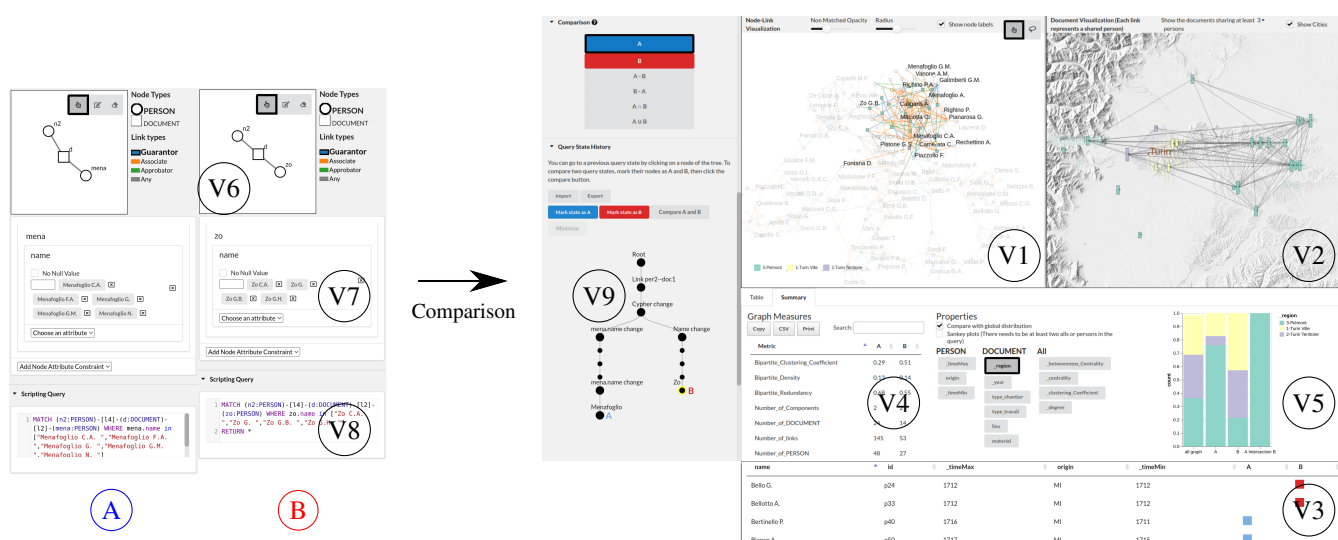 script (V8). The right part shows ComBiNet's global interface in* comparison *mode: (V1) Network visualization panel, (V2) Map of the geolocalized nodes, (V3) Table of persons, (V4) Graph measures comparison, (V5) Attribute distribution plots, and (V9) Provenance tree. The two visual queries on the left, select the "Menafoglio" family on the left, and the "Zo" family on the right, along with their construction contracts and close collaborators.*

Abstract

We present ComBiNet, a visualization, query, and comparison system for exploring bipartite multivariate dynamic social networks. Historians and sociologists study social networks constructed from textual sources mentioning events related to people, such as marriage acts, birth certificates, and contracts. We model this type of data using bipartite multivariate dynamic networks to maintain a representation faithful to the original sources while not too complex. Relying on this data model, ComBiNet allows exploring networks using both visual and textual queries using the Cypher language, the two being synchronized to specify queries using the most suitable modality; simple queries are easy to express visually and can be refined textually when they become complex. These queries are used for applying topological and attribute-based selection on the network. Query results are visualized in the context of the whole network and over a geographical map for geolocalized entities. We also present the design of our interaction techniques for querying social networks to visually compare the selections in terms of topology, measures, and attribute distributions. We validate the query and comparison systems by showing how they have been used to answer historical questions and by explaining how they have been improved through a usability study conducted with historians.

CCS Concepts

• **Human-centered computing** → **Visualization systems and tools;** • **Applied computing** → **Arts and humanities;**

1. Introduction

Social historians and sociologists aim to retrieve and study facts about a specific region and period of time that they focus on. Their work essentially relies on documents—such as marriage acts, census records, surveys, and business contracts—to gather information about the life of important actors that they explore in-depth or to draw conclusions on social aspects of groups in the society of a period and place. Instead of drawing conclusions from their gathered knowledge and interpretations of the documents, a more grounded approach [GS10] consists in constructing a social network from the documents and following a Social Network Analysis (SNA) methodology [Wet98]. For this, they need to encode their documents to extract the persons and other relevant information from the text and transfer it to a structured file or database. Social scientists can then explore, validate, or refute their hypotheses by observing and analyzing the network structure and the connectivity patterns between the entities of the resulting network. They also want to visually explore their data to generate new insights and hypotheses.

Currently, social scientists often model their datasets as simple networks where the nodes are the persons mentioned in the documents. Usually, *two* persons are then connected in the network when they appear in shared documents. This representation is easy to visualize and analyze but simplifies and distorts the information by hiding the documents that witness the relationships between the persons. Another approach consists in modeling the data as bipartite networks, where both documents and persons are represented as nodes that are connected when a document mentions a given person [Gra17, RVH14, SSM*17, SGL08].

In addition, historical documents include time and geospatial information corresponding to the date and location of the events they refer to. Documents often mention additional information on the persons, such as their gender, profession, and date of birth. These are often essential to understanding underlying social phenomena, as time, space, and social status play an important role in social dynamics. For these reasons, historical sources and the underlying social phenomena they refer to can be modeled well by *bipartite*, *multivariate dynamic* networks with *roles* [PDC*22, Cri08]. *Bipartite* means that both persons and documents (or events that are often witnessed by physical documents) are modeled as typed nodes. *Multivariate* means that the nodes and links can carry additional attributes. *Dynamic* means that time is a mandatory attribute of documents. Furthermore, a link created between a person's node and a document's node (when the person is mentioned in the document) has an associated link type that models the *role* of the person in the document/event. Additionally, documents can optionally carry a geographical location. This model unifies several social network models and allows accurate modeling of historical sources.

Many sophisticated tools exist to explore and analyze rich social networks. However, most of them either enforce a too simplistic network model, such as Gephi [BHJ09] and NodeXL [SSM*09], or do not enforce any data model and lead to very complex interfaces that are too demanding for most historians. Moreover, most network visual analytics tools provide limited interactions to query and explore richly encoded data.

In this article, we present a visual analytics system to explore and analyze Bipartite Multivariate Dynamic Social Networks, aimed at answering historical and sociological questions. We designed our tool based on four collaborations with social scientist colleagues. We first collected important questions they each had on their data and translated them into a network analysis perspective. The majority of the questions consisted in either finding specific patterns in the network or in comparing several subsets of the network in terms of network measures, attribute distributions and their overlaps. We thus focus on three high-level tasks: exploration, queries, and comparison of this type of network. Users can explore the network using two layouts: a node-link bipartite view showing the topology and a map layout showing the geolocation mentioned in documents.

We designed and implemented a new visual graph query system that allows building both topological and attribute-based queries. It is based on an interactive node-link representation to express the topological part, along with dynamic query widgets for the attribute values. Both are translated to textual queries performed by the Neo4j graph database [Neo] using its language *Cypher*. Most visualization systems offer dynamic queries to hide the complexity of query languages. However, using a rich data model, some queries are much easier to refine using scripting than interactive widgets. We implemented dynamic queries that also show the translated Cypher queries, and inversely, can translate Cypher queries into their visual forms. With that interface, social scientists can start building their queries with simple widgets and, if needed, complement them by editing the query, alone or with the help of power users. On top of that, they can easily copy and paste the textual query to share the current query and associated results with someone else or to start an analysis session from a previous result. ComBiNet also implements subgraph comparison techniques, allowing the comparison of network-related measures and attribute distributions between the entities returned by the queries. We validate the query and comparison system with a usability study, and we demonstrate ComBiNet can be used to answer sociological questions by describing in depth several real-world use cases.

After the related work section, we describe four use cases to motivate our data model and present our system ComBiNet, with the design of the visual query and comparison features. Finally, we present two use cases demonstrating the utility of our system, showing it can be used to explore complex historical data and answer complex historical questions using queries and comparisons. Our contributions are (1) the design and implementation of a graph query system, synchronizing the visual representation of the query and the associated script, (2) the design and implementation of visualization and interaction techniques aimed at comparing subgraphs, in terms of topology, attributes, time, and locations, (3) a usability study and two real-world use cases demonstrating the utility of the system in answering socio-historical questions.

2. Related Work

We describe how social networks have been modeled and for what goals, systems for network visual analytics, visual querying for networks, and for visual graph comparison. We also describe related work on provenance.

2.1. Social Networks Modeling

SNA aims to address sociological questions using mathematical methods based on graphs. It started by encoding social networks using standard graphs [Fre04], defined as $G = (V, E)$ with V a set of vertices and $E \subseteq V^2$ a set of edges. Social entities become vertices, usually called *nodes*, and relations become edges, usually called *links*. The SNA graph models have evolved to better take into account real properties of social networks, such as types of actors using labeled graphs, importance of actors or relations with weighted graphs, bipartite graphs later, to represent networks where relations only exist between two types of nodes, such as organization and employees, where the relations link employees to organizations but not employees to employees or organizations to organizations.

Temporal graphs (also called *dynamic* and *longitudinal*) are also important in SNA, with multiple models to associate time with vertices, edges, or to graph snapshots $G_1, G_2, \ldots, G_n$ at time $t_1, t_2, \ldots, t_n$, each graph sharing the same vertices [BBDW14].

Multivariate networks, i.e., graphs where vertices and edges can be assigned multiple "properties" or "attributes", possibly with an associated value, are less used in SNA. These attributes are often considered secondary, the emphasis of SNA being on the graph topology, its features, measures, and evolution.

In contrast to SNA, *Graph databases* are concerned with the concrete encoding of networks in computer files and memory, and less with their mathematical analysis. Popular graph databases such as Neo4j [Neo] can be queried and modified using domain-specific languages (called Cypher for Neo4j). Graph databases can model any kind of complex social network, but they do not provide guidance for modeling real social networks.

Modeling Social networks can also be done using the *Semantic Web* now called *Knowledge Graphs* [HBC*21], that rely on a graph representation called the "Resource Description Framework" (RDF). RDF models any graph and its attributes as a set of triples of the form *subject*, *predicate*, *object*. This representation can model a large variety of graphs, including social networks. However, like graph databases, it does not enforce one particular modeling and produces an important fragmentation of data models, which is detrimental to the design of graphical interactive tools.

Historians, demographers, sociologists, and anthropologists have been designing specific data models for their social graphs based on genealogy or, more generally, kinship [HHW11]. For genealogy, the standard GEDCOM [GED] format models a genealogical graph as a bipartite graph with two types of vertices: individuals and families. Both types can have attributes and be associated with *events*, such as birth, death, marriage, and many more. These events are dated. A genealogy graph is therefore encoded as a bipartite, multivariate, dynamic, social network with roles since individuals play particular roles in families. The PUCK software [HMSM] has extended the GEDCOM format to adapt to more flexible kinds of family structures for anthropological studies; it has also extended the types of relations and events to handle different kinds of ties between individuals or families. Due to its history and evolution, it is complex to understand but reflects the needs of social scientists.

2.2. Visual Analytics for Social Networks

In addition to mathematical analysis and data modeling, several systems have been developed to support the visual exploration of social networks. They all use visualization, but not all of them are interactive. Most of the SNA-oriented systems such as Gephi [BHJ09], Pajek [MB16], UCINET [BEF02], and NodeXL [SSM*09] visualize networks with node-link diagrams, using several layouts. More recent systems use alternative representations such as adjacency matrices [SMBP*17], TimeArcs [DPF16], and geographical maps for geolocated entities, and others [SMBP*17, VBP*21]. More specialized systems focus on visualizing bipartite networks: GeneaQuilts [BDF*10] visualizes genealogical trees as bipartite networks with a cascade of matrices. However, that visualization is specific to genealogies

Jigsaw [GLS14] is a visual analytics system based on documents, designed for intelligence analysis. Its data model is based on a collection of time-stamped documents. Each document contains text and identifies *named entities* such as persons, places, and organizations. Jigsaw provides a large set of visualization techniques to explore the documents in detail or as overviews, such as a node-link diagram of entities (a multi-mode network), a list view, and many more. Compared to PUCK, Jigsaw does not handle the *roles* of people in documents, it merely considers each mention of a person as a link with no finer precision.

PAOHVis [VBP*21] visualizes hypergraphs of documents and persons as dynamic hypergraphs. A hypergraph is a generalization of a graph where (hyper) edges can contain one or more vertices instead of exactly two. Internally, a hypergraph is represented as a bipartite graph with a document vertex connected to person vertices. Therefore, it also encodes social networks as bipartite dynamic social networks. It also supports *roles* (link types) and node types or groups. Yet, PAOHVis does not handle node or link attributes.

Recently, several visual analytics systems have been proposed to specifically explore historical datasets modeled as knowledge graphs [ZTC*22, ZWW*22]. This model enables the modeling of entities such as persons and events and various relation types, but the model can become very complex for social researchers.

2.3. Visual Graph Querying

Several scripting languages, such as R [R C14] and Python [vR95], have been extended to support the exploration of social networks using specialized libraries such as igraph [CN06] and NetworkX [HSS08]. However, social scientists are often challenged to use scripting languages and programming.

Finding and extracting a subgraph of interest in a bigger graph is an old problem in SNA. Constructing and querying a pattern from a graph requires knowledge of graph databases and query languages. To lower the complexity barrier, several visual graph query systems have been developed to allow analysts to rapidly build and refine their queries visually. GRAPHITE [CFT*08] and VERTIGo [CSIP21] allow specifying a graph query as a node-link diagram that the user creates interactively. Shadoan and Weaver [SW13] use a similar concept with hypergraphs to filter multidimensional data. Other systems, such as VIGOR [PHE*18]

only visualize the query after it has been written using a scripting language. However, these visual systems are limited to topological queries, including constraints on the vertex and edge types; they do not support constraints related to general attributes and time associated with vertices and edges.

2.4. Visual Graph Comparison

Few visual systems support comparison tasks on social networks. Gleicher et al. [GAW*11] propose a taxonomy of visual comparison designs for complex objects. They claim any visual comparison system can be classified into one (or a mix) of the following categories: juxtaposition, superposition, or explicit design.

Andrews et al. [AWW09] describe a technique to compare several graphs using a combination of juxtaposition and superposition techniques. The two candidate graphs are shown side by side, along with a third view composed of a fusion graph highlighting both the shared nodes along with the non-shared nodes with different colors. Freire et al. [FPSG10] describe the ManyNets system to compare many networks by using a table where each describes one graph and each column shows graph measures in terms of small visualizations, from simple bars to distributions, allowing the comparison of a large number of graphs. However, ManyNets does not visualize the networks per se (no layout shown) and do not take into account attributes, node types, or time. Hascoët and Dragicevic [HD12] describe a system to match and compare graphs using superposition, focusing on the topology, not taking into account attributes or time. Tovanich et al. [TPR*21] propose a visual analytics tool to compare multivariate, sometimes bipartite, dynamic graphs and find common structures. Yet, their tool does not handle roles and is designed for the specific task of matching a subgraph into a large graph.

2.5. Provenance

Provenance in the context of Visual Analytics consists in the logging of the sequence of actions of users on an interactive visualization system during analysis sessions. Collecting provenance information has proven to benefit users by providing them action recovery (undo/redo), collaborative, and reproducibility capabilities [RASC16]. For example, VisTrails allows users to reproduce their visual analyses by providing an executable history graph of their actions, [CFS*06] while GraphTrail provides provenance tools to ease collaborative analysis [DHRL*12]. Provenance can also be beneficial for visualization designers and researchers, as it gives them a tool to understand users' behaviors [BH19, BWD*19] and evaluate/improve visualization systems [RLB18]. All the reasons and concrete implementations of provenance are discussed in depth in Xu's survey [XOW*20].

3. Task Analysis and Design Process

We designed the ComBiNet tool in collaboration with historians; all their historical documents data fitted well the bipartite multivariate dynamic networks model. We first collected all the questions they had about their data and what they wanted to see in a visual interface. By analyzing the questions we leveraged tasks and requirements. We designed the interface from the requirements with continuous interactions with our collaborators.

3.1. Collaborations

We describe here four example projects coming from our collaborations, involving regular meetings and multiple interviews over two years; from these collaborations emerged our proposed network modeling. All these datasets are textual corpora constituted of historical documents mentioning people with complex relationships. They are well modeled by bipartite multivariate dynamic networks. We also list our collaborators' main questions and the graph queries extracting the information to start answering them. The complete answers involve visualizations of the query results that we describe in the next section. We categorized those according to four dimensions: global (G)/local (L) (do they want to categorize a group of nodes or retrieve specific persons/documents), if the question can be answered using the topology (T), and/or the attributes (A), and finally if a comparison (C) using several filters is needed or not (N).

1. Analysis of the social dynamics from **construction contracts in Italy in the 18th century (141 documents, 272 persons) [CR18]**. The corpus is made of contracts (manuscript documents) for different types of constructions in the Piedmont area in Italy. People are mentioned in three different roles: *Associates*, who participate in the construction; *Guarantors*, who bring financial guarantees; and *approbators*, who vouch for the guarantors. Along with the time and location of the construction site, documents have a construction type (military, religious, and civil), work type (big work, small work, reparation, transportation, etc.), and material (wood, stone, metal). People also have an origin attribute (the place they come from), manually extracted from the original documents.

   Question 1 Do approbators act as bridges compared to associates and guarantors? (G, T, C)
   Query 1.1 Request all approbators occurrences.
   Query 1.2 Request all associates and guarantors occurrences.
   Question 2 Are there pairs of people mutually guarantors to each other in different contracts? (G, AT, N)
   Query 2.1 Select pairs of people connected each to the two same documents, with a guarantor role and any other role.
   Question 3 Who are the persons of the extended Zo family? (G, AT, N)
   Query 3.1 Request all the persons of the Zo family and their N+2 ego network.
   Question 4 Compare the Menafoglio and Zo families in terms of contracts and activities. (G, AT, C)
   Query 4.1 Request all the persons of the Menafoglio family and the documents that mention them.
   Query 4.2 Request all the persons of the Zo family and the documents that mention them.
   Question 5 Who are the persons having the 3 roles? (G, AT, N)
   Query 5.1 Select persons with an associate, guarantor, and approbator role in 3 different documents.
   Question 6 What are the differences between Torino (Turin) and Torino surroundings according to the contracts? (G, AT, C)
   Query 6.1 Request all documents located in Torino, with the persons mentioned.
   Query 6.2 Request all documents located in the Torino area, with the persons mentioned.

2. Analysis of migrations from the **genealogy of a french family between the 17th–20th centuries (2053 events, 957 persons from a private source).** The corpus is made of family trees referring to several document/event types: birth and death certificates, marriage acts, military mobilization, and census report. The roles are different for each event type and consist of *chil-*

Main Tasks	Subtasks	Views	Constraints
Bipartite Graph Exploration	T1.1 See the structure of the network	V1	A node-link representation is expected. The geolocation of events has to be done according to the historical period.
	T1.2 Show the nodes attributes values and distributions	V1/V2/V4	
	T1.3 See the roles of the persons in the documents they appear	V1	
	T1.4 See the location of the different documents	V2	
	T1.5 Show the time of the documents	V1/V2/V4	
Apply filters to isolate subgraphs	T2.1 Filter topological patterns	V5	Filters must be visual and easy to apply.
	T2.2 Filter on attribute values	V6	
	T2.3 See the subgroup alone or in context of the network	V1/V2	
Compare several subgroups	T3.1 See the shared and exclusive entities	V1/V2	
	T3.2 Compare the node attribute distributions	V4	
	T3.3 Compare the subgraph measures	V3	

Table 1: Tasks to support during exploration, according to our expert collaborators, are split into 3 high-level tasks.

dren, father, mother for the birth events, *deceased* for the death event, *spouse* and *witnesses* for the marriages, and *family member*s for the census events.

Question 7 What is the trajectory of life for a given specific individual (birth, living, marriage, death)? (L, A, N)

 Query 7.1 Select one person and all her/his documents (to use the mentioned places).

Question 8 What is the trajectory of life for a family? (L, A, N)

 Query 8.1 Select birth certificates with the child, parents, and birthplace.

Question 9 Where are located the main migrations, and at which time do they occur? (G, A, N)

 Query 9.1 Select persons with a geolocated birth and death certificate.

Question 10 Are there differences in volume and location between migrations in the 18th and 19th centuries? (G, A, C)

 Query 10.1 Select persons with a geolocated birth and death certificate from the 18th century.

 Query 10.2 Select persons with a geolocated birth and death certificate from the 19th century.

Question 11 In the *Haute-Vienne* and *Côtes-d'Armor* administrative areas, are there cycles in living places every 10/20 years? (G, A, N)

 Query 11.1 Select persons with their census reports located in Côtes-d'Armor and Haute-Vienne.

Question 12 In the 19th century, was there an overall decrease in the social status and professions of persons in the dataset? (G, A, C)

 Query 12.1 Select persons in the first half of the 19th century with a profession mentioned.

 Query 12.2 Select persons in the second half of the 19th century with a profession mentioned.

3. Analysis of migrations from Spain to Argentina through the **marriage acts at Buenos Aires in the 17–19th centuries (1396 acts, 6731 persons) [Mou16].** The corpus is made of acts that mention the spouses and the witnesses of the wedding, which are the roles modeled by the links. The origin, date of birth, and parents' names are specified for both spouses.

Question 13 How are spouses and witnesses linked in their family network? (G, T, N)

 Query 13.1 Select marriages with spouses and witnesses, where one of the spouses and a witness have the same parents.

 Query 13.2 Select marriages with spouses and witnesses, where one of the spouses and a witness have the same grandparents.

Question 14 Who are the persons with 2 marriages with a long delay? (L, A, N)

 Query 14.1 Select persons in 2 marriages as husband or wife. Put a constraint on the difference of time in the marriages.

Question 15 Where are the persons marrying in Buenos Aires coming from? (G, A, N)

 Query 15.1 Select persons with a birth certificate located not in Buenos Aires.

4. Socio-political analysis of **migration of ethnic Germans from communist Romania to West Germany in the 20th century (ongoing work) [Dim20].** The corpus is made of administrative forms that mention persons requesting to migrate, along with the persons they want to join, and the administrative persons of the ministry in charge of the forms (3 roles). The family members of the aspiring migrant are also mentioned in the forms, with their respective dates of birth.

Question 16 What members of their family do emigrants typically join? (G, AT, N)

 Query 16.1 Select all migration documents with the emigrant and the person they are joining.

Question 17 What price does the emigrant have to pay, given their socio-economic profiles? (G, A, C)

 Query 17.1 Select people who are mentioned in a budget and a migration document.

3.2. Tasks Analysis

Most of the questions we collected from our collaborators could be answered by isolating a subgroup of entities and analyzing them in the context of the whole network, or by comparing two subgraphs, in terms of their entities, structure, and attribute distributions. From discussions with our collaborators and the analysis of their questions on their data, we elaborated a list of requirements for the visual interface, split into three main parts: 1) Exploration of the data, 2) Queries, and 3) Comparisons. The elaboration of the tasks was an iterative process, as we showed the interface to our collaborators several times in the development phase to get feedback. The tasks are described here and summarized in Table 1:

1. **Exploration of bipartite multivariate dynamic networks**. The visual interface must allow exploration of this specific type of graph, using every aspect of the data, i.e. its topology (T1.1), node attributes (T1.2), roles (T1.3), geolocation of the documents/events (T1.4) and time (T1.5). Common interactions such as selection and zooming are also needed for the exploration.

2. **Applying filters**. To answer their questions, users need to be able to apply filters to the data, to isolate specific groups of entities having specific behaviors or characteristics. To answer the

diversity of questions, they should be able to put constraints on every aspect of the data, i.e. the topology, the roles (T2.1), and the attributes (including time and geolocation) (T2.2). Access to provenance information can also help them in their query construction, by going to previous states and exploring different paths more easily (T2.3). Once they are satisfied with their query, they want to explore the results, usually in the context of the whole network (T2.4).

3. **Comparison of several subgraphs.** Users should be able to compare several subgraphs isolated after applying filters, to see the similarities and differences between groups of entities of interest. The system should be able to easily see the common and shared entities of the two subgraphs (T3.1), their respective place in the network, their structural differences (T3.2), and their different attribute distributions (T3.3).

4. The ComBiNet System

ComBiNet is designed to visualize, explore, and analyze social networks encoded as bipartite multivariate dynamic networks. When started, ComBiNet dynamically collects the node types, roles, subtypes, and attributes when reading the network from the database. The interface is constituted of four main panels, split into different views as shown in Figure 1: the query and comparison panel (V6, V7, V8, and V9), the bipartite visualization panel (V1), the map visualization panel (V2), and the query results panel (V3, V4, V5). Comparison features are incorporated in the same views with multiple comparison mechanisms, explained in subsection 4.3.

4.1. Visualizations

ComBiNet presents a social network with multiple visualizations highlighting different aspects of the data. The visualizations are linked when it makes sense.

V1: Bipartite Node-Link Diagram The bipartite node-link visualization panel shows the network using the DrL force layout from igraph [CN06] with overlap removal using D3 [BOH11]. Node-link representations are very common in social sciences [BHJ09, MB16] and were a specific request from our collaborators. In the context of our bipartite model, the persons are represented as circles and the documents/events as squares, while the roles are encoded as link colors. A link models the mention of a person in a document. This view provides an overview of the data by showing the structure of the network (T1.1) and the roles of the persons in their different documents (T1.2). The view provides pan & zoom and selection interactions for effective navigation. Nodes' labels are displayed (names of persons and ids of documents by default) on the canvas with an occlusion-free mechanism that hides nodes with a low degree when two or more node labels overlap.

V2: Map View The map visualization panel on the right shows an event-centric view, displaying only the geolocalized document/event nodes on a map. By default, only event nodes are shown. The system can create links between event nodes that share persons with a user-specified threshold. Persons are not directly shown in this view as they do not have a unique location. This map view presents a transformation of the bipartite graph, focused on

the geospatial information that is very important to social scientists (T1.3).

As we collaborate with historians who study different periods, we cannot use modern map backgrounds such as the default one provided by OpenStreetMap or Google Maps since many features are anachronistic (e.g., roads, administrative areas, borders). We, therefore, provide a map background with only these non-administrative features: elevation, lakes, rivers, and types of environment. We also show the most important cities as most of them existed in the past and provide landmarks. The map uses Natural Earth tiles and vector data [Nat].

The two views are coordinated: selecting/hovering an event node in the graph view highlights it on the map and vice versa, while hovering a person node highlights all its corresponding documents on the map, rapidly showing the person's events' locations.

V3: Entities Tables All the persons and the documents of the loaded dataset are listed in two separate tables, showing the attributes of the entities. This way, users can order the entities according to any attribute they want (T1.2). The tables are linked to the visualizations, meaning that selecting a row highlights the respective entity in the visualizations and vice-versa. The tables are exportable in CSV or PDF format, or directly in the clipboard.

V4: Graph Measures The Graph Measures view shows measures related to the network and gives insights into its structure to users (T1.1). We report simple measures like the number of persons, documents, links, and components, and more sophisticated bipartite network measures asked by our users that they can report for their analysis: the *bipartite centrality, bipartite clustering coefficient, and bipartite redundancy*. These measures are updated in real-time when filters and comparisons are applied.

V5: Attributes View All the attributes in the network are shown as buttons in the bottom right of the interface, sorted by their associated node type (person, document, and both). They can be quickly visualized by hovering over the button, producing two effects: it colors all the nodes on the two views according to their attribute values, and it shows a plot of the distribution of the selected attribute, as shown in Figure 1. By clicking on the button, the visual encoding and distribution remain selected. Users can follow a first exploration of their data by visually detecting correlations between attribute values and some groups of persons or between attribute values and some specific areas in the map view (T1.2, T1.4, T1.5).

4.2. Query Panel

The query panel allows users to easily build queries visually, with topological and attribute constraints. The visualization of the query is synchronized with the Cypher query sent to the database. Modifying one representation will update the other, allowing users to build a query visually and refine it in Cypher when appropriate. In this section, we describe all the features and interactions allowing ComBiNet to build a query and illustrate them with questions 2 and 6 of the use case #1. Our collaborator wants to *find the persons who are mutually Guarantor to each other in separate contracts (2)* and to know *how Torino and Torino's surroundings differ according to their contracts (6)*.
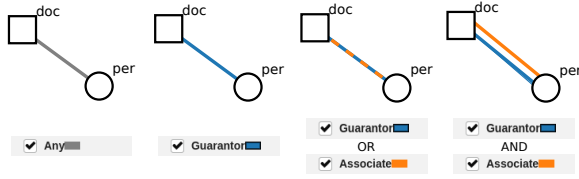
Figure 2: All link creation possibilities: Any link type (left), one selected link type, here guarantor (middle left), union of several link types (middle right), several links with different types (right)
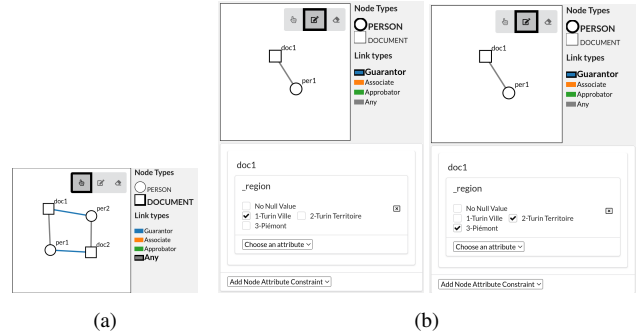

Figure 3: Visual queries created to answer questions 2 and 6 of our collaboration #1. (a) The visual query retrieves individuals who are mutual guarantors to each other in separate construction contracts. (b) The two visual queries retrieve the documents—along with the signatories—of Torino (left) and of Torino surroundings (right).

V6: Node-Link Dynamic Query The interactive node-link diagram allows for building a subgraph query graphically, which represents a topological constraint (T2.1).

The query subgraph is built and edited interactively. At each modification, the visual query is converted into a Cypher query and run in the database, which returns the results. All the matches are displayed in the entities tables (V3) and highlighted in the main visualization views (V1, V2). Three modes of interaction are available through the top-right menu: *selection, addition*, and *deletion*. The *selection* mode allows dragging the nodes in the panel, while the *addition* and *deletion* modes allow the following actions:

Node Creation: In *addition* mode, clicking on an empty area creates a new node. The node will be of the selected type from the legend on the right (Person, Document, or Any).

Node Deletion: In *deletion* mode, clicking on a node deletes it and its links.

Change Node type: In *selection* mode, clicking on a node opens a menu allowing to change its type.

Link Creation: In *addition* mode, clicking on a node and dragging the mouse to another node will connect the two with a link. Its type (color) will be the link type selected on the legend.

Link Deletion: In *deletion* mode, clicking on a link deletes it.

Change link type: In *selection* mode, clicking on a link opens a menu to change its type.

Users build concrete subgraphs with the same representation as in the bipartite graph view: a visual query is a graph template. Each role (link type) is rendered using a color (Figure 2 left). Users can also create untyped links using the *Any* value, which will match all the existing link types (Figure 2 left). Created links can be matched by different selected link types by checking several possible types for one link. These links are represented by a dashed line with the colors of the possible types (Figure 2 middle right). Several links with different types can also be created among two nodes to query a person with more than one role in the same event (Figure 2 right). When a node or link is created in the query, it is given an identifier starting with *per* for a person, *doc* for a document, *link* for a link, followed by a number. These identifiers are used in the attribute constraint panels and the textual query and can be changed through their textual representations.

To find persons who are mutual guarantors for our collaborator #1, we first create one person and two documents using the addition mode by clicking on the canvas. We then link the person node to the first document with a link that is not typed (Figure 2 left) and link it to the second document with a link typed *guarantor* (Fig-

ure 2 middle left). We then create a second person node and link it to the two documents with opposite link types. The resulting visual query is presented in Figure 3 (a). To answer question 6 (comparing Torino and Torino's surroundings), we start by requesting all the links in the graph, no matter the type, as shown in Figure 3 (b). The database then returns all the links in the graph with their attached nodes. Setting attribute constraints on the location of the constructions will then let us answer Question 2.

V7: Attribute Constraint Widgets Users can also add attribute constraints (T2.2) on the created nodes with the help of interactive widgets. An input button is created for each node and link identifier from the node-link query panel. It allows the creation of a dynamic query widget for any of its attributes. The widget design will vary according to the three possible attribute types: numeric, categorical, or nominal, as in the original dynamic queries [Shn94]:

1. **Numeric constraints** are modeled as range sliders, allowing the selection of a lower and upper bound to the filter.
2. **Categorical constraints** are modeled as a set of checkboxes. Each possible value has a corresponding checkbox.
3. **Nominal constraints** are modeled as text input, where the user can write any desired value. All the possible values are shown at the same time and filtered as the user writes.

For the categorical and nominal widgets, selecting several values will correspond to the union of the filters.

To answer our collaborator's question 6, we want to filter the documents located in Torino (*Turin* in French). For this, we first select the whole dataset by linking a person and document node with *Any* link. Then, we select the id *doc1* of the document of our visual node-link query and the *region* attribute. It will initialize a categorical widget including all the values found in the dataset for this attribute with associated checkboxes. We check the region of interest *1-Turin Ville* to select all the documents from this region. To select the documents of Torino surroundings, we can simply uncheck the *1-Turin Ville* value for the *region* attribute and check the two other values *2-Turin Territoire* and *3-Piemont*, which are areas corresponding to the surroundings of Torino. Both queries are represented in Figure 3 (b).
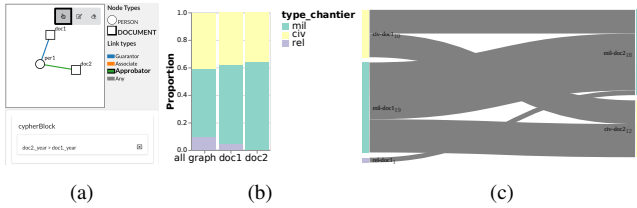
(a)   (b)   (c)

Figure 4: Two ways of showing the distribution of *type chantier* (type of works), a categorical attribute with three possible values "*rel*igious", "*mil*itary", and "*civ*ilian". (a) A query matching the contracts made by the same person (*per1*) as an *approbator* (green link to *doc2*) after being a *guarantor* (blue link to *doc1*) using the constraint (doc2._year > doc1._year). (b) Stacked bar chart for the matches, the earlier contract (*doc1*), and the older contract (*doc2*), and (c) Sankey diagram showing the first and second construction types in chronological order. The Sankey diagram reveals the value changes between the two documents: the guarantor who worked initially on religious switched to military work.

V8: Cypher Editor Users can build or modify a query using the Cypher query language with the Cypher text editor. This allows users to start creating a query visually and refining it in the text for complex constraints that cannot easily be represented in a visual form. The editor supports auto-completion, e.g., to help discover and spell the attribute names. The visual and textual representations are synchronized, meaning that changing one will update the other and update the results in the visualizations.

Query Results Each modification of the query, whether from the node-link dynamic query, the widgets, or the Cypher text boxes, update the two visualization panels (V1, V2), the entities tables (V3), the graph measures view (V5), and the attribute plots (V6). The nodes and links that do not match (are not retrieved by the query) are greyed out in V1 and V2, and are removed from the persons and documents tables (V3). A third table shows every found occurrence of the created pattern. Users can switch between tables in the table view with tabs. The graph measures are computed on the new graph formed by the union of all patterns found and updated on the graph measures view (V5). Since some measures can be long to compute, the values are computed iteratively and shown progressively [FFNS19] to avoid blocking the interface. The distribution plots in the attributes view (V6) are updated, showing the values of the entities of the new pattern, next to the global distributions.

Attributes Visualization When users select an attribute in the attributes view (V5), its distribution is visualized for the queried entities over the whole network. However, these plots show the aggregated values and we lose the potential value transitions between the query nodes. For example, Figure 4 shows a query to list the persons with the role of *approbator* in a contract after being a *guarantor* in another contract (using a time constraint). We may want to see if the locations or types of the two contracts are the same or if they change, case by case. Unfortunately, we lose this information with the aggregated plots. By checking the "Sankey" option on top of the distribution visualization, the plots are transformed into Sankey diagrams, giving information on how the attribute values relate between the nodes (person or event) of the same query.

A Sankey diagram showing the attribute distributions is particularly useful for queries where the nodes have intrinsic time relationships, such as birth certificates, marriage, or death certificates, where we know the order in which these events occurred. It is also useful for queries with user-defined time order constraints as in Figure 4.

V9: Provenance Tree Each change in the query panel is saved with the computed results so that the history of the query construction can be shown in the form of a provenance tree (T2.4), managed using the Trrack library [CGL20]. Each node of the tree represents a query change, with a description label like "New Link". It allows users to rapidly visualize the succession of filters applied with their refinements. At any moment, users can click on a tree node to go back to the previous state; allowing them to navigate in the exploration states. Hovering over a node shows a tooltip with its query visualization. If a change is done from a previous state, a new branch is created on the tree, allowing revisiting and refining explorations. Figure 1 shows the provenance tree at the bottom of the query panel.

4.3. Comparison

In addition to comparing the results of a query to the whole graph, ComBiNet allows comparing the results of two queries. Users select two query states in the provenance tree and mark them either as "A" or "B". Clicking on the button "Compare State A and B" compares them. The interface changes to *comparison mode*. Several buttons appear on top of the provenance tree: A, B, $A - B$, $B - A$, $A \cap B$, and $A \cup B$ for exploring the combinations of the two results of A and B in the two visualizations panels.

To answer several of the questions raised by our collaborators, we need to compare two subsets of the network. For question 6 of collaboration #1, we want to compare the works in Torino with the ones in Torino surroundings. For this, we can rename the provenance tree nodes which correspond to the two queries we made previously retrieving the works located in Torino and in Torino surroundings, with explicit names such as "Torino" and "Surroundings". We then mark them as A and B using the appropriate buttons. Clicking on the "Compare State A and B" will make the interface compare the two query results.

Topological Comparison In comparison mode, users can rapidly switch between the visual filters of (A) and (B) by hovering over their respective buttons on the comparison menu and thus compare the structure of the two resulting subgraphs (T3.1). Similarly, boolean comparison operations are available by hovering their respective buttons (Figure 1 top of V9), such as the intersection, union, and differences between the two filters. Moreover, the graph measures view (V4) allows comparing the different graph measures of the two subgraphs by showing them side by side (T3.3). Comparing these measures, such as the number of matched documents or the densities, is crucial for SNA.

Attribute-Based Comparison Comparing one or several attribute distributions between (A) and (B) is also useful for answering the historical questions of our users. In the attribute view of the results panel, hovering or clicking on an attribute name will show the distribution of this attribute in four contexts: the nodes of the whole
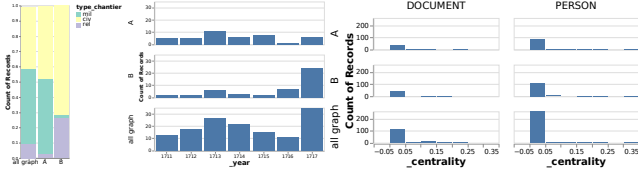
Figure 5: Distribution of two document attributes and one global attribute for the documents and signatories of Torino (A), Torino's surroundings (B), and the whole network.

Figure 6: Attributes distributions plots between the whole graph, the *Menafoglio* family (A), the *Zo* family (B), and $A \cap B$, for the *region,type_chantier, materiel type.*

graph, the queries (A), (B), and the currently selected Boolean operator (e.g., intersection or union). This allows users to compare attribute distributions between several subsets of interest (T3.2). For example, we can compare the attributes between the contracts of Torino and the ones of its close territory. We can also compare the persons who worked in Torino, in Torino's close territory, and both areas, by selecting the intersection operator. Figure 5 illustrates the comparison charts for different attributes. We can see that the types of construction sites differ between the two regions: the city of Torino clearly has a lot of military sites compared to the territory of Torino, which has almost none. This is the opposite for the number of religious sites, which are almost all localized in the territory of Torino. If we now look at the year distribution of the contracts, we can see a difference in the spike of the distributions. The majority of Torino's contracts occurred around 1713 and around 1717 for Torino's territory. We can also compare the profile of persons who collaborated at Torino and Torino's territory by selecting the intersection of those two queries. One of the questions the historian had was to know if those persons were a group with specific attributes and characteristics, or were inseparable from other persons working in the two areas. If we look at the betweenness centrality, on average, the values are higher for this group of people, meaning that the persons who work on the construction site at Torino and Torino's territory are clearly two distinct groups, and the persons collaborating in the two areas act as bridges between these groups. This visual demonstration was convincing and revealing for our users.

4.4. Implementation

ComBiNet is made of three components: a web visual interface, a python server, and a Neo4j graph database instance. The client interface is written in JavaScript using D3 [BOH11], Vega [SMWH16], and the Trrack library [CGL20]. The python server is written in Flask and interacts with the Neo4j instance for query processing. We implemented our Cypher parser with the ANTLR parser generator [PQ95].

5. Use Cases

In this section, we describe how our system has been able to specifically answer questions from two of our collaborations.

5.1. Construction sites in Piedmont (#1)

One of the main questions of our collaborator was to compare two families which he knew played a big role in the structure of the
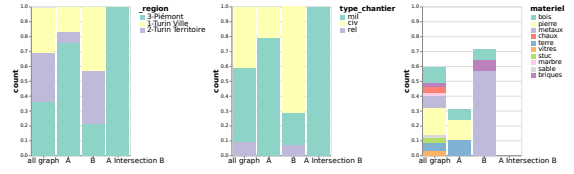
network: the *Menafoglio* and *Zo* families (question 4). He was interested in knowing if there were differences in specialization in type of contracts and area of work for the core members of these families, and to what extent they were collaborating. Moreover, he was very interested in characterizing the group of people collaborating with both families. To answer those questions, we first selected the core members of the *Menafoglio family*, by checking the people known by the historian, and their close neighbors. Looking at the bipartite view (see Figure 1), we can see that the group is pretty dense with people collaborating a lot between them. Looking at the map, we can clearly see that the family has been mostly active in Piedmont outside of Torino and Torino's close territory. We also have a first view of the attribute distribution of the persons in the group and their contracts. We then do the same query for the Zo family. We keep the same topological filter and replace the name filters with the core members of the Zo family known by the historian. The distribution plot of the *region* in Figure 1 shows that, contrary to the *Menafoglio*, the *Zo* have been more active in Torino and around (Figure 1 and 2 of the supplementary material show more details). The two groups can be compared using the *comparison mode* by selecting the two queries in the provenance tree. This opens the comparison menu to quickly navigate between the visual selection of (A), (B), and the set $A \cap B$ that interests our collaborator. The table showing the graph measures of the two subsets confirms what is shown visually: the *Menafoglio* group is more populated but less dense than the *Zo* family. Our user is then interested in comparing the distribution of several attributes between the two groups. We can clearly see in Figure 6 (middle) that the *Menafoglio* family is more specialized in military sites, while the *Zo* family is doing more civil constructions. This is confirmed by the *material* distribution that shows that the constructions of the *Menafoglio* are often using stones (*pierres*), whereas it is never the case for *Zo* constructions. Finally, the persons collaborating in the two groups have a betweenness centrality higher on average. This makes sense as they act as bridges linking the two families.

5.2. French Genealogy (#2)

We describe how ComBiNet allowed answering an important question of the use case #2: to detect the largest migrations across several generations, in which areas, and at what time they occurred (question 9 from collaboration #2). The map view shows at a glance (Figure 7) that the majority of events have taken place in three specific regions west, mid-north, and mid-south.

To find patterns of migrations within families, we first make a query representing a simple family by linking a person node to a
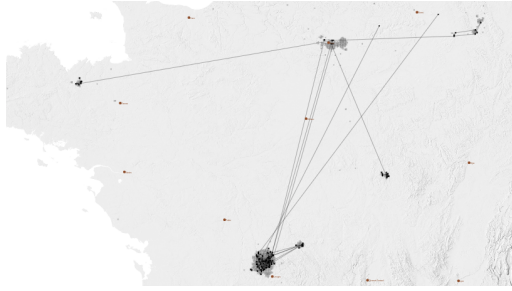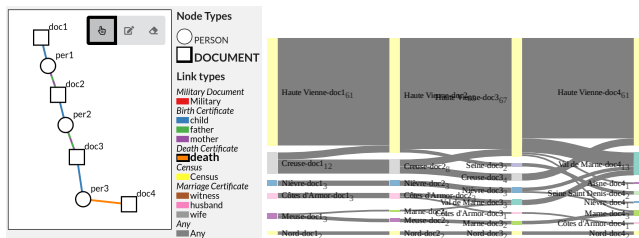
Figure 7: Migrations over three generations from use case #2.

(a) Visual query to find all 3-generation families

(b) Sankey diagram showing the birth and death places of people across generations

Figure 8: Migrations across départements over three generations

Figure 9: Sankey diagrams showing the migration of people in the 18th and 19th centuries, extracted from their birth and death places.

birth event, connected to the parents using a link of *father* or *mother* type. We repeat the process to the new parent node to add another generation. Finally, we connect the latest generation child with a death event, to have another date and location to compare to (see Figure 8a). This query returns every person with their parents and grandparents, along with their respective birth and death data for the latest person. We also create a constraint on the *département* attribute on the documents to only retrieve the events that have a non-null associated location. This request returns a subgraph of 64 persons and 88 documents. The user can now select the *département* attribute to create a Sankey diagram that shows the change of départements across the different generations of the families. Figure 8b shows that the majority of families are from *Haute-Vienne* (which can easily be confirmed by checking the map), and do not move much across generations. Our collaborator, however, detected interesting patterns of people moving from the département *Creuse* to *Haute-Vienne* across two generations. She refined the query by adding an attribute filter on this specific département using a widget. The table view then showed her who these migrants were and when it occurred. The bipartite visualization panel allowed exploring more in-depth this specific group of people.

Afterward, we answered question 10, to compare the migrations between the 18th and 19th centuries. She thought people started moving in the 19th century and wanted to confirm it. To answer this, we first created a query to retrieve the people with birth and death certificates from a specified département. We then applied a time filter on the death certificate node, first for the 18th century and then the 19th century, compared the two query results using comparison mode, and looked side by side at the Sankey graphs related to French *départements* (Figure 9). We can clearly see that
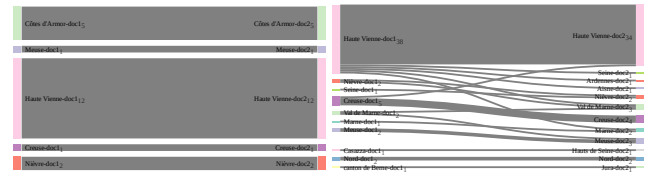
people do not move at all in the 18th century, while in the 19th century, even if the majority of people stay in the same place from their birth to their death, several people move to other départements.

6. Formative Usability Study

After demonstrating that ComBiNet can be used to answer questions raised by our collaborators, we performed a formative usability study with the two historians of use cases #1 and #2, and one expert in visualization, to see if the system was usable by the target users and improve its usability. We had three meetings with each and gave them control of the tool to see if they could use it to explore their data—the visualization expert used the interface with the dataset of construction contracts #1—perform queries, and comparisons. At the first meeting, we explained them the panels of the system and each feature. At each meeting, we asked them to speak aloud, commenting on their aims and actions. At the end of each session, we asked them for their general feedback and what other features they would like to have. We improved the system and made the changes asked by the users before setting up new appointments.

During the first meetings, all the users had difficulties using the query and comparison systems. We, therefore, enhanced the query system by adding a undo/redo feature, making the legend bigger (with the selected node and link types more contrasted), and by showing transparent nodes at the position of the mouse cursor in *addition* mode. It also led to redesigning the comparison system: at first, users had to click on a node of the query state provenance tree to compare its associated query to the current query state. Since users had trouble understanding how it was working, we changed it to force users first to mark the query state nodes they wanted to compare as "A" and "B" with the help of two buttons. Once two nodes are marked, a button "compare" appears and allows users to compare the two subgraphs. Making the comparison process more explicit helped users understand how it was working.

After performing a query, users wanted to retrieve the documents and persons found. They could do this by exploring the two views, but two users said it was not optimal. We, therefore, added person and document tables displaying the entities found in a query, enabling an easier exploration of the results. We also added the persistent selection of nodes across the two views and the tables, as we observed users had trouble finding persons and documents they selected on the network views in the tables. At the final meetings, the three users were able to perform exploration, queries, and comparisons.

6.1. Feedback

All three users liked the table views and were exploiting them to study in depth who were the person and documents found in their specific queries. Both historians liked the Sankey diagram of the attributes, allowing them to see the evolution of distributions and answering several of their questions. Our collaborator #2 was making sense of it by linking the migration patterns she was seeing in the Sankey diagram with specific persons of the dataset she knew in depth. She was also curious about other migration patterns she was not aware of and wanted to know who these persons were, the system allowing her to select them and explore deeper.

7. Discussion

Visualization Layouts and Encoding. Many visualization techniques exist to display multivariate and multilayers networks, which have been studied in various surveys [KPW14, MGM*19, NMSL19, HSS15]. In the current version, ComBiNet leverages a node-link diagram and a geographical map to explore the structural, attribute, and geolocated aspects of the data. Users can compare different time windows of the data and detect time-dependent patterns using the visual query system. The time aspect could, in the future, be shown with additional layouts [SRG*19, HMO22] or with the help of animation [BPF14]. Attributes (such as time) can also be encoded with color on the two visualizations. Since colors are also used to encode the link types by the query comparison, the number of colors could limit the system. Yet, none of the users complained.

Query Expressiveness. The visual query system currently allows finding occurrences of attributed subgraphs, with potential union operations on constraints (links and node attribute values can be set at one value or as a set of values). Being able to express attribute constraints (other than for labels and ids) and unions is new compared to other visual graph query systems. More complex constraints are then expressible using the Cypher editor, such as dependent constraints, e.g., if one node attribute value has to be greater or lower than another attribute value. The visual query system could be extended by introducing more complex time constraint capabilities, such as in [MWP*12].

Scalability. We assess the scalability in network size (number of nodes and links) concerning the cluttering and readability of the network visualizations. Our biggest dataset from #3 comprises 7212 nodes (4886 persons and 2326 events) and 7790 links after splitting the documents into birth and marriage event nodes. The system allows the exploration of networks of this size with a decent frame rate. ComBiNet allows navigating relatively large sparse graphs (thousands of nodes) with the node-link visualization using zoom & pan and filtering with the query system. It lets users focus on subsets of the data, one or two at a time. Comparing subgraphs two by two allowed historians to answer their questions in the presented case studies, but a more complex comparison of a larger number of subgraphs might be useful in other contexts.

Generalizibility. The system has been designed specifically for bipartite multivariate dynamic networks, which models well a diversity of historical sources we encountered via our collaborations: marriage acts, birth/death certificates, construction/work contracts,
census, and migrations forms. Moreover, bipartite multivariate dynamic networks can also be used to model other similar data types, such as scientific publications or thesis data. However, other kinds of historical textual data exist where documents can mention each other, such as in private letters for example. The model and interface would need to be slightly modified to take into account document-to-document links for these datasets. Bipartite networks are also used in various other disciplines, such as biology [KHT09]. ComBiNet could be extended to these other application domains, in particular by modifying the map view to show other location data related to the entities of the network or removing it altogether if it makes no sense for a particular domain.

8. Conclusion and Future Work

We presented ComBiNet, a system for exploring social networks modeled from historical textual sources aimed at social scientists. It relies on modeling data as bipartite, multivariate, dynamic social networks where persons are linked to documents or events using typed links that express roles. ComBiNet relies on this data model to let historians explore their data and then answer their socio-historical questions using 1) dynamic queries on the network structure and attributes to highlight groups of interests and 2) visual comparisons to contrast selected groups according to their structure, time, location, or any other attribute. The results can be visualized as a node-link diagram, a map, graph measures, and distributions of values for the attributes. We have shown that complex explorations and analyses were easy or possible to perform and validated our approach by first describing two use cases among many more projects we are collaborating with, and by performing a formative usability study showing that the system is usable by social scientists. Our collaborations with social scientists revealed that the process of modeling historical sources into an analyzable network is not easy. Modeling their data as bipartite multivariate dynamic networks allows social scientists to have a common representation to both clean up and analyze their data. By specifying a unifying data model and novel high-level visual and interactive tools for comparing topology, attributes, and time, we believe ComBiNet leads the way towards a new generation of highly interactive exploration tools applicable to wrangle and analyze a wide variety of real social networks modeled from textual sources.

References

[AWW09] ANDREWS K., WOHLFAHRT M., WURZINGER G.: Visual Graph Comparison. In *International Conference Information Visualisation* (July 2009), pp. 62–67. doi:10.1109/IV.2009.108. 4

[BBDW14] BECK F., BURCH M., DIEHL S., WEISKOPF D.: The State of the Art in Visualizing Dynamic Graphs. In *EuroVis - STARs* (2014), Borgo R., Maciejewski R., Viola I., (Eds.), The Eurographics Association. doi:10.2312/eurovisstar.20141174. 3

[BDF*10] BEZERIANOS A., DRAGICEVIC P., FEKETE J., BAE J., WATSON B.: GeneaQuilts: A System for Exploring Large Genealogies. *IEEE Trans. Vis. Comput. Graph. 16*, 6 (2010), 1073–1081. doi:10.1109/TVCG.2010.159. 3

[BEF02] BORGATTI S., EVERETT M. G., FREEMAN L. C.: UCINET 6 for Windows: Software for Social Network Analysis. Harvard, MA, Analytic Technologies, 2002. 3

[BH19] BATTLE L., HEER J.: Characterizing Exploratory Visual Analysis: A Literature Review and Evaluation of Analytic Provenance in Tableau. *Computer Graphics Forum 38*, 3 (2019), 145–159. doi:10.1111/cgf.13678. 4

[BHJ09] BASTIAN M., HEYMANN S., JACOMY M.: Gephi: An Open Source Software for Exploring and Manipulating Networks. In *ICWSM* (2009), Adar E., Hurst M., Finin T., Glance N. S., Nicolov N., Tseng B. L., (Eds.), The AAAI Press. URL: http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154. 2, 3, 6

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3 Data-Driven Documents. *IEEE Trans. Vis. Comput. Graphics 17*, 12 (Dec. 2011), 2301–2309. doi:10.1109/TVCG.2011.185. 6, 9

[BPF14] BACH B., PIETRIGA E., FEKETE J.-D.: GraphDiaries: Animated Transitions andTemporal Navigation for Dynamic Networks. *IEEE Trans. Vis. Comput. Graphics 20*, 5 (May 2014), 740–754. URL: http://ieeexplore.ieee.org/document/6658746/, doi:10.1109/TVCG.2013.254. 11

[BWD*19] BORS C., WENSKOVITCH J., DOWLING M., ATTFIELD S., BATTLE L., ENDERT A., KULYK O., LARAMEE R. S.: A Provenance Task Abstraction Framework. *IEEE Comput. Graph. Appl. Mag 39*, 6 (Nov. 2019), 46–60. doi:10.1109/MCG.2019.2945720. 4

[CFS*06] CALLAHAN S. P., FREIRE J., SANTOS E., SCHEIDEGGER C. E., SILVA C. T., VO H. T.: VisTrails: Visualization meets data management. In *SIGMOD '06* (Chicago, IL, USA, 2006), ACM, p. 745. doi:10.1145/1142473.1142574. 4

[CFT*08] CHAU D. H., FALOUTSOS C., TONG H., HONG J. I., GALLAGHER B., ELIASSI-RAD T.: GRAPHITE: A Visual Query System for Large Graphs. In *2008 IEEE International Conference on Data Mining Workshops* (Dec. 2008), pp. 963–966. doi:10.1109/ICDMW.2008.99. 3

[CGL20] CUTLER Z., GADHAVE K., LEX A.: Trrack: A Library for Provenance-Tracking in Web-Based Visualizations. In *2020 IEEE Visualization Conference (VIS)* (2020), pp. 116–120. doi:10.1109/VIS47514.2020.00030. 8, 9

[CN06] CSARDI G., NEPUSZ T.: The igraph software package for complex network research. *InterJournal Complex Systems* (2006), 1695. URL: https://igraph.org. 3, 6

[CR18] CRISTOFOLI P., ROLLA N.: Temporalités à l'œuvre dans les chantiers du bâtiment. *Temporalités 1*, 27 (2018). doi:10.4000/temporalites.4456. 1, 4

[Cri08] CRISTOFOLI P.: Aux sources des grands réseaux d'interactions. *Reseaux 152*, 6 (2008), 21–58. URL: https://www.cairn.info/revue-reseaux1-2008-6-page-21.htm. 2

[CSIP21] CUENCA E., SALLABERRY A., IENCO D., PONCELET P.: VERTIGo: a Visual Platform for Querying and Exploring Large Multilayer Networks. *IEEE Trans. Vis. Comput. Graphics* (2021), 1–1. doi:10.1109/TVCG.2021.3067820. 3

[DHRL*12] DUNNE C., HENRY RICHE N., LEE B., METOYER R., ROBERTSON G.: GraphTrail: Analyzing Large Multivariate, Heterogeneous Networks While Supporting Exploration History. In *CHI '12* (New York, NY, USA, May 2012), CHI '12, ACM, pp. 1663–1672. doi:10.1145/2207676.2208293. 4

[Dim20] DIMINESCU D.: 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, Mar. 2020). URL: https://hal.telecom-paris.fr/hal-02556007. 5

[DPF16] DANG T. N., PENDAR N., FORBES A. G.: TimeArcs: Visualizing Fluctuations in Dynamic Networks. *Computer Graphics Forum 35*, 3 (June 2016), 61–69. doi:10.1111/cgf.12882. 3

[FFNS19] FEKETE J.-D., FISHER D., NANDI A., SEDLMAIR M.: Progressive Data Analysis and Visualization (Dagstuhl Seminar 18411). *Dagstuhl Reports 8*, 10 (2019), 1–40. doi:10.4230/DagRep.8.10.1. 8

[FPSG10] FREIRE M., PLAISANT C., SHNEIDERMAN B., GOLBECK J.: ManyNets: An Interface for Multiple Network Analysis and Visualization. In *CHI '10* (New York, NY, USA, 2010), CHI '10, ACM, p. 213–222. doi:10.1145/1753326.1753358. 4

[Fre04] FREEMAN L. C.: *The Development of Social Network Analysis: A Study in the Sociology of Science.* Empirical Press, 2004. URL: https://books.google.fr/books?id=VcxqQgAACAAJ. 3

[GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization 10*, 4 (Oct. 2011), 289–309. doi:10.1177/1473871611416549. 4

[GED] GEDCOM: The Genealogy Data Standard. https://www.gedcom.org/. Last Access: July 22, 2022. 3

[GLS14] GÖRG C., LIU Z., STASKO J. T.: Reflections on the evolution of the jigsaw visual analytics system. *Inf. Vis. 13*, 4 (2014), 336–345. doi:10.1177/1473871613495674. 3

[Gra17] GRANDJEAN M.: Analisi e visualizzazioni delle reti in storia. L'esempio della cooperazione intellettuale della Società delle Nazioni. *Memoria e Ricerca, Rivista di storia contemporanea*, 2 (2017), 371–393. doi:10.14647/87204. 2

[GS10] GLASER B. G., STRAUSS A. L.: *The Discovery of Grounded Theory: Strategies for Qualitative Research*, 5. paperback print ed. Aldine Transaction, New Brunswick, 2010. 2

[HBC*21] HOGAN A., BLOMQVIST E., COCHEZ M., D'AMATO C., MELO G. D., GUTIERREZ C., KIRRANE S., GAYO J. E. L., NAVIGLI R., NEUMAIER S., NGOMO A.-C. N., POLLERES A., RASHID S. M., RULA A., SCHMELZEISEN L., SEQUEDA J., STAAB S., ZIMMERMANN A.: Knowledge graphs. *ACM Comput. Surv. 54*, 4 (jul 2021). URL: https://doi.org/10.1145/3447772, doi:10.1145/3447772. 3

[HD12] HASCOËT M., DRAGICEVIC P.: Interactive graph matching and visual comparison of graphs and clustered graphs. In *AVI '12* (2012), Tortora G., Levialdi S., Tucci M., (Eds.), ACM, pp. 522–529. doi:10.1145/2254556.2254654. 4

[HHW11] HAMBERGER K., HOUSEMAN M., WHITE DOUGLAS R.: Kinship Network Analysis. In *The Sage Handbook of Social Network Analysis*, Carrington J. S. . P. J., (Ed.). Sage Publications, 2011, pp. 533–549. URL: https://halshs.archives-ouvertes.fr/halshs-00658667. 3

[HMO22] HANDLER A., MAHYAR N., O'CONNOR B.: Clioquery: Interactive query-oriented text analytics for comprehensive investigation of historical news archives. *ACM Transactions on Interactive Intelligent Systems (TiiS)* (2022). 11

[HMSM] HAMBERGER K., MOMON C., SAVOIA E., MENEZES T.: PUCK: Program for the Use and Computation of Kinship data. https://www.kintip.net/. Last Access: July 22, 2022. 3

[HSS08] HAGBERG A. A., SCHULT D. A., SWART P. J.: Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conference* (Pasadena, CA USA, 2008), Varoquaux G., Vaught T., Millman J., (Eds.), pp. 11 – 15. 3

[HSS15] HADLAK S., SCHUMANN H., SCHULZ H.-J.: A Survey of Multi-faceted Graph Visualization. *Eurographics Conference on Visualization (EuroVis) - STARs* (2015), 20 pages. Artwork Size: 20 pages ISBN: 9783905674996 Publisher: The Eurographics Association. URL: https://diglib.eg.org/handle/10.2312/eurovisstar.20151109.001-020, doi:10.2312/EUROVISSTAR.20151109. 11

[KHT09] KLAMT S., HAUS U.-U., THEIS F.: Hypergraphs and Cellular Networks. *PLoS computational biology 5*, 5 (2009), e1000385. 11

[KPW14] KERREN A., PURCHASE H., WARD M. O. (Eds.): *Multivariate Network Visualization: Dagstuhl Seminar # 13201, Dagstuhl Castle, Germany, May 12-17, 2013, Revised Discussions*, 1st ed. 2014 ed. No. 8380 in Information Systems and Applications, incl. Internet/Web, and HCI. Springer International Publishing : Imprint: Springer, Cham, 2014. doi:10.1007/978-3-319-06793-3. 11

[MB16] MRVAR A., BATAGELJ V.: Analysis and visualization of large networks with program package pajek. *Complex Adaptive Systems Modeling 4*, 1 (Apr. 2016). doi:10.1186/s40294-016-0017-8. 3, 6

[MGM*19] MCGEE F., GHONIEM M., MELANÇON G., OTJACQUES B., PINAUD B.: The State of the Art in Multilayer Network Visualization. *Computer Graphics Forum 38*, 6 (Sept. 2019), 125–149. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13610, doi:10.1111/cgf.13610. 11

[Mou16] MOUTOUKIAS Z.: Buenos aires, port between two oceans: mobilities, networks, stratifications (2nd half of the 18th century). *E-Spania—Revue Électronique d'Études Hispaniques Médievales 25* (2016). 5

[MWP*12] MONROE M., WONGSUPHASAWAT K., PLAISANT C., SHNEIDERMAN B., MILLSTEIN J., GOLD S.: Exploring point and interval event patterns: Display methods and interactive visual query. *University of Maryland Technical Report* (2012). 11

[Nat] Natural Earth. https://www.naturalearthdata.com/. Last Access: July 22, 2022. 6

[Neo] Neo4j graph data platform. https://neo4j.com/. Last Access: July 22, 2022. 2, 3

[NMSL19] NOBRE C., MEYER M., STREIT M., LEX A.: The State of the Art in Visualizing Multivariate Networks. *Computer Graphics Forum 38*, 3 (June 2019), 807–832. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13728, doi:10.1111/cgf.13728. 11

[PDC*22] PISTER A., DUFOURNAUD N., CRISTOFOLI P., PRIEUR C., FEKETE J.-D.: From Historical Documents To Social Network Visualization: Potential Pitfalls and Network Modeling. In *VIS4DH 2022 - 7th Workshop on Visualization for the Digital Humanities* (Oklahoma City, United States, Oct. 2022). URL: https://hal.inria.fr/hal-03784532. 2

[PHE*18] PIENTA R., HOHMAN F., ENDERT A., TAMERSOY A., ROUNDY K., GATES C., NAVATHE S., CHAU D. H.: VIGOR: Interactive Visual Exploration of Graph Query Results. *IEEE Trans. Vis. Comput. Graphics 24*, 1 (Jan. 2018), 215–225. doi:10.1109/TVCG.2017.2744898. 3

[PQ95] PARR T. J., QUONG R. W.: ANTLR: A predicated-LL (k) parser generator. *Software: Practice and Experience 25*, 7 (1995), 789–810. 9

[R C14] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL: http://www.R-project.org/. 3

[RASC16] RAGAN E., ALEX E., SANYAL J., CHEN J.: Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Trans. Vis. Comput. Graphics 22*, 1 (Jan. 2016). doi:10.1109/TVCG.2015.2467551. 4

[RLB18] REN D., LEE B., BREHMER M.: Charticulator: Interactive Construction of Bespoke Chart Layouts. *IEEE Trans. Vis. Comput. Graphics 25*, 1 (2018), 789–799. 4

[RVH14] ROSSI F., VIALANEIX N., HAUTEFEUILLE F.: Exploration of a large database of French notarial acts with social network methods. *Digital Medievalist 9* (July 2014), 2013. URL: https://hal.archives-ouvertes.fr/hal-01053673. 2

[SGL08] STASKO J. T., GÖRG C., LIU Z.: Jigsaw: Supporting Investigative Analysis through Interactive Visualization. *Inf. Vis. 7*, 2 (2008), 118–132. doi:10.1057/palgrave.ivs.9500180. 2

[Shn94] SHNEIDERMAN B.: Dynamic Queries for Visual Information Seeking. *IEEE Softw. 11*, 6 (Nov. 1994), 70–77. doi:10.1109/52.329404. 7

[SMBP*17] SERRANO MOLINERO V., BACH B., PLAISANT C., DUFOURNAUD N., FEKETE J.-D.: Understanding the Use of The Vistorian: Complementing Logs with Context Mini-Questionnaires. In *Visualization for the Digital Humanities* (Phoenix, United States, Oct. 2017). URL: https://hal.inria.fr/hal-01650259. 3

[SMWH16] SATYANARAYAN A., MORITZ D., WONGSUPHASAWAT K., HEER J.: Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. Vis. Comput. Graphics 23*, 1 (2016), 341–350. 9

[SRG*19] STOIBER C., RIND A., GRASSINGER F., GUTOUNIG R., GOLDGRUBER E., SEDLMAIR M., EMRICH B., AIGNER W.: *netflower:* Dynamic Network Visualization for Data Journalists. *Computer Graphics Forum 38*, 3 (June 2019), 699–711. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13721, doi:10.1111/cgf.13721. 11

[SSM*09] SMITH M. A., SHNEIDERMAN B., MILIC-FRAYLING N., RODRIGUES E. M., BARASH V., DUNNE C., CAPONE T., PERER A., GLEAVE E.: Analyzing (social media) networks with NodeXL. In *Proceedings of the Fourth International Conference on Communities and Technologies, C&T 2009, University Park, PA, USA, June 25-27, 2009* (2009), Carroll J. M., (Ed.), ACM, pp. 255–264. doi:10.1145/1556460.1556497. 2, 3

[SSM*17] SHAFIE T., SCHOCH D., MANS J., HOFMAN C., BRANDES U.: Hypergraph Representations: A Study of Carib Attacks on Colonial Forces, 1509-1700. *Journal of Historical Network Research* (Oct. 2017), 52–70 Pages. doi:10.25517/JHNR.V1I1.6. 2

[SW13] SHADOAN R., WEAVER C.: Visual Analysis of Higher-Order Conjunctive Relationships in Multidimensional Data Using a Hypergraph Query System. *IEEE Trans. Vis. Comput. Graphics 19*, 12 (Dec. 2013), 2070–2079. doi:10.1109/TVCG.2013.220. 3

[TPR*21] TOVANICH N., PISTER A., RICHER G., VALDIVIA P., PRIEUR C., FEKETE J.-D., ISENBERG P.: VAST 2020 Contest Challenge: GraphMatchMaker: Visual Analytics for Graph Comparison and Matching. *IEEE Comput. Graph. Appl. Mag* (2021), 1–1. doi:10.1109/MCG.2021.3091955. 4

[VBP*21] VALDIVIA P. R., BUONO P., PLAISANT C., DUFOURNAUD N., FEKETE J.-D.: Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. *IEEE Trans. Vis. Comput. Graphics 27*, 1 (Jan. 2021), 1–13. doi:10.1109/TVCG.2019.2933196. 3

[vR95] VAN ROSSUM G.: *Python tutorial*. Tech. Rep. CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995. 3

[Wet98] WETHERELL C.: Historical Social Network Analysis. *International Review of Social History 43*, S6 (Dec. 1998), 125–144. doi:10.1017/S0020859000115123. 2

[XOW*20] XU K., OTTLEY A., WALCHSHOFER C., STREIT M., CHANG R., WENSKOVITCH J.: Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum 39*, 3 (June 2020), 757–783. doi:10.1111/cgf.14035. 4

[ZTC*22] ZHANG W., TAN S., CHEN S., MENG L., ZHANG T., ZHU R., CHEN W.: Visual Reasoning for Uncertainty in Spatio-temporal Events of Historical Figures. *IEEE Trans. Vis. Comput. Graphics* (2022), 1–1. URL: https://ieeexplore.ieee.org/document/9695348/, doi:10.1109/TVCG.2022.3146508. 3

[ZWW*22] ZHANG W., WONG J. K., WANG X., GONG Y., ZHU R., LIU K., YAN Z., TAN S., QU H., CHEN S., CHEN W.: CohortVA: A Visual Analytic System for Interactive Exploration of Cohorts Based on Historical Data. *IEEE Trans. Vis. Comput. Graphics* (2022), 1–11. URL: https://ieeexplore.ieee.org/document/9912359/, doi:10.1109/TVCG.2022.3209483. 3