Visual Queries on Bipartite Multivariate Dynamic Social Networks

A. Pister[1,2] C. Prieur[3,2], JD. Fekete[1]

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

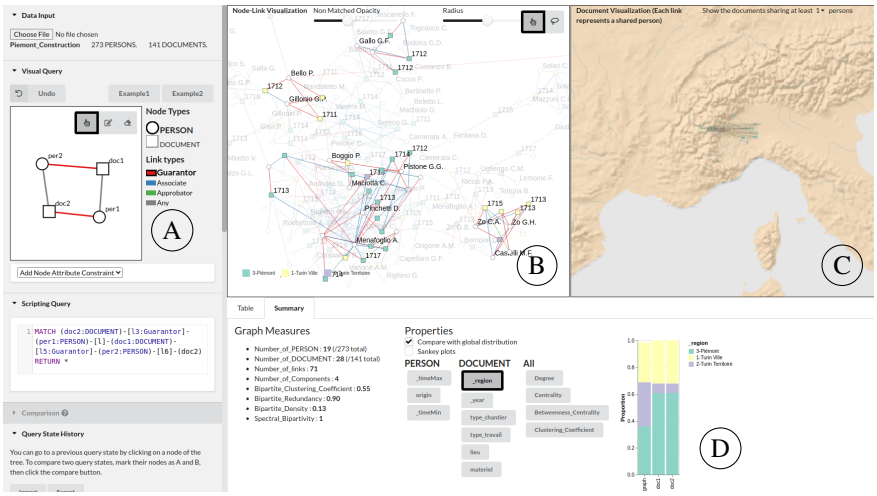

Figure 1: *Proposed interface, used to search for people who are mutually guarantors to each other in construction contracts [CR18], retrieved from dynamic visual queries. (A) shows the visual query the user created. (B) shows the bipartite node-link view and (C) the document map view. (D) shows the query results panel where one attribute is selected, showing its distribution.*

Abstract

We present a new system for the visualization, visual query, and comparison of social networks modeled as bipartite multivariate dynamic graphs. Historians and sociologists study collections of documents —such as marriage acts, birth certificates, or contracts—from a region and time of interest and can transform them into Bipartite Multivariate Dynamic Social Networks to follow an in-depth analysis. However, few social network analysis tools are designed for this type of network, and existing ones rarely provide enough interactions to answer complex sociological questions. We present a new visual analytics system allowing to explore this type of network. In particular, we designed a visual query system specifically for this data model, allowing social scientists to easily apply filters on the topology and attributes of a network. We demonstrate how the interface coupled with the query system can be used to answer sociological questions with one real-world use case based on construction contracts.

1. Introduction

Historians and sociologists aim at characterizing the social relationships between people of a period and region of interest. For this, they rely on a corpus of documents—which can be birth certificates, marriage acts, work contracts, or migration forms—which mention people and their social ties. From this corpus, they construct social networks on which they can follow a Social Network Analysis (SNA) [Fre04]. This type of dataset is best modeled using Bipartite Multivariate Dynamic Social Networks, where both documents and persons are modeled as nodes, containing additional information extracted from the original sources, such as the document time and location [Bor09]. A link represent a mention of a person in a document and encode the *role* of the person in the document (husband, wife, or witness in a marriage act for example). However, most visual analytics tools for social network either enforce a simple person-to-person model [BHJ09] or no model at all [SGL08], and lack complex interactions needed by social sci-

DOI: 10.2312/evp.20221115

entists. We thus present a new visual analytics system to explore Bipartite Multivariate Dynamic Social Networks, based on a visual query system that allows social scientists to easily filter and characterize specific groups of persons and patterns of interests.

2. Visual Analytic System

Our proposed tool allows exploring the data with two coordinated views shown in Figure 1: the node-link view and the map view. The node-link view allows users to explore the topology of the graph, while the map allows seeing the geolocation of the documents. Two documents are linked if they share at least one person in common. This way, the map allows to rapidly see in which areas people are interacting. The two views are coordinated and selecting a person or a document in the bipartite view highlights her/him in the map view and vice versa. Moreover, every attribute in the data can be selected to show a plot of its distribution. When selected, the nodes are colored in the two views according their attribute value.

3. Visual Query

Our visual query system is similar to VERTIGo [CSIP21] and [SW13] for topological queries, but enforces the bipartite network model and allows more complex attribute constraints. It is based on the Neo4j database [neo] and the Cypher language. In contrast to other visual query systems that hide the raw script request, we show both representations (visual and textual) and enforce a one-to-one synchronisation between the two. This way, users can start to build a query visually, and modify it later textually for very complex constraints. Users can build two types of constraints:

Topological: An interactive node-link diagram let users construct patterns visually. They can build documents or person nodes by clicking on the canvas, and links by clicking and dragging between nodes. They can choose any of the roles of the dataset when creating a link, along an *ANY* link type, representing a link which can be any of the existing role.

Attributes: Widgets let create attribute constraints on the created nodes, using the dynamic queries principle [Shn94]. The widgets designs differ according to the attribute type: text input for nominal attributes, checkboxes for categorical and double sliders for numerical ones. These simple designs allow social scientists to easily make attribute constraints.

When modifying the query visually or textually, the results are directly highlighted in the two views, and summary results are shown along an occurrence table and attribute distribution plots.

4. Use Cases

We elaborated our system in collaboration with historians and sociologists who have a variety of datasets: marriage acts, birth certificates, construction contracts, work contracts and migration forms. The questions they had on their data often consisted in extracting groups of people or finding document and persons with specific patterns, and studying their characteristics in respect of the rest of the network, or to compare two groups of nodes between them (see supplementary material).

We describe here a use case in collaboration with one historian

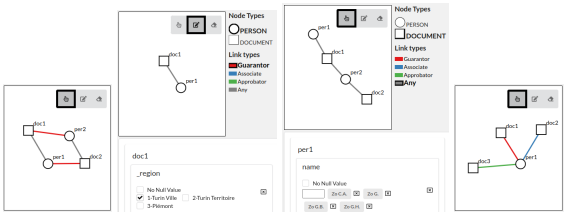

Figure 2: *Queries to answer our collaborator's four questions.*

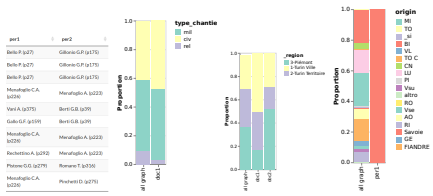

Figure 3: *Summary results of the four queries: Table view of the occurrences for question 1 (left). Distributions of the attributes for the questions 2, 3, and 4.*

who wants to analyze a dataset made of construction documents in Piemont area in Italy in the 18th century [CR18]. The dataset is made of 141 documents and 272 persons. People can be mentioned in a document under 3 roles: *Associates*, *Guarantors* and *Approbators*. Here are some of the questions he had that we collected from him, which can be answer with queries:

1. Find the pairs or persons who are mutually guarantor to each other in two separate construction documents.
2. Find the persons who are collaborating with the Zo family.
3. What are the characteristics of the construction documents in Torino, and who are the people working there.
4. Who are the persons with 3 different roles: *Guarantor*, *Associate* and *Approbator*.

Figure 2 shows the visual representations of the constructed queries to find the patterns one of our collaborator is interested in. After he constructed the queries, he was able to explore the results in the two views, as shown in Figure 1. Figure 3 shows the results of the four queries with an occurrences table (left) or attribute distributions (right). It shows that the constructions in Torino are mostly military (*mil*) and civil (*civ*). The Zo family collaborates with other persons mostly in Torino (but not only). Finally, the persons who have the 3 different roles only come from Biela (*BI*).

5. Conclusion

We present a new visual analytics for Bipartite Multivariate Dynamic Social Networks constructed from historical sources. It allows to explore the topology of the network and the localization of the documents, along node attribute distributions, including time. Moreover, the interface includes a new visual query system. Using it, social scientists can isolate specific patterns or groups of entities and look at their positions in the network, along with the values of their attributes compared to the global distributions. We demonstrate the power of the query system with one real-world use case, showing that it can be used to answer sociological questions.

References

[BHJ09] BASTIAN M., HEYMANN S., JACOMY M.: Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009* (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. 1

[Bor09] BORGATTI S.: Social Network Analysis, Two-Mode Concepts in. *Computational Complexity: Theory, Techniques, and Applications* (Jan. 2009). doi:10.1007/978-0-387-30440-3_491. 1

[CR18] CRISTOFOLI P., ROLLA N.: Temporalités à l'œuvre dans les chantiers du bâtiment. *Temporalités*, 27 (jun 2018). doi:10.4000/temporalites.4456. 1, 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. Graph.* (2021), 1–1. doi:10.1109/TVCG.2021.3067820. 2

[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. 1

[neo] Neo4j graph data platform. Last access: Nov. 18, 2021. URL: https://neo4j.com/. 2

[SGL08] STASKO J., GÖRG C., LIU Z.: Jigsaw: supporting investigative analysis, 2008. 1

[Shn94] SHNEIDERMAN B.: Dynamic Queries for Visual Information Seeking. *IEEE Softw. 11*, 6 (Nov. 1994), 70–77. doi:10.1109/52.329404. 2

[SW13] SHADOAN R., WEAVER C.: Visual Analysis of Higher-Order Conjunctive Relationships in Multidimensional Data Using a Hypergraph Query System. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2070–2079. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi:10.1109/TVCG.2013.220. 2