Using Dynamic Hypergraphs to Reveal the Evolution of the Business Network of a 17th Century French Woman Merchant

Paola Valdivia, Paolo Buono, Catherine Plaisant, Nicole Dufournaud, Jean-Daniel Fekete *Senior Member, IEEE*

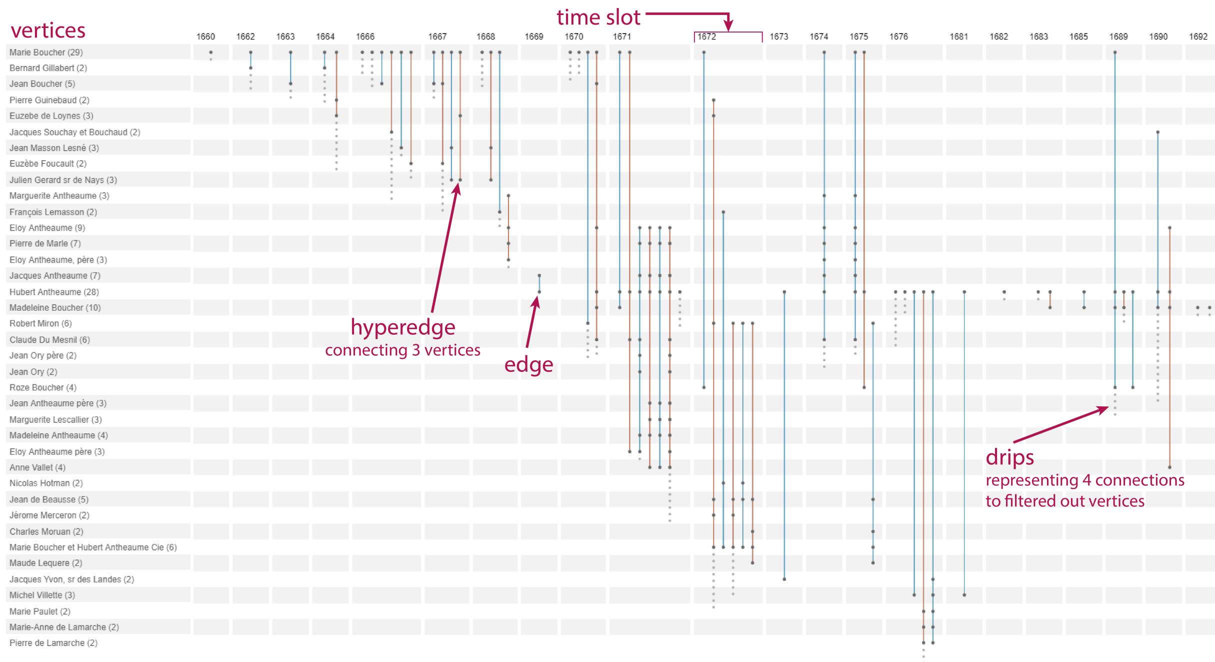

Figure 1: A Parallel Aggregated Ordered Hypergraph visualization extracted from 59 legal documents. Time runs from left to right with discrete time slots representing the network at that time. Each parallel vertical line is an hyperedge, connecting two or more people. People are represented by parallel horizontal bars, with all names aligned on the left. Dots mark connections. In this view the people that appear in only one document have been hidden from the list of people and their existence is only hinted with drips, i.e. smaller gray dots at the lower end of hyperedges.

Abstract

Many digital humanity use cases require the analysis of relationships between entities (e.g. people or countries). The analysis of those relationships is particularly difficult when these relations change over time. A common representations of such relations is through graphs, which connect pairs of entities. However, in the real world relationships are often more complex and can be better described using hypergraphs (where edges can connect more than two entities). In this paper, we present a digital humanity case study of the analysis of people mentioned in 16th and 17th Century legal documents, modeled as a dynamic hypergraph. We use a new representation called Parallel Aggregated Ordered Hypergraph. Our prototype implementation of Parallel Aggregated Ordered Hypergraph, and the benefits of the PAOH representation are discussed.

Keywords: dynamic graph, hypergraph, history, visual analytics, case study

1 Introduction

In many social science and history studies people are mentioned in dated documents, such as contracts, diary entries, and justice decisions [5, 10, 12, 19, 24]. The acquisition and curation process may be very tedious: documents have to be discovered in archives, transcribed and annotated (often manually because transcription and entity extraction algorithms do not work well with ancient documents and languages). Datasets are then carefully analyzed and studied in great depth and detail by humanists over long periods of time, as the research progresses. As a result, those studies generate moderately large datasets containing a few hundred people, with rarely more than a dozen people connected in each document.

In this paper, we focus on a case study of an historian studying a collection of historical documents describing business agreements (contracts) between different people over the years [8]. Each contract involves two or more people. The historian needs to understand the business network and how it evolved over time.

Using classic network analysis techniques found in the literature, the network is represented as a graph, and a contract between three entities is represented as three edges, making it unclear whether they corresponded to a single contract, or two or three individual contracts. The network overall growth - or decrease - over time is fairly easy to see, but as time passes by, it can be difficult to tell if connections with old business partners remain when new business partners come into play.

In order to more truthfully represent the network, the data can be modeled as a dynamic hypergraph. In an hypergraph a single hyperedge can involve two or more people, in contrast with traditional graphs where a relation would strictly be between two people. Such need exists also in other contexts, such as co-authorship networks, where a paper can be written by multiple authors; alliances between countries, where multiple countries could be part of the same alliance.

This paper describes the initial case study of our Parallel Aggregated Ordered Hypergraph (PAOH) visualization, an extension of the poster [21], which is to our knowledge is the first interactive technique for dynamic hypergraphs. PAOH shows data aggregated on a series of time intervals, called time slots. In Figure 1, the time slots correspond to years, starting with 1660. The vertices of the hypergraph (here, people) are encoded as parallel horizontal bars and edges (here, documents) are encoded as parallel vertical lines connecting the vertices. A dot marks the connection between documents and people.

We start by providing background information on the case study in Section 2, then discuss hypergraph visualization and related work in section 3, describe the technique in more details in section 4, and conclude in section 5.

2 TRADE NETWORK IN THE 17^{th} CENTURY

We worked closely with a professional historian studying the role and power of a non-married women, Marie Boucher, merchant living in the 17th century in Nantes, France [8]. The focus of the study was to understand Marie Boucher's life strategy, and an important part of the historian's task was to understand the complex trade relations that Boucher had based on the contracts discovered in multiple archives.

The dataset gathered by Dufournaud et al, is composed of 59 commercial contracts related to the life of Marie Boucher. The contracts mention 90 persons overall. One contract, characterized by a date, can be modeled by one hyperedge, mentioning several persons. Modeling this database as a graph results in 488 edges, one for each relation between a pair of persons. Using hyperedges it requires only 59 hyperedges.

This dataset had already been studied by Dufournaud et al. using traditional graphs [8]. We first quickly describe the tools they originally used, then describe how PAOH differs and can improve the analysis.

The initial graph analysis required two separate representations: a node-link diagram and a TimeArc representation. The node-link diagram of Figure 2-left shows the bipartite graph that had been used to reveal the connections between documents and persons, but it does not show any temporal information. The squares represent documents, while circles represent people. The size of a vertex reveal the cardinality of entities (people or documents). Figure 2-right shows the TimeArc diagram that was used to reveal how connections between persons changed over time. Unfortunately here the documents connecting people are not visible. The color of arcs can indicate the type of document but it does not disambiguate between two contracts of the same type between two of the three people or a single contract among the same three people.

To understand the evolution of Marie Boucher's business connections required switching back and forth between the two representations. In Figure 2-left we see that Mary is connected with Hubert Antheaume. The existence of the contracts is visible, the bipartite node-link does not show when such contracts were signed, or that the relationship between Marie and Hubert started around 1670. To know such information the historian must look to the other visualization in Figure 2-right.

In contrast, the new PAOH representation can show all the information in a single display (see Figure1). Each hyperedge is a document. We can easily see that Marie's connection to Hubert started in 1670 and that she had seven contracts with him. Figure 1 reveals that Hubert is as active in commerce as Marie, he had a contract one year before starting a business with her and had his own contacts network. Hubert signed seven contracts with Marie and with other people, in particular, in 1674 and 1675 they signed two contract with mostly the same people except for six different persons who appear only once, so they are represented as ”drips” (see section 4.2).

Compared to the original technique used in Figure 2, the analysis with PAOH requires only one visualization and can be done more accurately. Our historian collaborator was able to understand the PAOH representation after a very brief description, and found the representation very clear and more accurate. For example reading the original second visualization, Dufournaud et al. had made the assumption that two persons connected the same year appeared in only one contract. However, looking at the PAOH representations in Figure 1, it appears that e.g. in 1667, Marie Boucher had two contracts with Jean Boucher (line 3) and two others with Julien Gérard Seigneur de Nays (line 9). With no overlap and no line crossing the display can be read even without any interaction (e.g. when printed.)

Our collaborator explained that three main phases had been identified in the lengthy prior analysis (of the graphs and the text of the documents): 1) an initial phase from 1660 to 64 with mostly French trading; 2) a second phase with cross atlantic trade from 1666 to 68, and 3) a third expansion phase until 1675 when Marie disappears from the records (until a 1689 mention of her being deceased). Those phases were more clearly apparent in the PAOH display that represent time and connections simultaneously. She noted that the PAOH representation provided a good narrative support, and would be very useful to explain and communicate the findings.

We are now working with other historians who agreed to test and provide feedback after using PAOH with their own datasets. By addressing historian needs and datasets we will validate and improve our approach more extensively.

We will now discuss hypergraph visualization and related work in section 3, before describing the design of the PAOH representation in more details.

3 HYPERGRAPH VISUALIZATION

Formally, hypergraphs can be described as follows. A *hypergraph* $G = (V,H)$ where H is a set of h edges that can connect at least two vertices. $h \in H$ is called a *hyperedge*. $h \in \mathscr{P}(V)$, where $\mathscr{P}(V)$ is the the set of all sets of V (a.k.a *power-set* of V). A dynamic hypergraph is a set of hypergraphs over discrete time steps, $G_{i\in[1,n]} = (V,H_i)$, where all the hypergraphs H_i share the same vertices.

Hypergraphs have mostly been studied by modeling them as standard networks [13, 18] or bipartite networks [14, 15]. To directly visualize hypergraphs, designers have used visualization techniques designed for sets [1] like Kelp Diagrams [7], or using variations of node-link diagrams [20] where vertices belonging to the same hyperedge are outlined or colored differently, which works for only a very small number of hyperedges. Others, visualized hyperedges as groups [23].

After developing PAOH we found an early drawing using a similar an approach in an archaelogy book in 1997 [2]. This static one-time illustration shows a small dataset composed of eight territories that are the vertices of the hypergraph. Each vertical line describe a battle connecting territories (see Figure 3). Attacking territories are shown as black dots and defending ones are white dots. If a line connects more than two territories with the same color, then they are allied in the battle. In Figure 3, in 660 A.D., Mercia was allied of West Saxon and they attacked the Isle of Wight that was allied of Britons, both defending. Simultaneous battles are also visible if two vertical lines connecting territories are visible right next to each other on the timeline. An example happened approximately around 673 A.D. Each battle is placed on a continuous horizontal timeline, using a stream metaphor instead of a series of discrete time slots as our technique does. The stream metaphor has some benefits but often leads to overlaps, which are avoided in our Parallel Aggregated Ordered Hypergraph design by aggregating in time slots.

A visually related approach, used to visualize traditional graphs,

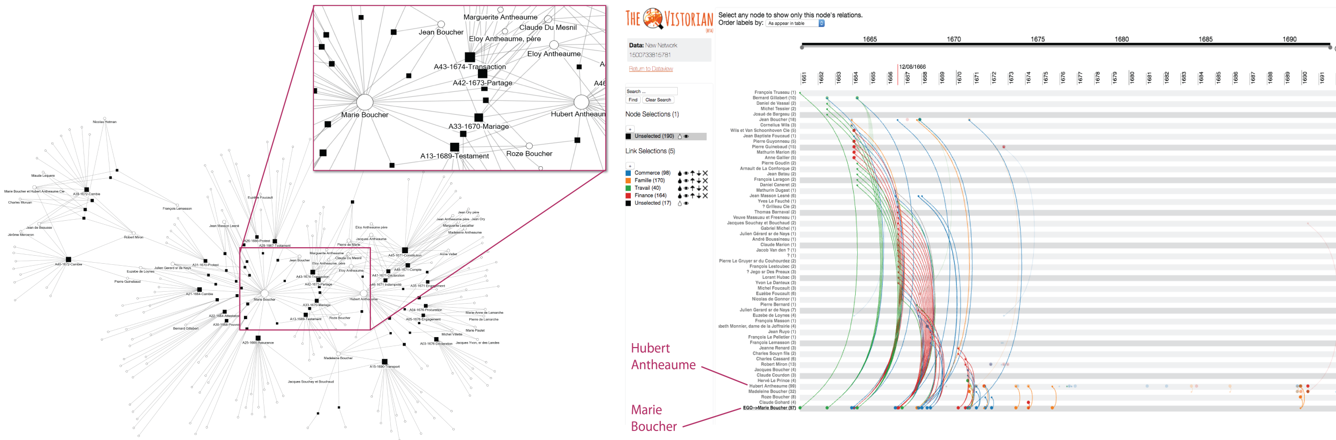

Figure 2: The two representations of the bipartite network original used by Dufournaud et al. [8] linking persons to documents over time. Left shows the static bipartite node-link diagram connecting the persons and documents in which they are mentioned, made with NodeXL [11]. Right shows the connections between the persons over time using TimeArcTrees, made with the Vistorian tool [3]. The new PAOH representation shows all that information in a single display, see Figure 1.

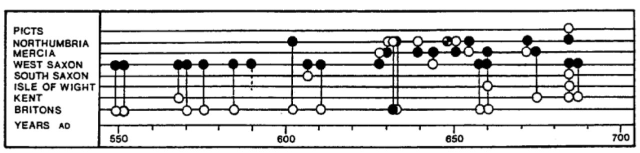

Figure 3: Arnold's graph representation as vertical connections

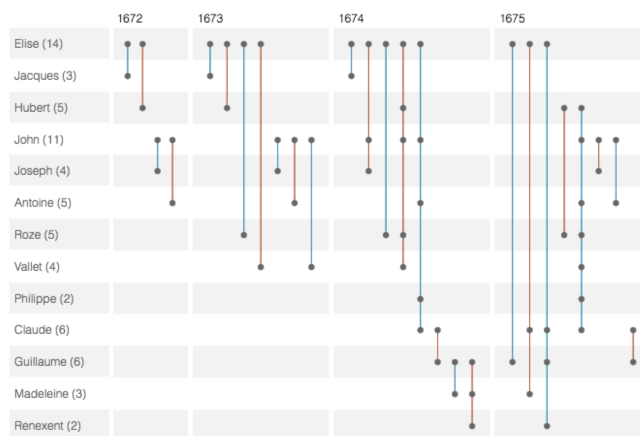

Figure 4: Visualization of a small database sample of business relationships evolving over time for a group of thirteen people using PAOH dynamic hypergraph.

is the Biofabric technique [17]. It does not consider the evolution of the network topology over time. Van den Elzen et al. [22] uses a technique similar to Biofabric and adds the evolution of time as a stream of events (but still does not model hypergraphs). Some other similar techniques, which address hypergraphs, only support a few vertex and do not consider time [6, 16]. PAOH is an extension of [21], which also display dynamic hypergraphs.

4 Parallel Aggregated Ordered Hypergraph

We now describe the Parallel Aggregated Ordered Hypergraph (PAOH) technique in more details. To have more readable figures for the paper we use a sample imaginary dataset (see Figure 4). Time flows from left to right in a series of time slots separated by small white gaps. Each time slot corresponds to an interval of time. Vertices are represented as parallel horizontal bars and hyperedges as parallel vertical lines connecting vertices, which are emphasized with dots.

The sample is composed of thirteen vertices, representing people; four time slots related to years from 1672 to 1675, and 27 hyperedges corresponding to business relationships among people during that period. Each hyperedge connects all vertices involved in the relationships; for example, in 1674 (third time slot of Figure 4) the rightmost hyperedge is a contract among Guillaume, Madeleine and Renexent.

Time representation: PAOH manages the time in different ways. According to the dataset, PAOH can aggregate data in: time points, where each hyperedge has a specific position in the timeline; time intervals, where each hyperedge belongs to a time interval; or time slots, where each hyperedge is labeled with the name of a time slot. All the connections occurring in a time slot are visible inside the horizontal bounds allocated to that time slot; this means that the width of the time slot varies to fit all its hyperedges. If multiple occurrences of an hyperedge occur within a time slot, they appear separated. For example, if two contracts have been signed among the same three people in a time slot, PAOH shows two different hyperedges. The user will distinguish the contracts through the description that appears when the user move the mouse over a hyperedge.

Entity Ordering: Entities have a position fixed vertically in the visualization. The order of the entities determines the visual appearance of the dynamic hypergraph. The position is stable over time, this means that the best position meets a global optimization and not a local one. Vertex order determines the length of the hyperedges and highly impacts the comprehension of the graphs. Different orderings help performing different tasks, for example

- Original order: Some application domains use a canonical vertex order that practitioners are trained to understand; this is the default ordering.
- Chronological order: vertices are sorted according to their chronological appearance. The names involved in the early contracts move to the top, and the names appearing later are down below.
- Degree: vertices with higher number of connections (degree)

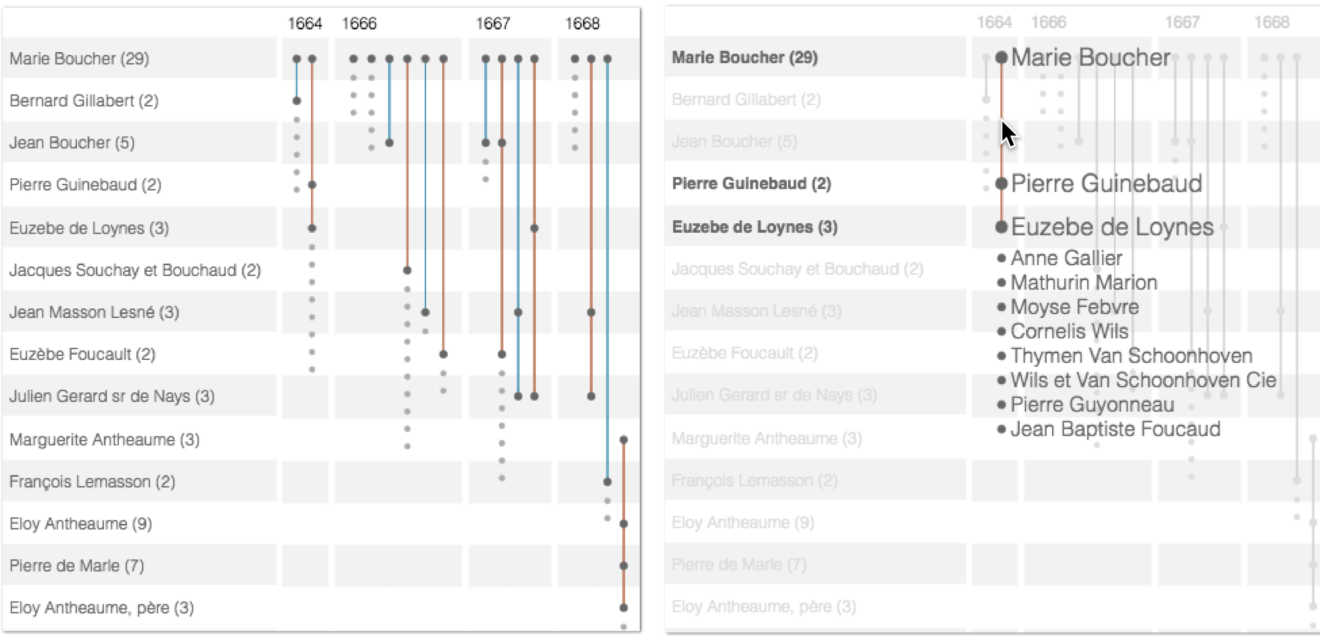

Figure 5: A subset of Marie Boucher's network. The vertices of people who signed only 1 or 2 contracts have been filtered out. Missing people are visible at the lower end of the hyperedges by a drip line of smaller gray dots. In the image on the right, the tooltip reveals all the names—using a smaller font size for the names of filtered out people.

through all the time slots appear towards the top.

- Optimal Leaf: vertices are positioned close-by, according to an adjacency similarity [4], in order to reduce the length of the edges.

Edge Ordering Edge ordering is related to the overall order of entities, but in a time slot a set of hyperedge is displayed chronologically. Within each time slot the ordering of the edges can be changed, to improve legibility. For example the edges can be ordered by length or in chronological order (which moves older connections to the left of the time slot and recent ones to right.)

4.1 Interaction

As always, interaction allows users to customize the display. Visual attributes of the PAOH representation can be modified. For example, the color of the dots or the color of the background of the horizontal lines can used to indicate what group entities belong to. The hue of lines can be make to alternate between blue and orange, which makes it a bit easier to follow long hyperedges.

Highlighting allows users to review connections, e.g. users can hover over a name to highlight, in bold, all its connections, and all the names connected to it. Hovering over a vertical line (i.e. a contract) highlights the vertex line itself and all the names involved in the contract (right of Figure 5).

To focus the analysis on a subset of the network, users can select names and then filter out all the vertices that are not connected to the selected names. For example, in Fig. 6, Joseph has been selected and the filter is applied to show only his network composed of three people.

4.2 Hypergraph Simplification

To compact the display vertices with low degrees can be aggregated [9]. For examples users can choose to remove names that appear in a single contract. They are removed from the list of names on the left, but a hint that some names are missing in the hyperedges is provided in the form of "drips" i.e. smaller gray dots appended below the hyperedge lines. See Figure 5. The names still appear clearly in the tooltip when hovering on the hyperedge, as shown on the right of Figure 5.

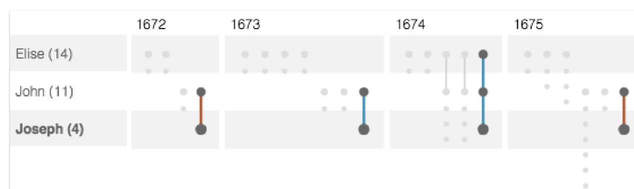

Figure 6: Filters reveal that Joseph's ego-network is composed of four contracts with John and, among these, one in 1674 also with Elise.

5 Conclusion

We described a novel way to read historical data and see insights that may be rapidly gained from using the PAOH representation. The Parallel Aggregated Ordered Hypergraph technique visualizes dynamic hypergraphs and allow users to follow the evolution of the network topology over time. Some interaction features and visualization simplification operations to explore larger hypergraphs are also addressed. Dynamic hypergraph visualization is still in its infancy, and can be applied in many cases of the digital humanities field to understand and analyze the evolution of relationships between entities. We hope our paper will stimulate additional work in this area.

Acknowledgments

We thank our colleagues from the University of Maryland HCIL and from INRIA Aviz , and the students of IVU lab from the University of Bari , for their early feedback and suggestions.

References

[1] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. The state-of-the-art of set visualization. *Computer Graphics Forum*, 35(1):234–260, 2016. doi: 10.1111/cgf.12722

[2] C. J. Arnold. *An Archaeology of the Early Anglo-Saxon Kingdoms*. Routledge, second ed., 1997.

[3] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. *Computer Graphics Forum*, 2015. doi: 10.1111/cgf.12615

[4] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17(suppl 1):S22–S29, 2001.

[5] T. Clement, A. Don, C. Plaisant, L. Auvil, G. Pape, and V. Goren. 'something that is interesting is interesting them': Using text mining and visualizations to aid interpreting repetition in gertrude stein's the making of americans. In *Proceedings of the Digital Humanities Conference*, pp. 40–44, 2007.

[6] F. Dias, M. R. Mansour, P. Valdivia, J. Cousty, and L. Najman. Watersheds on Hypergraphs for Data Clustering. In J. Angulo, S. Velasco-Forero, and F. Meyer, eds., *Mathematical Morphology and Its Applications to Signal and Image Processing*, pp. 211–221. Springer International Publishing, Cham, 2017.

[7] K. Dinkla, M. J. van Kreveld, B. Speckmann, and M. A. Westenberg. Kelp diagrams: Point set membership visualization. *Computer Graphics Forum*, 31(3pt1):875–884, 2012. doi: 10.1111/j.1467-8659.2012.03080.x

[8] N. Dufournaud, B. Michon, B. Bach, and P. Cristofoli. *L'analyse des réseaux, une aide à penser : réflexions sur les stratégies économique et sociale de Marie Boucher, marchande à Nantes au XVIIe siècle*, pp. 109–137. Presses universitaires de Bordeaux, 2017.

[9] C. Dunne and B. Shneiderman. Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3247–3256. ACM, 2013.

[10] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2013.

[11] D. Hansen, B. Shneiderman, and M. A. Smith. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2010.

[12] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On close and distant reading in digital humanities: A survey and future challenges.

[13] H. Kang, C. Plaisant, B. Lee, and B. B. Bederson. Netlens: iterative exploration of content-actor network data. *Information Visualization*, 6(1):18–31, 2007.

[14] P. Kapec. Visualizing software artifacts using hypergraphs. In *Proceedings of the 26th Spring Conference on Computer Graphics*, SCCG '10, pp. 27–32. ACM, New York, NY, USA, 2010. doi: 10.1145/1925059.1925067

[15] M. Latapy, C. Magnien, and N. D. Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31 – 48, 2008. doi: 10.1016/j.socnet.2007.04.006

[16] C. K.-S. Leung, C. L. Carmichael, and E. W. Teh. Visual analytics of social networks: mining and visualizing co-authorship networks. In *International Conference on Foundations of Augmented Cognition*, pp. 335–345. Springer, 2011.

[17] W. J. Longabaugh. Combing the hairball with BioFabric: a new approach for visualization of large networks. *BMC Bioinformatics*, 13(1):275, 2012. doi: 10.1186/1471-2105-13-275

[18] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404–409, January 2001. doi: 10.1073/pnas.021544898

[19] J. Novak, I. Micheel, M. Melenhorst, L. Wieneke, M. Düring, J. G. Morón, C. Pasini, M. Tagliasacchi, and P. Fraternali. Histograph–a visualization tool for collaborative analysis of networks from historical social multimedia collections. In *Information Visualisation (IV), 2014 18th International Conference on*, pp. 241–250. IEEE, 2014.

[20] J. Paquette and T. Tokuyasu. Hypergraph visualization and enrichment statistics: how the egan paradigm facilitates organic discovery from big data. *Proc. SPIE*, 7865:78650E–78650E–18, 2011. doi: 10.1117/12.890220

[21] P. R. Valdivia, P. Buono, and J.-D. Fekete. Hypenet: Visualizing Dynamic Hypergraphs. EuroVis 2017 - 19th EG/VGC Conference on Visualization, June 2017. Poster.

[22] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk. Dynamic Network Visualization with Extended Massive Sequence Views. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1087–1099, Aug 2014. doi: 10.1109/TVCG.2013.263

[23] C. Vehlow, F. Beck, and D. Weiskopf. The State of the Art in Visualizing Group Structures in Graphs. In R. Borgo, F. Ganovelli, and I. Viola, eds., *Eurographics Conference on Visualization (EuroVis) - STARs*. The Eurographics Association, 2015. doi: 10.2312/eurovisstar.20151110

[24] R. Vuillemot, T. Clement, C. Plaisant, and A. Kumar. What's being said near "martha"? exploring name entities in literary text collections. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pp. 107–114. IEEE, 2009.