



Exploring dynamic multilayer graphs for digital humanities

Stefan Bornhofen, Marten Düring

► To cite this version:

Stefan Bornhofen, Marten Düring. Exploring dynamic multilayer graphs for digital humanities. Applied Network Science, 2020, <10.1007/s41109-020-00295-x>. <hal-03019519>

HAL Id: hal-03019519

<https://hal.science/hal-03019519v1>

Submitted on 23 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

RESEARCH

Open Access

Exploring dynamic multilayer graphs for digital humanities



Stefan Bornhofen^{1*}  and Marten Düring²

*Correspondence: sb@eisti.eu

¹Laboratoire ETIS, CY Cergy Paris
Université, ENSEA, CNRS, UMR8051,
33 Boulevard du Port, 95000 Cergy,
France
Full list of author information is
available at the end of the article

Abstract

The paper presents Intergraph, a graph-based visual analytics technical demonstrator for the exploration and study of content in historical document collections. The designed prototype is motivated by a practical use case on a corpus of circa 15.000 digitized resources about European integration since 1945. The corpus allowed generating a dynamic multilayer network which represents different kinds of named entities appearing and co-appearing in the collections. To our knowledge, Intergraph is one of the first interactive tools to visualize dynamic multilayer graphs for collections of digitized historical sources. Graph visualization and interaction methods have been designed based on user requirements for content exploration by non-technical users without a strong background in network science, and to compensate for common flaws with the annotation of named entities. Users work with self-selected subsets of the overall data by interacting with a scene of small graphs which can be added, altered and compared. This allows an interest-driven navigation in the corpus and the discovery of the interconnections of its entities across time.

Keywords: Visual analytics, Network visualization, Dynamic multilayer networks, Digital humanities

Introduction

In recent years, vast quantities of the human cultural records have been digitized, further described with metadata and made available in the form of collections. Such collections within the fields of cultural heritage and digital humanities typically consist of digitized multimedia objects with a strong bias towards unstructured text, metadata of various levels of detail and completeness, and often a layer of named entity annotations. Today, most scholars in the humanities and related disciplines rely on keyword search and faceted search to retrieve relevant content. Any analysis of such collections needs to be based on an understanding of the underlying rationale for the creation of collections, how they are organized and to be able to retrieve relevant content in an exploratory manner (van Ham and Perer 2009; Brown and Greengrass 2006). In this paper we present Intergraph, a technical demonstrator for the exploration of such collections based on named entity linking and collection-inherent metadata as well as the results from preliminary user evaluations. Intergraph was designed to utilise multilayer network visualizations to

support non-technical users in the exploration of historical document collections. More specifically, it helps to answer the following questions: How is a given named entity (person, institution or location) represented in a collection? Who appears with whom? How does this change over time? How can we compare the coverage of entities? Answers to these questions can help to detect patterns in the collection such as biases in the composition of a collection (in our case for example gaps in the coverage of specific entities), or to highlight unexpected links between them.

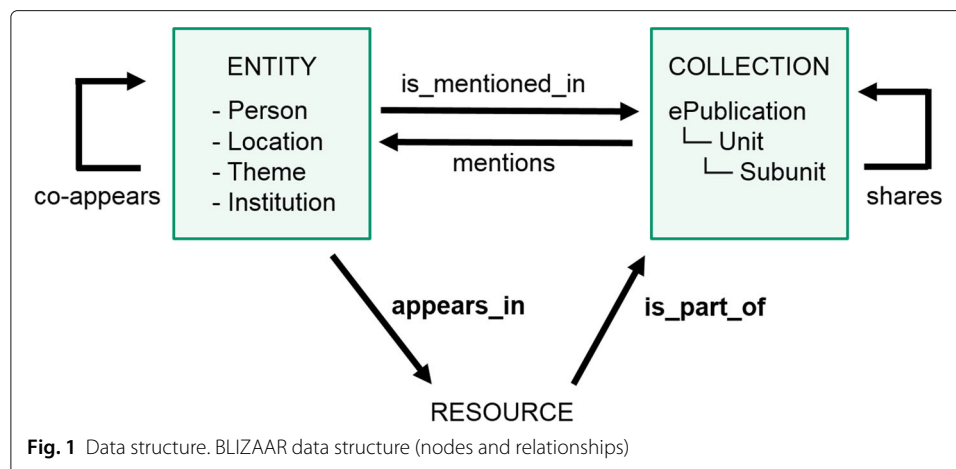
Intergraph was developed as one of three technical demonstrators during the BLIZAAR project, a French-Luxembourgish research project dedicated to develop novel visualizations of dynamic multilayer graphs (<https://blizaar.list.lu>). BLIZAAR concentrated on a use case in biology and a principal use case in history. The targeted users in both cases were scholars with no or very limited experience with data analysis in general and network analysis and visualization in particular.

The following section describes the principal use case in more detail, with its dataset and user requirements. “*State of the art*” section reviews the state of the art in the visualization of dynamic multilayer networks. “*Intergraph*” section presents the framework designed in response to the use case, and provides an overview of its main features. “*User test results*” section gives an account of the user tests conducted. Finally, “*Conclusion & future work*” section concludes the paper with a discussion and prospects for future work.

Dataset and requirements

The data is derived from resources on the European integration process since 1945 collected by the Centre Virtuel de la Connaissance sur l'Europe (CVCE) (<https://www.cvce.eu>), a former research and documentation center which in 2016 was integrated into the University of Luxembourg. The CVCE created a multilingual collection of approximately 25,000 digitized documents organized in 29 hierarchically structured thematic corpora. The documents differ significantly in nature: they include newspaper articles, diplomatic notes, personal memoirs, audio interview transcripts, cartoons and photos, all with descriptive captions. The histogram project (Guido et al. 2016) processed a subset of circa 15,000 of these documents with named entity recognition (NER) and -disambiguation and stored links between entities and documents in a Neo4j graph database. This dataset was made available to and further processed by the BLIZAAR project for the development of more advanced graph exploration prototypes.

Figure 1 shows the BLIZAAR data structure with its nodes and relationships. Firstly, resources are part of one or more collections, from the highest logical unit of thematic corpora (ePublications) down to the corresponding hierarchical units and subunits; this is modeled by the “*is_part_of*” relationship. Secondly, named entities (people, locations, themes, institutions) have been extracted using named entity recognition software such as YAGO (Max-planck-institut fur informatik: YAGO) and TextRazor (The natural language processing API). This process enabled the generation of the “*appears_in*” relationship. Entities “*co-appear*” in resources, and collections “*share*” resources, by bipartite network projection (Latapy et al. 2008; Zweig and Kaufmann 2011). Finally, a collection “*mentions*” an entity, and accordingly the entity “*is_mentioned_in*” the collection, if the collection contains at least one resource where the entity appears. Table 1 gives an idea of the size of the dataset.



A previous study by the BLIZAAR project on visual analytics requirements for research in digital cultural heritage was published in McGee et al. (2016). The authors suggest that the data structures supporting the analysis of a complex digital corpus, in which people, organizations, places, multimedia documents and document collections are connected across time, should best be modeled as a dynamic multilayer network. As a matter of fact, concerning the given dataset:

- Nodes can be considered on at least three layers, a resource layer, an entity layer and a collection layer, and they have different relationships;
- Nodes on the two latter layers have subtypes which can be treated as extra layers: entities are people, locations, institutions or themes, collections are ePublications, units or subunits;
- Resources are time-stamped by their historical publication date. The network therefore changes depending on the studied time period. Moreover, distinct time slices can also be regarded as defining separate layers.

The definition of layers in this data model deliberately remains flexible (node types, node subtypes, time periods), and may depend on the user's research question and vision of the data. Also, note that compared to the framework of Kivela's universal model of multilayer networks (Kivela et al. 2014), the BLIZAAR data does not possess multiple types of relationship between nodes. This network dimension has therefore not been considered for the rest of this paper.

In view of the concrete dataset and the CVCE (subsequently the University of Luxembourg) acting as a collaborator and a stakeholder, the BLIZAAR project is an instance of problem-driven visualization research. It has therefore been conducted in the spirit

Table 1 BLIZAAR dataset magnitudes

Item	Approximate count
Resources	15,000
Collections	4,000
Identified entities	36,000
Entity appearances	300,000
Entity co-occurrences	7,000,000

of a design study, which is defined as *a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines* (Sedlmair et al. 2012). The collaboration was organized as follows: A kick-off workshop with four CVCE domain experts helped to assess technical skill levels and to determine the research priorities for historians and to develop user stories. This was followed by monthly exchanges with a primary domain expert (also the second author of this paper). One intermediate and one final evaluation with four CVCE domain experts validated the user stories and provided feedback on the usability of the demonstrator.

In the kick-off workshop, the domain experts identified content retrieval and insights concerning the representation and interconnections of entities in a corpus as main objectives. Due to the heterogeneity of the documents in the corpus and lacking information on the nature of links between entities, the data was not considered usable for the reconstruction and quantitative analysis of a historical social network without significant manual annotation which was beyond the scope of the project. In addition, domain experts did not have a strong background in advanced data analysis but highly value at least a basic comprehension of the inherent logic of the tools they work with. These needs were expressed in the following user stories:

(1) Content overview: “I would like to have an overview of how a specific person/institution/location is represented in the corpus and of the other entities with whom they are mentioned. This helps me to decide which documents I want to study in greater detail.”

(2) Query knowledge expansion: “I am interested in a topic but simple keywords are not suitable to retrieve relevant documents. Starting with my limited knowledge of the topic I want to receive suggestions for promising contents and additional keywords which can guide my exploration.”

(3) Explore search results: “I am interested in a broader topic but am overwhelmed by the very large number of diverse search results. I want to be able to dissect and organize these results and understand how they are related to each other and their attribute values.”

(4) Entity comparison: “I want to compare specific entities (persons, institutions, locations, but also collections) in the corpus to get a better understanding of their presence in the corpus. I want to study how the contexts in which they appear change over time. I want to explore links between the entities I compare.”

Within the BLIZAAR project, one or more of these user stories were targeted by different prototypes (<https://blizaar.list.lu>). The design of the Intergraph demonstrator put special focus on (1) Content overview and (4) Entity comparison. To implement these user stories within a graph-based environment, the following user requirements have been identified:

- 1 Create subgraphs of understandable size and complexity
- 2 Set up multiple graphs for the sake of comparison or contrasting
- 3 Observe temporal changes
- 4 Filter for node and edge properties
- 5 Maintain a straightforward link to all connected resources for further study
- 6 Compensate for errors in named entity linking, e.g. duplicates

A specific problem related to the automatic generation of the network is data imperfections. Most commonly we observe fragments which were wrongly identified as entities, duplicate entities which have not been disambiguated correctly, and entities which have been disambiguated wrongly and linked to homonyms (the politician “Robert Schuman” vs. the composer “Robert Schumann”). The effort required to rectify all of the above-mentioned flaws is too costly and therefore unrealistic for this and comparable corpora. Functionalities moderating the flaws were therefore considered to be the most promising strategy in this case.

State of the art

Network visualizations offer a unique way to understand and analyze complex data by enabling users to inspect and comprehend relations between individual units and their properties. Some scientific fields have been using network visualizations for a long time, most notably systems biology where purpose-built visualizations have been developed for more than twenty-five years (Mendes 1993; Shannon et al. 2003; Pavlopoulos et al. 2008).

Interactive network visualizations have been used in and around the digital humanities sphere to make datasets accessible for exploration and research (Jänicke et al. 2015; Düring 2019; Jessop 2008; Boukhelifa et al. 2015; Düring 2013; Windhager et al. 2018) inasmuch as they offer novel search and discovery tools which enhance well-established techniques such as faceted search and keyword search. Examples include stand-alone applications based on letter exchanges (Warren et al. 2016), to explore bibliographic data (SNAC; Verhoeven and Burrows 2015), collections of documents based on unstructured text (Guido et al. 2016; Moretti et al. 2016), often with a strong element of decentralised and collaborative data curation. To avoid costly manual relationship extraction, network data is often generated automatically based on named entity recognition (NER), existing document metadata or other data extracted from unstructured text and inferred relations between them. With its focus on the exploration of automatically enriched unstructured texts, Intergraph bears closest resemblance to Guido et al. (2016) and Moretti et al. (2016) but adds new functionality for the exploration of multilayer networks on multiple canvases.

Dynamic networks represent evolving relationships between entities that evolve over time (Beck et al. 2017). A small number of tools are readily available for dynamic graph visualization, such as Gephi (Bastian et al. 2009) or Commetrix (Trier 2008) which are arguably the two most prominent solutions. These applications are under continuous development and have a large user community, but they are not adapted to visualizing multiple layers and therefore cannot properly meet the specificities of multilayer data. As a matter of fact, recent research suggests that multilayer graphs allow for more complexity in the exploration of historical data (McGee et al. 2016; Valleriani et al. 2019; van Vugt 2017; Grandjean 2019). In multilayer networks, subnetworks are considered on independent layers, but they can also interact with each other (Kivelä et al. 2014). Multilayer networks can have multiple types of node (Ghani et al. 2013), with different attributes (Kerren et al. 2014; Nobre et al. 2019) and different types of relationships (Singh et al. 2007). Tulip (Auber 2004) is a powerful graph visualization framework capable of embracing the complexity of multilayer data, however its configuration requires expertise in programming and network analysis and has a substantial learning curve for the user to obtain the intended visualization.

The particularities of networks being both dynamic and multilayer have recently come to the attention of network science due to their importance for real-world applications. Multilayer networks open up new opportunities for the interactive exploration of (historical) datasets but also require novel types of data visualization (Rossi and Magnani 2015). In recent years, two collaborative European projects, Plexmath (2012–2015) (https://cordis.europa.eu/project/rcn/105293_fr.html) and Multiplex (2012–2016) (https://cordis.europa.eu/project/rcn/106336_en.html), were entirely dedicated to this topic, allowing to design a number of novel visualization methods (<http://www.mKivela.com/pymnet>, <https://github.com/sg-dev/multinet.js>) (De Domenico et al. 2014; Piškorec et al. 2015). The tools published in the scope of these two projects have been primarily designed to demonstrate concepts and to illustrate universal approaches to dynamic multilayer graph visualizations. Given the lack of ongoing development and active support, their usability for the BLIZAAR use case given in this paper is problematic. A complete survey on current visualization solutions for multilayer graphs and their features, can be found in Ghoniem et al. (2019).

Based on the BLIZAAR use requirements identified in the previous section, three feature sets were found to be essential for any adopted solution:

- 1 the swift creation of subgraphs from a larger dataset to support an iterative exploration workflow;
- 2 follow-up functionalities beyond selection, layout rearrangement or camera movement. In our case, this includes querying, generating, and juxtaposing related subgraphs and layers;
- 3 a flexible layer model which allows the user to switch between, and even to combine models (e.g. node type layers vs. time slice layers).

To our knowledge there is no tool available which contains all three feature sets and allows their combination in exploratory workflows. This motivated the development of the Intergraph visualization platform which we describe in the following section.

Intergraph

Intergraph offers a novel approach to exploring digital humanities corpora by means of an iterative search and discovery workflow. The demonstrator has been designed to meet all user requirements specified in the previous section. Written in javascript, Intergraph runs in a web browser and communicates with a node.js server which queries the data from a Neo4j database. The front-end client renders the graphs using the Three.js graphics library.

Given the size of the BLIZAAR dataset, an overall visualization of the corpus is neither suitable nor desirable for exploration. Instead, users are rather interested in creating and inspecting subnetworks with entities relevant to their current research. The main idea of Intergraph is therefore to begin the exploration from one or more known start nodes. Following the expand-on-demand principle (van Ham and Perer 2009), the user will encounter new relevant nodes and pursue their exploration by conveniently creating additional graphs stemming from the existing ones. This path of exploration yields a sequence of linked subgraphs (user requirement 1). Depending on the query and the users understanding of the data, a new graph may be used and looked upon as a complementary layer, or as a complementary graph of an existing layer.

Figure 2 shows a general screenshot of the Intergraph interface. Graphs can be dynamically added to and deleted from the scene. Following the VisLink approach (Collins and Carpendale 2007), they are rendered on free-floating planes which can be arbitrarily translated, oriented and scaled using familiar transformation widgets. Depending on the user's tasks and preferences, the scene can be viewed from a 2D or a 3D perspective. The default 2D view is known to be most effective for visual data exploration and analytics, since 3D visualizations tend to suffer from occlusion, overlapping and distortion, and they often require increased viewpoint navigation to find an optimal perspective (Shneiderman 2003). However, 3D scenes allow users to stack multiple planar graph layers in space and to create so-called “2.5D” visualizations, which can be useful for understanding complex networks (Ware 2001).

Since target users are not experts in network science and since the data itself does not lend itself to reconstruct meaningful social networks, Intergraph forgoes advanced graph concepts. Network analysis functions, metrics and algorithms like clustering coefficients or betweenness have not been considered for implementation in this prototype. On these grounds, Intergraph also uses standard visual encoding. Node colors reflect the node type and sizes indicate the number of underlying resources. A click on a node or edge gives immediate access to these resources (user requirement 5). New graphs are typically produced by querying ego-networks of existing nodes, i.e. subgraphs linking a node with its immediate neighbors, via easy-to-communicate operations such as:

- All entities co-appearing with a given entity
- All collections mentioning a given entity
- All entities mentioned in a given collection
- All collections sharing resources with a given collection

If the same node appears on two or more graphs of the scene, coupling edges (Kivelä et al. 2014) are drawn (see Figs. 2 and 3). This user-driven network generation approach was partly inspired by “citation-chaining”, one of the most commonly used search strategies for literature among historians (Ellis 1989; Buchanan et al. 2005). Intergraph applies the citation chaining principle to documents and the entities mentioned in

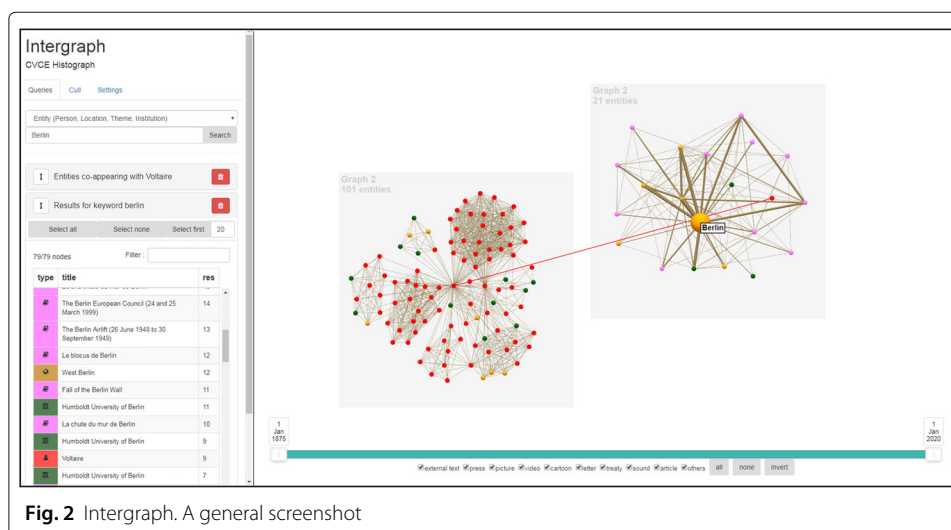
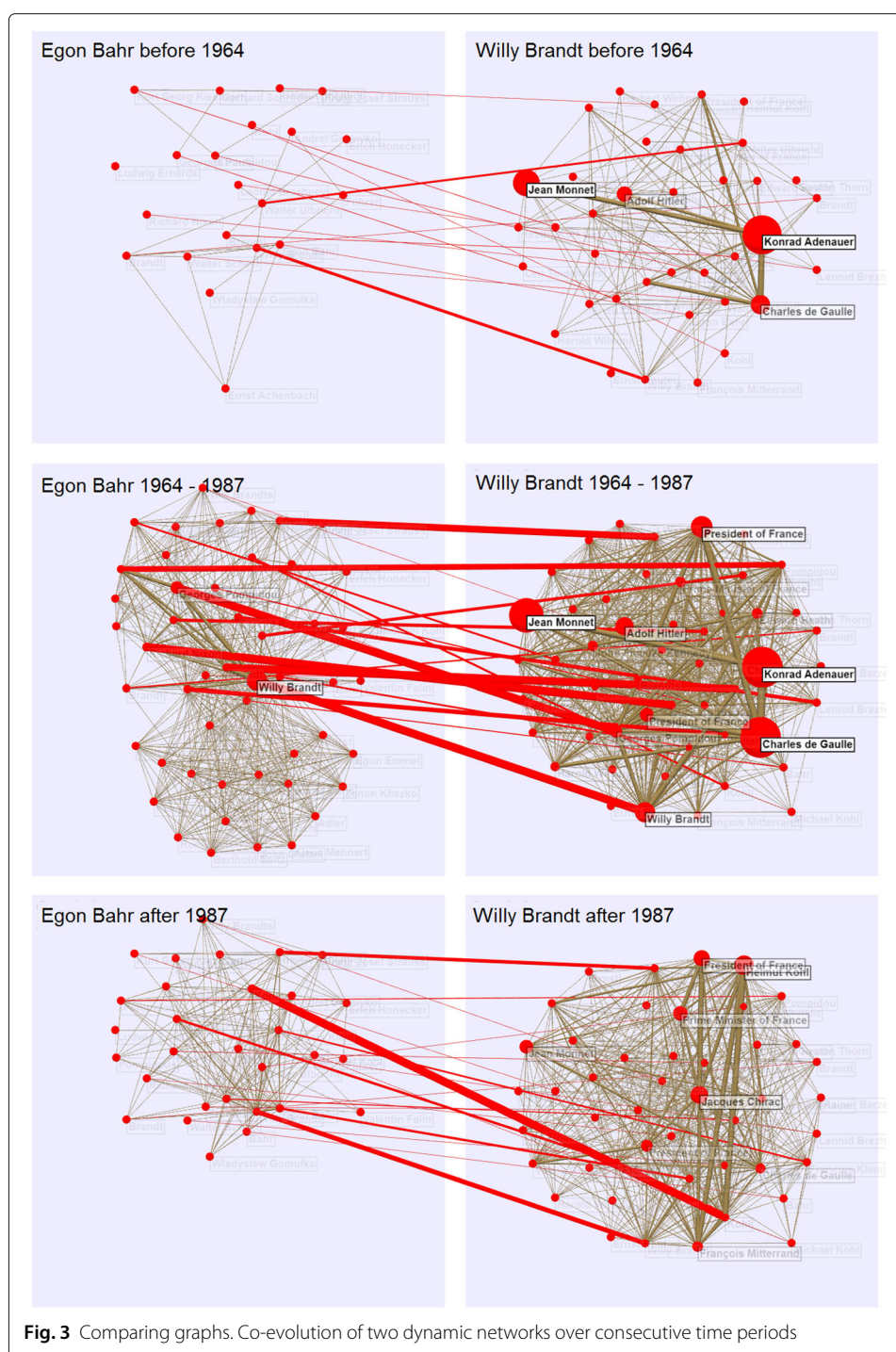


Fig. 2 Intergraph. A general screenshot



them. This allows users to create their own interest-driven search and discovery paths across the dataset.

With regard to data imperfections, one of the most frequently encountered flaws with entity disambiguation in the CVCE dataset are entity duplicates, i.e. multiple recognized entities where in reality only one was actually mentioned. For example, named entity recognition yielded three separate nodes for “East Berlin”. If the user wants to consider

these three nodes as one in order to create an ego-network, it is possible to multi-select a number of nodes and to query “All entities co-appearing with a given entity (union)”, meaning that the result will be the list of nodes co-appearing with at least one of the selected nodes. The user can then define a unique group node for “East Berlin” and draw a meaningful graph (user requirement 6). It is also possible to query “All entities co-appearing with a given entity (intersection)”. This operation returns the list of entities co-appearing in the corpus with all selected nodes and can be used to merge multiple nodes, for example if understood as representatives of a social group.

The results of new queries first appear in the form of a table in the left pane. This first kind of visualization, itemizing only the nodes without the edges, may in some cases already be sufficient to work with. The table lets users decide whether it is worth generating the graph or whether they prefer to recompile the list of nodes, if there are missing nodes or nodes which should be excluded from the graph. A graph of a given node table, or part of it, can be generated on demand and is added to the canvas on the right side of the interface.

The scene can be submitted to a filter which operates on resource type and time period (user requirement 4). Subgraphs of a given resource type can provide a better understanding of its distribution within the corpus. Subgraphs considering the resources within a specific time window allow the user to assess the relevance and interconnections of entities during a specific period. The user can shift the time window and obtain an animated representation of the dynamic graph (user requirement 3). If time-to-time mapping, i.e. animation, is not convenient to analyze the evolution of a network over time, time-to-space mapping is also possible. For this purpose the user can clone and “freeze” a graph of the scene, meaning that its current filter is fixed. Using this method, several graphs with the same nodes but distinct time periods can be juxtaposed (2D) or superimposed (3D) in space (Beck et al. 2017) (user requirement 2).

Figure 3 illustrates how Intergraph addresses the three challenges as specified in the state of the art by offering interactive subgraph creation, multiple operations to define layers and the free organization of these layers in the canvas. We take as example Willy Brandt, former Chancellor of the Federal Republic of Germany, and his advisor Egon Bahr. The scene shows a set of subgraphs which have been incrementally constructed. Starting with a search for the two nodes “Willy Brandt” and “Egon Bahr”, we use the node-level follow-up operation “persons co-appearing with” and the graph-level operations “clone”, “freeze” and “time filter” to create two ego networks at three consecutive time periods (before 1964, 1964–1987 and after 1987). The flexible layer concept allows organizing these six layers freely, in this case to reveal how the co-occurrence networks evolve and overlap.

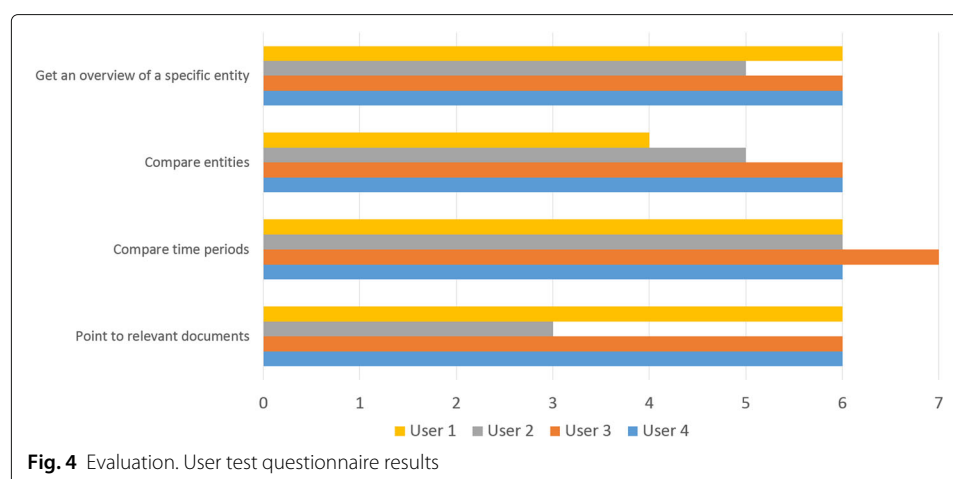
User test results

To assess the usability of Intergraph, two user evaluation sessions were held. The first to validate the user stories and to provide feedback on an earlier version of Intergraph, the second to evaluate the current version. The tests were conducted with a group of four scholars, all of whom were former CVCE employees. The selection criteria were familiarity with the underlying corpus on European integration and with the application of digital tools and methods. These criteria were applied to ensure that users could turn their attention to the interaction with the prototype with only minimal reminders of the underlying

data model and content, and also that they were qualified to judge the pertinence of the output. For the second evaluation, evaluators were asked to submit in advance a list of five persons or intuitions and optionally up to three time periods between 1945 and 2009 they wished to explore in greater depth under the precondition that they had expert knowledge about their presence in the CVCE corpus.

To compensate for the relative unfamiliarity with the software they were to test, the scholars were reminded of BLIZAAR's research objectives and received a circa 10-min demonstration of the Intergraph core functionalities. After the presentation of the platform, they were invited to use Intergraph for themselves and to begin their session with an elementary keyword search for an entity they knew was mentioned in the corpus. From this starting point, they were free to perform more synoptic tasks, such as finding relevant collections and resources, searching for co-appearing entities and comparing their corresponding networks, in order to obtain a comprehensive overview of how the investigated element is represented, positioned and linked in the corpus. Throughout the session, users were encouraged to continuously verbalize their train of thoughts and actions, in line with the thinking aloud approach (Boren and Ramey 2000). Following the 45-min testing period, users were asked to give verbal feedback and to complete a questionnaire.

In their verbal feedback, users appreciated the ease of navigating through the corpus, the flexibility and freedom to combine different elements, the links across canvases, the management of duplicate entities as well as the ability to drill down to the underlying resources. With regard to the added value in their research workflows, they highlighted the ability to detect unexpected relationships between entities (in this case between a bank and politicians) and the ability to contribute to a global assessment of the collection and its composition. Critical remarks addressed the absence of additional inter-layer links across canvases (e.g. multiple entities mentioned in the same ePublication), the obscurity of the frequent use of the context menu triggered by a right-click, and long loading times for graphs with more than 100 nodes due to performance limitations. Users were invited to fill in a questionnaire and to quantify the utility of Intergraph on a scale of 1 to 7 with regard to all aspects of the initially defined user story. As a result, Fig. 4 shows a high general acceptance. In one case, Intergraph did not produce a number of documents the evaluator expected to retrieve, which explains the low mark below average. Most notably, all users declared in the questionnaire that for the given user story they would prefer to



use Intergraph over the other available tools (CVCE homepage search and CVCE backend search).

Conclusion & future work

This paper presented Intergraph, a visual analytics platform designed for effective navigation through the content of digital humanities corpora by non-experts. The work is inspired by recent advances in the visualization of dynamic multilayer networks, and has been enhanced and optimized for humanities scholars with no or very little skills in data analysis and visualization and their subject-specific workflows. The user tests conducted showed a high acceptance of the demonstrator with respect to the original requirements.

The user tests did however also reveal a number of challenges. Our evaluation setup sought to minimize the hurdles involved with an ad-hoc evaluation of a new tool and uncommon methods of search and discovery based on named entities and visualizations but could not remove them. Users require significant time to learn how to operate especially prototype software and to adjust long established research workflows. Most crucially however, automated data extraction and visualization-based exploration need to be understood well and their potential can only be explored by extensive experimentation and adoption to a new domain which was beyond the scope of this project.

Given the exploratory nature of Intergraph, future work will concentrate on additional ways of suggesting related nodes and creating pertinent graphs out of existing ones, for example by applying recommendation algorithms (Bobadilla et al. 2013). The multi-layer character of the data should be leveraged by adding other types of interlayer edges, such as those indicating the “mentions” relationship between collections and entities. For the time being, it is also impossible to visualize more than one type of link within a given layer, which precludes for example the possibility to visualize registered family or friendship relationships between people in addition to their coappearance relationship. Moreover, data imperfections may cause significantly skewed results and therefore need to be taken into account. Such imperfections may stem from the automated processing of digitized materials by methods such as Optical Character Recognition (OCR) or Named Entity Recognition (NER), but also from manual curation of metadata, for example, as well as intrinsic ambiguities in the source material. Since data cleaning is typically too costly, we observe a strong need for systems which enable users to cope with data-inherent imperfections. Intergraph’s functionality to merge duplicate nodes is a step in this direction. Another promising direction is the critical assessment of the composition of datasets and the potential of visualization to reveal its inherent biases. These requirements present interesting challenges and opportunities for the design of innovative tools for visual analytics.

Finally, it is important to observe that the used data model is highly generic. Entities identified in collections of time-stamped resources are likely to be found in a huge number of digital corpora. It is intended to open the existing platform to other datasets, and for this purpose the authors are currently working on extending Intergraph and its functionalities to two additional historical databases:

- Regesta Imperii (Kuczera 2019) (<http://www.regesta-imperii.de>): a collection of records of the Roman-German kings and emperors, as well as of the popes from the Middle Ages;

- Romans 1by1 (Varga 2017) (<http://romans1by1.com>): a population database registering the people which are attested in Greek and Roman epigraphic sources.

By this means, we hope that Intergraph can evolve into a valuable visualization and exploration tool for many scholars working in the field of digital humanities.

Abbreviations

CVCE: Centre Virtuel de la Connaissance sur l'Europe; NER: Named Entity Recognition; OCR: Optical Character Recognition

Acknowledgements

The authors would like to thank the master student Samuel Guillaume for his excellent contribution to the implementation of the Intergraph platform. The authors would furthermore like to express their gratitude to the journal editors and reviewers for their highly constructive feedback on earlier versions of this article.

Authors' contributions

Stefan Bornhofen is responsible for the design and implementation of the Intergraph software. Marten Düring is responsible for the development of the use case and evaluation. Both authors read and approved the final manuscript.

Authors' information

Stefan Bornhofen studied mathematics and computer science at the University of Mainz, Germany, and obtained a PhD in Computer Science from Paris-Sud University, France, in 2008. He holds a teaching and research position at CY Tech engineering school in Cergy near Paris, and is the head of a Master's program specializing in computer graphics, computer vision and human-computer interaction.

Marten Düring studied history at the universities of Augsburg and Manchester, and received a PhD degree in contemporary history from the University of Mainz, Germany in 2015. He is Assistant Professor in Digital History at the Luxembourg Centre for Contemporary and Digital History.

Funding

This research was funded by the ANR grant BLIZAAR ANR-15-CE23-0002-01 and the FNR grant BLIZAAR INTER/ANR/14/9909176.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to copyright restrictions but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratoire ETIS, CY Cergy Paris Université, ENSEA, CNRS, UMR8051, 33 Boulevard du Port, 95000 Cergy, France.

²Luxembourg Centre for Contemporary and Digital History, 11 Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg.

Received: 24 February 2020 Accepted: 29 July 2020

Published online: 26 August 2020

References

- Auber D (2004) Tulip - A Huge Graph Visualization Framework. *Graph Draw Softw*:105–126. https://link.springer.com/chapter/10.1007/978-3-642-18638-7_5#citeas
- Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. *Int AAAI Conf Weblogs Soc Media*:361–362. <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Beck F, Burch M, Diehl S, Weiskopf D (2017) A taxonomy and survey of dynamic graph visualization. *Comput Graph Forum* 36:1
- Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. *Knowl Based Syst* 46:109–132
- Boren T, Ramey J (2000) Thinking aloud: reconciling theory and practice. *IEEE Trans Prof Commu* 43(3):261–278
- Boukhelifa N, Giannakis E, Dimara E, Willett W, Fekete J-D (2015) Supporting Historical Research Through User-Centered Visual Analytics. In: Bertini E, Roberts JC (eds). *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, pp 1–5. <https://diglib.org/handle/10.2312/eurova.20151095.001-005>
- Brown SC, Greengrass M (2006) RePAH: a user requirements analysis for research portals in the arts and humanities. *Humanities Research Institute Online*, Sheffield University, Leicester
- Buchanan G, Cunningham SJ, Blandford A, Rimmer J, Warwick C (2005) Information Seeking by Humanities Scholars:218–229
- Collins C, Carpendale S (2007) VisLink: Revealing Relationships Amongst Visualizations. *IEEE Trans Vis Comput Graph (Proc IEEE Conf Inf Vis (InfoVis))* 13(6):1–6
- De Domenico M, Porter MA, Arenas A (2014) MuxViz: a tool for multilayer analysis and visualization of networks. *J Complex Netw* 3:159–176
- Düring M (2013) HNR Bibliography. *Historical Network Research*. Retrieved September 1, 2018 from <http://historicalnetworkresearch.org/bibliography>

- Düring M (2019) Networks as Gateways. Gleanings from applications for the exploration of historical data. In: Kerschbaumer F, von Keyserlingk L, Stark M, Düring M (eds). *The Power of Networks. Prospects of Historical Network Research*. Routledge Publishers, Abingdon
- Ellis D (1989) A behavioural approach to information retrieval system design. *J Doc* 45(3):171–212
- Ghani S, Kwon B. C., Lee S, Yi J. S., Elmqvist N (2013) Visual analytics for multimodal social network analysis: A design study with social scientists. *IEEE Trans Vis Comput Graph* 19(12):2032–2041
- Ghoniem M, McGee F, Melançon G, Otjacques B, Pinaud B (2019) The State of the Art in Multilayer Network Visualization. *Comput Graph Forum* 38(6):125–149
- Grandjean M (2019) A Conceptual Framework for Multilayer Historical Networks. *Sharing the Experience: Workflows for the Digital Humanities*. DARIAH-CH Proceedings, Neuchâtel
- Guido D, Wieneke L, Düring M (2016) *histograph*. Graph-based exploration, crowdsourced indexation. CVCE, Luxembourg
- Jänicke S, Franzini G, Cheema MF, Scheuermann G (2015) On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. In: *Eurographics Conference on Visualization (EuroVis) - STARS*. The Eurographics Association. <https://diglib.org/handle/10.2312/eurovisstar.20151113>
- Jessop M (2008) Digital visualization as a scholarly activity. *Lit Linguist Comput* 23(3):281–293
- Kerren A, Purchase HC, Ward MO (2014) Multivariate network visualization. *Lect Notes Comput Sci* 8380:8380
- Kivelä M, Arenas A, Barthélemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. *J Complex Netw* 2(3):203–271
- Kuczera A (2019) Die “Regesta Imperii” im digitalen Zeitalter. Das Regest als Netzwerk von Entitäten. *Das Mittelalter* 24:157–172. <https://doi.org/10.1515/mial-2019-0011>
- Latapy M, Magnien C, Vecchio ND (2008) Basic notions for the analysis of large two-mode networks. *Soc Netw* 30(1):31–48
- Max-planck-institut für informatik: YAGO A high quality knowledge base. <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago>. Accessed 17 Aug 2020
- McGee F, Düring M, Ghoniem M (2016) Towards Visual Analytics of Multilayer Graphs for Digital Cultural Heritage
- Mendes P (1993) GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput Appl Biosci* 9(5):563–571
- Moretti G, Sprugnoli R, Menini S, Tonelli S (2016) Knowledge-Based Systems ALCIDE : Extracting and visualising content from large document collections to support humanities studies. *Knowl-Based Syst* 111:100–112
- Nobre C, Streit M, Meyer M, Lex A (2019) The State of the Art in Visualizing Multivariate Networks. *Comput Graph Forum* 38(3):807–832
- Pavlopoulos GA, O'Donoghue SI, Satagopam VP, Soldatos T, Pafilis E, Schneider R (2008) Arena3D: visualization of biological networks in 3D. *BMC Syst Biol* 2:104
- Piškořec, Sluban, Šmuc (2015) MultiNets: Web-Based Multilayer Network Visualization. In: *Machine Learning and Knowledge Discovery in Databases*. Springer, Cham. pp 298–302
- Rossi L, Magnani M (2015) Towards effective visual analytics on multiplex and multilayer networks. *Chaos Solitons Fractals* 72(0):68–76
- Sedlmair M, Meyer M. D., Munzner T (2012) Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans Vis Comput Graph* 18(12):2431–2440
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504
- Shneiderman B (2003) Why not make interfaces better than 3D reality? *IEEE Comput Graph Appl* 23(6):12–15
- Singh L, Beard M, Getoor L, Blake M (2007) Visual mining of multimodal social networks at different abstraction levels. In: *Information Visualization, 2007. IV '07. 11th International Conference*. IEEE, Zurich. pp 672–679
- SNAC Social Networks and Archival Context. <http://socialarchive.iaht.virginia.edu>. Accessed 17 Aug 2020
- The natural language processing API. <https://www.textrazor.com>. Accessed 17 Aug 2020
- Trier M (2008) Research Note - Towards Dynamic Visualization for Understanding Evolution of Digital Communication Networks. *Inf Syst Res* 19:335–350
- Valleriani M, Kräutli F, Zamani M, Tejedor A, Sander C, Vogl M, et al. (2019) The emergence of epistemic communities in the ‘Sphaera’ corpus: mechanisms of knowledge evolution. *J Hist Netw Res* 3:50–91. <https://doi.org/10.25517/jhnr.v3i1.63>
- van Ham F, Perer A (2009) Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest. *IEEE Trans Vis Comput Graph* 15(6):953–960
- van Vugt I (2017) Using Multi-layered Networks to Disclose Books in the Republic of Letters. *J Hist Netw Res* 1(1):25–51
- Varga R (2017) Romans 1 by 1 v.1.1. New developments in the study of Roman population. *Digit Class Online* 3(2):44–59
- Verhoeven D, Burrows T (2015) Aggregating Cultural Heritage Data for Research Use: The Humanities Networked Infrastructure (HuNI). In: *Metadata and Semantics Research*. Springer, Cham. pp 417–23
- Ware C (2001) Designing with a 2 1/2D Attitude. *Inf Des J* 10(3):171–182
- Warren C, Shore D, Otis J, Wang L, Finegold M, Shalizi C (2016) Six Degrees of Francis Bacon. A Statistical Method for Reconstructing Large Historical Social Networks. *Digit Human Quart* 10(3)
- Windhager F, Federico P, Schreder G, Glinka K, Dörk M, Miksch S, Mayr E (2018) Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. *IEEE Trans Vis Comput Graph* 25:20
- Zweig KA, Kaufmann M (2011) A systematic approach to the one-mode projection of bipartite graphs. *Soc Netw Anal Min* 1(3):187–218

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.