

# Interactive Networks for Digital Cultural Heritage Collections - Scoping the Future of histoGraph

Marten Düring<sup>1</sup>, Lars Wieneke<sup>1</sup>, and Vincenzo Croce<sup>2(✉)</sup>

<sup>1</sup> CVCE, Sanem, Luxembourg

{marten.duering, lars.wieneke}@cvce.eu

<sup>2</sup> Engineering Ingegneria Informatica spa, Rome, Italy

vincenzo.croce@eng.it

**Abstract.** Network visualizations are powerful albeit suggestive means to represent and explore complex relations between entities (people, words, places...) and have become increasingly popular in the humanities. In this article we briefly introduce the functionality of histoGraph, a technical demonstrator for the network-based analysis of photographs, survey current applications of network visualizations in the cultural heritage domain, and outline histoGraph's future development as an open source tool for the visualization of cultural heritage data.

## 1 Introduction

In the humanities the concept of truth is a fuzzy and relies on a more or less stable foundation of what is true. Network visualization are powerful albeit suggestive means to represent and explore complex relations between entities (people, words, places...) with great potential for the humanities and the cultural heritage sector. In this article we introduce the functionality of histoGraph, a demonstrator which exploits the surplus value of human touch for the identification of identities in historical image collections through a hybrid crowd-sourcing approach.

We distinguish two perspectives on network visualizations in the (Digital) Humanities: visualizations can be used to illustrate specific insights based on existing knowledge or to explore data and to discover something that is not yet known. Within this larger concept of visual analytics we also identify two perspectives: one that stipulates the idea of a holistic or “bigger” picture, so that we can gain insight by combining different information into one image (seeing the forest for the trees) and one that focuses on identifying the peculiar in a massive amount of information.

We propose to build a bridge between the two: an analytical tool to identify peculiarities and on the other hand an authoring tool for visual storytelling. This would offer us an interesting cross-connection with the idea of enhanced publication as it is understood by the Driver project [1]. Usually scientific publications contain read-only material that users can see but not act upon. What if we have a diagram that users can click on to manipulate the underlying dataset? histoGraph could offer a visualisation with an explanation but also let users select and then modify the analytical tool, to draw their own conclusions, to agree with the author or to challenge his findings.

## 2 The Tool in Digital Humanities

histoGraph was developed by the FP7-funded project CUBRIK [2] which focused on advanced multimedia search technologies. The current version creates a social network of people who appear in photos related to the history of European integration and automatically enriches the network with relevant sources based on keyword queries in full text. A demo is available online [3]. To create the network, faces in the historical photos need to be identified, a very challenging task for machines. Humans and machines therefore share this work and the input from one improves the performance of the other.

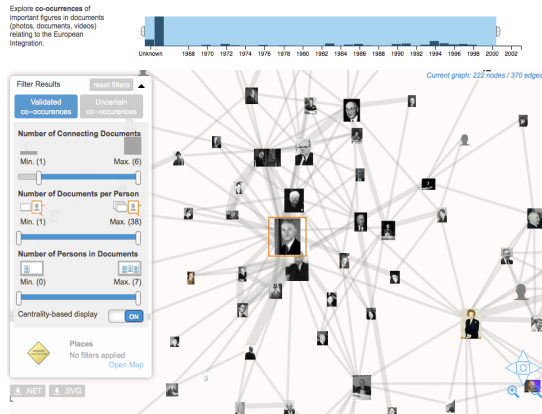
histoGraph introduces an effective interface to access collections of historical sources and to discover links among and entities within them. The integration of human expertise and machine computation opens up new possibilities to create a new type of applications. So far however, this potential remains largely untapped because of the significant requirements for such projects: The implementation and integration of advanced algorithms, for example for the identification of faces, requires specialized know-how and users from the humanities are challenged with defining unprecedented tasks for methods which haven't even emerged yet. histoGraph combines new approaches to engage the public to commit to humanities research, to facilitate exchange between users and to help us reach our audiences.

CUBRIK integrated research in computer science, the design of human-computation tasks, data visualization, social engineering and the humanities. Multimedia search is still a challenge and the CUBRIK project approached it from many different angles: Human-Machine-Interaction, face identification in photos and videos, copyright issues, gamification, content exploration and user communication. Alongside an app for exploring and searching fashion, histoGraph is one of two demos which implement the different modules developed by CUBRIK.

histoGraph is based on a collection of more than 3000 images which represent the main events and actors in the history of European integration hosted by the CVCE. To prepare the photos for the network, we use an image indexation pipeline which detects the location of individual faces in the photographs. A crowd of "click-workers" with no specific training double-checks whether the algorithms detected faces correctly or whether it missed some. In the next step, an automatic face recognition process is triggered that associates each of the now verified faces with a list of ten possible identities. This list of candidates is then disseminated for example through Twitter to a crowd of experts who vote for and comment on their preferred identity. The image metadata, for example the names of persons, the time or the place where an image was taken as well as contextual information about associated historical events can be reviewed by expert users and delegated to a crowd of specialists on the history of European Integration for review.

Based on the co-occurrence of persons in images, a social network is calculated which links individual persons with each other (**Fig. 1**. Screenshot of histoGraph's network visualization tool.). Connections gain in strength the more often persons appear together in an image. Users can interact with histoGraph in different ways, e.g. a click on a node leads to an ego-network of the selected person and a click on an edge displays

documents, which mention both actors (**Fig. 2**). This feature is powerful since it guides users back to the primary sources on which a tie is based on and thereby makes it easier to understand what a tie and a node represent. Many of the documents stored in our collection come with a date of creation. This allows us to filter the network so it only displays connections of documents created within certain time spans.



**Fig. 1.** Screenshot of histoGraph's network visualization tool



**Fig. 2.** Screenshot of histoGraph's tie contextualization

As introduced above a challenge for histoGraph and the Digital Humanities in general is the conception of truth. Scientists can rely on a more or less stable foundation of what is true, experiments can be replicated and measured precisely. In the humanities the concept of truth is far more complex: It is based on the insight, that there is no neutral or objective way to study human environments. The way in which questions are asked, when, by whom, how data is selected to answer them, by what means this data is analyzed and finally the way in which the results of such analyses are communicated and received all challenge the idea of “one truth”. Persons may, for example, change careers, their home countries might be renamed or they choose to go by different names. This means that at different points in time there might be more than one “true” answer to the simple question “Who is this?”. In order to represent the discursive nature of truth in the humanities we make use of a community-driven tool for question answering, similar to stackoverflow.com [4]. Users have the opportunity to answer questions and thus benefit from the knowledge within the expert crowd.

### 3 Conclusion

We envision that histoGraph fills this niche and becomes a general purpose context exploration and storytelling tool for research and teaching in cultural heritage, the humanities and journalism. Crucially, histoGraph will maintain its current ability to process photos and will also become capable of processing text documents and meta-data. For future development of histoGraph we propose the following use cases: Network creation based on images, network creation based on entity co-occurrences, network annotation and storytelling. This will require the following components:

1. Redevelopment of the image processing components with open source solutions and enhanced usability
2. Development of a component which imports collections and uses language processing technology to automatically produce visualizations of co-occurring entities such as persons, institutions, places.
3. Development of browsing histories, dynamically updated node lists as well as logs of reversible user actions as implemented in Open Refine [5] for better orientation.
4. Development of an annotation component which allows users to manually link nodes and ties to (third party) source documents and to provide additional textual information
5. Redevelopment of the contextualization feature which visualizes such links to source documents for any node and tie
6. Development of map-based visualizations of data and hybrid map+network visualizations.

These are ambitious goals, which will require time, funding and a number of case studies for the development of prototypes. Network visualizations have unrivalled powers to make complex relations comprehensible – we hope that histoGraph will allow authors and audiences to make the most of this potential.

### References

1. Driver project. <http://www.driver-repository.eu/>
2. CUBRIK project. <http://www.cubrikproject.eu/>
3. histoGraph demonstrator. <http://histograph.eu/>
4. Stack Overflow. <http://stackoverflow.com/>
5. Open Refine. <http://openrefine.org/>