CULTURE DIGITALI

INTERSEZIONI   FILOSOFIA   ARTI   MEDIA

- TESTO
- ARTI
- FILOSOFIA
- CONTENUTI
- INTELLIGENZA
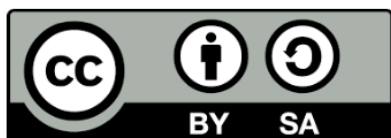
PROCEEDINGS

Fabio Ciracì, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Culture digitali. Intersezioni: filosofia, arti, media. Preceedings della 11ª conferenza nazionale, Lecce, 2022. Fabio Ciracì, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Digital cultures. Intersections: philosophy, arts, media. Preceedings of the 11th national conference, Lecce, 2022.

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato Scientifico di AIUCD 2022.

All the paper published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2022 Scientific Committee.

Il programma della conferenza AIUCD 2022 è disponibile online all'indirizzo/ The AIUCD 2022 conference program is available online all'apposito indirizzo

http://aiucd2022.unisalento.it http://conference.unisalento.it/ocs/index.php/aiucd2022/index/pages/view/programma

Comitato Scientifico:

General Chair: Fabio Ciracì (Università del Salento)

Local Chair: Mario Bochicchio (Università del Salento, Università di Bari)

Membri Comitato Scientifico: Marina Buzzoni (Presidentessa AIUCD, Uni. Venezia), Federico Boschetti (Ric. ILC-CNR); Federico Meschini (Uni. Tuscia); Roberto Rosselli Del Turco (Uni Torino); Rachele Sprugnoli (Ass. Ric. Univ. Cattolica); Donato Malerba (Università Bari);

Luca Bandirali, Daniela Castaldo, Francesco Ceraolo, Stefano Cristante, Domenico M. Fazio, Manolita Francesca, Marco Mancarella, Pietro Luigi Iaia, Massimiliano Rossi, Grazia Semeraro, Franco Tommasi, Luigi Patrono (Università del Salento)

Membri del Comitato di programma: Mario Bochicchio (Local Chair), Luca Bandirali, Daniela Castaldo, Marco Mancarella, Pietro Luigi Iaia, Federica Epifani (Responsabile Comitato di Programma), Ilenia Colonna, Patrizia Miggiano; Carola Gatto; Giulia Miglietta; Marco Giannotta; Alessia De Blasi, Isabella Hernandez.

Direttori di Area: Luca Bandirali; Mario Bochicchio; Fabio Ciracì; Roberto Rosselli Del Turco; Marco Mancarella; Grazia Semeraro.

Segreteria del Convegno: Dott.ssa Silvia Gravili
Resp. tecnico: Carlo Tafuro; web design: Dr.ssa Paola D'Amico; comunicazione: Dr.ssa Loredana De Vitis

Enti organizzatori / Organizing institutions:

AIUCD;

Università del Salento: Centro interdipartimentale in Digital Humanities in collaborazione con i corsi di laurea in Filosofia, DAMS, Beni Culturali e Digital Humanities; ISUFI, Scuola Placetelling.

Università degli Studi Aldo Moro, Dipartimento di Informatica

Sponsor

Regione Puglia; Provincia di Lecce; Città di Lecce; CINI – Consorzio Universitario Nazionale per l'Informatica; SFI-Società Filosofica Italiana; AFC - Apulia Film Commission, Teatro Pubblico Pugliese; Argo Software.

Lista dei revisori - List of the reviewers

Agnese Addone; Tommaso Agnoloni; Luca Bandirali; Nicola Barbuti; Andrea Bellandi; Armando Bisogno; Mario Alessandro Bochicchio; Andrea Bolioli; Federico Boschetti; Dominique Brunato; Paolo Buono; Dino Buzzetti; Marina Buzzoni; Luigi Catalani; Francesco Ceraolo; Daniele Chiffi; Simona Chiodo; Fabio Ciotti; Ilenia Colonna; Christian D'Agata; Elisa D'Argenio; Riccardo De Biase; Manuela De Giorgi; Daniela De Leo; Salvatore De Masi; Pierpaolo Del Coco; Angelo Mario Del Grosso; Francesca Di Donato; Giorgio Maria Di Nunzio; Federica Epifani; Daniela Fogli; Claudio Forziati; Greta Franzini; Francesca Frontini; Emiliano Giovannetti; Edmondo Grassi; Fabiana Guernaccini; Barbara Guidi; Pietro Luigi Iaia; Benedetta Iavarone; Fahad Khan; Maurizio Lana; Angelica Lo Duca; Donato Malerba; Marco Mancarella; Tiziana Mancinelli; Chiara Mannari; Valentina Marangi; Cristina Marras; Federico Meschini; Patrizia Miggiano; Giulia Miglietta; Paolo Monella; Giovanni Morrone; Serge Noiret; Deborah Paci; Antonio Pascucci; Enrico Pasini; Luigi Patrono; Igor Pizzirusso; Simone Rebora; Massimiliano Rossi; Daniela Rotelli; Enrica Salvatori; Eva Sassolini; Daria Spampinato; Rachele Sprugnoli; Enrico Terrone; Francesca Tomasi; Francesco Tommasi; Sara Tonelli; Gennaro Vessio; Marco Salvatore Zappatore.

Visualizing the genetic process of literary works

Valentina Pallacci[1], Andrea Benedetti[2], Tommaso Elli[3], Elena Spadini[4],

Michele Mauri[5], Daniel Maggetti[6]

[1]Politecnico di Milano, Italy - valentina.pallacci@mail.polimi.it
[2]Politecnico di Milano, Italy - andrea.benedetti@polimi.it
[3]Politecnico di Milano, Italy - tommaso.elli@polimi.it
[4]University of Lausanne, Switzerland - elena.spadini@unil.ch
[5]Politecnico di Milano, Italy - michele.mauri@polimi.it
[6]University of Lausanne, Switzerland - daniel.maggetti@unil.ch

ABSTRACT

The paper presents the preliminary results of a research aimed at designing a visual model to represent the genetic process of literary works, namely the network of texts (such as drafts, diary notes, and previously published materials) used by an author to develop a new work. In the paper, the visual model is described highlighting the main design choices, the design process, and its assessment with a domain expert. Both the underlying data model and the visual model aim to be general enough to offer possibilities for application to other cases of scholarly research.

TALK

1.   INTRODUCTION

In this paper we introduce the preliminary results of a research aimed at designing a visual model for studying the genetic process of literary works. The focus of the paper is the design process carried on in collaboration between communication designers specialized in data visualization, and scholars of the project "Gustave Roud. *Œuvres complètes*"[1].

Gustave Roud (1897-1976) was a Swiss poet, photographer, and translator, also active in arts and literary criticism. The genesis of Roud's works is grounded in his diary and is characterized by the reuse of already published materials. For the study of genetic processes, scholars consider multiple sources and examine the relationships between them. A particularly relevant concept in the fields of genetic criticism and *filologia d'autore* is the one of genetic dossier: a group of plans, sketches, drafts, and clear copies that testify to the project of a literary work ([7];[6]); they are identified by scholars interpreting the documents, and can be represented as a node in the network that stands between documents and publications.

Before the beginning of the collaboration with communication designers, scholars of the project "Gustave Roud. *Œuvres complètes*" developed a data model for literary genesis in the form of an OWL 2 ontology[2]. The data model is aimed at storing and representing the literary genesis in the form of a conceptual network of documents and publications ([5]). Presenting the data model is beyond the scope of this paper, which instead focus on the visual model.

In what follows, we pursue the design process (section 2), briefly introduce the assessment activity (section 3) and present future works in the conclusions (section 4).

2.   DESIGN PROCESS

The aim of the collaboration between the two partners is the creation of a model that enables the visual reading of the genesis of Roud's works. The visual model is built on top of the aforementioned data model and could be adapted to the study of other genetic processes. The design process was divided in three key moments: (section 2.1) definition of design requirements, (section 2.2) visualization process, and (section 3) validation of a preliminary outcomes with a domain expert.

---

[1] University of Lausanne, 2017-2021 (https://p3.snf.ch/Project-157970).

[2] The project ontology is available at. The project makes use of the framework DSP (DaSCH Service Platform, https://dsp.dasch.swiss/): DSP provides a general ontology, on top of which each project defines its own ontology. The part of the ontology devoted to literary genesis is documented at https://gen-o.github.io/ and at https://github.com/gustaveroudproject/geneticNetworksDataViz.

2.1 DESIGN REQUIREMENTS

Designers used structured interviews ([9]) to inform their design process. The method is commonly used to gather a rich and holistic understanding about the topic of a work of visualization ([4]). The interviews focused on Roud's production and on the work of the domain experts, that is the data modeling process and preliminary data representations (sketches and hand-made network visualizations). In addition to interviews, designers autonomously explored the data using an open-source tool aimed at the visual exploration of network graphs ([1]). The most important finding of these research activities regards the nature of the genetic dossiers: they support the understanding of the genesis of a work, but, at the same time, they are interpretations of scholars, which could potentially evolve with the sedimentation of their knowledge about Gustave Roud and with the growth of the author's archive.

The interviews and the exploration of the data lead to the identification of four design requirements for the visual model, which should:

R1. be designed for an audience of domain experts;

R2. explicit the role of genetic dossiers as an interpretive layer;

R3. differentiate between typologies of objects (diary entries, drafts, books and periodicals), clusters (genetic dossiers, and works separated into parts) and relationships (used in, part of, rewritten in);

R4. mediate between the complexity of the data model and the legibility of the networks.
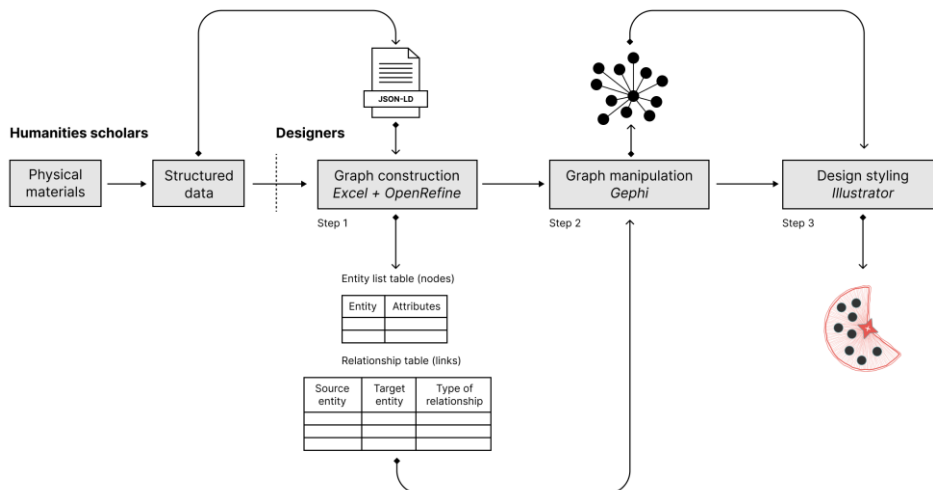
2.2 VISUALIZATION PROCESS

Figure 1. Schematization of the visualization process in relationship to the work done by humanities scholars. The process unfolds in three major steps, each with its own output.

The visualization process unfolds in three steps (fig. 1). Firstly, it was necessary to convert the data into a tabular format (step 1) to fit the selected tool for network visualization ([1]). Every genetic network constitutes of a JSON-LD file produced by scholars[3], and designers defined a process to convert them into two tables: one for nodes (publications, manuscripts, periodicals and their parts) and one for links (relationships among nodes, like rewriting and reuse). By taking into account the outcomes of preliminary explorations, designers decided to consider genetic dossiers not as nodes, but as sets of nodes, the center of which would be the resulting publication. Having a separate way of defining the genetic dossiers was crucial in simplifying the reading of the network (R4); in addition, it helped readers in understanding the existence of two layers: a factual one with publications and manuscripts, and an interpretive one with genetic dossiers (R2, fig. 2). This operation was implemented by bypassing the genetic dossiers nodes and directly connecting used materials to the publication that resulted from them.

---

[3] The data in JSON-LD are available at https://github.com/gustaveroudproject/geneticNetworksDataViz.
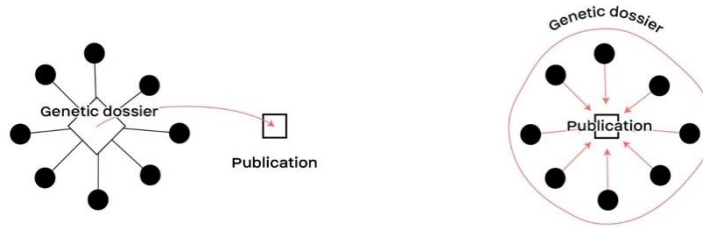
Figure 2. The shift from a node to an enclosure, to represent genetic dossiers.

Successively, designers spatialized the genetic networks using a force directed layout algorithm (step 2), which simulates a physical system in order to calculate the position of the elements of the networks ([8]). In the simulation, nodes repulse or attract each other according to the links that connect them. In addition, designers set a stronger pull between the nodes belonging to the same genetic dossier; they achieved this result by incrementing the weight value of the corresponding links. This kind of direct manipulation resulted in a compact positioning of the elements of the genetic dossiers, that are visually separated from nodes connected through different kinds of connections. This spatialization allowed researchers to perform a visual network analysis ([10]) and to identify clusters and outliers (R3), and observing the overall shape of the network (fig. 3).
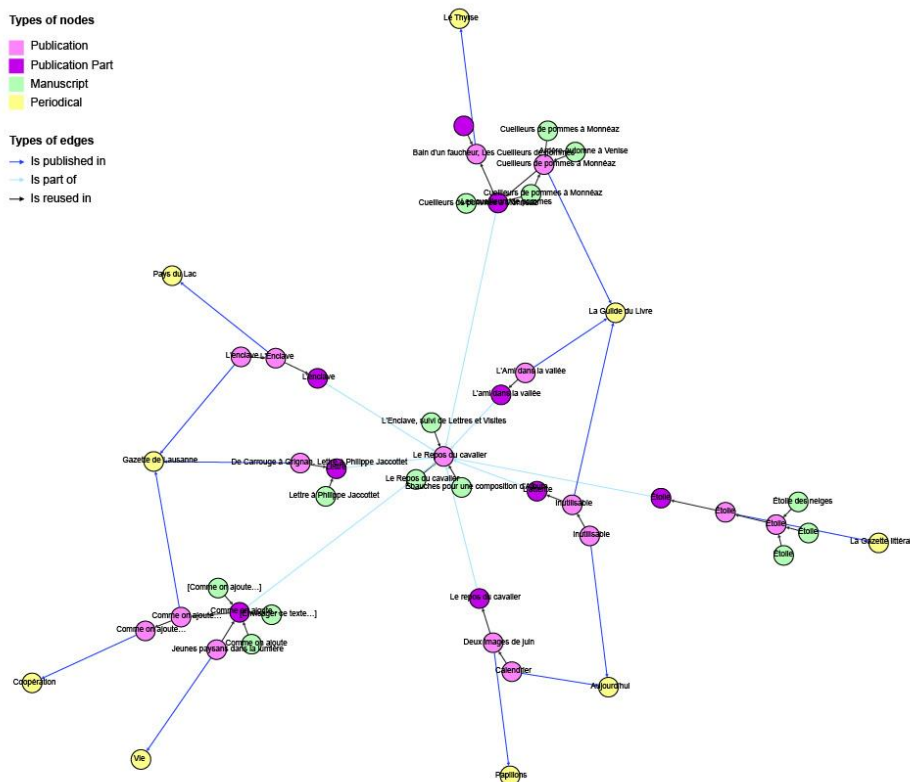

Figure 3. The result of the spatialization using *Force Atlas 2* (in this case the genesis of Roud's work *Le Repos du Cavalier*) is a visualization in which the quality of linkages determines the position of nodes. Different colored nodes identify different kinds of entities (books, articles, etc…) and different colored links identify different kinds of relationships (rewriting, reuse, publications…).

Lastly, designers reworked the network visualizations using a vector-based editing software (step 3). During this activity they applied a visual embellishment that employs a visual metaphor, following a consolidated strategy for improving comprehension and memorability of visualizations ([2]). Gustave Roud was often inspired by nature, countryside, or stars and, after having analyzed three different options, astral maps emerged as a suitable metaphor for supporting the creation of a visual language: they represent objects of different nature (stars), and they embed the human interpretations of stars arrangements (constellations). The metaphor inspired the form of nodes and connections (fig. 4) and allowed for the

integration of a layer of domain experts' interpretations (fig. 5). The genetic dossiers, indeed, were treated as an additional layer of the visualization, one that groups elements.
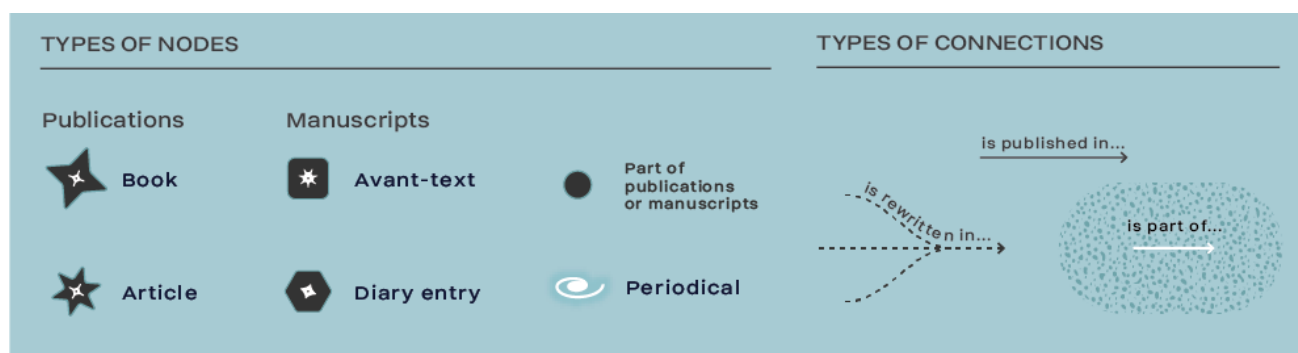

Figure 4. The legend of the visualization, divided between nodes and connections, with resemblances with stars, nebulae, and galaxies. Each typology of entity is represented by a different glyph, and each typology of connection is stylized in a different manner.

Figure 5. Representations of genetic dossiers. In the visualization, genetic dossiers are treated as an interpretive layer: they surround documents that converge in a publication.

3.   PRELIMINARY ASSESSMENT

Before starting the semi-automatic production of a larger number of visualizations, designers and scholars deemed important to assess the functioning of the visual model. The assessment activity is intended to inform another design iteration and to identify meaningful aspects on which to further work.

3.1   METHODS

The design outcomes were assessed with one of the scholars, creator of the data model and knowledgeable about the overall goals of the project, but who was never involved in any visualization activity. The scholar will be referred to as domain expert. The assessment uses three genetic networks[4] (one of them in fig. 6) and is based on a talk-aloud protocol ([3]) and on a task-oriented approach. The domain expert was asked to read aloud the network and enunciate aloud any comment that might occur. The domain expert was then asked to address specific tasks: (1) identify the central work of the visualization, (2) count the total amount of genetic dossiers, (3) identify the documents included in each dossier and the resulting publication, and (4) identify the periodicals in which the poet published.

3.2   OUTCOMES

The domain expert was able to read the visualizations almost entirely (R1, R3 and R4) and managed to distinguish genetic dossiers from other elements (R2). The result suggests that the design needs refinements but is overall solid. By looking at the networks, the domain expert was able to formulate hypothesis about the works genesis, especially on the role of periodicals in the selection of already published materials to reuse. The domain expert also found stimulating the possibility to move from the observation of the genesis of a single work, to the observation of general trends in Roud's way of working. Possible improvements that emerged are to: (1) **improve the structure of the genetic dossiers**, making it easier to identify its center and the elements belonging to them; (2) **differentiate between hierarchies of genetic dossiers**, since the expert identified genetic dossiers related to the main work and those related to further reuses; (3) **use chronology to clarify the reading order of the network**.

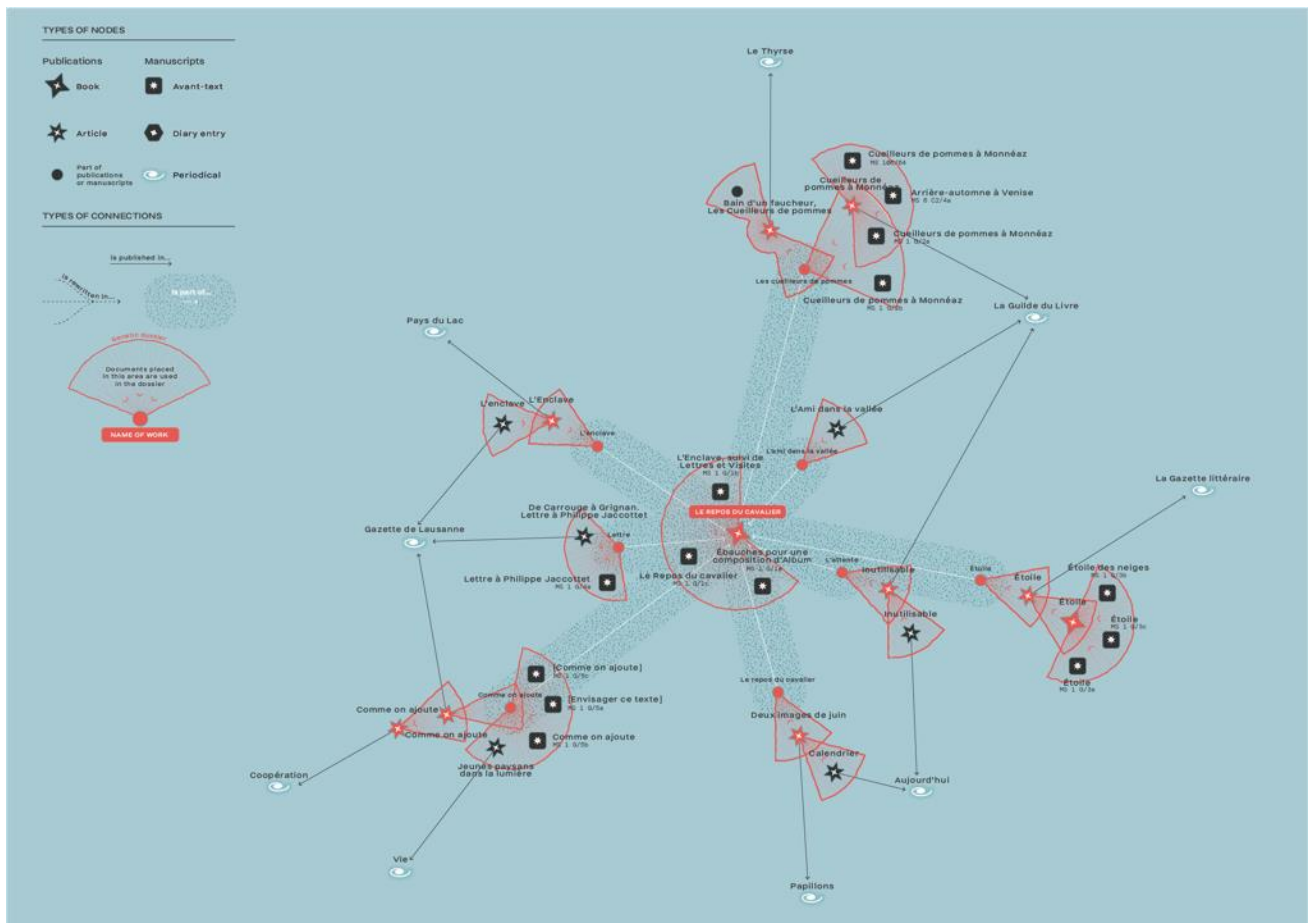[4] All design outcomes can be consulted at: https://bit.ly/roud-aiucd22

Figure 6. The network for *Le Repos du Cavalier*, the final result of the visualization process, and one of the design products that were assessed with the interview of a domain expert.

4. CONCLUSIONS

This paper documents on-going research for the design of a visual model aimed at visualizing the genetic process of literary works. The presented results will be further developed before being included into the digital edition of "Gustave Roud. *Œuvres complètes*". In the forthcoming design iterations, researchers will further develop the visual model by improving the representation of genetic dossiers, and by employing chronology to clarify the reading of the network. After having addressed those aspects, designers will dedicate some effort to the semi-automation of the visualization pipeline: convert the data, spatialize the network nodes, and partially apply the visual embellishment. The activity will likely result into an online and open-source code notebook[5] to serve as a tool for other scholars, or to be forked and being the basis for contributions and collaborations. The final visualizations can be made available online or be printed to directly serve in the archive of the author in combination with its catalog; the presence of the visualizations will help in guiding researchers in examining the documents. The most generalizable aspects of the research are the data model and the visualization pipeline, that can be both applied beyond the specific case to other scholarly studies.

REFERENCES

[1] Bastian, M., S. Heymann, and M. Jacomy. 2009. "Gephi: An Open Source Software for Exploring and Manipulating Networks." In *Proceedings of the International AAAI Conference on Web and Social Media*, 3:361–62.

[2] Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. "Gephi - The Open Graph Viz Platform." 2009. https://gephi.org/.

[3] Bateman, Scott, Regan Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. "Useful Junk?: The Effects of Visual Embellishment on Comprehension and Memorability of Charts." In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 2573. Atlanta, Georgia, USA.

[4] Boren, Ted, and Judith Ramey. 2000. "Thinking Aloud: Reconciling Theory and Practice." *IEEE Trans. Profess. Commun.* 43 (3): 261–78.

---

[5] For examples see https://observablehq.com/

[5] Carpendale, Sheelagh. 2008. "Evaluating Information Visualizations." In *Information Visualization*, A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, 4950:19–45. Berlin, Heidelberg: Springer Berlin Heidelberg.

[6] Christen, Alessio, and Elena Spadini. 2019. "Modeling Genetic Networks. Gustave Roud's Œuvre, from Diary to Poetry Collections." *Umanistica Digitale,* 7 (December).

[7] De Biasi, A. 2004. "Toward a Science of Literature: Manuscript Analysis and the Genesis of the Work." In *Genetic Criticism: Texts and Avant-Textes*, J. Deppman, D. Ferrer, and M. Groden, 26–68. Philadelphia: University of Pensylvania Press.

[8] Grésillon, Almuth. 1994. *Éléments de Critique Génétique: Lire Les Manuscrits Modernes.* Paris: Presses universitaires de France,.

[9] Jacomy, Venturini, Heymann, and Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE* 9 (6): e98679.

[10] Seidman, Irving. 2006. *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. 3rd ed. New York: Teachers College Press.

[11] Venturini, Tommaso, Mathieu Jacomy, and Pablo Jensen. 2021. "What Do We See When We Look at Networks: Visual Network Analysis, Relational Ambiguity, and Force-Directed Layouts" *Big Data & Society,* 8 (1).