The Art of Relations

Santini, Cristian

cristian.santini@fiz-karlsruhe.de
FIZ Karlsruhe – Leibniz-Institut für
Informationsinfrastruktur, Germany; Institute for Applied
Informatics and Formal Description Methods (AIFB),
Karlsruhe Institute of Technology, Germany

Garay, Nele

usjpw@student.kit.edu
Karlsruhe Institute of Technology, Germany

Posthumus, Etienne

etienne.posthumus@partners.fiz-karlsruhe.de
FIZ Karlsruhe – Leibniz-Institut für
Informationsinfrastruktur, Germany
ORCID: 0000-0002-0006-7542

Sack, Harald

harald.sack@fiz-karlsruhe.de
FIZ Karlsruhe – Leibniz-Institut für
Informationsinfrastruktur, Germany; Institute for Applied
Informatics and Formal Description Methods (AIFB),
Karlsruhe Institute of Technology, Germany
ORCID: 0000-0001-7069-9804

Introduction

Networks are everywhere. The brain is a network of neurons, proteins interact in cellular networks, and social networks shape our lives. Network science, i.e. "the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena" (National Research Council, 2005), provides a methodological framework to understand the complex interconnections that drive the behaviour of these systems, unveiling fundamental principles that govern their structure, dynamics, and function. Despite the fact that the concept of social networks became popular nowadays with the advent of social media, the representation of social dynamics in networks of relations provides invaluable insights into the way communities have formed and thrived throughout history. This research focuses on the realm of art-history and draws upon the seminal work "The Lives of the Artists" (1550) by Giorgio Vasari in order to examine the potential of network science to analyse this fundamental historical resource, by leveraging State-Of-The-Art (SOTA) Natural Language Processing (NLP) techniques.

Motivation

"The Lives of the Artists" is considered as the first art-historical biography ever written, consisting of a narration about lives, endeavours and social engagements of Italian and European artists living from the XIIIth to the XVIth century. This book, originally published by the Italian painter in 1550, has not only been considered for centuries as the main bibliographic source for the history of Italian and European visual arts (painting, sculpture and architecture), but it also allows to trace, through the eyes of this historian, the web of social influences and collaborations which shaped the environment in which artists worked at that time. In this context, network theory can play a major role: artists' proactiveness towards collaboration or their degree of influence in Vasari's reportage could be analysed with purely mathematical methods.

As discussed by Franco Moretti (2013), computational methods are opening scenarios for the theory of literature - and generally for the whole humanities - which were previously unimaginable. Quantitative methods from network theory can be employed to interpret cultural heritage documents not only formally but in order to grasp large-scale informative patterns, in an approach that Moretti calls "Distant Reading". In the realm of network analysis and art-history, Kienle (2017) showcases promises and challenges of graph theory for this domain. The article argues for the integration of quantitative approaches with established theories, emphasising the potential of network analysis to reveal marginalised actors and transnational histories, while cautioning against reinforcing social hierarchies if data principles are not critically examined. Lincoln (2020) discusses the deployment of computational network analysis by art historians across periods and cultures, emphasising the appeal of network rhetoric. It highlights the impact of network analysis in characterising individual positions and showing temporal changes, recognizing the challenges of integrating these methods into art-historical interpretation. Bishop (2018) instead critically addresses digital technology and the computational turn in contemporary art history. Divided into two parts, it connects issues with "digital art history" to neoliberal metrics and suggests a critical deployment of "distant reading" in analysing contemporary art, providing an alternative perspective on the role of digital methods in the field. Aim of our research is to exploit the rich bibliographic material contained in Vasari's magnum opus in order to represent the social context narrated in "The Lives of the Artists" with network theory, thus offering a novel approach to historiographers to dig deeper into historical resources and get quantitative features in order to support their study.

Methodology

This paper presents a novel approach for the extraction of social networks from historical books by exploiting Indices

of Names, sections of books which list in which page persons or other named entities are referenced. This approach leverages a mixture of data mining and natural language processing techniques in order to automatically identify these references from unstructured data. In our research, we use as source a publicly available English translation of "The Lives of the Artists" (Vasari, 2018) which provides an Index of Names of the artists at the end of the book. In a first data collection phase, a scraping algorithm converts the Index of Names in the web page into a JSON file in which artists names are keys and the list of pages in which these names are mentioned are the values of these keys. Furthermore, the book sections are divided into paragraphs and stored in a CSV file, where each paragraph is associated with a page number.

In a second step, SOTA NLP techniques are used to identify references to persons in the text. The paragraphs stored in this CSV are processed with a Named Entity Recognizer (Yamada et al., 2020) in order to spot entities which the algorithm classifies as persons. Moreover, paragraphs with the same page number are put together and processed with a Coreference Resolver (Otmazgin et al., 2020) in order to cluster pronouns and names which are presumed to belong to the same entity. Results from these two algorithms are combined in order to identify all pronouns and names which refer to entities categorised as persons and group them. Finally, the Index of Names is used to resolve these clusters to specific artists by looking in the index to names present in a certain page. In order to build a social network out of textual references, artists are connected if pronouns or nouns referring to them are occuring in the same paragraph.

The use of textual co-occurrences is of big interest for literary studies, even though it is not unproblematic (Bourgeois et al., 2015). Even though the co-occurrence of two entities in a paragraph is not a sufficient condition to infer some kind of relation among the two, it can be an index of their interconnectedness inside a document. This is because two entities which tend to co-occur together in several textual sections, such as a paragraph, are likely to share some kind of relation in the document. By using Pointwise Mutual Information (PMI) (Church et al., 1990), a measure from information theory, it is possible to estimate the degree of association for a suspected relation between two artists based on their paragraph co-occurrence. PMI is a measure of association for two events A and B which acquires a positive value if two events occur together more than would be expected by their independent probability. In this way, a positive PMI for the co-occurence of two artists inside paragraphs demonstrates a relation among the two which is not given by pure chance and can be used to infer a connection, which will be therefore represented as a weighted edge between two nodes in a social network.

Results

The social network extracted from "The Lives of the Artists" is designed in order to provide a heuristic tool for art-

historians and digital librarians who aim to enrich any digital edition of this art-historical document. By making use of social networks, researchers can dig deeper into the historiography of Western Art from the XIIIth to the XVIth century. With our method, 10 networks are obtained, one for each volume of the English translation of Vasari, which can be visualised online (Santini et al., 2023a). Additionally, by using Linked Open Data (LOD) vocabularies, Vasari's social network is also available as a Knowledge Graph (KG) (Santini et al., 2023b) in order to provide a valuable instrument for the Semantic Web community. By using links between artists in our KG and Wikidata entities, provided with the *rdfs:seeAlso* property, it is possible to enrich the information present in Wikidata with results obtained from our network analysis, such as centrality scores or PMI values of a pair of artists.
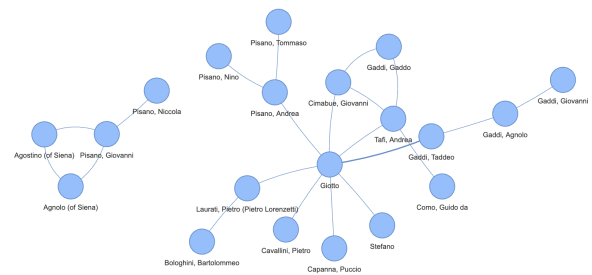

Figure 1: Portion of the extracted social network for the first volume of "The Lives of The Artists" (Vasari, 2008). Giotto's network and distinct components are clearly visible.

Furthermore, it is possible to see how high PMI values in the network are often linked to artistic collaborations, such as that of Piero di Cosimo and Cosimo Rosselli or that of Leonardo da Vinci and Verrocchio. Moreover, by looking at the centrality of certain nodes, we noticed how these are heavily influenced by the role which artists covered in Vasari's reportage of historical events. For example, the most active and influential artists, such as Michelangelo, Giotto, Da Vinci and Tiziano have high betweenness centrality, showing their central role in shaping the art scene of this period and being the masters of several other protagonists of the book.

Conclusion

In conclusion, this paper aims to provide a generalizable methodology to extract social networks from historical biographies of artists by using as case study the seminal work of Giorgio Vasari. By exploiting entity extraction techniques in tandem with an Index of Names, our approach shows how to extract a social network out of in-text co-occurrences of entities. Main contribution of our work is to make available results from our network analysis and provide interactive network visualisations for all 10 volumes of "The Lives of the Artists": this network will allow rese-

archers in art-history to address relevant questions for the understanding of this book in a new multidisciplinary way by leveraging network theory in art-historical studies.

Finally, by converting Vasari's social network into a KG, in the end, network analysis is integrated with powerful exploratory tools and a query language, i.e. SPARQL, which allows domain-experts to better envision how computational techniques can reshape theories and methodologies not only in art-history but in the whole humanities domain. In order to favour the reproducibility of the studies carried in this research, all the data collected and processed to support the statements of this paper are made available on Zenodo (Santini, 2023a) (Santini, 2023b).

Bibliographie

Bourgeois, Nicolas, Marie Cottrell, Stéphane Lamassé, and Madalina Olteanu. 2015. "Search for Meaning Through the Study of Co-Occurrences in Texts". In Advances in Computational Intelligence, 578–91. Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-19222-2_48.

Bishop, Claire. 2018. "Against Digital Art History". International Journal for Digital Art History, no. 3 (July). https://doi.org/10.11588/dah.2018.3.49915.

Church, Kenneth Ward, and Patrick Hanks. 1990. 'Word Association Norms, Mutual Information, and Lexicography'. Computational Linguistics 16 (1): 22–29.

Kienle, Miriam. 2017. "Between Nodes and Edges: Possibilities and Limits of Network Analysis in Art History". Artl@s Bulletin 6 (3). https://docs.lib.purdue.edu/artlas/vol6/iss3/1.

Lincoln, Matthew D. 2020. "Tangled Metaphors: Network Thinking and Network Analysis in the History of Art". In The Routledge Companion to Digital Humanities and Art History. Routledge.

Moretti, Franco. 2013. "Distant Reading". Verso Books.

National Research Council. 2005. "Network Science". The National Academies Press. https://doi.org/10.17226/11516.

Otmazgin, Shon, Arie Cattan, and Yoav Goldberg. 2022. "F-Coref: Fast, Accurate and Easy to Use Coreference Resolution". In Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: System Demonstrations, 48–56. Association for Computational Linguistics. https://aclanthology.org/2022.aacl-demo.6.

Santini, Cristian. 2023. "Data for Vasari Social Network". https://doi.org/10.5281/zenodo.8395369

Santini, Cristian. 2023. "Results for Vasari Social Network". https://doi.org/10.5281/zenodo.8395425

Santini, Cristian, Etienne Posthumus, and Harald Sack. 2023. "Vasari Social Network". https://ISE-FIZKarlsruhe.github.io/vasari_network(accessed: November 13, 2023).

Santini, Cristian, Etienne Posthumus, and Harald Sack. 2023. "Vasari Knowledge Graph" https://ise-fizkarlsruhe.github.io/vasari_network/rdfs/vasari-kg.ttl(accessed: November 13, 2023).

Vasari, Giorgio. 2008. "Lives of the Most Eminent Painters Sculptors and Architects, Vol. 01 (of 10)Cimabue to Agnolo Gaddi". Translated by Gaston du C. De Vere. https://www.gutenberg.org/ebooks/25326. (accessed: November 13, 2023).

Yamada, Ikuya, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. "LUKE: Deep Contextualized Entity Representations with Entity-Aware Self-Attention". In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 6442–54. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.523.