Network Analysis Report on the Digital Humanities Research Field at the University of Bologna

Collaborators:

Carla Menegat, Digital Humanities and Digital Knowledge, 0001112209

Chiara Martina, Digital Humanities and Digital Knowledge, 0001139532

Ekaterina Krasnova, Digital Humanities and Digital Knowledge, 0001121358

Lucrezia Pograri, Digital Humanities and Digital Knowledge, 0001119705

Rubens Fernandes Jr., Digital Humanities and Digital Knowledge, 0001120856

1. Introduction and Context

The field of Digital Humanities (DH) is evolving rapidly, bringing together interdisciplinary research efforts that combine computational methods with humanities scholarship. The University of Bologna has long been a prominent hub for interdisciplinary research, fostering collaborations and advanced projects in areas such as the Semantic Web, Information Science, Knowledge Organization, Digital Scholarly Editing, Digital Philology, Intertextuality, Creative Industries, and Linked Open Data. This study examines Bologna's pivotal role in advancing social science methodologies within DH, with a particular emphasis on the development of its academic environment. We focus specifically on dynamics among members of the Digital Humanities Advanced Research Centre (DH.ARC) and the faculty of the Digital Humanities and Digital Knowledge (DHDK) master's degree.

Collaboration is essential to academic research, influencing knowledge production through joint projects, co-authorship and shared keywords. This report provides an analysis based on the following networks: (1) the co-authorship network, (2) the project collaboration network, and (3) the keyword co-occurrence network. The aim is to gain a comprehensive understanding of academic interaction and research influence within the field.

2. Problem and Motivation

The Digital Humanities community at the University of Bologna brings together researchers from multiple disciplines, including Linguistics, Cultural Heritage, Textual Scholarship, Library/Archive Science, Information Science, and Computer Science. Understanding the structural differences between different collaboration modes – e.g., co-authorship, research

projects, and thematic connections – is crucial for optimizing academic productivity and knowledge dissemination.

This study investigates how researchers within DH.ARC research center and the DHDK Master's program collaborate internally and externally. We examine co-authorship patterns, project collaborations, and their impact on research output. Key research questions include:

- How do co-authorship and project collaborations compare within DH community?
- Do DH.ARC scholars collaborate extensively with external researchers?
- Are project-based collaborations reflected in co-authorship patterns, or do they exist separately?
- What recurring keyword patterns can be identified in the field, and how do they reveal dominant research themes and emerging trends within the Digital Humanities scientific community at the University of Bologna?

To understand the dynamics of DH research network in Bologna, we analyze three primary structures: (1) the co-authorship network, based on publications from the timeframe 2017-2023[1]; (2) the projects collaboration network, based on active DH.ARC-funded projects[2]; and (3) the keyword co-occurrence network. By examining trends in co-authorship, citations, and research keywords, we aim to uncover how ideas spread, communities form, and influential topics emerge within Bologna's DH ecosystem. This analysis provides insights into the evolution of DH research at the University of Bologna and its global collaborations.

3. Datasets

To create the datasets, we consulted DH.ARC platform to identify researchers, DHDK professors, and relevant collaborators. We then manually compiled two JSON datasets: the first, `authors`, containing general data on the participants, and the second, `projects`, documenting the research initiatives conducted at the center.

`Authors`

- ORCID_ID: ORCID unique identifier of the participant.
- name: Name of the participant.
- type: Role of the participant on DH.ARC/DHDK.
- position: Academic/professional status of the participant.

[1] The establishment of the Digital Humanities and Digital Knowledge Master's course, which boosted the development of the existing DH.ARC research center, as evidenced by the project release dates on the website: https://centri.unibo.it/dharc/en/research/projects-at-dh-arc. Additionally, the first academic year of the DHDK course dates to 2017, as indicated in the Course Structure Diagram webpage: https://corsi.unibo.it/2cycle/DigitalHumanitiesKnowledge/course-structure-diagram.

[2] See more: https://centri.unibo.it/dharc/en/research/projects-at-dh-arc.

- area: Field of expertise of the participant.
- affiliation: Institution the participant is associated with.

`Projects`

- Project id: Numeric ID assigned to each project.
- Name: Name of the project.
- Release year: Year when the project was public (if there were two, the last one).
- Description: Short explanation of the project's purpose.
- Participants: List of participants involved in the project.

We used the Current Research Information System (CRIS) platform to analyze the participant's publications and extract the necessary data published between 2017 (the year of DH.ARC's foundation) and 2023, exporting it in BibTeX format. The extracted information was then structured into a JSON dataset, `publications`, following the schema:

`Publications`

- Type: The category of the publication (e.g., Conference Paper, Journal Article).
- Author: A list of contributing researchers.
- Title: The title of the publication.
- Year: The year of publication.
- Book title / Journal / Volume: The conference proceedings or journal where the work was published.
- Keywords: Relevant topics and research areas covered in the publication.
- URL/ DOI / ISBN: Unique identifiers or links for accessing the publication.
- Pages: The page ranges within the publication source.

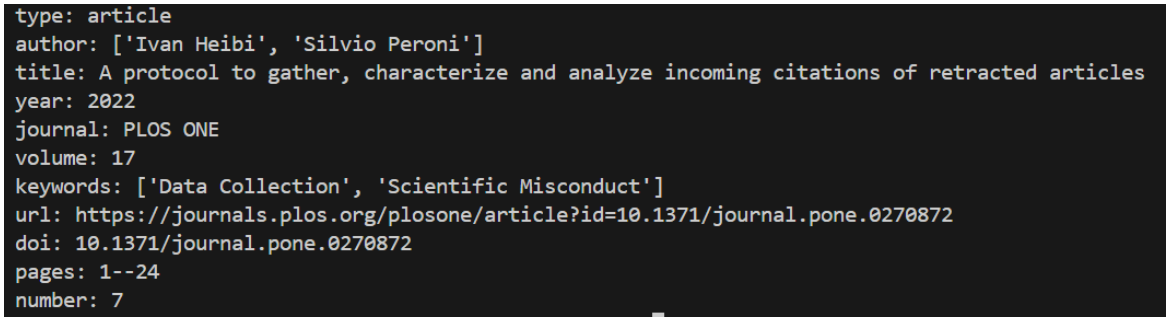


Figure 1: Attributes of publications dataset.

Within `publications`, to ensure accuracy and consistency, we automated and manually corrected data discrepancies, including different encodings of special characters, variations on the writing of author names and keywords, and the removal of entries lacking essential metadata.

Additional clarification on keywords processing: Keywords were split by commas and semicolons for uniformity, and the most common keywords in Italian, such as *Intelligenza Artificiale*, *Filologia italiana, Filologia digitale,* among others, were translated into English to maintain a standardized vocabulary. Also, the *Ontology* keyword entries have been transformed into *Ontologies*. Some overly general keywords - *Digital Humanities*, *Software,* and *Computer Science* - were filtered out. Each publication entry has been uploaded into the graph using its unique identifier, selected from the first available entry among these attributes: DOI, URL, ISBN, ID, or title. Publications lacking any of these required attributes were automatically excluded. The Python script used for network creation is available at this link.

Finally, we divided the JSON file by year to enable a structured and chronological analysis of publication trends. The datasets were processed using Python, NetworkX, Pandas, and Matplotlib.

We used Gephi to produce the visualizations with some personal style differences (specially colors), but in general the configurations were: Force Atlas 2 for layout (with the use of Noverlap, Label Adjust and OpenOrd for better visualization in the two biggest networks); in statistics we run all the metrics representations (Average Degree, Average Weighted Degree, Network Diameter, Graph Density, Modularity, Eigenvector Centrality and Average Clustering Coefficient); we use Modularity visualization with colors to identify different communities; and adjusted the gravity to have a better visualization in the different sizes networks. For the visualizations of publications by year, color gradations were used for all computed metrics.

4. Validity and Reliability

This study reflects the structure of Digital Humanities research collaborations at the University of Bologna, focusing on DH.ARC and the DHDK program. However, since the analysis includes both internal and external collaborators, some authors might not appear in all datasets, which could limit the overlap between different collaboration networks. Additionally, the study has the hypothesis that project-based collaboration should translate into publication collaboration, which may not always be the case. The results are reliable and reproducible, as they were generated using open-source tools and structured datasets and all the scripts and their resulting CSV and graph files are accessible in an open repository.

5. Measures and Results

To analyze the structure of the Digital Humanities research network at the University of Bologna, using the three datasets we established three networks and a subset to analyze:

1. The Complete Co-Authorship Network

2. The Filtered Co-Authorship Network (DH.ARC and DHDK researchers only, subset of the previous)
3. The Project Collaboration Network
4. The keywords network

The centrality metrics used are:

- Degree Centrality: Measures the number of direct connections an author has
- Betweenness Centrality: Determines how often a researcher serves as a bridge between distinct groups
- Closeness Centrality: Reflects how quickly a researcher can reach all others in the network
- Eigenvector Centrality: Evaluates influence by considering both an author's connections and the importance of their collaborators
- Clustering Coefficient: Assesses the local cohesiveness of collaboration

5.1 Complete Co-Authorship Network Analysis

This network (based on data from *publications*) consists of 1,090 researchers and 7,820 collaborations, representing all academic publications from 2017 to 2023 involving at least one researcher from DHDK or DH.ARC. Individual publications were not included.

- Average degree centrality: 0.011 (max: 0.129)
- Average closeness centrality: 0.201
- Average clustering coefficient: 0.90

These results indicate that while some researchers have extensive collaborations, many only have a few co-authors. The high clustering coefficient suggests strong connectivity within smaller research groups.

The bar graph shows the distribution of author collaborations by number of collaborations. Just a little group of researchers have more than 30 collaborations, reflecting that this little group is more stablished in their academic life.

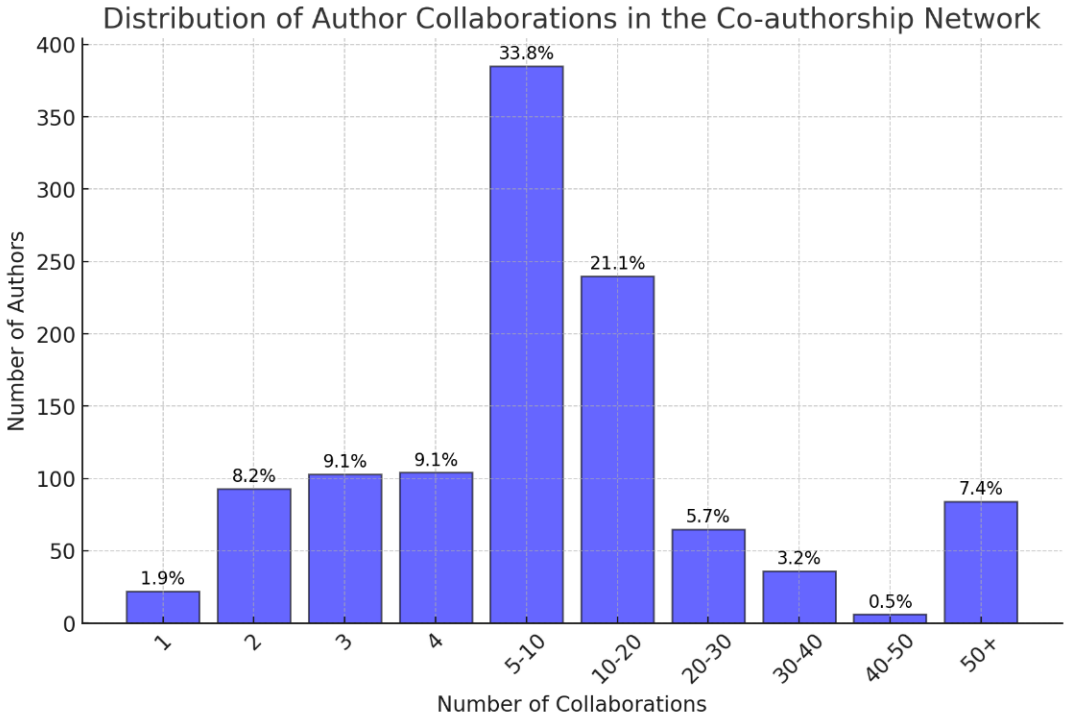


Figure 2: Bar Chart of the distribution of authors collaborations by numbers of collaboration

In this network, the most published authors include Aldo Gangemi, Valentina Presutti, Silvio Peroni, and Michela Milano, among others, all of them Professors in the University. This ranking reflects the researchers who have the most extensive publication collaborations across the entire network. Also, the high number of authors with less than ten collaborations (62,1% of the network) shows that this network is very young, and the groups are not totally established.

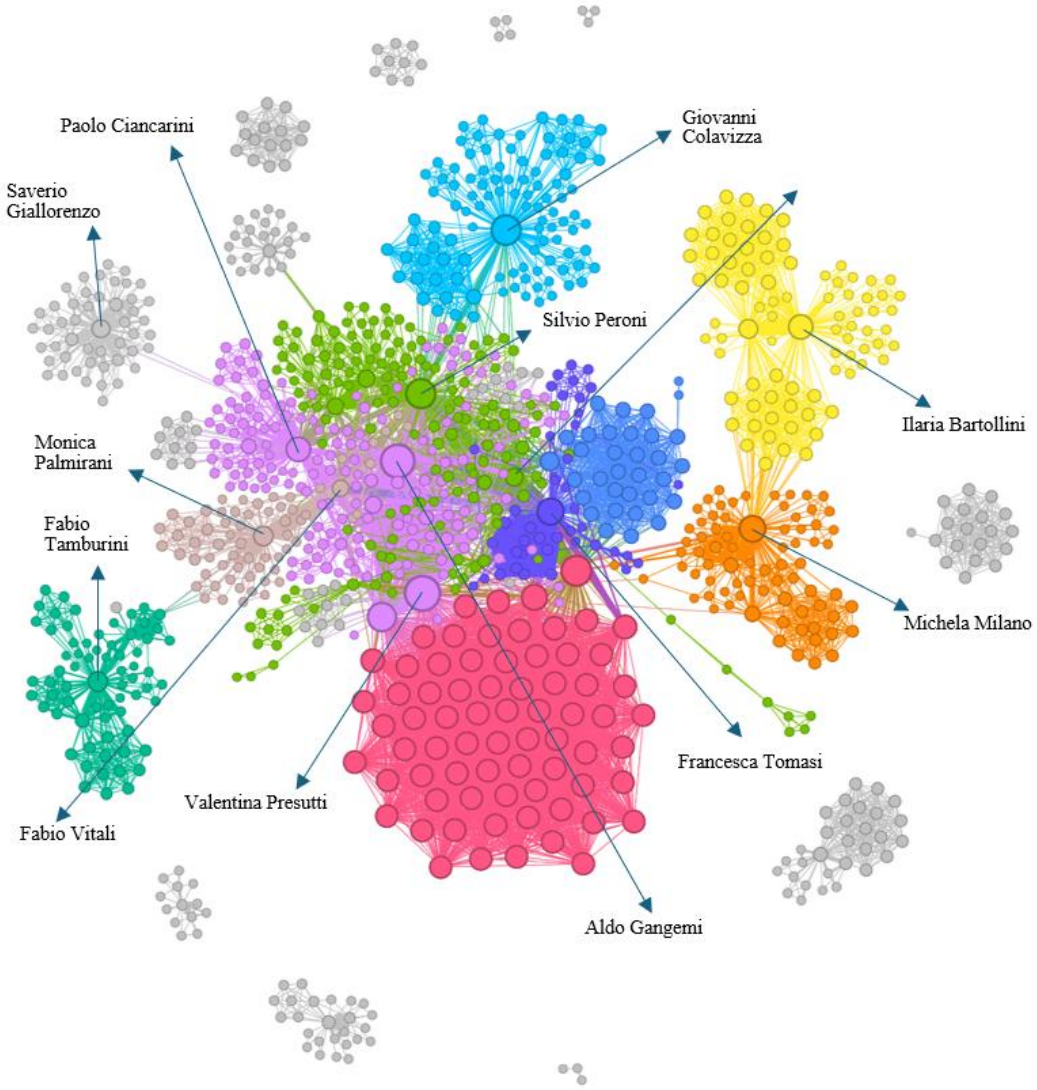
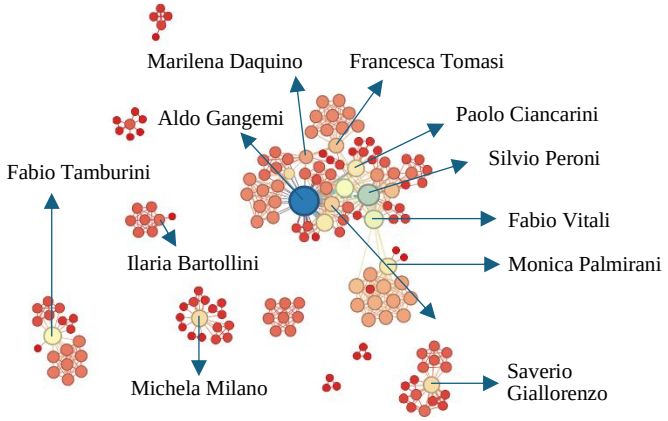


Figure 3: Spatial distribution by degree centrality in the co-authorship network with color identification of sub-communities of researchers and circle size of degrees.

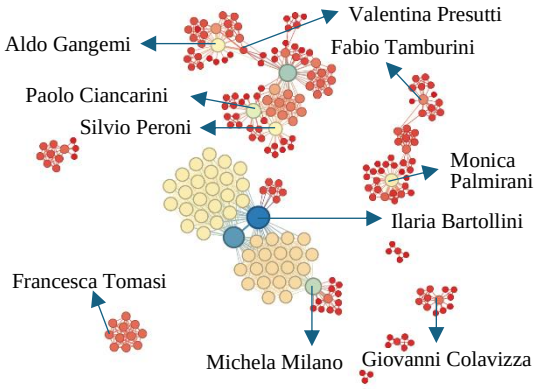
5.1.1 Analysis of co-authorship over the years

The following networks also use the *publications* dataset, divided by year for visualization and analysis purposes. The key modularity participants shown in *Figure 3* are also highlighted in these networks.

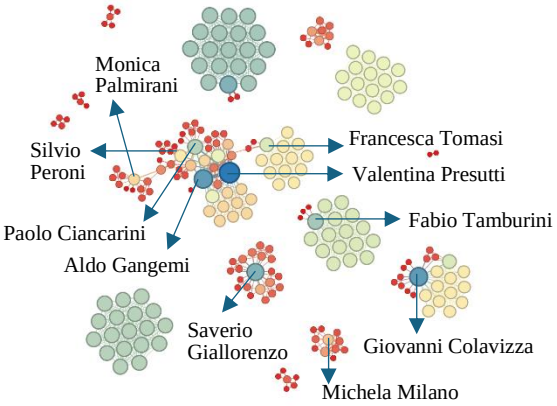
2017



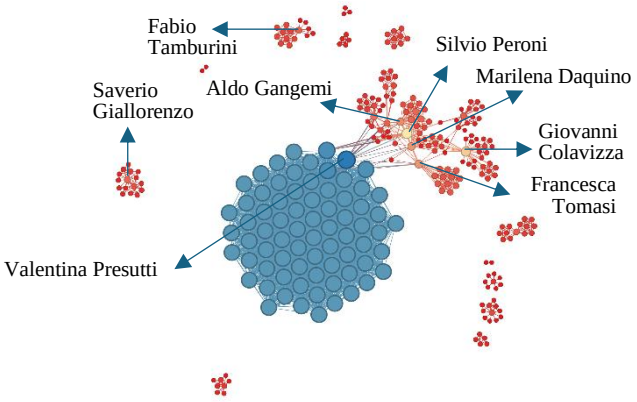
2018



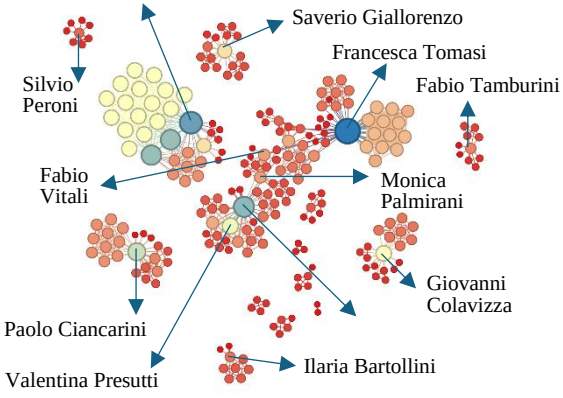
2019



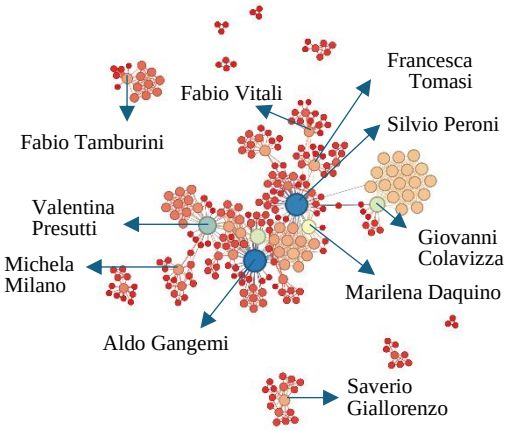
2020



2021



2022



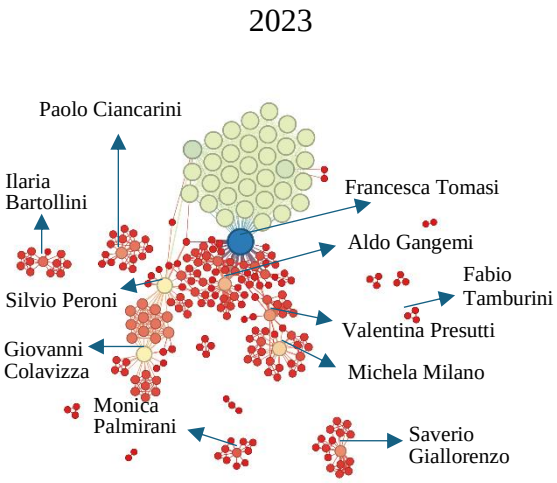


Figure 4: Spatial distribution by degree centrality in the co-authorship network per each year with color gradation for publications from red (lower quantity) to blue (higher quantity).

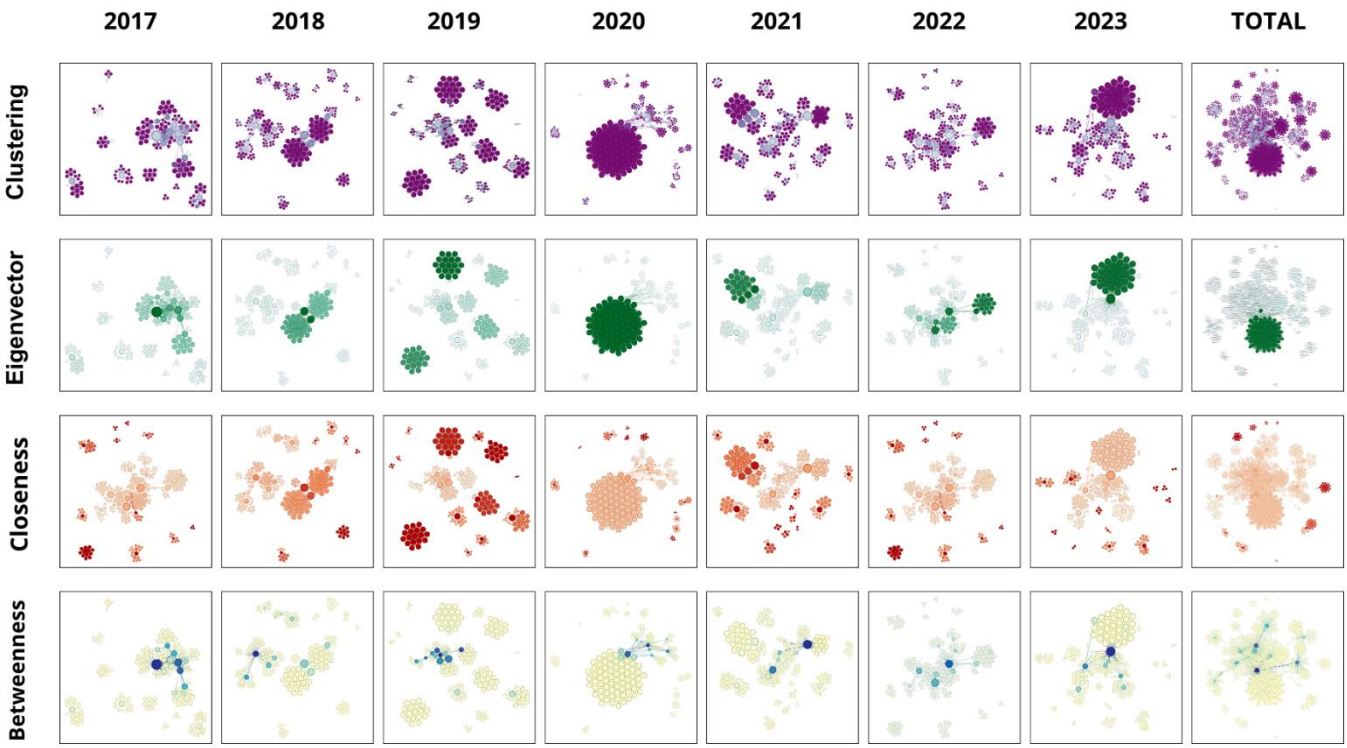


Fig.5 Spatial distribution of metrics used for co-authorship analysis.

The metrics based on the relative influence of nodes (*table 1*) reveal functions participants perform which might not be immediately apparent in the graphical representation. Among them, only the names in *betweenness* are ranked, while in the other metrics, the names share the same value.

	Avg. Degree	Clustering	Eigenvector	Closeness	Betweenness
2017	6,571	0,907	Aldo Gangemi	Fabio Tamburini Michela Milano, Saverio Giallorenzo	Aldo Gangemi Silvio Peroni Fabio Vitali
2018	9,537	0,927	Ilaria Bartolini	Francesca Tomasi Giovanni Colavizza Ilaria Bartolini	Paolo Canciarini Aldo Gangemi Angelo di Iorio
2019	9,272	0,916	Daniele Donati	Ilaria Bartolini Fabio Tamburini Giovanni Colavizza	Valentina Carriero Francesca Tomasi Fabio Vitali
2020	23,532	0,936	Valentina Presutti	Ilaria Bartolini Saverio Giallorenzo Michela Milano	Silvio Peroni Valentina Presutti Giovanni Colavizza
2021	7,195	0,919	Michela Milano	Fabio Tamburini, Giovanni Colavizza Saverio Giallorenzo	Francesca Tomasi Aldo Gangemi Fabio Vitali
2022	6,838	0,917	Giovanni Colavizza	Saverio Giallorenzo Ilaria Bartolini Fabio Tamburini	Silvio Peroni Aldo Gangemi Valentina Presutti
2023	8,925	0,909	Francesca Tomasi	Fabio Tamburini Ilaria Bartolini Paolo Bonora	Francesca Tomasi Silvio Peroni Valentina Presutti

Table.1 Metrics used for co-authorship analysis.

We can verify *Fabio Vitali*'s betweenness role as an intermediary (a *broker*, in anthropological means) who, despite not having many co-authored publications, performs a significant part in linking the groups; The recurrence of Francesca Tomasi, Silvio Peroni, Aldo Gangemi, and Valentina Presutti as the most connected participants in the publication network forms a core conglomerate. These actors also exhibit high relative centrality, as indicated by the Eigenvector metrics. However, in some years, individuals who made numerous publications in more isolated groups but interacted with this core, such as Michela Milano and Giovanni Colavizza, showed greater relative influence. Finally, through closeness, we observed that some participants belong to isolated groups, where interactions within the network occur through intermediaries, but they do not directly collaborate with the core participants, such as Saverio Giallorenzo, Ilaria Bartolini, and Fabio Tamburini.

Overall, the analysis of the co-authorship network from 2017 to 2023 shows a significant shift in academic collaborations, particularly among those who are not part of the core conglomerate participants.

Until 2019, publications were distributed in small clusters of authors, with few connections between them, suggesting more localized and specialized collaborations. In 2020, we observed a drop in productivity among DHDK professors due to pandemic restrictions. However, there was an event in publications among international students, some of whom were participants that generated significant centrality, which influenced the “complete metrics” (2017-2023). After 2021, the network became fragmented again, with greater diversification of

connections between authors, reflecting a gradual return to more distributed research activities. In 2023, a higher concentration of authors was observed compared to the previous two years, a phenomenon that can be explained by several factors: first, the natural delays in academic publishing led to research started in previous years becoming more visible only in 2023; second, the post-pandemic recovery may have encouraged large-scale collaborations, with a greater involvement of researchers in joint projects.

5.2 Filtered Co-Authorship Network Analysis

This network consists of 57 authors, representing project researchers who have co-authored publications and are professors of DHDK or members of DH.ARC.

- Average degree centrality: 5.89
- Average betweenness centrality: 0.0019
- Clustering coefficient: 0.49

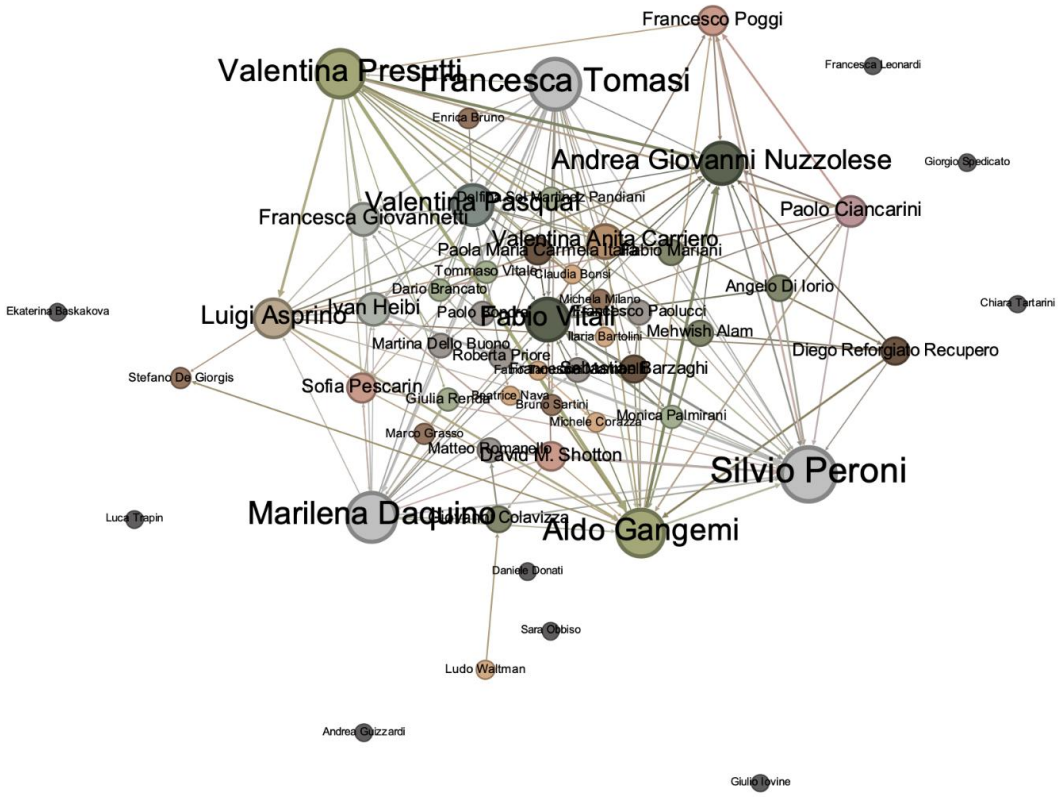


Figure 6: Spatial distribution by degree centrality of the co-authorship network, with color identification of sub-communities of researchers.

Compared to the complete network, the filtered network is more centralized, with fewer researchers dominating collaborations. The lower clustering coefficient reflects smaller, more isolated groups. Also, we can see that a small group of individuals, around 5,7%, do not collaborate with in this subset. This reflects the character of multidisciplinary in the DH field.
In this filtered network, the rankings shift slightly, with Silvio Peroni, Francesca Tomasi, and Marilena Daquino emerging as top authors. This change in rankings highlights how the focus on project-related publications alters the collaboration landscape since these three individuals collaborate largely with each other.

5.3 Projects Collaborations Network Analysis

This network consists of 56 authors, representing project researchers who have co-authored publications. The data for this analysis is localized in the `project` file and was gathered from DH.ARC website.

- Average degree centrality: 0.099
- Average betweenness centrality: 0.014
- Average clustering coefficient: 0.48

The data is very close to the subset of authors in Co-authorship, indicating that a lot of publications are derived from the projects that DH.ARC developed in the period.

The project network (*Figure 7*) is more selective, with many researchers engaging in only a small number of partnerships. The lower clustering coefficient compared to the co-authorship networks indicates that project teams are more fragmented. Also, we can see that people in the group of DH.ARC collaborate with each other on a more permanent basis and the group of Professors of DHDK that were not in DH.ARC do not have collaborations in this approach of the networking.

In the filtered network, rankings shift slightly, with *Silvio Peroni, Francesca Tomasi, Marilena Daquino, Valentina Presutti, and Aldo Gangemi* emerging as the most connected researchers. This change highlights how focusing on project-related publications alters the collaboration landscape.

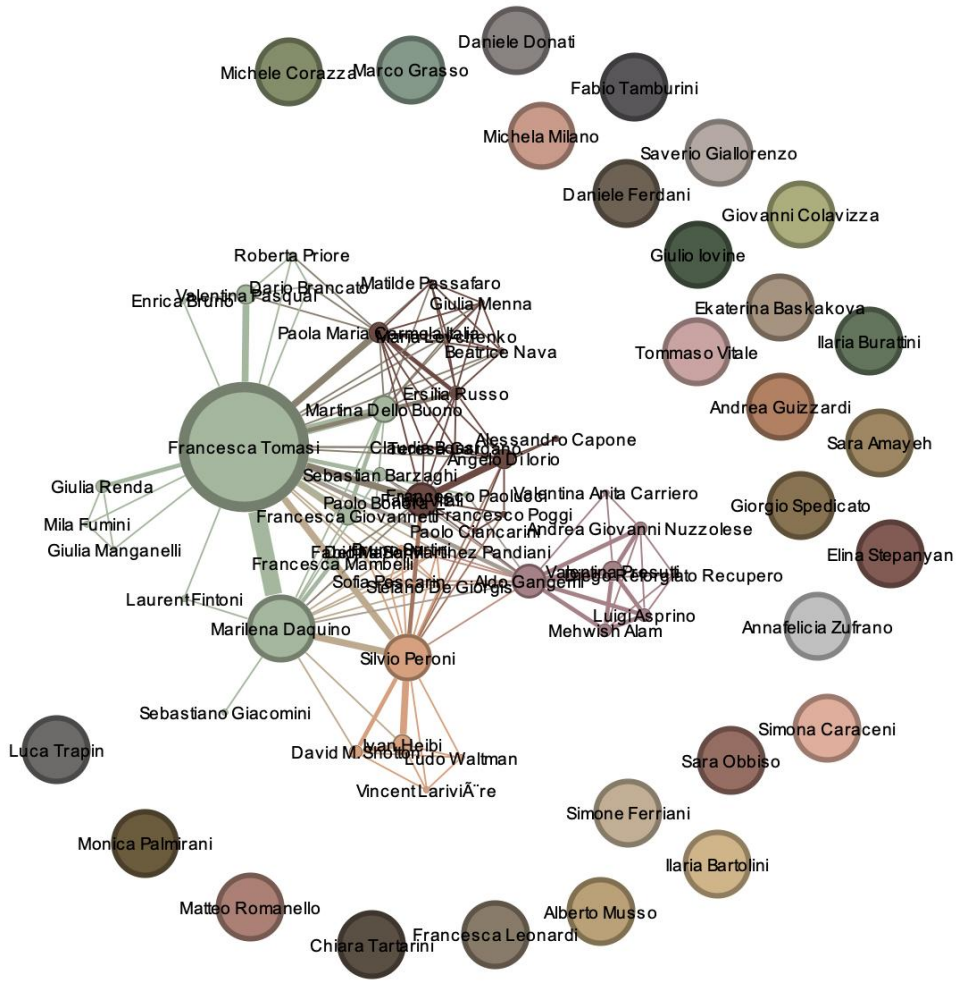


Figure 7: Visualization of the network of projects collaboration with names of researchers.

5.4 Comparative Analysis of Co-authorship and Projects Networks

5.4.1 Network Size and Density

The complete co-authorship network (1,090 researchers, 7,820 collaborations) is significantly larger than the filtered project collaboration network (56 researchers, 152 collaborations), indicating that academic publications encompass a much broader network of scholarly interactions. In contrast, project-based collaborations are more selective, involving a smaller, more focused group of researchers.

5.4.2 Collaboration Intensity

- Complete co-authorship network: 14.34 collaborations per author.
- Filtered co-authorship network: 5.43 collaborations per author.

This comparison highlights that researchers in the broader academic network tend to have more co-authors, while project-based collaborations involve fewer interactions, likely due to thematic or institutional constraints. Probably some students that engage in the projects for short internships do not participate in the publications.

5.4.3 Clustering and Network Structure

- Complete co-authorship network: Highly interconnected (clustering coefficient: 0.90)
- Filtered co-authorship network: Less connected (clustering coefficient: 0.48)

The significantly higher clustering coefficient in the complete network suggests dense local communities, where researchers frequently collaborate within well-connected clusters. The lower clustering coefficient in the filtered network indicates **smaller, more isolated groups**, suggesting that project-based research teams tend to form **specialized, less interconnected sub-networks** rather than broad co-authoring relationships.

5.4.4 Centralization and Key Researchers

- Complete co-authorship network: Decentralized (max degree centrality: 0.129)
- Filtered co-authorship network: More centralized (max degree centrality: 0.099)

The complete network exhibits a more distributed collaboration pattern, whereas the filtered network is more centralized, with fewer individuals dominating research collaborations. This suggests that project-related publications rely on a core set of key researchers rather than evenly distributed contributions.

5.4.5 Network Overlap

Only 56 researchers from the filtered project network appear in the complete co-authorship network. This indicates that some researchers involved in project collaborations do not engage in academic publications within the dataset's timeframe. This suggests that certain DH.ARC researchers focus on non-publication-based research activities or administrative roles, while others are more actively engaged in scholarly output.

5.4.6 Influential Researchers Across Networks

Several researchers, such as Silvio Peroni and Francesca Tomasi, consistently appear among the most connected authors across networks. Marilena Daquino ranks highly in both filtered publications and project collaborations, reflecting her strong involvement in collaborative research. The overlap between these key researchers indicates that highly active scholars contribute both to academic publications and project collaborations, though their levels of involvement may vary.

5.5 Keyword Co-occurrence Analysis

The keyword co-occurrence network (Fig. 8 and 9) is implied for analyzing the relationships between keywords in academic publications within researchers and collaborators involved in the Digital Humanities field at the University of Bologna. It helps uncover patterns in how research topics are interrelated and identifies key concepts that dominate literature. Additionally, it provides insights into emerging research trends and helps visualize the thematic structure of the publication set. The resulting graph consists of **2,953 nodes** and **8,444 edges**.

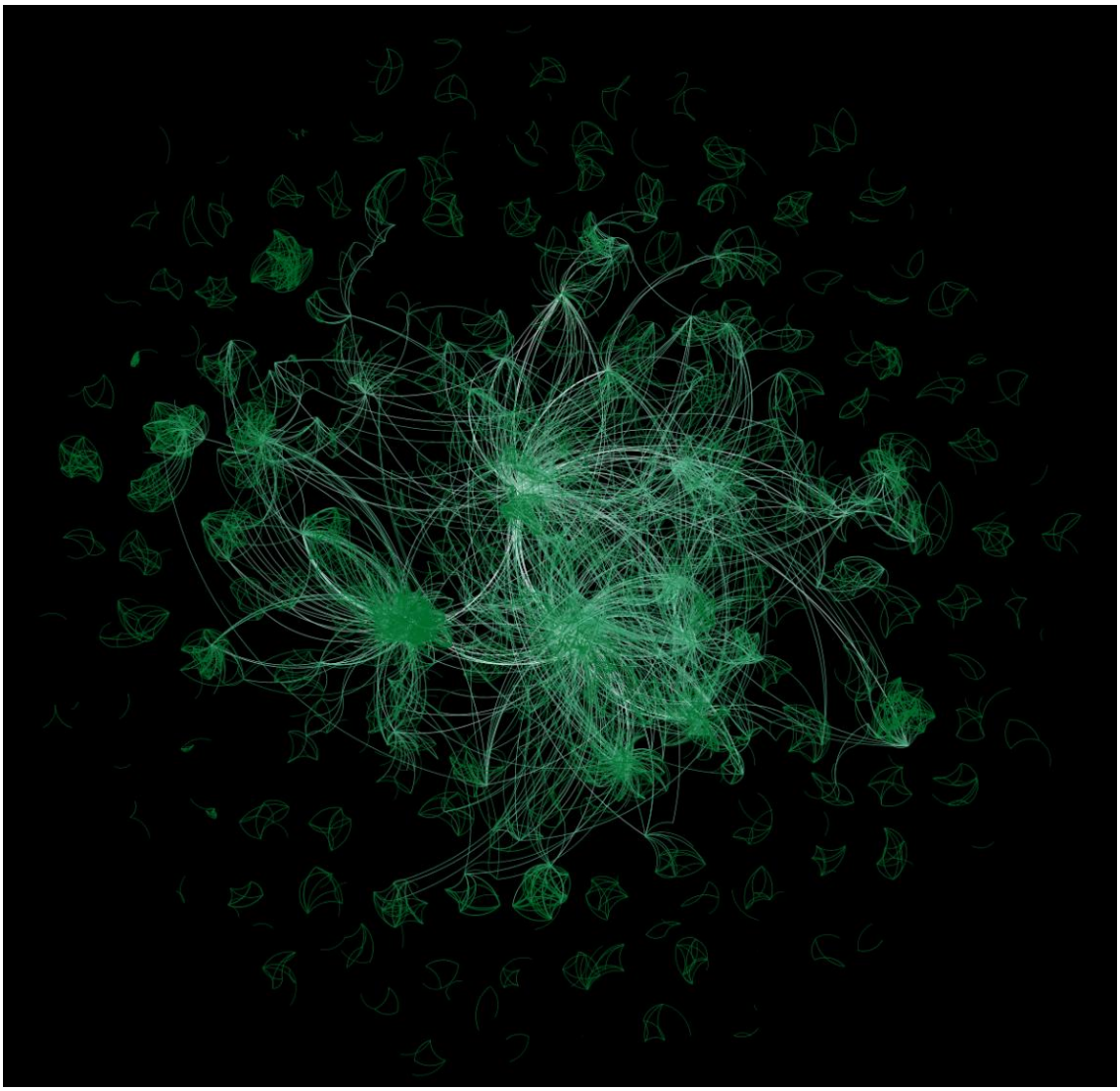


Figure 8: Visualization of the keyword co-occurrence network.

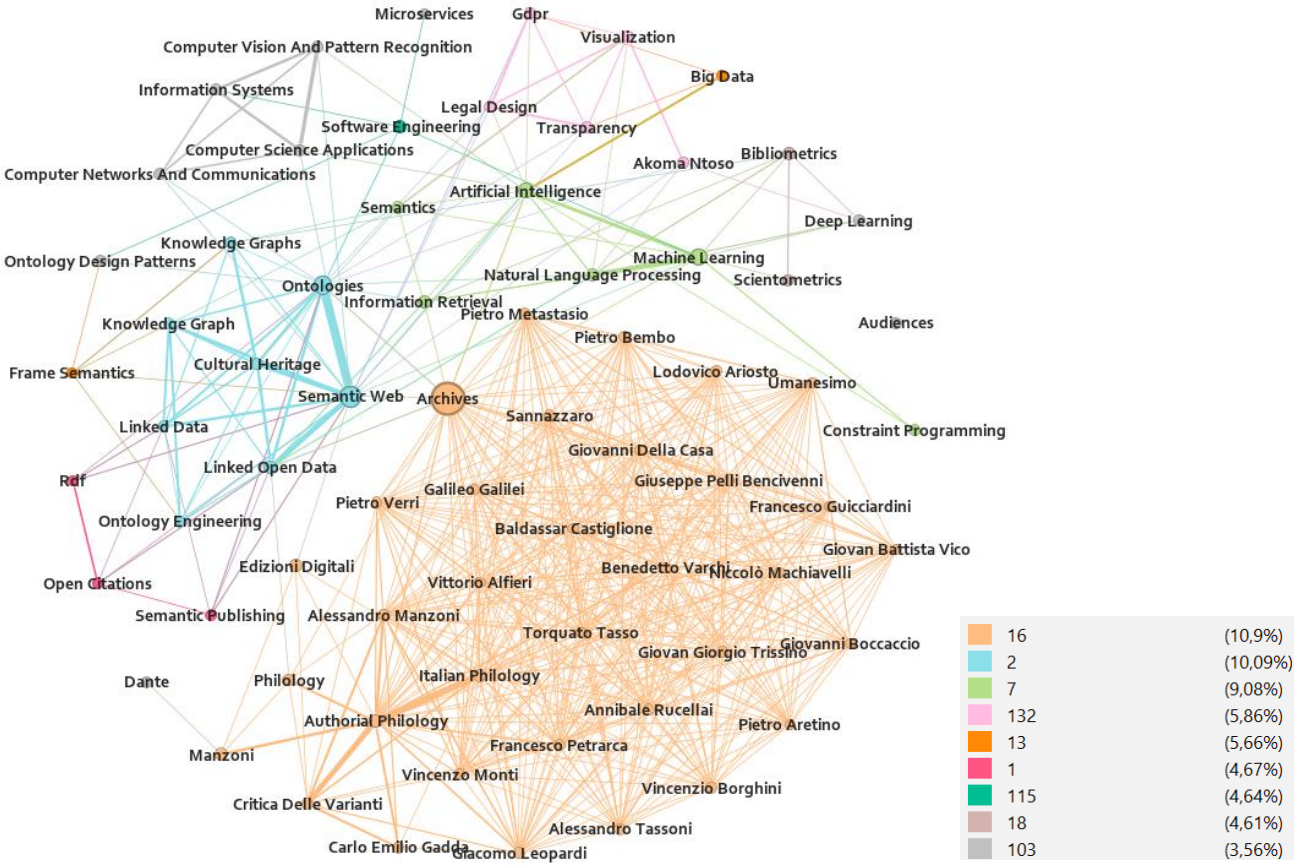


Figure 9: Keyword co-occurrence network filtered by degree range 30 (Gephi). Color differentiated by modularity and nodes size ranked by betweenness centrality.

5.5.1 Nodes and Edges

The network is as a **weighted, undirected heterogeneous graph** $G = (V, E, W)$ where:

- $\boldsymbol{V}$: Set of nodes (keywords and papers).
- $\boldsymbol{E}$: Set of edges, including:
  - **Keyword co-occurrence**: An edge $(k_i, k_j) \in E$ is created between two keywords k_i and k_j if they appear in the same article, with the weight increasing based on their co-occurrence frequency.
  - **Article-keyword links**: An edge $(a, k) \in E$ connects an article a to each of its associated keywords k, with a fixed edge weight of 1.
- $\boldsymbol{W}$: Weight function ($w: E \rightarrow R^+$) representing the strength of co-occurrence.

The graph is mathematically described using an adjacency matrix $A \in R^{|V| \times |V|}$, where:

$$A_{ij} = \begin{cases} w_{ij} & \text{if there is an edge between nodes } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases}$$

Each **node** represents either a keyword or a publication. Keywords are extracted from the metadata of publications and undergo a cleaning and standardization process, as mentioned earlier. Articles are connected to their associated keywords.

An **edge** is established between two keywords if they co-occur within the same publication. The weight of this edge increases each time the two keywords appear together in a different publication, thereby reflecting the strength of their relationship based on repeated co-occurrence.

- **Edge Weight**: If two keywords k_i and k_j co-occur in a publication, the weight of the edge between them is incremented:

$$w_{ij} = w_{ij} + 1 \qquad \text{for each additional co-occurrence.}$$

- **Article-Keyword Edge**: Each article node a connects to its associated keywords with a weight of 1:

$$w_{ak} = 1 \qquad \text{for } k \in \text{keywords of article } a.$$

Each publication is thus represented as a node connected to its associated keywords. These edges, with a uniform weight of 1, signify a single instance of keyword association.

The resulting network structure is based on the frequency and strength of keyword co-occurrence. Keywords that frequently appear together form stronger connections, as reflected in the edge weights. This approach emphasizes the thematic clustering within the publication dataset, enabling deeper insights into the interconnectedness of research topics.

5.5.2 Metrics

Each selected network measure serves a specific purpose in understanding different aspects of research interconnections:

- **Average Degree (5.72):** The network is relatively sparse (low edge-to-node ratio), meaning that most keywords are not densely interconnected, but rather form smaller clusters or communities.
- **Degree Centrality:** This metric identifies the most interconnected research areas. The prominence of *Semantic Web* (0.4286) indicates a strong institutional focus on knowledge representation through Linked Open Data and ontologies. High scores for *Archives* (0.3571) and *Ontologies* (0.3333) confirm an emphasis on digital preservation and structured knowledge organization. The significant centrality of *Authorial Philology* (0.3095) and *Critica Delle Varianti* (0.2381) underlines the enduring role of textual scholarship. The inclusion of literary figures such as *Manzoni*, *Leopardi*, *Alfieri*, and *Monti* highlights a focus on italian literary heritage.

- **Betweenness Centrality:** This measure identifies keywords that act as bridges between different research domains. *Archives* (0.3955) emerges as the most critical connector, linking diverse topics. The high betweenness of *Semantic Web* (0.2379) and *Ontologies* (0.1770) reflects their role in integrating knowledge across disciplines. The presence of *Artificial Intelligence* (0.1097) and *Machine Learning* (0.1047) suggests that computational methodologies are becoming crucial intermediaries. Similarly, *Philology* (0.0698) and *Authorial Philology* (0.0548) demonstrate the centrality of textual studies in bridging traditional humanities and digital methods.
- **Closeness Centrality:** This metric highlights key hubs that efficiently connect different research areas. *Semantic Web* (0.5833), *Archives* (0.5753), and *Ontologies* (0.5526) appear as major nodes, reinforcing the institution's focus on digital and archival methodologies. The presence of *Machine Learning* (0.5185), *Artificial Intelligence* (0.5185), and *Linked Open Data* (0.5250) indicates a computational shift in research approaches, further supported by *Natural Language Processing* (0.4468) and *Knowledge Graphs* (0.4330).
- **Modularity (0.8441):** The relatively high modularity score suggests that research at Bologna is structured into thematic clusters, each centered around a specific subdomain, such as philology, computational humanities, and archival studies.
- **Network Density (0.0019):** The low density indicates that while strong thematic clusters exist, the network lacks high interconnectivity, resulting in some topics being more isolated. Despite the presence of key central nodes, the graph remains fragmented into distinct communities. The relatively high **clustering coefficient (0.5521)** suggests that within these subfields, concepts are strongly interconnected.

All results are recorded in external CSV files, which are openly available in the repository of this report.

5.5.3 Community Detection

The **Louvain method** was used to detect communities within the network, helping to identify clusters of strongly connected keywords across publications. These clusters represent thematic groups, shedding light on key research areas within the Digital Humanities field in Bologna.

Louvain was chosen for its efficiency, scalability, and strong performance in detecting well-defined communities in large networks. Given the size of the keyword co-occurrence network, Louvain's ability to optimize modularity quickly while handling large-scale data made it a suitable choice.

Initially, the algorithm was applied to the keyword co-occurrence network on a full-scale, revealing a total of 148 communities. To refine the analysis and highlight dominant patterns, a

subgraph was created by filtering nodes based on degree centrality, retaining only those with a score above 0.5.

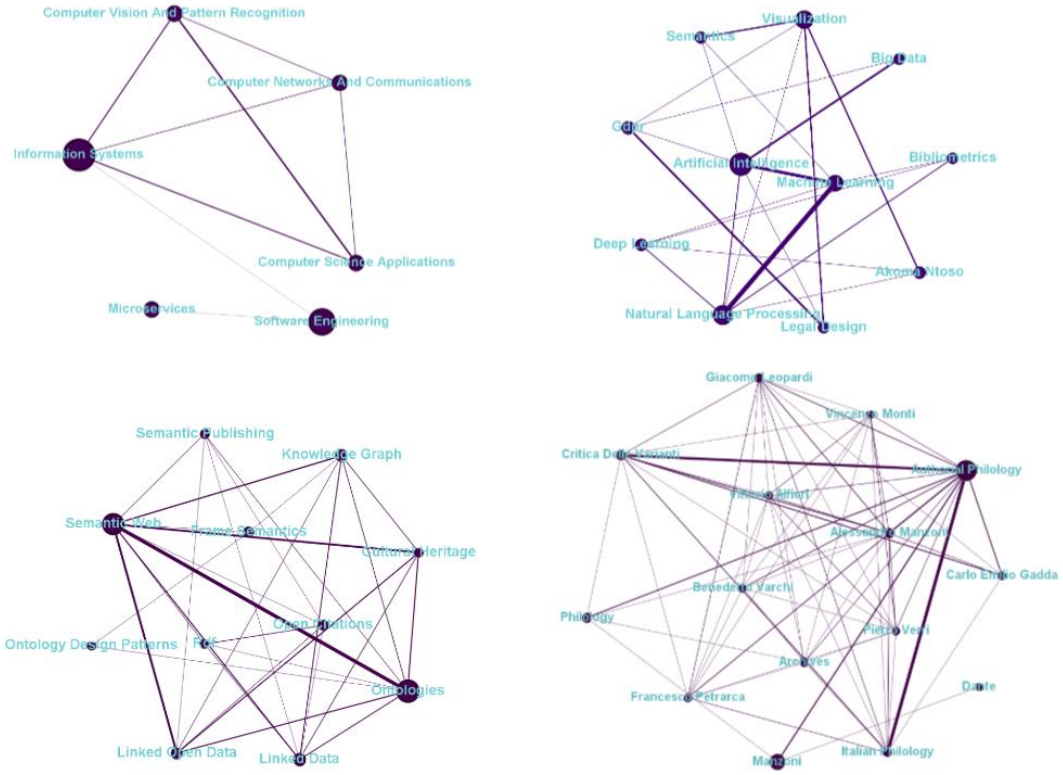


Figure 10: The four communities identified in the keyword co-occurrence subgraph using the Louvain method.

These include (Fig. 10):

- **Top Left Community:** This community focuses on the technical aspects of Digital Humanities, emphasizing computer science disciplines such as computer vision, network communications, and software engineering. It highlights the integration of advanced computational techniques and software development practices within the Digital Humanities framework.

  Key Nodes: Computer Vision, Pattern Recognition, Computer Networks, Information Systems, Computer Science Applications, Software Engineering, Microservices.

- **Top Right Community:** A community centered around Artificial Intelligence and Data Science, with a strong emphasis on Machine Learning, Deep Learning, and Natural Language Processing. It also includes Bioinformatics and Visualization, indicating a focus on data-driven approaches and the application of AI in various fields, including

Biology and Legal Design. Notable is the growing intersection of computational methods with humanities research.

Key Nodes: Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, Semantics, Visualization, Bioinformatics, Biological Metrics, Akoma Ntoso.

- **Bottom Left Community:** This community revolves around the Semantic Web and Knowledge Representation. It emphasizes semantic publishing, ontology design, and the use of linked open data to enhance Digital Humanities research, particularly in the context of cultural heritage. This underscores the prominence of Knowledge Representation in the research landscape.

  Key Nodes: Semantic Web, Semantic Publishing, Knowledge Graph, Ontology Design Patterns, Linked Open Data, Linking Data, Digital Humanities, Cultural Heritage.

- **Bottom Right:** A cluster more humanities-oriented that focuses on Literary Studies and Philology, particularly Italian Literature. It emphasizes textual analysis and the continued importance of traditional literary studies within the Digital Humanities domain is reflected by the inclusion of influential literary figures such as Alessandro Manzoni and Giacomo Leopardi.

  Key Nodes: Authorial Philology, Giacomo Leopardi, Alessandro Manzoni, Carlo Emilio Gadda, Archives, Francesco Petrarca, Dante, Pietro Verri, Benedetto Varchi, Vittorio Alfieri.

6. Conclusions

Considering the limitations of this report, many details of our analysis could not be explored. Here we sought to provide a synthesis of the comparative analyses of the networks we produced.

1. The project network is more centralized than the publication network, with a few researchers maintaining most connections.
2. Collaboration is more selective in projects, with many researchers engaging in only a small number of partnerships.
3. Project-based research has fewer intermediaries, suggesting that research leadership is concentrated in key figures.
4. Clustering is lower in projects, indicating that researchers in funded collaborations do not interact as broadly as they do in publications.
5. Academic publishing fosters more widespread and diverse collaborations, while project-based research tends to be more hierarchical and selective.
6. There are variations in researcher rankings across different networks, highlighting the diverse ways researchers contribute to the field.

7. The comparison between networks offers insights into how collaborations and influences work. However, when did inside a unique network, like was did in the segmentation by years in the co-authorship network can offer interesting results. It is very obvious the impact of Covid-19 pandemic in the production, but also, is possible to see how this very young field is evolving in the cycle of academic publications.
8. Looking to keywords network we can see that an individual that has no significance in the other networks is the center of a subset of this network, suggesting that this individual does not publish or conduct projects in collaboration but has a high influence on other researchers works.
9. Some patterns indicate the dynamics of the field: the interdisciplinary character, as also the young researchers with less collaborations than the top publishers but in positions of brokers (intermediaries between) communities inside the network, indicating a very innovative characteristic of the production in the field.
10. The relatively high average degree of the keywords network suggests that most research topics are interconnected, and multiple keywords tend to appear together within the same publications. So, despite the multidisciplinary nature of DH field, we can say that there is a significant vocabulary shared. This is a very strong indication of identity in a field of knowledge.
11. The high degree centrality of some themes in the keywords network (Semantic Web (0.4286), Archives (0.3571), Ontologies (0.3333), Authorial Philology (0.3095), Critique of the Variants (0.2381)) shows the concentration of efforts in the network analysis, indicating at same time, the prospectives of recruitment of new members and the more stablished studies. This shows that the communities around these themes are more consolidated, even when they do not have a big collaborative representation in the Co-Authorship Network (the case of Authorial Philology and Critique of the Variants).
12. Some of the most significative trends are strongly connected with collaborative work. This indicates the strategy of diverse groups (very common in DH) and the scarcity of researchers that can perform multiple roles or that have multiples areas of knowledge.
13. Still the high level of Modularity (0.5698) at same time indicates that DH in University of Bologna is very far away from being defined as uniform field, reflecting a different approach in relation to other university centers that concentrate on a specific subfield. If we see that Community structure suggests topic specialization. The low density 0.1672 of the network corroborates this finding.

These findings reveal significant differences in how researchers collaborate across different academic environments. While project participation may encourage academic publishing, it does not always translate into broad publication networks. Some project researchers remain isolated in publications, while a few key individuals dominate both research projects and academic publishing.

Understanding these differences can help universities, funding institutions, and researchers optimize collaboration strategies for both academic publishing and project-based research activities. The varied contributions of researchers across different collaboration types underscore the importance of considering multiple facets of academic engagement when evaluating research impact and productivity.

7. Critique

The study demonstrates significant potential for analysis, revealing insights that extend beyond the scope of a brief report. The approach of applying a comprehensive set of metrics to explore various outcomes is valuable, although it became evident that for understanding collaboration dynamics in projects and publications, metrics such as degree centrality, betweenness centrality, and clustering coefficient proved most effective.

A notable challenge emerged in visualizing and analyzing the extensive co-authorship network through static graphical representations, as exemplified in Graph 2. The complexity of large networks multiplies analytical possibilities, suggesting the need for more focused examinations of specific researcher communities within the network. For instance, investigating whether different patterns emerge among groups from diverse disciplinary backgrounds could yield valuable insights, given the network's multidisciplinary nature.

To enhance the study's ability to address the research problems more comprehensively, several approaches could be considered:

1. Gather additional data on researchers' disciplinary backgrounds and career stages to provide context for collaboration patterns.
2. Implement dynamic network visualization tools to better represent and explore large-scale networks.
3. Conduct temporal analysis to track how collaboration patterns evolve over time.
4. Incorporate qualitative data through interviews or surveys to understand motivations behind collaboration choices.
5. Expand the comparative analysis to include partial network examinations, focusing on specific subgroups or thematic clusters.

These enhancements would allow for a more nuanced understanding of the complex dynamics within Digital Humanities research collaborations, potentially revealing insights that are not immediately apparent in the current analysis.