# Network Analysis of DH.ARC – Digital Humanities Research Centre Collaboration Dynamics

Report group members:

Lucrezia Pograri, DHDK, ID number: 0001119705

Name2 Surname2, Degree Programme, ID number (matricola)

## Introduction

The context of this study lies in the field of Digital Humanities, specifically focusing on the collaboration dynamics within the DH.ARC - Digital Humanities Advanced Research Centre at the University of Bologna. Digital Humanities is an interdisciplinary area of study that combines computational methods with humanities research. This project aims to explore how researchers and professors within DH.ARC collaborate, utilizing network analysis to examine key relationships and structures within the research centre.

## Problem and Motivation

The main problem addressed in this project is understanding the **structure of academic collaborations** and **identifying influential individuals** within the network. Understanding these patterns is important for identifying key researchers, bridging gaps in communication, and fostering new collaborations. Theoretical contributions include better insight into the dynamics of academic collaboration, while practical implications involve enhancing strategies for fostering interdisciplinary research.

The main problem addressed in this study is understanding how collaboration is structured among members of DH.ARC, including identifying central figures, key influencers, and the overall cohesion of the group. By examining these dynamics, we aim to uncover insights into the research centre's collaborative environment, providing valuable information for optimizing future research initiatives. Theoretical contributions include understanding the interconnectedness of researchers within a highly interdisciplinary field, while practical outcomes might include recommendations for enhancing collaboration and identifying potential areas for increased synergy.

The main contributions of the project are:

* **Identification of central figures** in the research community.
* **Detection of collaborative clusters** to understand thematic groupings.
* Insights into the **role of individuals who bridge subgroups**.

## Datasets

The dataset was gathered from publicly available sources, such as the DH.ARC website, University of Bologna publications, and academic profiles (e.g., ORCID, ResearchGate). The data includes co-authorship, shared research projects, and formal supervision relationships. Tools used include Python for data handling (e.g., Pandas for data manipulation) and Gephi for visualizing and computing network measures. The data represents nodes as individual researchers and edges as co-authorship or participation in joint research projects.

### Data handling

The data was stored and manipulated using **Python** with libraries like **Pandas** for data processing.

### Computing measures

We used **NetworkX**, a Python library, to compute various network measures such as centrality and clustering coefficients.

The dataset can be structured as follows:

1. **Nodes**

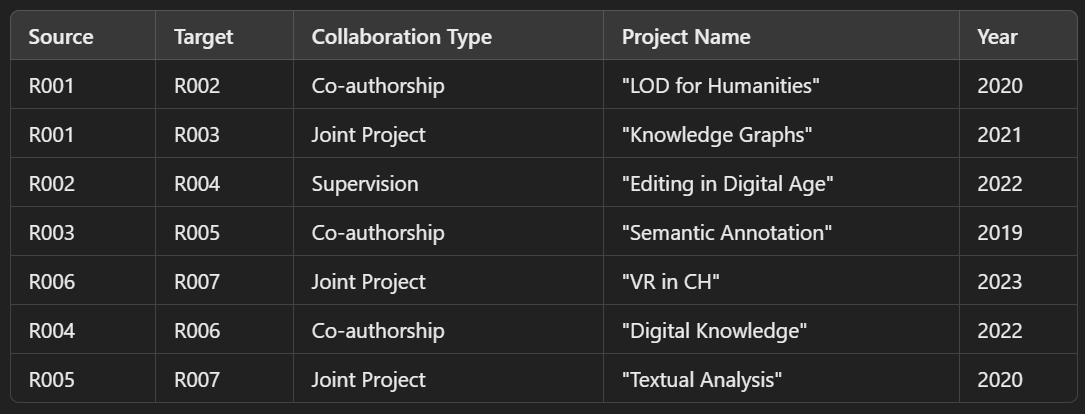
* **Node Type**: Researcher
* **Attributes**:
  + **ID**: Unique identifier for each researcher (e.g., R001, R002, etc.)
  + **Name**: Full name of the researcher (e.g., "Silvio Peroni")
  + **Research Field**: Primary research area within Digital Humanities (e.g., "Semantic Web," "Digital Scholarly Editing")
  + **Position**: Academic position (e.g., "Professor," "Postdoc," "Ph.D. Student")
  + **Affiliation**: Affiliation within DH.ARC (e.g., "DH.ARC Researcher," "Visiting Scholar")
  + **Publication Count**: Number of published research articles (e.g., 15, 30)

Immagine che contiene testo, schermata, numero

Descrizione generata automaticamente

1. **Edges**

* **Edge Type**: Collaboration
* **Attributes**:
  + **Source**: ID of the source node (e.g., R001)
  + **Target**: ID of the target node (e.g., R002)
  + **Collaboration Type**: Type of collaboration (e.g., "Co-authorship," "Joint Project," "Supervision")
  + **Project Name**: Name of the project or publication (e.g., "LOD for Humanities")
  + **Year**: Year of collaboration (e.g., 2021)



1. **Additional Tables (Optional)**

* **Publications**: A list of publications with details like title, year, authors, and field.
* **Projects**: A list of ongoing or past projects, with participating researchers and funding information.
* **Events Participation**: Researchers participating in conferences, seminars, and workshops (e.g., "DH Conference 2023").

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

### Step 1: Definition of the Adjacency matrix

You can create adjacency matrices for your nodes (researchers and their relationships).

In Python, we created a dictionary to represent the relationships, which will then be used to generate an adjacency matrix.

### Step 2: Build the NetworkX Graph

Define Researchers and Relationships: A list of researchers is defined, and an adjacency matrix is initialized to zeros.

Fill the Adjacency Matrix: The collaborations are defined in a list of tuples, and the corresponding cells in the matrix are set to 1 to indicate a connection.

Create the Graph: The graph is created from the adjacency matrix using nx.from\_numpy\_matrix, and nodes are relabeled to the actual researcher names.

Visualize the Graph: The graph is drawn using matplotlib for visualization.

## Validity and Reliability

**Validity**: The model is based on actual collaboration data, which is a close representation of academic interactions. However, informal collaborations not captured in publications or projects, or non-documented contributions may limit the scope of the network.

**Reliability**: The use of standard tools like NetworkX and well-defined data sources makes the study reproducible. However, some biases might persist due to incomplete or unavailable data.

## Measures and Results

* **Degree Centrality**: Measures how connected a node (individual) is by counting the number of connections. It helps identify **key figures** who are highly active in collaborations.
* **Betweenness Centrality**: Measures the extent to which a node lies on the shortest paths between other nodes. It indicates **individuals who act as bridges** between different subgroups, enabling the flow of ideas.
* **Clustering Coefficient**: Measures the degree to which nodes tend to form tightly knit groups. This helps identify **collaborative sub-communities** within the network.
* **Core-Periphery Analysis**: Identifies core members of the network who are central to the functioning and peripheral members who are less engaged in collaborations.

The gathered data was represented as a monomodal, undirected graph, where nodes represent researchers and edges represent collaboration. The application of these measures revealed that **Silvio Peroni and Francesca Tomasi** have high degree and betweenness centrality, suggesting they are key connectors in the network, while clustering coefficients identified thematic research clusters.

## Conclusion

The analysis reveals a **well-structured network** of researchers within DH.ARC with distinct core and periphery dynamics. Central figures in the network play an essential role in fostering collaboration across various thematic areas, while smaller subgroups are formed based on specific research interests. There is evidence of **strong sub-communities**, suggesting thematic specialization among researchers. These insights suggest a cross-disciplinary joint efforts and they can help optimize future collaboration efforts by recognizing influential members and fostering more inclusive interactions.

## Critique

The project provides a useful overview of collaboration patterns, but it may not fully represent the **informal interactions** that contribute significantly to academic growth. Additionally, including **citation data** could have offered insights into influence beyond direct collaborations. In future research, incorporating data from **conferences or workshops** could provide a more comprehensive understanding of the relationships within the Digital Humanities field. Applying other measures, like **homophily** or **triad census**, could further reveal patterns of preference for collaboration and social dynamics within the network.

While the analysis effectively highlights key collaborators and the general structure of the research network, the data collection could be expanded to include informal collaborations and other communication channels, such as workshops and seminars. This would provide a more comprehensive picture of collaboration dynamics. Additionally, alternative measures like eigenvector centrality could be applied to better understand the influence of researchers beyond direct connections.