NETWORK ANALYSIS WITH PALLADIO

Unveil Hidden Connections

Ruby Martinez

# Introduction

Network analysis and visualization software have become crucial tools for transforming our understanding of culture, history, and society in the constantly changing field of digital humanities. These resources give scholars, researchers, and educators a lens to examine complex relationships, patterns, and narratives. The following will walk you through using Palladio and will serve as a complement to the recording.

# INTRODUCTION TO NETWORK ANALYSIS

## What is network analysis?

Network analysis is a computational method for studying the relationships between interconnected objects or entities. These entities, known as nodes, can represent a multitude of things—ranging from individuals in a social network to molecules in a biological system, or even concepts in a knowledge graph. The relationships or interactions between these nodes are depicted as edges or links, forming a network.

In 'The Network Turn' (2020), it is argued that network analysis, being a quantitative method capable of studying diverse phenomena, serves as a bridge between the arts, humanities, natural, and social sciences by uniting researchers from different disciplines. While historians traditionally relied on textual narratives to reconstruct events, the evolution of network analysis allows for the visualization and analysis of connections between individuals, organizations, and ideas. Düring (2015) effectively illustrates this evolution using Palladio, demonstrating the connections between individuals who aided Jews during the Holocaust. These applications demonstrate the potential of network analysis to illuminate the intricate webs of human interaction that shape history, setting the state for further exploration in the following section.

## Applications of Network Analysis

Network analysis can map and visualize complex relationships between people, ideas, texts, or other objects. For example, network analysis can be used to visualize connections between characters in a novel, the relationships between scholars in a field of study, or the links between websites on the internet.

*Example:* [*The Six Degrees of Francis Bacon*](http://www.sixdegreesoffrancisbacon.com/)

* This project aims to map the social network of Early Modern Britain from the 1500s to the 1700s and centers on the influential figure of Francis Bacon. This project utilizes network analysis to visualize the connections and relationships between philosophers, writers, poets, artists, scientists, etc. providing insights in the special, intellectual, and political landscape of that period.

Network analysis can also be used to identify influential nodes and communities within a network. For example, network analysis can be used to identify the “most important” people in a social network, the most influential papers in a field of study, or the most popular websites on the internet.

*Example:* [*Open Syllabus Project*](https://www.opensyllabus.org/4)

* This project collects and analyzes syllabi from various academic disciplines, allowing researchers to explore the network of knowledge and connections between courses, instructors, and institutions. In this example, network analysis techniques are applied to identify patterns in syllabus content, curricular trends, and the flow of knowledge across disciplines.

It can also analyze the flow of information or influence through a network. For example, network analysis can be used to trace the spread of ideas through a network, the diffusion of innovations through a population, or the dissemination of news through a media.

*Example:* [*Mapping the Republic of Letters*](http://republicofletters.stanford.edu/4)

* This project delves into the world of 18th-century European correspondence networks, examining the exchange of letters among scholars, philosophers, and literary figures. Network analysis helps reveal the interconnectedness of these individuals, the dissemination of ideas, and the formation of intellectual communities.

## Key Concepts

But at its most basic level, network analysis is essentially mapping relationships and connections between elements. And as evidenced in the previous projects,

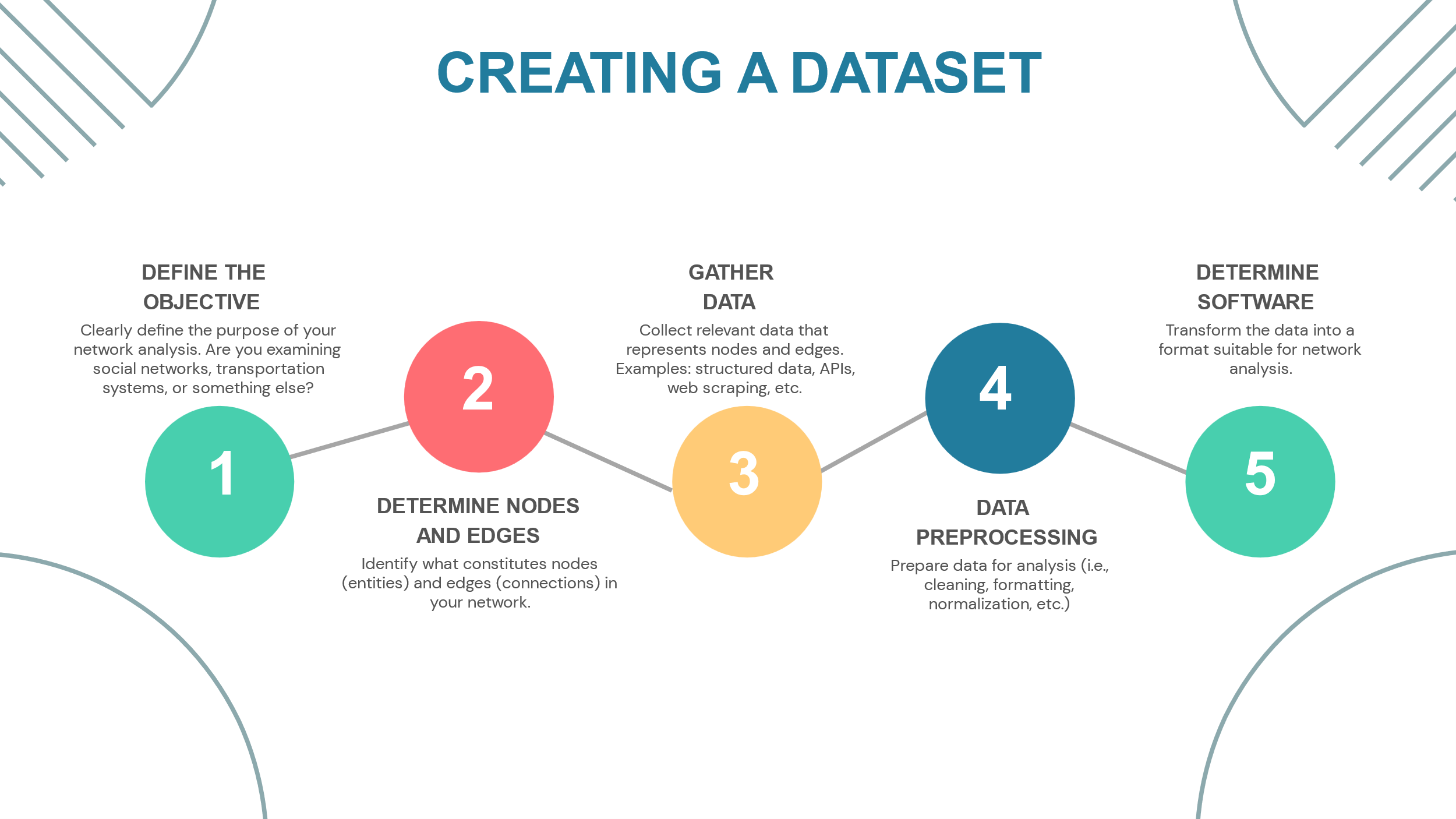
* Networks are everywhere. Networks are not just found in the natural world, but also in the social, cultural, and economic worlds.
* Networks are complex. Networks are often made up of many different nodes and edges, which can make them difficult to understand.
* Networks are dynamic. Networks are constantly changing as new nodes and edges are added and existing nodes and edges are removed.
* Networks are important. Networks play a role in a wide variety of phenomena, from the spread of disease to the evolution of language.

## Terminology

* Nodes – represent the entities we are interested in studying, such as people, organizations, ideas, or websites.
* Edges – represent connections between nodes.
* Degree – the degree of a node is the number of edges that are connected to it.
* Centrality – the measure of how important a node is to a network.

# CREATING A DATASET

This section will guide you through the process of creating a dataset that effectively captures the relationships and attributes of interest for your research.



The above image is a sample workflow for creating a dataset. Each step will have questions to reflect on as you go through the workflow. In addition, I will also provide how I used this workflow to create the sample dataset (iPRES Conference Proceedings). Steps of the workflow include: 1. Define the Objective; 2. Determine the Nodes and Edges; 3. Gather Data; 4. Data Preprocessing; and 5. Determine Software.

## Sample Dataset

Get to know the sample dataset.

### Data Source

* International Conference on Digital Preservation (iPRES) conference proceedings
* Data Period: 2022 – 2023

### Data Elements

* Author
* Affiliation
* Country
* Paper Title
* Topics
* Submission Year

## Define the Objective

Before diving into network analysis, establishing clear objectives is crucial to guide the process and derive meaningful insights. Consider the following aspects:

* What is the purpose of the network analysis?
* What questions do you want to answer?
* What relationships and attributes are essential to understanding your research topic?

Reflecting on these questions ensures that the exploration and interpretation of the network align with the broader goals of your research or study.

### Define the Objective – Sample Dataset

After reflecting on the questions above, I asked myself what I wanted to learn from the iPRES Conference Proceedings. This conference is the only international conference on digital preservation that first began in 2004. The big picture for me is exploring research connections and trends within the community.

But before I can formulate any research questions, I needed to identify data elements that are available via the proceedings, which is also important for our next step. Like most conference proceedings the data that is publicly available are elements like *Author*, *Affiliation*, *Country*, *Paper* *Title*, *Topics*, *Submission* *Year*, etc. Now that I have an idea of the data I’m working with, some of what I wanted to explore are co-authorship patterns. How are authors connected through co-authorship networks?

## Determine Nodes and Edges

In constructing a network, defining the nodes and edges is foundational to representing the entities and their relationships. Here are questions to consider:

* What are the core identities or individuals that will be represented as nodes in your network?
* What types of relationships or interactions do you want to capture between these entities?
* How will you represent these relationships as edges in your network?

Reflection on these questions ensures a thoughtful approach to defining nodes and edges within your network, laying a strong foundation for a comprehensive and meaningful representation of your data.

### Determine Nodes and Edges – Sample Dataset

Nodes represent the entities we are interested in studying, such as people, organizations, ideas, or websites. Within the context of our sample dataset. Here are some sample nodes: Author, Paper, and Affiliation can all represent nodes.

Edges represent connections between nodes. Here are sample edges.

* Author-Paper Relationship: Create edges to represent the relationship between authors and papers they've written. Each edge connects an author node to a paper node, indicating that the author has contributed to that paper.
* Co-authorship: Establish edges between author nodes to denote co-authorship. If two authors have collaborated on the same paper, create an edge between their respective nodes to illustrate their collaborative relationship.

This representation allows analysis of relationships between papers, authors, collaborations, and potentially insights into co-authorship networks, paper citations, author affiliations, etc.

## Gather Data

Efficiently gathering relevant data is pivotal for robust network analysis. Consider the following aspects:

* What are the most appropriate data sources for your research?
* What combination of primary and secondary sources will you use?
* What data collection methods will you employ?

Reflection on these questions will help in devising a comprehensive data collection strategy, ensuring the acquisition of relevant and reliable data for your network analysis.

### Gather Data – Sample Dataset

The data was sourced from conference proceedings by accessing respective conference websites and platforms. Extracting the information was a manual process, involving significant copy-and-paste actions. The compiled data was organized into an Excel spreadsheet, categorizing it under data fields such as year, paper ID, title, and more.

## Data Preprocessing

Preparing the data for analysis involves organizing, cleaning, and transforming it into a suitable format. Reflect on these aspects:

* How will you organize and structure your data?
* What data cleaning techniques will you use to identify and address errors, inconsistencies, or missing values?

Reflecting on these questions and considerations helps in establishing a robust data preprocessing pipeline, ensuring that the data is refined, consistent, and ready for meaningful analysis within your network framework.

This is a more common step when you’re working with data that wasn’t created by you. This involves:

* Cleaning – remove duplicates, handle missing values, and standardize formats.
* Formatting – ensure consistency in how nodes and edges are represented.
* Normalization – Normalize data if necessary for uniformity.

***Please note that some of these steps are entirely optional.***

### Data Preprocessing – Sample Dataset

* An example of data cleaning for this dataset was when authors had a paper in both the 2022 and 2023 conference proceedings but only included their middle initial in one year. The computer is going to interpret this as two separate authors even though we know it’s one.
* Formatting also comes into play here, especially if you are using platforms that are case sensitive. Some authors use diacritics in their name, but when I extracted that data from the conference proceedings excel tried to normalize this on its own and messed with the formatting.
* Lastly, some examples of normalization can look like removing professional titles from names (i.e. Dr., Mr., Mrs.), if you’re using dates you want to ensure the same date format is being used across the dataset as well.

## Determine Software to Use

Selecting the right software is pivotal for conducting effective network analysis. Consider these key reflections:

* What tool(s) would be appropriate based on your data and research objectives?
* What are the key factors to consider when choosing network analysis software?

Considering these reflections aids in making an informed decision regarding the choice of software for network analysis, ensuring it aligns seamlessly with your research objectives and facilitates the execution of your analytical processes. Additional aspects you’ll want to consider is whether your data is compatible with the software, how easy is it to use, is it compatible with other tools. How much time and effort can you invest in this project?

### Determine Software to Use – Sample Dataset

Since we’re going to be working with a small dataset, I wanted to use Palladio because of its intuitive features that will be discussed in the next section.

The next steps involve transforming the data into the necessary format compatible with the chosen tool and uploading it. As you become more familiar with the software, you may discover alternative ways to structure your data, potentially refining your analysis to yield more nuanced and specific results.

# GETTING TO KNOW PALLADIO

This section will introduce you to the fundamentals of Palladio, equipping you with the knowledge and skills to navigate its interface and embark on your network analysis journey.

## What is Palladio?

Palladio, a web-based data visualization tool and widely used in the DH community. Its intuitive interface and comprehensive features make it an invaluable tool for exploring complex relationships and uncovering hidden patterns.

Functionality includes exploration of connections, relationships, and patterns within datasets. Some of the key features include:

* Interactive interface: allows dynamic manipulation and visualization of data.
* Customizable visualizations: offers flexibility in creating tailored network maps and visual representations.

## The Four Views of Palladio

Palladio offers four distinct views—Map, Graph, List, and Gallery. Each view provides a unique perspective and allows users to comprehend data from various angles.

### Map View

Displays geospatial data as either a single point, or as lines connecting two points.

### Graph View

Conceptual web connecting data points and attributes (i.e., nodes and edges)

### List View

A table view, in which the full dataset can be filtered through simple data processing parameters.

### Gallery View

Images and metadata associated with data points can be viewed and filters in a display grid.

## 

## Importing Data

This section will guide you through the process of importing data into [Palladio](https://hdlab.stanford.edu/palladio/).

### Getting Started

Data Verification: Ensure the data is in a format compatible with Palladio. Supported formats include CSV, TSV, and Excel spreadsheets. In this case, the sample dataset is already saved in CSV format.

Create a New Project: You will launch Palladio in your web browser and click start. You will be directed to this page and instructed to load your data. You can do this by copying and pasting your spreadsheet direct from excel. Or drag and drop your file into the empty box.

A screenshot of a computer

Description automatically generated

Once you’ve done this the box will no longer be blank. You can do a quick glance through to make sure everything looks good before clicking LOAD.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedAfter clicking **LOAD**, you will be shown a high-level view of your data, which Palladio calls **DIMENSIONS**. Most of the dimensions will auto fill. for example, **YEAR** auto fills to date. **PAPER ID** auto fills to number, in this instance, even though the **PAPER ID** is a number we are not using this dimension as numerical data and needs to be updated to be read as **TEXT** instead.

# SAMPLE VISUALIZATIONS

This section will showcase a selection of sample visualizations, highlighting the versatility and power of Palladio’s visualizations tools.

## A screenshot of a computer Description automatically generatedSample Visualization 1 – Graph View

This visualization uses Palladio’s **Graph View**, using standard settings in the image below. Palladio uses the concepts of **SOURCE** and **TARGET** which directly corresponds to the concepts of nodes.

In Palladio, when you're working with a dataset that involves connections or relationships between entities (e.g., ‘**Paper ID’**, ‘**Country’**), the 'source' represents the origin or starting point of a relationship, while the 'target' signifies the endpoint or destination of that relationship. In network analysis terms, these ‘sources’ and ‘targets’ become nodes, representing individual entities, while their connections form edges, symbolizing relationships.

A close-up of a network

Description automatically generated

For instance, in this sample visualization, each ‘**Paper ID**’ and ‘**Country’** functions as a node, while the connections between them represent edges. The visualization depicts the relationship between paper submissions and countries, revealing prominent representation by the United States and the United Kingdom in iPRES 2022. Furthermore, closer inspection reveals collaborative efforts between countries on specific papers as you zoom in.

## Sample Visualization 2 – Map View

The second visualization utilizes the **Map View** and showcases the countries of where the authors are affiliated. The size of the nodes is determined by how much that country is referenced in the dataset. Similar to our last visualization the US and the UK make up the bulk of that representation for the conference. It will be interesting how the nodes change if we were to work with a dataset that consisted of more than two iterations of the conference.

A screenshot of a computer

Description automatically generated

A map of the world

Description automatically generated

Please note mapping countries requires geospatial data (latitude and longitude). The data underwent geocoding using the [Geocoding by Awesome Table](https://workspace.google.com/marketplace/app/geocode_by_awesome_table/904124517349) extension in Google Sheets, enabling Palladio to map out country affiliations.

## Sample Visualization 3 – Table View

A screenshot of a computer

Description automatically generatedThe final visualization uses the TABLE VIEW feature. This visualization presents keywords associated with submission organized by country. This feature is incredibly useful to quickly consolidate data, enabling the identification of trends within topics.

A screenshot of a computer

Description automatically generated

The **TABLE VIEW** provides a comprehensive overview of keywords aligned with specific countries, facilitation quick and efficient analysis of thematic trends across submissions. Its tabular presentation enhances accessibility and aids in identifying patterns or common themes within a dataset.

# SOME ETHICAL CONSIDERATIONS

Ethical challenges

* Privacy concerns: Network analysis can be used to collect and analyze sensitive personal data. Scholars need to be careful to protect the privacy of individuals when using network analysis.
* Power and influence: Networks can be used to map and analyze power relationships. Scholars need to be aware of the potential for network analysis to be used to reinforce existing power structures or to identify and target marginalized groups.
* Transparency and accountability: Scholars need to be transparent about their use of network analysis and accountable for the results of their research.

Despite these challenges, network analysis is a powerful tool that can be used to make significant contributions to the digital humanities. Scholars who are aware of the challenges and take steps to address them can use network analysis to gain new insights into a wide variety of digital humanities topics.

# Summary

To summarize, networks are everywhere, they are complex and dynamic, but most important they are important. I hope this workshop was able to equip you with understanding the basics of network analysis and its application in digital humanities research. I also hope that you leave this workshop with the confidence to create or work with a dataset, but more importantly the confidence to use tools that are designed to visualize and analyze your data. It makes for a fun way to unveil those hidden connections and apply them to your research goals.

# References

* Ahnert, R., Ahnert, S. E., Coleman, C. N., & Weingart, S. B. (2020). *The Network Turn: Changing Perspectives in the Humanities*. Cambridge University Press.
* Düring, M. (2015). From hermeneutics to data to networks: Data extraction and network visualization of historical sources. *The Programming Historian*.
* "Francis Bacon Network [2, 1562-1626, 61-100%]." Six Degrees of Francis Bacon. <http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min_confidence=60&type=network>
* Giovanna Ceserani. (2015). Interactive *Visualizations for British Architects on the Grand Tour in eighteenth-century Italy* [Palladio Components, HTML, CSS, JavaScript, JSON, and Markdown files]. Stanford Digital Repository. Available at <http://purl.stanford.edu/ct765rs0222>
* Humanities + Design Lab at Stanford University. (2023). Palladio. Version 1.2.4. [Web]. <https://hdlab.stanford.edu/palladio/>.
* The Open Syllabus Project. (2023). Retrieved from <https://www.opensyllabus.org/>.