

Art Analysis through Network Analysis of WikiArt Data

Lameya Islam*

* Faculty of Science & Engineering, Åbo Akademi University

1. INTRODUCTION

The digital transformation of cultural heritage offers unprecedented access to art collections, allowing insights into artistic relationships, influence, and community dynamics. This project aims to explore a dataset derived from WikiArt, an online art encyclopedia, to analyze and understand relationships within the art world, including influences, nationalities, movements, and institutions. Specifically, this study answers the following questions:

1. Who are the most influential artists?
2. Which art movements are the most impactful?
3. Which institutions have played a significant role in art history?
4. Which nationalities have the largest representation among artists?
5. What are the largest communities within the art network?

The analysis involves preprocessing, constructing a network model, applying network analysis techniques, and visualizing findings to uncover patterns and communities within the art world.

2. METHODOLOGY

2.1 PREPROCESSING AND EXPLORATORY DATA ANALYSIS

Data Description: The dataset comprises four files:

- **artists.csv:** Contains artist details, including nationality, number of artworks, and active years.
- **relationships.csv:** Includes artist relationships—those they influenced and were influenced by, alongside associated institutions and schools.
- **institutions.csv:** Lists institutions with details about their location.
- **schools.csv:** Contains information about schools associated with various artists.

Steps in EDA:

1. **Data Loading and Inspection:** Each file was loaded into a pandas DataFrame and inspected for structure and completeness.
2. **Missing Values Analysis:** Missing values were identified and handled, with special focus on columns like nation, influenced_by, and influenced_on.
3. **Data Distribution:**
 - a. **Nationality:** Visualized to determine the countries with the largest representation.

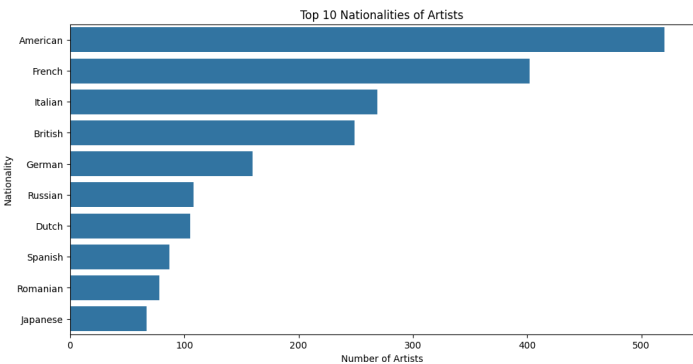


Figure 1: Distribution of nationalities of artists

- b. **Movements:** Counted and analyzed to identify popular movements.

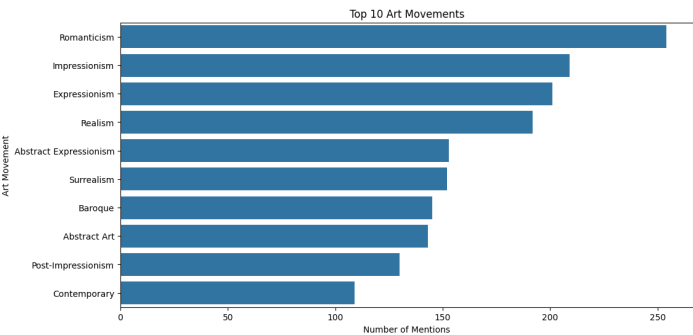


Figure 2: Distribution of movements

- c. **Artwork Counts:** Processed from the totalworkstitle column in artists.csv, extracting only numeric values to represent the number of artworks for each artist.

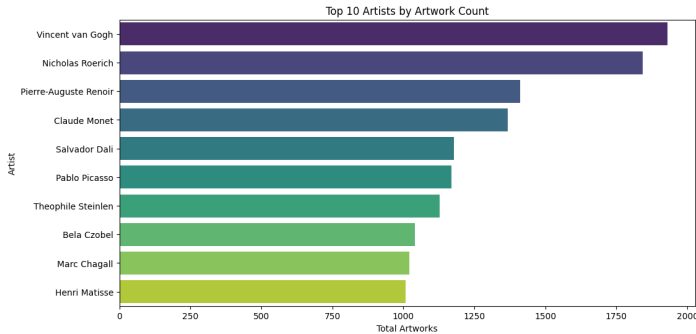


Figure 3: Distribution of artworks of artists

2.2 NETWORK CREATION AND ANALYSIS

Network Representation:

1. **Nodes:**
 - a. **Artists:** Each artist from artists.csv was added as a node with attributes like nationality and total artwork count.
 - b. **Institutions and Schools:** Nodes were added from institutions.csv and schools.csv.
2. **Edges:**
 - a. **Influences:** Directed edges were added from relationships.csv where one artist influenced another (influenced_on and influenced_by).
 - b. **Affiliations:** Edges were added to represent artist affiliations with institutions and schools.

Considering the main goal of this work, analysis has been performed on most influential artists, institutions and movements.

Analysis on Most Influential Artists:

- The artists were separated based on influenced_on and influenced_by data.
- After a few data processing, both data has been concatenated.
- The centrality measure 'PageRank' has been used to find the most influential artists.
- For unique visualization, 'pyvis' has been used.

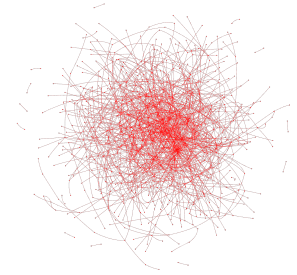


Figure 4: Most Influential Artists

Also, search option can be performed here by preferences to see which artist has most influence.

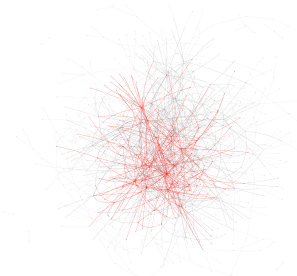


Figure 5: Most Influential Artists (After Search: Pablo Picasso)

Analysis on Most Influential Institutions:

- The institutions were separated based on influenced_on and influenced_by data.
- After a few data processing, both data has been concatenated.
- The centrality measure 'PageRank' has been used to find the most influential institutions.
- For unique visualization, 'pyvis' has been used.

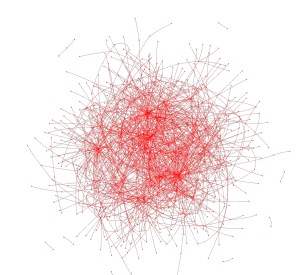


Figure 6: Most Influential Institutions

Also, search option can be performed here by preferences to see which institution has most influence.

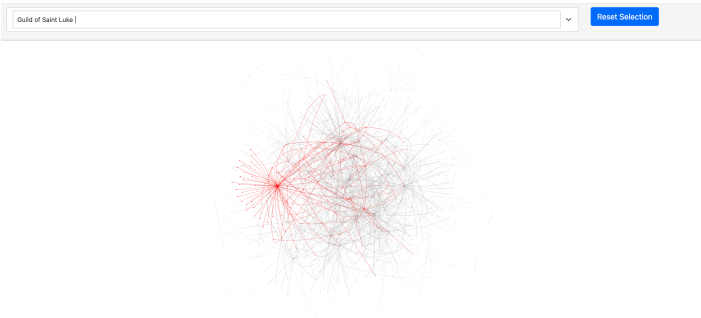


Figure 7: Most Influential Institutions (After Search: Guild of Saint Luke)

Analysis on Most Influential Movements:

- The movements were separated based on influenced_on and influenced_by data.
- After a few data processing, both data has been concatenated.
- The centrality measure ‘PageRank’ has been used to find the most influential artists.
- For unique visualization, ‘pyvis’ has been used.

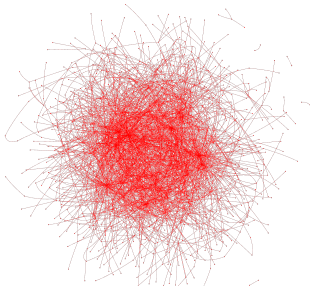


Figure 8: Most Influential Movements

Also, search option can be performed here by preferences to see which movement has most influence.

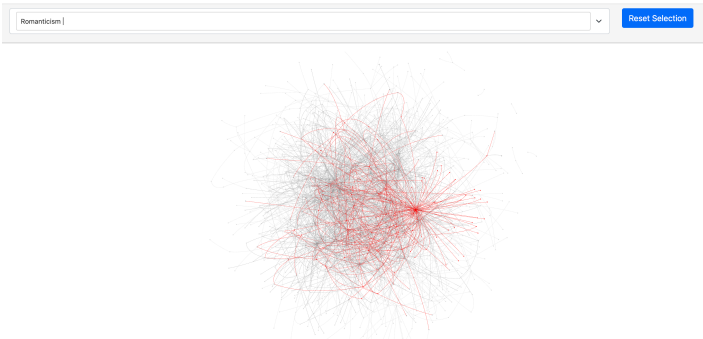


Figure 9: Most Influential Movements (After Search: Romanticism)

2.3 METHODS FOR NETWORK ANALYSIS

To answer the research questions, various network analysis methods were applied:

- **Centrality Measures:** Degree and betweenness centrality were used to identify the most influential nodes (artists, movements, and institutions).
- **Community Detection:** The **louvain** and **k-clique** clustering algorithm was applied to identify tightly connected communities.
- **Clustering Coefficients:** Used to explore local clusters and identify nodes central to community formation.

3. RESULTS

3.1 MOST INFLUENTIAL ARTISTS

Using **PageRank (Centrality Measure)**, the top influential artists were identified. High degree centrality indicates artists with many connections (either influencing others or being influenced), while high betweenness centrality highlights artists acting as bridges between different artist communities.

Artists	Centrality
Charles Gibbons	0.0077
Michel Majerus	0.0071
Paul Nash	0.0065
Jackson Pollock	0.0060
Oleksandr Murashko	0.0058

Table 1: Top 5 Influential Artists by Centrality Measure

3.2 MOST INFLUENTIAL INSTITUTIONS

Institutions that had high centrality were noted for their influence. These institutions, like the Royal Academy of Fine Arts and the Guild of Saint Luke, served as important nodes in the network.

Institutions	Centrality
Gluid of Saint Lake	0.0192
UCL Slade School of Fine Art	0.0138
Ecole des Beaux-Arts	0.0129

Table 2: Top 3 Influential Institutions by Centrality Measure

3.3 MOST INFLUENTIAL MOVEMENTS

Art movements with high centrality were identified as the most influential. These movements, such as Romanticism and Impressionism, were linked with multiple artists and played key roles in connecting different styles and periods.

Institutions	Centrality
Romanticism	0.0111
Impressionism	0.0097
Expressionism	0.0095

Table 2: Top 3 Influential Movements by Centrality Measure

3.4 NATIONALITY WITH MAJORITY REPRESENTATION

The nationality distribution revealed that most artists originated from countries with prominent art histories, such as the United States, France, Italy, and the UK.

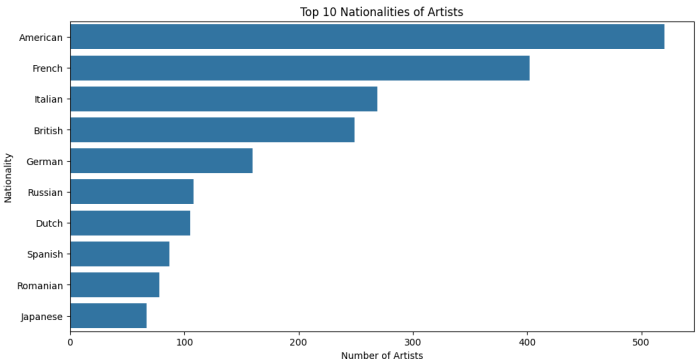


Figure 10: Distribution of Artists by Nationality

3.5 LARGEST COMMUNITY IN THE NETWORK

There are 5 largest communities, and the size of each community is different.

Communities	Size
Community 1	293
Community 2	175
Community 3	114
Community 4	100
Community 5	85

Table 3: Largest Communities and their size

3.6 NETWORK CLUSTER

The louvain and k-clique community detection revealed significant clusters, representing interconnected groups of artists and institutions. Large communities were found around influential

movements and styles, showing clusters with similar stylistic or temporal affinities.

Artist Influence Network Clusters

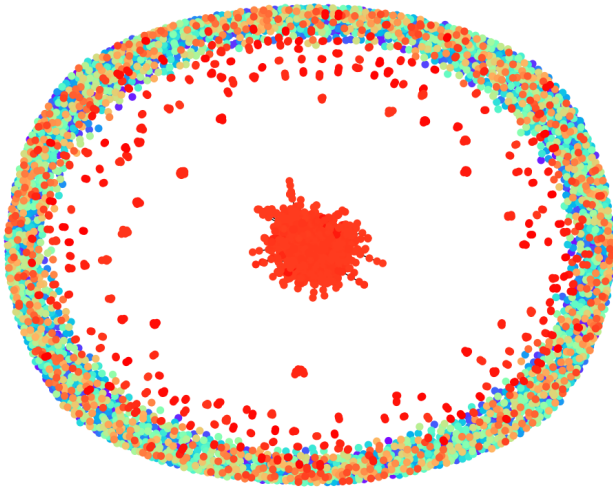


Figure 11: Artist Influence Network Clusters

4. CONCLUSION

This analysis successfully identified influential artists, movements, and institutions within the WikiArt dataset, revealing clusters that provide insights into the structure of artistic influence. High centrality artists and institutions often act as hubs or bridges

within the art network, connecting different movements and geographical areas. Nationalities like the United States, France, and Italy were prominently represented, reflecting their historical significance in art.

Challenges and Future Work

- **Challenges:** Some challenges included managing missing data in influence relationships and balancing the large size of the network for readable visualization.
- **Future Work:** Future analyses could use dynamic network analysis to observe changes in influence over time or explore other clustering methods for deeper community detection insights.

The project demonstrates the utility of network analysis in uncovering patterns of influence and relationships in art history, offering a foundation for more nuanced studies on cultural evolution and interconnectedness.

REFERENCES

- [1] <https://networkx.org/documentation/latest/tutorial.html>