

Mini Project II – Art Analysis

1. Introduction

With evolving network analysis, we can now glimpse into the analysis of complex relationships between various connection arts may have. This assignment leverage data from WikiArt, to construct and analyze a network that represents relationships between artists, institutions, movements, and geographical locations. The aim is to uncover answers to certain questions to uncover patterns and entities among the art works.

2. Methodology

2.1 Preprocessing and EDA

Data Overview: The analysis utilized four main datasets from WikiArt:

- **artists.csv:** Contains detailed information on artists, such as their unique IDs, names (title), nationalities (nation), URLs, and total number of artworks produced (totalWorksTitle).
- **relationships.csv:** Describes various relationships between artists, such as friendships and influences, and details affiliations with schools and institutions.
- **institutions.csv:** Lists art institutions, including their names (title), cities (city), and countries (country).
- **schools.csv:** Provides information about art schools and movements, with URLs and names (title).

Exploratory Analysis: Important data distributions were mapped using preliminary EDA.

- **Nationalities:** The distribution of artists' nationalities showed that most artists came from European countries and the United States.

```
=== Data Distributions ===

Artists Nationality Distribution:
nation
American    520
French       402
Italian      269
British      249
German       160
Russian      108
Dutch        105
Spanish       87
Romanian      78
Japanese     67
```

Figure: Nationalities

- **Artworks:** The analysis showed a few artists created a substantial number of works. Most other contributed less.
- **Institutions:** Location review highlighted cities like Paris, New York, and London in distribution of institutions.

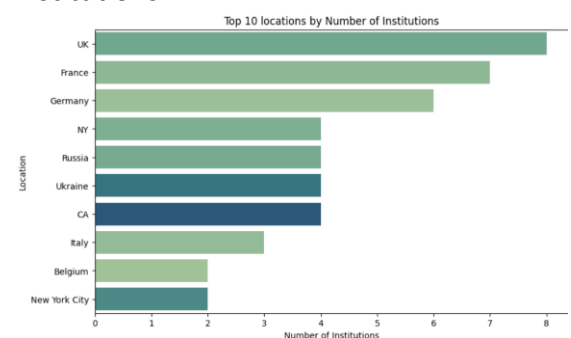


Figure: Institutions

Missing Value Treatment: Detecting and handling missing values was crucial for completeness of the dataframes.

- Each dataset was inspected for missing values, and the total count

of missing data points per column was recorded.

- Columns with missing data were cleaned by replacing NaN with default values .

```
=== Missing Values in DataFrames ===  
Missing Values in artists_df:  
nation      32  
year         1  
dtype: int64  
  
Missing Values in relationships_df:  
movements   40  
type         1  
dtype: int64  
  
Missing Values in institutions_df:  
city         2  
country      2  
dtype: int64  
  
No missing values in schools_df
```

Figure: Missing Values

Conversion of Data Types: Some columns in the relationships.csv dataset (friends, influenced_by, influenced_on) were stored as strings representing lists.

2.2 Network Creation

A directed graph (DiGraph) was used to show network whereas the nodes showed artists, institutions and schools and the edges represent relationships.

- Each artist node was with attributes like title, nation, and totalWorksTitle. Separate nodes were created for each institution and school.

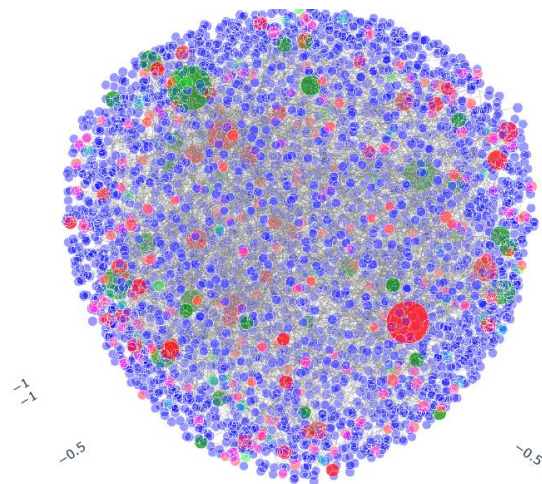


Figure: Interactive visualization

- Directed edges connected influencing artists to the influenced. Bidirectional edges were used for mutual friendships between artists. Edges connected artists to affiliated institutions and schools.

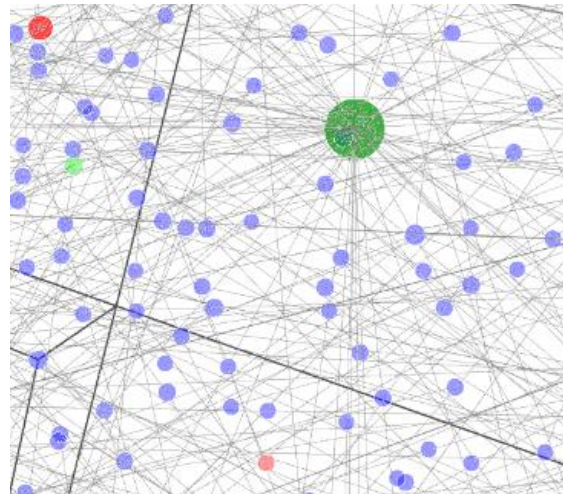


Figure: Nodes and Edges

2.3 Analysis Techniques

Centrality Measures:

- Degree Centrality: Measured the number of direct connections of each node.

- Betweenness Centrality: Quantified the importance of nodes as intermediaries in the network.
- Eigenvector Centrality: Evaluated nodes based on their connections to other highly connected nodes.

Community Detection: The Louvain method was used for community detection. It identified clusters within the network. The analysis allowed grouping nodes that are more interconnected with each other than with the rest of the network.

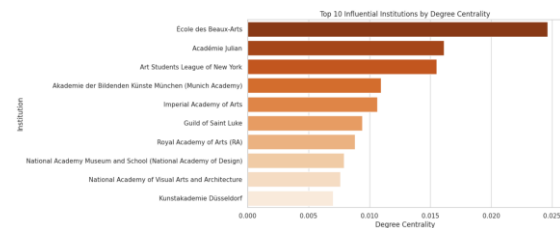
3. Results

3.1 Most Influential Artists

The analysis showed the most influential artists according to different centrality measures:

- Degree Centrality identified artists with the highest number of connections. Artists such as Pablo Picasso, Claude Monet, and Henri Matisse ranked highly.
- Betweenness Centrality highlighted artists who served as bridges between disparate art groups. For instance, Jackson Pollock and Paul Gauguin became important linkages, impacting different eras and styles.
- Artists such as Édouard Manet and Amedeo Modigliani, who were not only influential but also had connections to other powerful people, were featured in Eigenvector Centrality.

3.3 Most Influential Institutions



3.4 Concentration of Artists by Nationality

France, the United States, Italy, and Russia were among the top countries with the most artists.

3.5 Biggest Communities in the Network

The Louvain method revealed the largest communities within the art network:

Community 1 included many artists from the Impressionist and Post-Impressionist periods.

Community 2 reflected the modern art movement, including influential figures like Picasso and Matisse.

Other communities corresponded to movements like Surrealism and Expressionism, highlighting how artists within each movement were interconnected through influence and shared ideas.

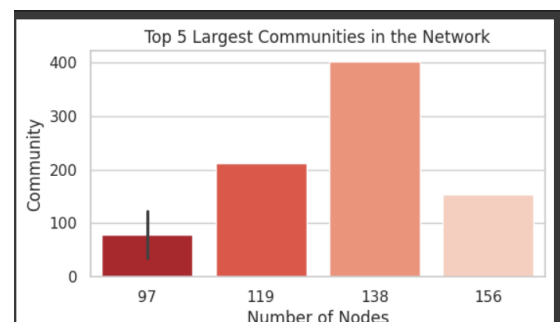


Figure: Communities

4. Conclusions

Challenges and Insights:

The structured format of the datasets made loading and preprocessing relatively straightforward. However, handling relationships with multiple connections was really challenging and was difficult to implement. It required careful data cleaning and connecting logic to execute properly.

Complexities in Network Analysis:

Analyzing centrality and understanding the influence of artists required selecting the right metrics to draw conclusions. The community detection was complex, requiring to constantly focus to connect groups and communities.

Results and Expectations:

The results confirmed known historical patterns, such as the prominence of artists like Picasso and Monet and the importance of movements like Cubism and Impressionism.

The identification of major communities reflected the interconnected nature of specific movements and periods, shedding light on how artists influenced one another within their circles.

To conclude, this assignment, while difficult, enabled successful analysis of complex networks.