Network Analysis of Art Influences

Mini-Project III

Mahira Ibnath Joytu

1. Introduction:

The cultural heritage sector is undergoing a digital transformation, with museums, galleries, and institutions digitizing their collections to increase accessibility and engagement. However, transitioning from physical to virtual spaces requires more than simple replication—it demands innovative approaches to convey the interconnected nature of art and history.

Artists and movements have historically influenced one another, creating a rich network of relationships. Understanding these connections is crucial for appreciating the evolution of art and the broader cultural context.

This project aims to explore and analyze data derived from WikiArt, an extensive online art encyclopedia that contains rich information about artists, art movements, and institutions. Through this analysis, I seek to create a network that represents these relationships and use network analysis techniques to extract meaningful insights. The goals of this analysis are to:

a. Identify the most influential artists in the network.

b. Discover which art movements have had the greatest influence.

c. Identify which institutions have been most pivotal in shaping art history.

d. Determine which nationalities have the highest concentration of artists.

e. Uncover the largest and most interconnected communities within the network.

To achieve these goals, the project will utilize a combination of data preprocessing, network creation, and various network analysis techniques such as centrality measures, community detection, and clustering. By visualizing and interpreting the results, I hope to illuminate the complex and multifaceted nature of the art world and its interconnected web of relationships.

Understanding these relationships can provide valuable insights not only for historians and art scholars but also for the public interested in learning more about the forces that have shaped the world of art over time. This project serves as an example of how digital tools and data science can be leveraged to uncover hidden patterns and offer new perspectives in the analysis of historical and cultural information.

2. Methodology:

2.1. Data Collection:

The data used for this project comes from WikiArt, a comprehensive online art encyclopedia known for its extensive repository of information about artists, their works, movements, and associated institutions. To conduct a robust analysis of the art world, I utilized the four primary CSV datasets gained from this resource:

2.1.1 artists.csv:

This file contains detailed information about individual artists. Key columns include:

- *artistUrl*: A unique URL identifier for each artist.
- *id*: A unique numeric ID assigned to each artist.
- *image*: A link to an image representing the artist.
- *nation*: The nationality of the artist.
- *title*: The name of the artist.
- *totalWorksTitle*: The total number of artworks produced by the artist.
- *year*: The active years or the period during which the artist was most productive.

2.1.2 institutions.csv:

This dataset contains information about art institutions that have influenced or housed notable artists. Key columns include:

- *city*: The city where the institution is located.
- *country*: The country of the institution.
- *title*: The name of the institution.
- *url*: A link to the institution's page on WikiArt.

2.1.3 relationships.csv:

The most complex dataset, this file maps out relationships between different artists and the entities they are associated with. Key columns include:

- *artistUrl*: The unique URL identifier for each artist.
- *friends*: Other artists closely associated with this artist.
- *influenced_by*: A list of artists or movements that influenced this artist.
- *influenced_on*: A list of artists or movements that this artist influenced.
- *institution*: The art institutions the artist was affiliated with.
- *movements*: The art movements the artist was part of or influenced by.
- *school*: The painting schools the artist was associated with.
- *type*: Specifies whether the row represents an artist or a collection.

2.1.4 schools.csv:

Contains information about painting schools that played a significant role in shaping artistic techniques and styles. Key columns include:

- *title*: The name of the painting school.
- *url*: A link to the school's page on WikiArt.

2.2 Data Preprocessing:

Effective data preprocessing is essential for ensuring the quality and accuracy of any analysis, particularly when dealing with large and complex datasets like those sourced from WikiArt. The data preprocessing phase involved several key steps to clean, transform, and prepare the data for network analysis.

2.2.1 Loading and Initial Inspection

I started by loading the four datasets—artists.csv, institutions.csv, relationships.csv, and schools.csv—into Python using pandas. After loading, I conducted an initial inspection to understand the data structure and content by:

- Checking column names using columns() function of Pandas
- Checking for missing values using isnull().sum()

2.2.2 Handling Missing Values

To maintain data integrity and ensure seamless analysis, I handled missing values as follows:

- ***artists.csv***: Missing values in the ***nation*** and ***year*** columns were replaced with 'N/A' to indicate unknown information
- ***institutions.csv***: Missing values in ***city*** and ***country*** columns were filled with 'N/A'.
- ***relationships.csv***: Columns such as ***friends, influenced_by, influenced_on, institution,*** and ***school*** were filled with an empty string ' ' to denote missing data without disrupting parsing.

2.3 Network Creation

To represent the relationships between artists, movements, and institutions, I constructed a directed graph using ***networkx***. The graph G was built as follows:

- **Node Creation**: Each unique artist, movement, school, and institution was added as a node with an appropriate type attribute (e.g., 'artist', 'movement', 'school', 'institution').
- **Edge Creation**: Edges were added to represent various relationships:
  a. ***Influence Relationships***:
     If an artist was influenced by others, nodes were created (if not already present), and directed edges were added from the influencing entity to the artist *(relationship='influenced_by').*
  b. If an artist influenced others, edges were added from the artist to the influenced entities (*relationship='influences'*).
- ***Affiliations:*** For artists associated with institutions, nodes representing these institutions were created, and edges were added from the institution to the artist (*relationship='affiliated_with'*).

Each row in ***relationships.csv*** was iterated through to parse and add relevant nodes and edges:

- The influenced_by and influenced_on columns were split into lists and processed to add relationships.

- Special handling ensured movements and schools were tagged correctly by checking if their identifiers contained keywords like 'artists-by-art-movement' or 'artists-by-painting-school'.

This approach resulted in a comprehensive directed graph, G, containing **3,227** nodes and **7,629** edges, encapsulating the network of influences and affiliations in the dataset.

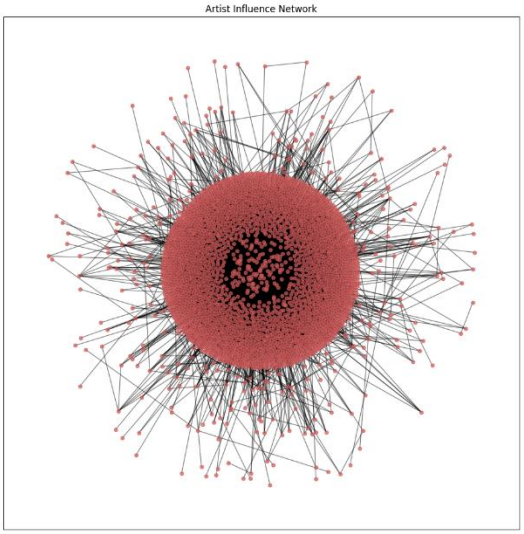

Fig 1: Created network of Artist Influence

2.4 Network Analysis Techniques

To explore the structure and dynamics of the art network, the following network analysis methods were applied:

1. Degree Centrality: Degree centrality was calculated to identify nodes with the most direct connections. This measure highlights influential artists, movements, or institutions by their number of direct relationships. The top five nodes with the highest degree centrality were determined and analyzed for their role within the network.

2. Betweenness Centrality: Betweenness centrality was used to find nodes that act as critical connectors or bridges within the network. Nodes with high betweenness are essential for the flow of information or influence between different parts of the network. This analysis helped identify key figures that facilitate interactions between distinct groups or sub-networks.

3. PageRank: PageRank, a measure originally developed by Google, was employed to rank nodes based on their relative importance within the network. This metric considers both the number and quality of connections, allowing for the identification of nodes that hold significant influence due to their connections to other prominent nodes.

4. Community Detection: The greedy modularity algorithm was applied for community detection to uncover clusters of closely connected nodes. Communities in the network often represent groups of artists, movements, or institutions with shared characteristics or mutual influences. The analysis revealed the number of distinct communities and identified the largest one, providing insights into how the art world clusters into influential subgroups.

5. Clustering Coefficient: The clustering coefficient was calculated to understand the degree of interconnectedness among a node's neighbors. This measure reflects how tightly a node's direct connections form a cluster. The overall average clustering coefficient of the network was evaluated to assess its general cohesiveness. Additionally, nodes with the highest individual clustering coefficients were identified to highlight regions of dense interconnection.

-1: A perfect negative linear relationship. When one increases, the other decreases.

0: No linear relationship between the two variables.

3. Results

3.1 Most Influential Artists

Using degree centrality as a measure, the most influential artists in the network were identified. These artists have the highest number of direct connections, signifying their influence across the network. The top 10 most influential artists are:

- Pablo Picasso (25 connections)
- Paul Cézanne (23 connections)
- Caravaggio (21 connections)
- Rembrandt (21 connections)
- Titian (17 connections)
- Gustave Courbet (17 connections)
- Nicolas Poussin (16 connections)
- Diego Velázquez (16 connections)
- Michelangelo (15 connections)
- Raphael (15 connections)

These results align with known art history, highlighting artists who played pivotal roles in influencing others and contributing to major art movements.

3.2 Most Influential Movements

The analysis identified the most influential art movements using centrality measures, which reflect the number of connections and the extent of influence each movement has within the network. The top 10 most influential movements are as follows:

1. **Cubism** (11 connections)
2. **Byzantine Art** (9 connections)
3. **Expressionism** (6 connections)
4. **Surrealism** (6 connections)
5. **Impressionism** (5 connections)
6. **Dada** (5 connections)
7. **Gothic Art** (3 connections)
8. **Post-Impressionism** (3 connections)
9. **Dutch Golden Age** (3 connections)
10. **Native Art** (3 connections)

3.3 Most Influential Institutions

The analysis of institutions based on centrality measures identified the most influential art institutions that have nurtured or shaped the trajectories of many artists. The top 10 most influential institutions are:

- Unnamed Institution (centrality: 1.673)
- École des Beaux-Arts
- Académie Julian
- Art Students League
- Akademie der Künste
- Imperial Academy of Arts, Saint Petersburg
- Guild of Saint Luke
- Royal Academy
- National Academy
- National Academy of Visual Arts and Architecture

3.4 Nationalities Concentrating the Majority of Artists

The distribution of artists by nationality reveals which countries have historically been centers of artistic activity. The top 10 nationalities with the highest number of artists are:

- American: 520 artists
- French: 402 artists
- Italian: 269 artists
- British: 249 artists
- German: 160 artists
- Russian: 108 artists
- Dutch: 105 artists
- Spanish: 87 artists
- Romanian: 78 artists
- Japanese: 67 artists

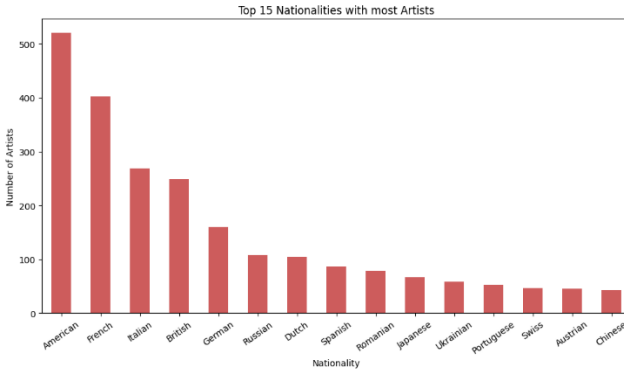

Fig 2: Top 15 Nationalities

3.5 Largest Communities in the Network

Community detection analysis revealed a significant finding which is:

Largest community size: 2019 members

A sample of members from this community includes: Benjamin Cañas, Fern Coppedge, Ottone Rosai, Romul Nuțiu, and Paul Brach. This community likely represents a highly interconnected group of artists, showcasing how artistic influences and collaborations form dense clusters within the broader network.

3.5 Network Analysis

3.5.1 Top Nodes by Degree Centrality: Institutions were highly central, with an unnamed institution having the highest degree centrality (1.6733), followed by École des Beaux-Arts and Académie Julian. Pablo Picasso was the most influential artist by this measure (0.0127), showcasing his strong influence in the network.

3.5.2 Top Nodes by Betweenness Centrality: The unnamed institution also had the highest betweenness centrality (0.9291), indicating its pivotal role as a bridge within the network. Paul Cézanne (0.0077) and Gustave Courbet (0.0049) were the top artists, suggesting their importance in connecting different parts of the network.

3.5.3 Top Nodes by PageRank: The unnamed institution again ranked highest (0.4135), emphasizing its importance within the network. Ancient Greek Painting and contemporary artists like Damien Hirst had notable PageRank values, highlighting a diverse influence across eras.

3.5.4 Community Detection: A total of 57 communities were detected, with the largest community comprising 1992 members. This indicates a significant interconnected group, reflecting shared themes and relationships within the art network.

3.5.5 Clustering Coefficient: The network's average clustering coefficient was 0.0733, suggesting relatively sparse local connections. However, specific artists such as Apelles, Ancient Greek Pottery, and Andrei Rublev had perfect clustering coefficients (1.0), indicating that their immediate neighbors formed complete subgraphs.

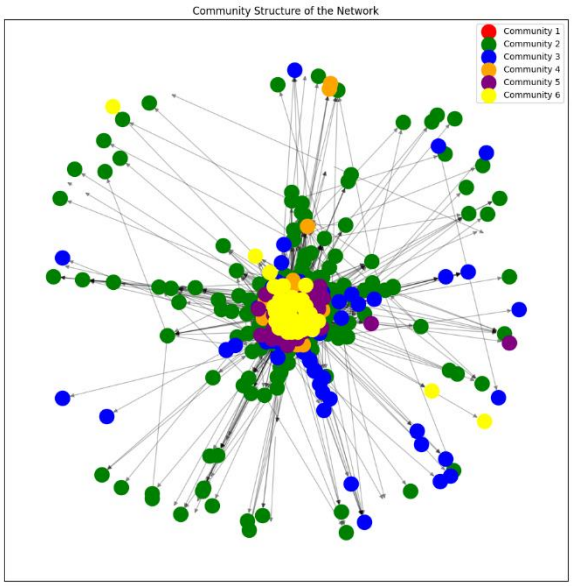

Fig 3: Community Structure of the Network

4. Key Findings

The network analysis provided comprehensive insights into influential artists, movements,

institutions, nationalities, and community structures within the art world.

4.1 Influential Artists and Movements

The most influential artists identified through degree centrality include Pablo Picasso, Paul Cézanne, Caravaggio, and Rembrandt. These figures are known for their groundbreaking contributions that shaped movements such as Cubism and Baroque art, impacting generations of artists. The top movements identified were Cubism, Byzantine Art, Expressionism, and Surrealism, showcasing both modern and historical influences. Cubism, in particular, stands out for its role in redefining artistic form and perspective.

4.2 Influential Institutions

Key institutions such as École des Beaux-Arts, Académie Julian, and the Art Students League emerged as central to the development of major art movements and the nurturing of influential artists. These institutions have historically been hubs of innovation, shaping both regional and global art trends.

4.3 Nationalities Concentrating the Majority of Artists

The distribution analysis revealed that the United States, France, and Italy have the highest number of artists, highlighting their historical roles as cultural epicenters. This finding aligns with the emergence of influential movements like Impressionism and Abstract Expressionism.

4.4 Community Structure

Community detection uncovered a large network of interconnected artists, with the largest community comprising 2019 members. This suggests a dense web of shared influences and affiliations, emphasizing the collaborative nature of the art world. The presence of both prominent and lesser-known artists within this community indicates the diverse sources of artistic inspiration.

4.5 Interpretation

The results underscore the interconnected and multifaceted nature of the art world, where influential figures and movements are interwoven with educational and cultural institutions. Artists like Picasso and movements such as Cubism highlight how revolutionary ideas can redefine art. The network also showcases the continued relevance of older movements and the significant impact of formal training provided by key institutions.

These findings illustrate that art evolves through complex relationships influenced by historical, cultural, and educational structures, providing a deeper understanding of the global art network.

5. Conclusion

This project successfully analyzed the network of artists, movements, and institutions using WikiArt data, uncovering patterns that highlight the interconnected nature of the art world. The identification of key figures like Pablo Picasso and influential movements such as Cubism reaffirmed their pivotal roles in shaping art history. The analysis also underscored the significance of institutions like École des Beaux-Arts and Art Students League in fostering artistic development.

The findings showed that regions like the United States and France have been major contributors to global art, aligning with historically significant periods of artistic innovation. Community analysis revealed densely connected groups of artists, highlighting the collaborative nature of art, where both renowned and lesser-known figures contribute to shared themes and movements.

Challenges included data inconsistencies and accurately mapping relationships. Future work could involve integrating weighted edges for influence strength and expanding data sources for a more comprehensive view.

Overall, this project emphasized the value of network analysis in exploring the complex web of art history, offering new insights into how artists, movements, and institutions interact to shape the evolution of art. Understanding these networks deepens appreciation for the collective forces that drive artistic innovation.