# Exploring the Networks of Artistic Influence

Mete Harun Akcay

## I. Introduction

Art stands as one of humanity's most profound and enduring achievements. Across centuries, artists have shaped and been shaped by a vast network of institutions, artistic movements, and mutual influences. To understand this complex web of influence more comprehensively, network analysis methods offer powerful tools for uncovering relationships and identifying key figures and movements within the art world.

This report seeks to analyze and reveal influential figures and structures within art history, utilizing data from WikiArt, the largest visual art encyclopedia in the world. Specifically, the goal of this paper is to answer the following research questions:

- Which were the most influential artists?
- Which were the most influential movements?
- Which were the most influential institutions?
- Which nationalities concentrate the majority of artists?
- Which are the biggest communities in the network?

Through preprocessing and exploratory data analysis (EDA) methods, a network graph was created, complemented by interactive visualizations that allow users to adjust the influence levels dynamically. Ultimately, this report discusses the most influential concepts within the art network and identifies the largest communities, providing insights into the interconnected landscape of artistic influence.

## II. Methodology

The data used for this project consists of four CSV files, which contain information scraped from WikiArt. These files contain information on artists, their relationships, institutions, and schools. All of them were converted to data frames, and after a preprocessing process, they were combined in a single dataframe.

### A. Preprocessing and EDA

*1) Artists:* Artists dataframe had columns named artistUrl, id, image, nation, title, totalWorksTitle and year. There were 32 cells in nation column which were empty, and only one empty cell in year column. Since these numbers are negligible compared to 2996 rows in the data frame, they were removed.

One of the insights that can be extracted from the dataset is the total number of artworks created by each artist. The bar chart in Figure 1 displays the top 10 artists ranked by the total number of artworks attributed to them. Notably, Vincent van Gogh leads with nearly 2,000 works, followed closely by Nicholas Roerich. Artists like Pierre-Auguste Renoir, Claude Monet, Salvador Dalí, Pablo Picasso and Theophile Steinlen also feature prominently, each with over 1,000 works.

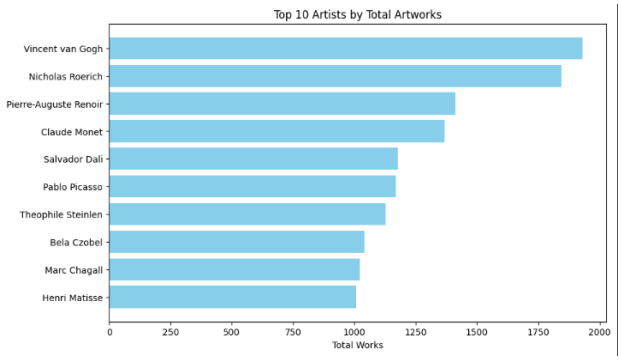


Fig. 1: Top 10 Artists by Total Artworks

The chart given in Figure 2 below illustrates the distribution of artists' active periods by century. The data reveals a notable concentration of artists who began their careers in more recent centuries, particularly from the 18th century onwards. The largest number of active artists is recorded in the 20th century, followed by the 19th century, making more than 80% of the artists in the dataset when combined.

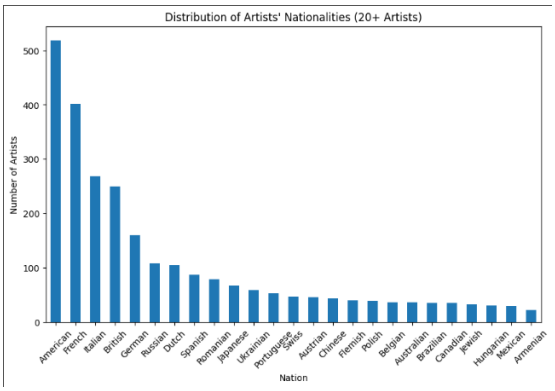


Fig. 2: Distribution of Artists' Nationalities

An observation can be made when the nationality and artwork columns are analysed. Figure 3 shows the nations having more than 20 artists in the dataset. Americans, with more than 500 artists dominate the list, followed by French. Other countries that have significant number of artists in the dataset are Italy, Britain and Germany. However, there is a significant difference when the nationalities are ranked based on the average number of artworks by artists. According to Figure 4, Russians, with average 140 artworks per artist almost double the second nationality, French. The list is followed by Ukrainian and Spanish. Americans, despite having the most artists in the dataset, do not seem as productive as the

other significant nationalities, having 50 artworks per artist in average.

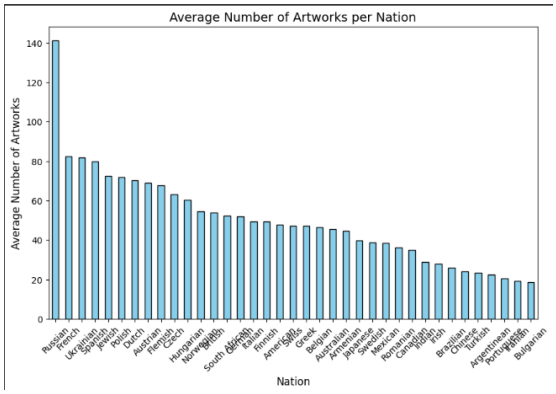


Fig. 3: Average Number of Artworks per Nation

*2) Relations:* Relations data frame contain relations for each artist; thus, it has the same number of rows as the artists data frame (2996). The following table (Table 1) shows the columns of this data frame along with their missing values. The row with missing type value was removed; however, the others were kept. The reason is that these missing values did not stem from errors in data collection or inconsistencies but rather represent cases where certain information was simply not applicable. For instance, some artists may not have documented friendships, influences, or affiliations with institutions or schools. To handle these cases consistently, missing values in such columns were modified as empty lists. This data frame provides valuable information for the influence level of the entities; however, a more detailed examination of these relationships will be presented in the network analysis section. Thus, in order to avoid repetition, the results will not be discussed in this section.

| Column Name | Missing Values |
|---|---|
| artistUrl | 0 |
| friends | 2580 |
| influenced_by | 2512 |
| influenced_on | 2637 |
| institution | 2362 |
| movements | 40 |
| school | 1966 |
| type | 1 |

TABLE I: Missing Values of Relations

*3) Institutions:* This data frame had the following columns: city, country, title and url. There were two rows without the city and country information, which were removed. The bar chart given in Figure 4 illustrates the distribution of art institutions across the top 5 countries by count. Each bar represents a country, with different colors indicating the cities within each country where these institutions are located.

France and the UK have the highest number of institutions, with 7 and 8 institutions respectively. The institutions in France are all located in Paris, while those in the UK are distributed across cities like London and East Sussex, as indicated by the color-coded legend. Germany follows with a diverse spread across multiple cities such as Berlin, Dresden, Düsseldorf, and Frankfurt, showcasing a more decentralized distribution of institutions. Russia and New York (NY) have fewer institutions, with Moscow and St. Petersburg in Russia and New York City in the United States.

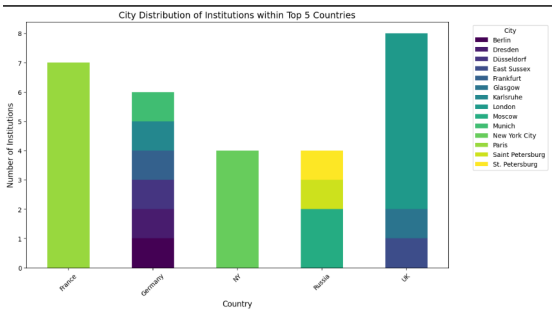


Fig. 4: City Distribution of Institutions within Top 5 Countries

*4) Schools:* Schools data frame had 220 rows and only 2 columns, which were title and url. There were no missing values; thus, no operation was done on this data frame.

### *B. Network Creation and Analysis*

After the preprocessing was completed, four data frames were combined into a single data frame named final_df. This data frame included all the information needed to answer the research questions of this paper. Then, a directed graph was constructed to represent relationships between artists, movements, institutions, nationalities and schools. Each entity type was represented as a unique node in the graph, with edges indicating various types of relationships, as outlined below:

- Artists: Each artist in the dataset was added as a node. Directed edges between artist nodes were created to represent influence relationships, where an edge from artist A to artist B signifies that artist A influenced artist B.
- Movements: Artistic movements were added as separate nodes, and edges were created from each artist to the movements that they were involved.
- Institutions: They were also represented as nodes. Each artist associated with an institution was connected by a weighted directed edge from the artist to the institution, showing the number of artists each artist influenced.
- Nations: Each nationality was represented as a node. Then, an edge between each artist and his/her nationality was added.
- Schools: Similarly, schools were added as nodes, and edges were established from each artist to the school(s) they were associated with.

The resulting graph was a complex network containing multiple layers of relationships between artists, movements, institutions, nations and schools. It had 3700 nodes and 10595 edges in total.

## III. Results

### A. Which were the most influential artists?

To determine the most influential artists in the dataset, out-degree centrality was calculated for each artist node in the network. Figure 5 shows the top 10 most influential artists. In order to emphasize the influence level, node sizes scaled accordingly. In addition, as it can be seen from the color bar, nodes were represented with different colors.

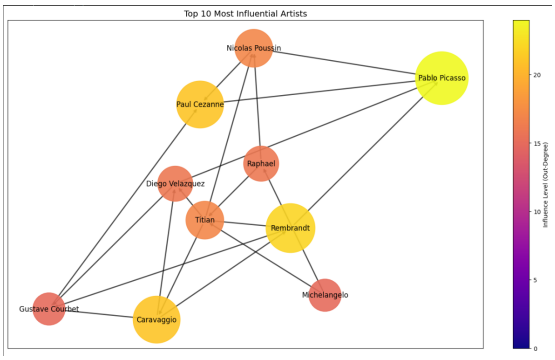


Fig. 5: Top 10 Influential Artists

From the analysis, Pablo Picasso stands out as the most influential artist in this dataset, as shown by his larger, yellow-colored node, indicating both high out-degree centrality and extensive influence on other artists. Other highly influential artists include Paul Cézanne, Caravaggio, and Rembrandt, whose relatively large node sizes and color intensities demonstrate their significant impact within the network. The exact number of influences of the most influential artists are given below in Table 2.

| Artist | Influence Level |
|---|---|
| Pablo Picasso | 24 |
| Paul Cezanne | 22 |
| Caravaggio | 21 |
| Rembrandt | 21 |
| Titian | 17 |
| Gustave Courbet | 17 |
| Nicolas Poussin | 16 |
| Diego Velazquez | 16 |
| Michelangelo | 15 |
| Raphael | 15 |

TABLE II: Top 10 Most Influential Artists

### B. Which were the most influential movements?

Determining the influence of movements was done by calculating the in-degree of each movement node, which represents the number of artists influenced by that movement. A higher in-degree indicates a larger number of artists connected to that movement, signifying a greater level of influence in the artistic network.

As can be seen from the Figure 6, Romanticism, Impressionism, and Expressionism emerged as some of the most influential movements, influencing 254, 212, and 203 artists, respectively. Realism and Abstract Expressionism also show high influence, though to a slightly lesser extent, highlighting their pivotal roles in the art world.

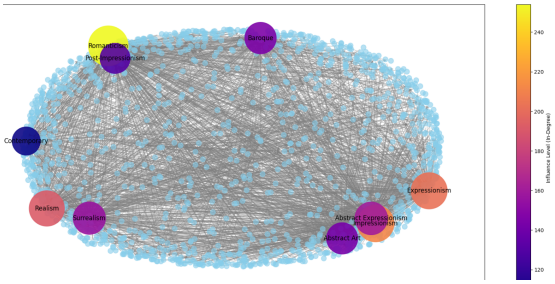


Fig. 6: Top 10 Influential Movements

The blue nodes in the background represent artists, densely clustered around these influential movements. This clustering reflects the broad appeal and adoption of certain movements by numerous artists. Additionally, the proximity of movements like Impressionism, Abstract Expressionism, and Abstract Art suggests interconnectedness, possibly indicating stylistic or historical progression, where certain artists may have contributed to or evolved through multiple movements. Similarly, Romanticism and Post-Impressionism appear to influence a similar group of artists, as indicated by their spatial proximity and overlapping connections in the network.

### C. Which were the most influential institutions?

Influence of institutions was calculated using the similar logic used in movements' influence calculation, by summing the in-degree values of each institution, which represents the number of artists influenced by the artists associated with each institution. Again, in this visualization, larger node sizes and warmer colors indicate institutions with higher in-degrees, signifying greater influence.

As seen in the Figure 7, Cole Des Beaux Arts is by far the most influential institution, with 187 associated artists. It is followed by Acad Mie Julian, with 104 artists, and Art Students League, with 97 artists.

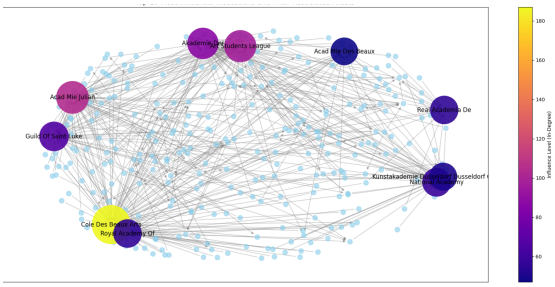


Fig. 7: Top 10 Influential Institutions

### D. Which nationalities concentrate the majority of artists?

The top nations, as already identified in Figure 2, have been visualized here in Figure 8 with a focus on their interconnected influence within the artistic network. This visualization not only highlights the influential nations but also reveals patterns

in how nations influenced each other. Unsurprisingly, geographically and culturally proximate nations tend to overlap in their artistic influence. For instance, Japan and Russia show a degree of overlap, which reflects the historical cultural exchanges in Asia. Similarly, the Spanish, Italian, and Dutch clusters align, which is expected given their shared European cultural landscape and artistic exchanges over the centuries. Although causality cannot be extracted from the figure, with a little historical knowledge, the proximity of American artists to both British and French artists can be explained by the fact that American art evolved with significant input from British and French traditions.

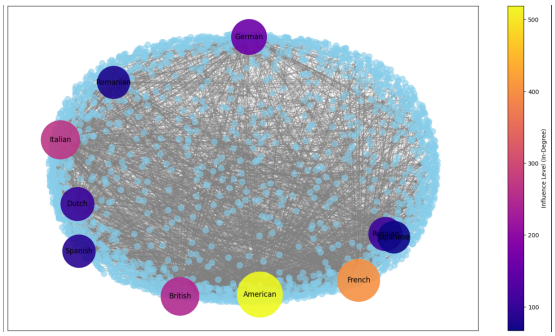


Fig. 8: Top 10 Nations

### *E. Which are the biggest communities in the network?*

In order to detect the biggest communities, the graph was first converted to an undirected form to facilitate the process, focusing on the connections between artists, movements, institutions, nations, and schools. The Louvain algorithm was then applied to partition the graph into communities based on the network's modularity, which measures the strength of division of a network into clusters.

Each detected community was then categorized to include artists, movements, institutions, nations, and schools, grouping nodes based on these types. By sorting communities by size, the top three largest communities were identified for further analysis. Table 3 lists the top three largest communities along with their counts for each entity.

| Category | Community 16 | Community 2 | Community 3 |
|---|---|---|---|
| Artists | 518 | 344 | 260 |
| Movements | 38 | 13 | 8 |
| Institutions | 21 | 3 | 10 |
| Nations | 7 | 7 | 9 |
| Schools | 28 | 30 | 27 |
| Total Entities | 612 | 397 | 314 |

TABLE III: Top 3 Largest Communities in the Network

Moreover, these communities were visualized, with each node type color-coded for clarity:

- Artists are represented in light blue.
- Movements are represented in orange.
- Institutions are represented in green.
- Nations are represented in purple.
- Schools are represented in red.

These visualizations reveal the structural composition of each community, highlighting the interconnectedness within and between different types of entities in the artistic network. The communities show clusters where artists of specific movements, institutions, nations, or schools are densely interconnected, reflecting shared influences or historical affiliations.

Community 16 (Figure 9) emerges as the largest, with a total of 612 entities. It includes 518 artists, 38 movements, and 21 institutions. Notable members in this community are artists such as Jackson Pollock and Edward Hopper, along with institutions like the National Academy and the Art Students League.

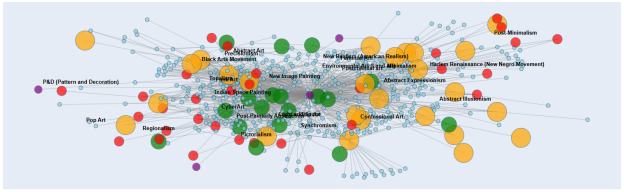


Fig. 9: Community 16

Community 2 (Figure 10) follows, containing 397 entities. It features artists such as Michelangelo and Raphael and is connected to historical movements like the Early Renaissance and Baroque, as well as institutions like the Guild of Saint Luke.

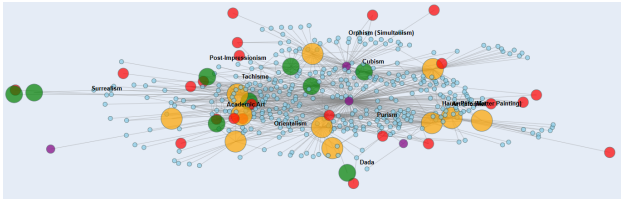


Fig. 10: Community 2

Community 3 (Figure 11) is slightly smaller, with 314 entities. This community reflects a strong German and Expressionist influence, including artists like Wassily Kandinsky and Franz Marc, and institutions such as the Kunstakademie Düsseldorf.

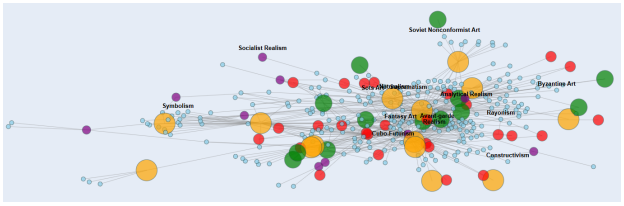


Fig. 11: Community 3

### *F. BONUS: Interactive Visualization*

In addition to the provided network analysis, an interactive system was implemented to allow users to explore the network by selecting node types (e.g., artists, movements, institutions, nations) and setting a minimum degree threshold to filter

influential nodes. This customization feature enables users to focus on particular entities within the network based on their influence level.

For instance, the provided figure below (Figure 12) demonstrates the network when the node type is set to ”artist” and the minimum degree is set to 6. This configuration displays only those artists who have an in-degree centrality above 6. As seen, artists like Pablo Picasso, Paul Gauguin, and Vincent van Gogh are featured, reflecting their significant impact on other artists.

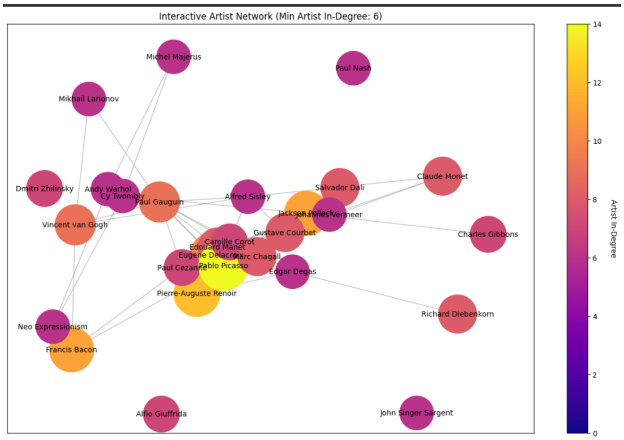


Fig. 12: Interactive Visualization (type = ’artist’, degree = 6)

## IV. Conclusion

This project successfully implemented a comprehensive network analysis on the influence within the art world, examining artists, movements, institutions, nations, and schools, as well as identifying significant communities within this interconnected landscape. Through the combination of data preprocessing, exploratory data analysis, and network construction, the project identified and visualized key influential figures and structures. These results were then visualized and discussed in detail, revealing insights into artistic influence patterns across artists, institutions, and art movements.

Working with such a large and complex network presented certain challenges. The large volume of data made it difficult to visualize the entire network without overwhelming the graphical output, which necessitated the implementation of filtering options, such as the minimum-degree threshold, to focus on the most influential nodes. Additionally, the creation of an interactive system proved technically challenging, as the author of this paper was not familiar with ipywidgets package of Python. However, with the assistance of ChatGPT for coding support, these interactive elements were successfully incorporated, providing a flexible and engaging way to explore the network’s structure.

Overall, the results largely aligned with the initial expectations, confirming that well-known artists and institutions held significant influence and that geographically proximate nations and historical artistic movements shared close connections. Future work could focus on expanding the dataset to include more recent or underrepresented artists, as well as exploring other network analysis techniques, to examine how influence patterns evolve over time.

The complete code and interactive elements of this project can be accessed in the following Colab notebook: Mini Project 3 – Art Analysis.ipynb.