Exploring Artist-Exhibition Relationships at MoMA

Albert Garcia Uría

Introduction

The Museum of Modern Art (MoMA) has played a fundamental role in shaping modern and contemporary art. This project uses Network Analysis to explore relationships between artists and exhibitions at MoMA. A bipartite network is constructed linking artists and exhibitions, allowing us to identify participation patterns, detect influential artists, and analyze the evolution of connections over time. Additionally, the network is segmented by decades to study the erosion and fragmentation of connectivity over the years.

Research Questions

- What network metrics (degree, betweenness, density, diameter, components) reveal co-occurrence patterns among artists in MoMA exhibitions?
- How are nationalities and genders distributed among the detected communities in the network?
- What structural changes (in cohesion, fragmentation, and density) are observed when segmenting the network by decades?

Objectives

- Construct and analyze a bipartite network between artists and MoMA exhibitions.
- Calculate and compare network metrics to identify co-occurrence patterns among artists.
- Analyze the distribution of nationalities and genders both in the global network and in detected communities.
- Evaluate the structural evolution of the network, considering metrics such as density, diameter, and components, when segmenting it by decades.
- Apply clustering techniques to temporal subnetworks and analyze network fragmentation in each period.

Data and Preprocessing

This section involves loading, cleaning, and preparing the data, selecting only the relevant columns for analysis.

Library Loading

```
library(readr)
library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)
```

```r
library(igraph)
library(ggraph)
library(plotly)
library(ggforce)
library(ggrepel)
library(patchwork)
library(forcats)
```

Data Loading

```r
directors <- read_csv2("MoMADirectorsDepartmentHeads.csv", locale = locale(encoding = "windows-1252"))
exhib <- read_csv("MoMAExhibitions1929to1989.csv", locale = locale(encoding = "windows-1252"))
```

Data Cleaning and Selection

Selecting only relevant columns

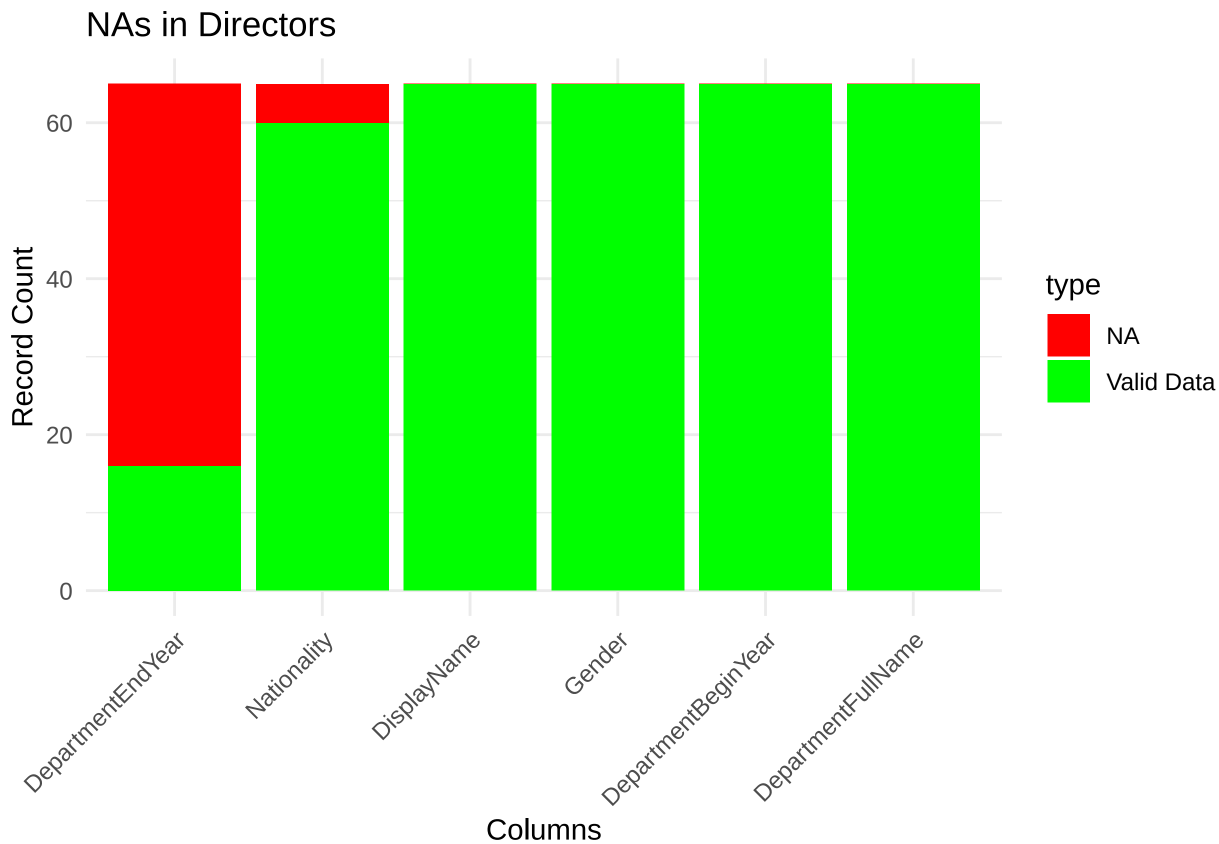
```r
directors <- select(directors, DisplayName, Gender, Nationality, DepartmentBeginYear, DepartmentEndYear

exhib <- select(exhib, ExhibitionID, ExhibitionTitle, ExhibitionBeginDate, ExhibitionEndDate, Constituen
```

Converting dates to the proper format

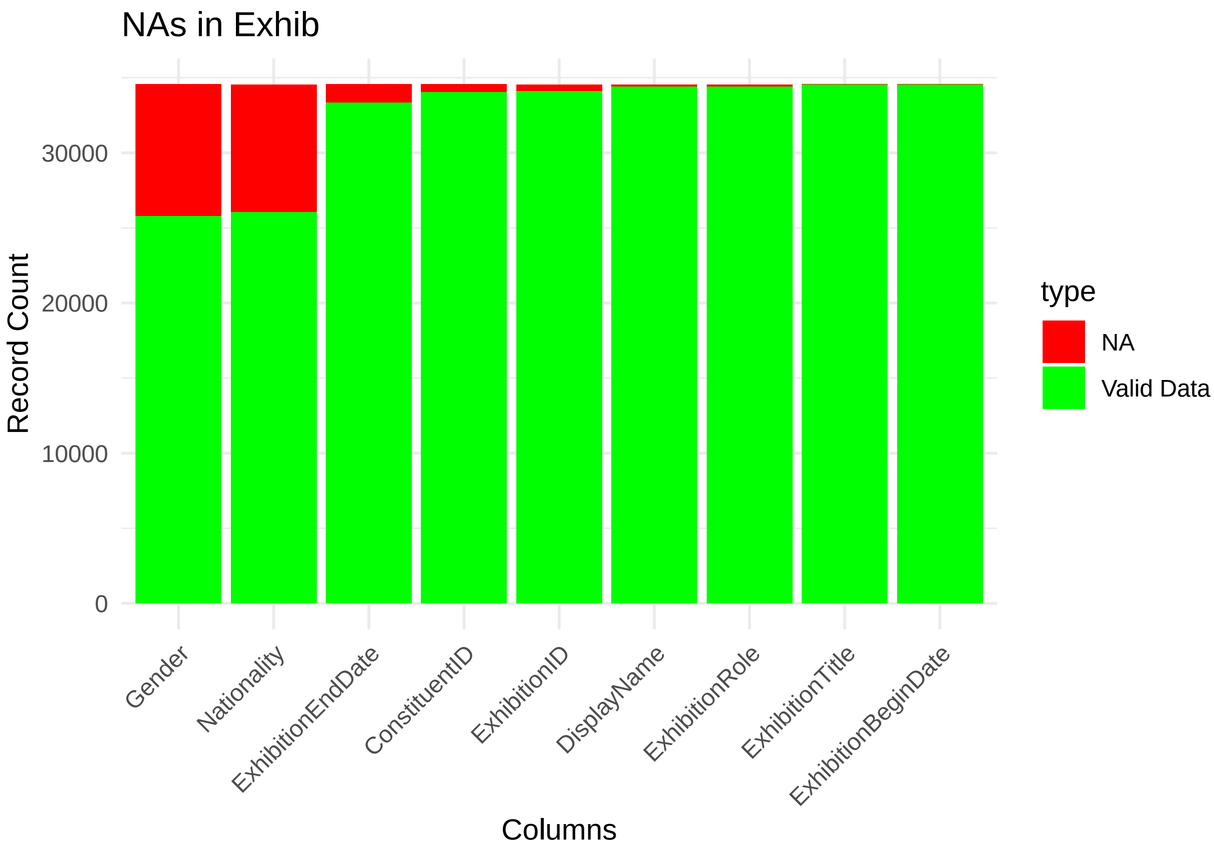
```r
exhib$ExhibitionBeginDate <- mdy(exhib$ExhibitionBeginDate)
exhib$ExhibitionEndDate <- mdy(exhib$ExhibitionEndDate)
```

Missing Values Detection

```r
visualize_na <- function(dataset, title = "Total and NA per column") {
  na_counts <- colSums(is.na(dataset))
  df_plot <- data.frame(
    column = names(na_counts),
    NA_count = as.numeric(na_counts)
  )
  total_count <- nrow(dataset)
  df_plot$total <- total_count
  df_plot$Valid_data <- total_count - df_plot$NA_count
  df_long <- pivot_longer(df_plot, cols = c("Valid_data", "NA_count"), names_to = "type", values_to = "
  df_long$column <- factor(df_long$column, levels = df_plot$column[order(-df_plot$NA_count)])
  plt <- ggplot(df_long, aes(x = factor(column), y = value, fill = type)) +
    geom_bar(stat = "identity") +
    labs(title = title, x = "Columns", y = "Record Count") +
    scale_fill_manual(values = c("Valid_data" = "green", "NA_count" = "red"), labels = c("NA", "Valid Da
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
  return(plt)
}

visualize_na(directors, "NAs in Directors")
```




```
visualize_na(exhib, "NAs in Exhib")
```



Since DepartmentEndYear has more than half of the total missing values, I decided to drop this column to preserve more data for analysis before removing all rows with NA values. In Exhibit, we can drop the rows with the remaining values because roughly the same number of values NA are missing from the Gender and Nationality columns.

Deleting columns with many missing values

```
directors <- directors %>% select(-DepartmentEndYear) %>% drop_na
exhib <- exhib %>% drop_na
```

Network Analysis

Network Analysis is used to model relationships between artists based on their co-occurrence in exhibitions. The network structure consists of: - **Nodes**: Artists. - **Edges**: Links between artists who have appeared together in at least one exhibition.

Building the Artist Co-occurrence Network

Below is an example in R that performs several important operations. First, the relevant columns from the exhib dataset (in this case, ExhibitionID, DisplayName, Nationality, and Gender) are selected, and rows containing missing values are removed. Then, a list of edges connecting two artists who participated in the same exhibition is created, and from this list, an undirected graph is constructed. Finally, gender attributes are assigned to the nodes of the graph using the available information, setting the value "Unknown" for those cases where such information is unavailable.

```
temp <- exhib %>% select(ExhibitionID, DisplayName, Nationality, Gender) %>% drop_na()
```

```r
edgelist <- merge(temp, temp, by = "ExhibitionID") %>%
  filter(DisplayName.x != DisplayName.y)

graph <- graph_from_data_frame(edgelist[, c("DisplayName.x", "DisplayName.y")], directed = FALSE)

df_nodes <- unique(temp[, c("DisplayName", "Gender")])
V(graph)$Gender <- df_nodes$Gender[match(V(graph)$name, df_nodes$DisplayName)]

V(graph)$Gender[is.na(V(graph)$Gender)] <- "Unknown"
```

Network Metrics Calculation

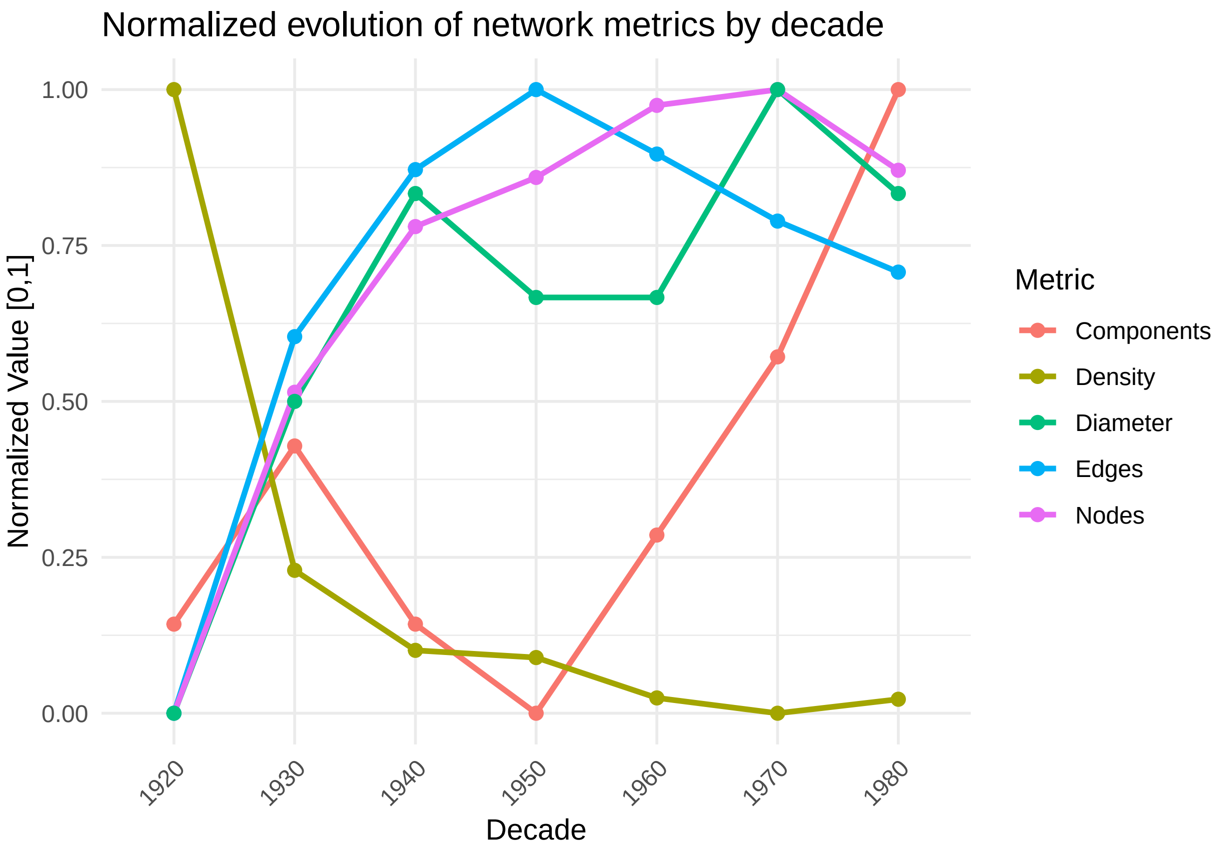
```r
centrality <- data.frame(
  Artist = V(graph)$name,
  Degree = degree(graph),
  Betweenness = betweenness(graph)
)
```

Decade-long network erosion analysis

To study the evolution of the network, the dataset is segmented by decade. This process allows us to observe how the network fragments and the loss of connectivity by filtering exposures from different periods.

```r
exhib$Decade <- floor(year(exhib$ExhibitionBeginDate) / 10) * 10

results <- data.frame(
  Decade = integer(),
  Nodes = integer(),
  Edges = integer(),
  Density = numeric(),
  Diameter = numeric(),
  Components = integer()
)

for (decade in sort(unique(exhib$Decade))) {
  sub_exhib <- filter(exhib, Decade == decade)

  sub_temp <- sub_exhib %>%
    select(ExhibitionID, DisplayName) %>%
    drop_na()

  sub_edgelist <- merge(sub_temp, sub_temp, by = "ExhibitionID") %>%
    filter(DisplayName.x != DisplayName.y)

  sub_graph <- graph_from_data_frame(sub_edgelist[, c("DisplayName.x", "DisplayName.y")],
                                     directed = FALSE)

  nodes <- vcount(sub_graph)
  edges <- ecount(sub_graph)

  density_val <- if (nodes > 1) edge_density(sub_graph) else NA

  diameter_val <- if (nodes > 1) diameter(sub_graph) else NA
```

```r
  comp_val <- if (nodes > 0) components(sub_graph)$no else NA

  results <- rbind(results, data.frame(
    Decade = decade,
    Nodes = nodes,
    Edges = edges,
    Density = density_val,
    Diameter = diameter_val,
    Components = comp_val
  ))
}

# Normalize the results
results_norm <- results %>%
  mutate(across(c(Nodes, Edges, Density, Diameter, Components),
                ~ (. - min(.)) / (max(.) - min(.)),
                .names = "{.col}_norm"))

results_long <- results_norm %>%
  pivot_longer(cols = ends_with("_norm"),
               names_to = "Metric",
               values_to = "Value") %>%
  mutate(Metric = recode(Metric,
                         "Nodes_norm" = "Nodes",
                         "Edges_norm" = "Edges",
                         "Density_norm" = "Density",
                         "Diameter_norm" = "Diameter",
                         "Components_norm" = "Components"))

plt <- ggplot(results_long, aes(x = factor(Decade), y = Value, group = Metric, color = Metric)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(title = "Normalized evolution of network metrics by decade",
       x = "Decade",
       y = "Normalized Value [0,1]") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
  plt
```



Nodes: Represent the number of artists in the network for a given decade. An increase in the number of nodes indicates that MoMA exhibited more creators during that period, while a decrease reflects fewer artists present.

Edges: Reflect the co-occurrences of artists in shared exhibitions. An increase in edges signals that exhibitions were organized with several artists coinciding, while a decrease suggests fewer interactions or more specialized shows.

Components: Measure how many disconnected subgroups exist within the network. A high number of components implies greater fragmentation, with isolated or barely overlapping exhibitions; conversely, fewer components indicate greater cohesion.

Density: Indicates the proportion of connections that are activated relative to the total possible. A rise in density implies exhibitions with numerous artists in common, whereas a decrease reveals lower coincidence and more segmented exhibitions.

Diameter: Represents the longest distance between two artists within the main component. A high diameter suggests that the network is more dispersed and some nodes are far apart, while a low diameter indicates that the artists are relatively close to each other.

The evolution of these metrics suggests that over the decades, MoMA has experienced phases of both higher and lower cohesion in its exhibitions. Decades with more nodes and edges (and relatively high density) point to massive or highly attended exhibitions, while periods with more components or a higher diameter indicate a certain fragmentation of the network (exhibitions more isolated from one another). These findings are related to the erosion of the network when filtering by decades: when exhibitions are very diverse and do not share as many artists, the network fragments, increasing the number of components and the diameter, whereas if there are exhibitions that bring together many common artists, fragmentation is reduced and density increases.

Network Centrality Analysis by Decade

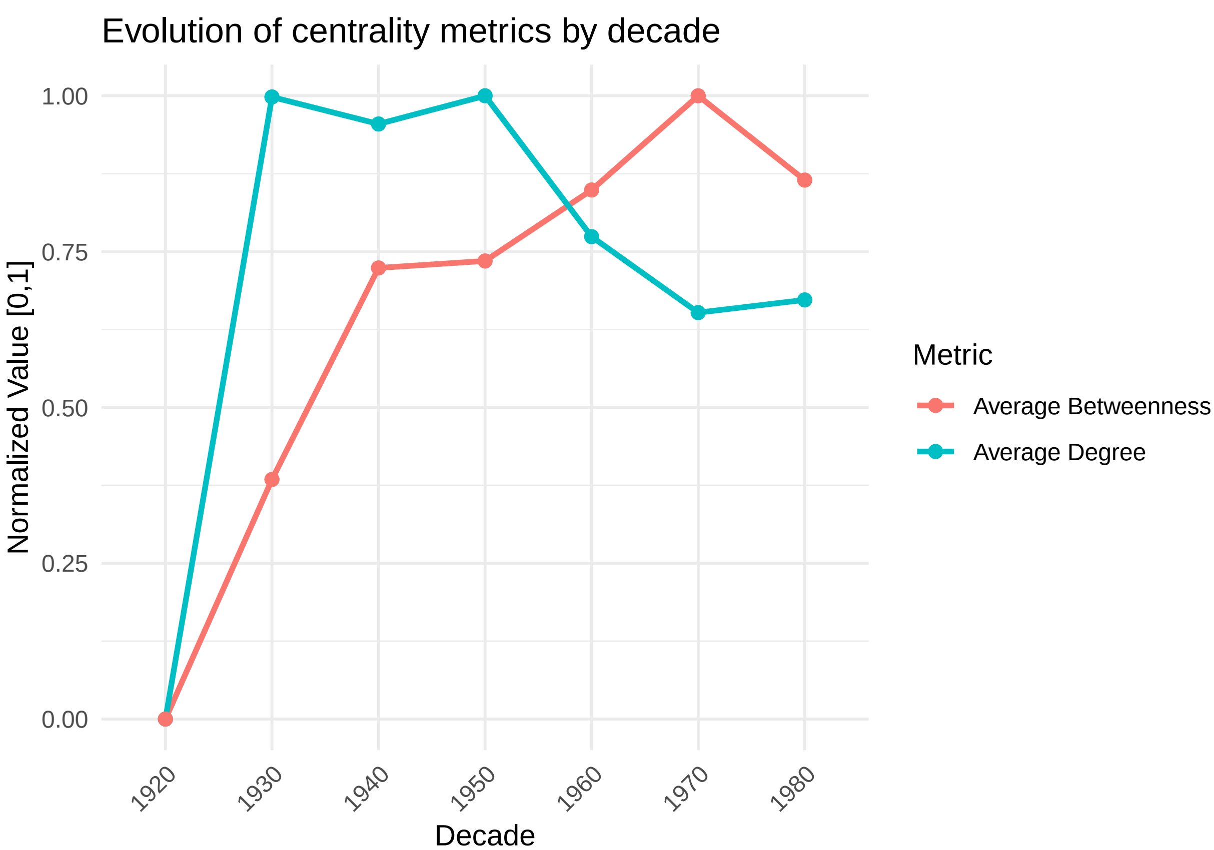
In this section, the evolution of centrality metrics in the artist co-occurrence network over the decades is evaluated. Two key indicators are analyzed:

- **Average Degree**: Represents the average number of connections that each artist has in the exhibitions of a given decade. A higher degree suggests that, on average, artists have more links, which can be interpreted as greater integration or collaboration in the exhibitions.
- **Average Betweenness**: Measures an artist's ability to act as a bridge between others in the network, calculating the frequency with which a node appears on the shortest path between pairs of nodes. A high value indicates that the artist plays a crucial role in connecting different subgroups within the network.

Both metrics are **normalized** on a scale from 0 to 1 to facilitate their comparison in a single interactive graph, clearly showing the evolution of the network's centrality over time.

```r
results_cent <- data.frame(
  Decade = integer(),
  AvgDegree = numeric(),
  AvgBetweenness = numeric()
)

# For each decade, we build the co-occurrence network and calculate the centrality metrics.
for (decade in sort(unique(exhib$Decade))) {
  sub_exhib <- filter(exhib, Decade == decade)

  sub_temp <- sub_exhib %>%
    select(ExhibitionID, DisplayName) %>%
    drop_na()

  sub_edgelist <- merge(sub_temp, sub_temp, by = "ExhibitionID") %>%
    filter(DisplayName.x != DisplayName.y)

  sub_graph <- graph_from_data_frame(sub_edgelist[, c("DisplayName.x", "DisplayName.y")],
                                     directed = FALSE)

  if (vcount(sub_graph) > 0) {
    avg_degree <- mean(degree(sub_graph))
    avg_betweenness <- mean(betweenness(sub_graph))
  } else {
    avg_degree <- NA
    avg_betweenness <- NA
  }

  results_cent <- rbind(results_cent, data.frame(
    Decade = decade,
    AvgDegree = avg_degree,
    AvgBetweenness = avg_betweenness
  ))
}

# Normalize
results_cent_norm <- results_cent %>%
  mutate(across(c(AvgDegree, AvgBetweenness),
              ~ (. - min(.)) / (max(.) - min(.)),
```

```
                .names = "{.col}_norm"))

results_cent_long <- results_cent_norm %>%
  pivot_longer(cols = ends_with("_norm"),
               names_to = "Metric",
               values_to = "Value") %>%
  mutate(Metric = recode(Metric,
                         "AvgDegree_norm" = "Average Degree",
                         "AvgBetweenness_norm" = "Average Betweenness"))

plt_cent <- ggplot(results_cent_long, aes(x = factor(Decade), y = Value, group = Metric, color = Metric)
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(title = "Evolution of centrality metrics by decade",
       x = "Decade",
       y = "Normalized Value [0,1]") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

plt_cent
```



Between the 1920s and 1930s, a significant increase is observed in both the average degree and average betweenness. This indicates that, starting in 1930, artists appear more frequently alongside others (higher degree) and some of them begin to play a more central role as bridges in the network (higher betweenness).

From the 1930s to the 1960s, both metrics fluctuate slightly but remain at relatively high levels. This behavior suggests a consolidated network, with exhibitions bringing together well-connected artists and the presence of nodes that act as essential intermediaries for overall cohesion.

Towards the 1970s, average betweenness reaches its peak, indicating the existence of key artists who decisively link subgroups of the network. At the same time, the average degree also rises, reflecting a period of exhibitions with numerous co-occurrences of artists.

Finally, in the 1980s, a slight decline in both metrics was observed. This decline could be related to greater specialization of exhibitions or curatorial changes that reduced the overlap between artists, causing the network to lose some of its connectivity and bridging nodes.

Network Clustering by Decade

Clustering is a technique that allows nodes in a network to be grouped into communities or clusters, with the aim of identifying connection patterns and underlying structures. In this case, the Fast Greedy algorithm is applied to detect communities in the network for each decade.

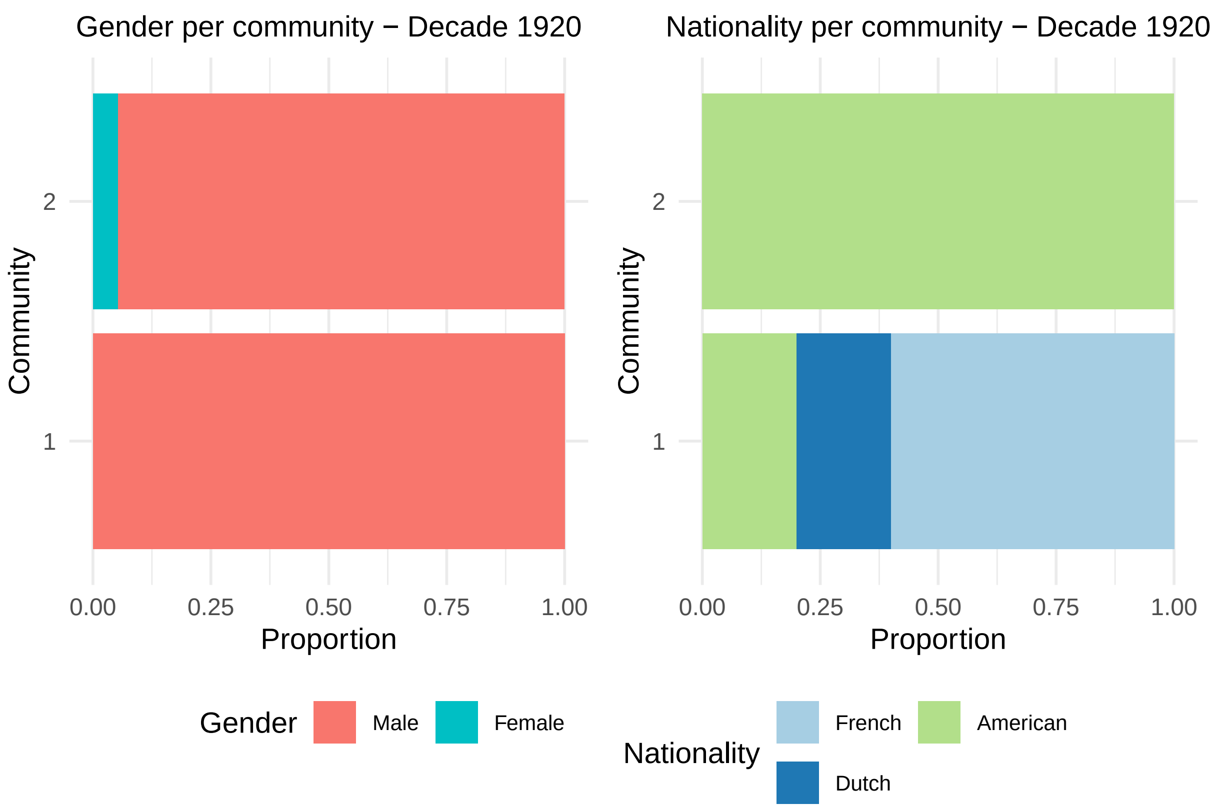
```r
get_graph_by_decade <- function(decade, data = exhib) {
  sub_exhib <- filter(data, Decade == decade)
  sub_temp <- sub_exhib %>%
    select(ExhibitionID, DisplayName, Nationality, Gender) %>%
    drop_na()

  sub_edgelist <- merge(sub_temp, sub_temp, by = "ExhibitionID") %>%
    filter(DisplayName.x != DisplayName.y)

  g <- graph_from_data_frame(sub_edgelist[, c("DisplayName.x", "DisplayName.y")],
                             directed = FALSE)

  df_nodes <- unique(sub_temp[, c("DisplayName", "Gender", "Nationality")])
  V(g)$Gender <- df_nodes$Gender[match(V(g)$name, df_nodes$DisplayName)]
  V(g)$Nationality <- df_nodes$Nationality[match(V(g)$name, df_nodes$DisplayName)]

  # Aply clustering with cluster_louvain
  clusters <- cluster_louvain(g)
  V(g)$community <- membership(clusters)
  V(g)$deg <- degree(g)

  return(g)
}

for(dec in sort(unique(exhib$Decade))) {
  g_dec <- get_graph_by_decade(dec)
  filename <- paste0(dec, ".gml")
  write_graph(g_dec, file = filename, format = "gml")
}
```

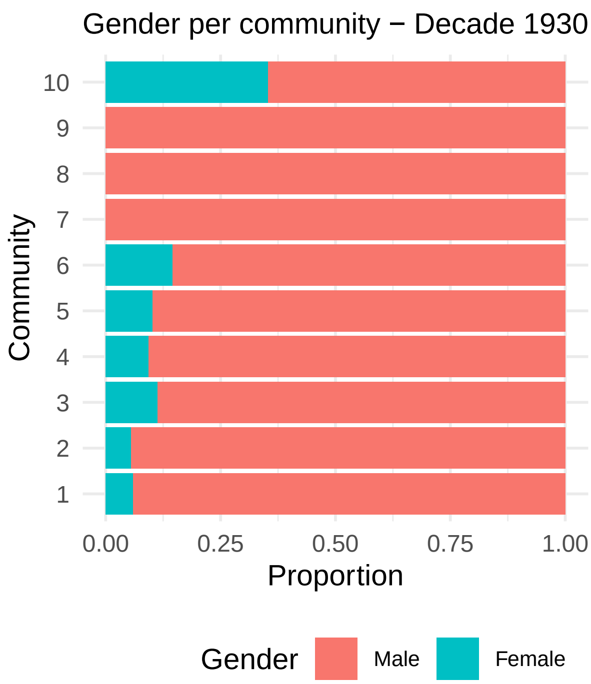
Clustering Analysis by Decade

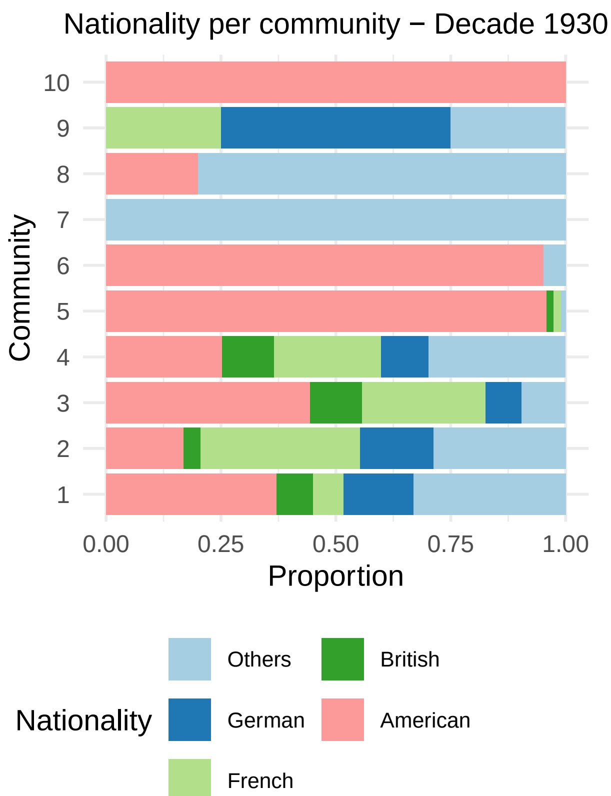
In this analysis, I decided to generate stacked bar charts to visualize the distribution of attributes such as gender and nationality in each **community** in my network, making sure to eliminate null values and convert the "community" variable into a factor for proper clustering. I also chose to group nationalities that represented less than 3% of the total into the "Others" category to avoid cluttering the chart with unrepresentative categories. I then loaded the corresponding graph in GML format for each decade (from 1920 to 1980), generated charts for both attributes, and combined them into a single design, thus facilitating comparison over time and allowing me to analyze the evolution and erosion of communities in the network.

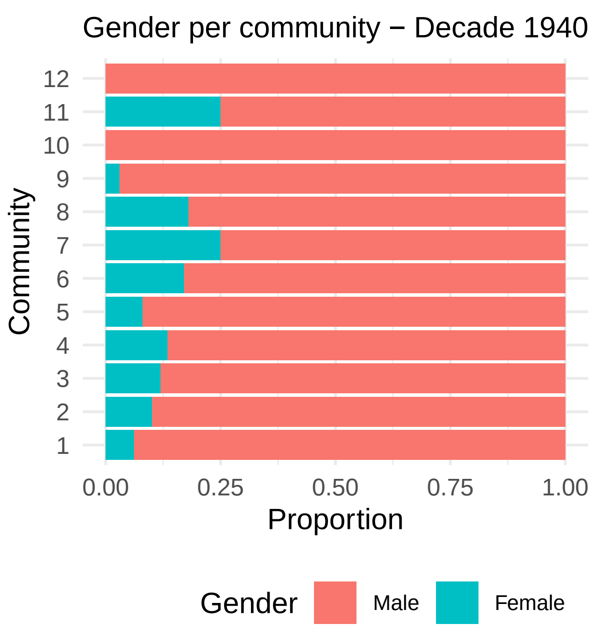
```
plot_attribute_by_community <- function(g, attribute, title_suffix = "") {
  df <- as_data_frame(g, what = "vertices")

  df <- df %>% filter(!is.na(!!sym(attribute)) & !!sym(attribute) != "NULL")

  df$community <- factor(df$community)

  if (attribute == "Nationality") {
    df_counts <- df %>%
      group_by(Nationality) %>%
      summarise(count = n(), .groups = "drop") %>%
      mutate(percentage = count / sum(count))

    major_nations <- df_counts %>% filter(percentage >= 0.03) %>% pull(Nationality)
    df <- df %>%
      mutate(Nationality = ifelse(Nationality %in% major_nations, Nationality, "Others"))
  }
  df_counts <- df %>%
    group_by(community, !!sym(attribute)) %>%
    summarise(count = n(), .groups = "drop")

  colnames(df_counts)[2] <- "attribute_value"

  df_counts <- df_counts %>%
    arrange(community, desc(count)) %>%
    mutate(attribute_value = fct_rev(factor(attribute_value)))

  palette_nationality <- scale_fill_brewer(palette = "Paired")

  p <- ggplot(df_counts, aes(x = community, y = count, fill = attribute_value)) +
    geom_bar(stat = "identity", position = "fill") +
    coord_flip() +
    labs(x = "Community",
         y = "Proportion",
         fill = attribute,
         title = paste(attribute, "per community", title_suffix)) +
    theme_minimal() +
    theme(plot.title = element_text(hjust = 0.5, size = 11),
          axis.text = element_text(size = 9),
          legend.position = "bottom",
          legend.text = element_text(size = 8)) +
    guides(fill = guide_legend(ncol = 2))

  if (attribute == "Nationality") {
    p <- p + palette_nationality
  }

  return(invisible(p))
}

decades <- seq(1920, 1980, by = 10)
for (year in decades) {
```

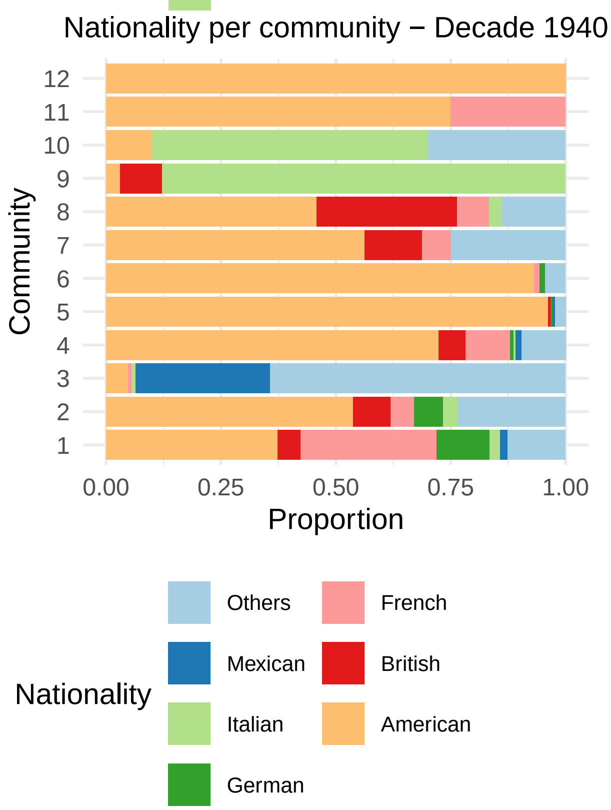
```
  filename <- paste0(year, ".gml")
  g <- read_graph(filename, format = "gml")

  p_gender <- plot_attribute_by_community(g, "Gender", title_suffix = paste("- Decade", year))
  p_nat <- plot_attribute_by_community(g, "Nationality", title_suffix = paste("- Decade", year))

  combined <- (p_gender | p_nat) +
    plot_layout(guides = "collect") &
    theme(legend.position = "bottom")
  print(combined)
  output_filename <- paste0("plots/", year, ".png")
  invisible(ggsave(output_filename, plot = combined, width = 12, height = 6, dpi = 300))
}
```

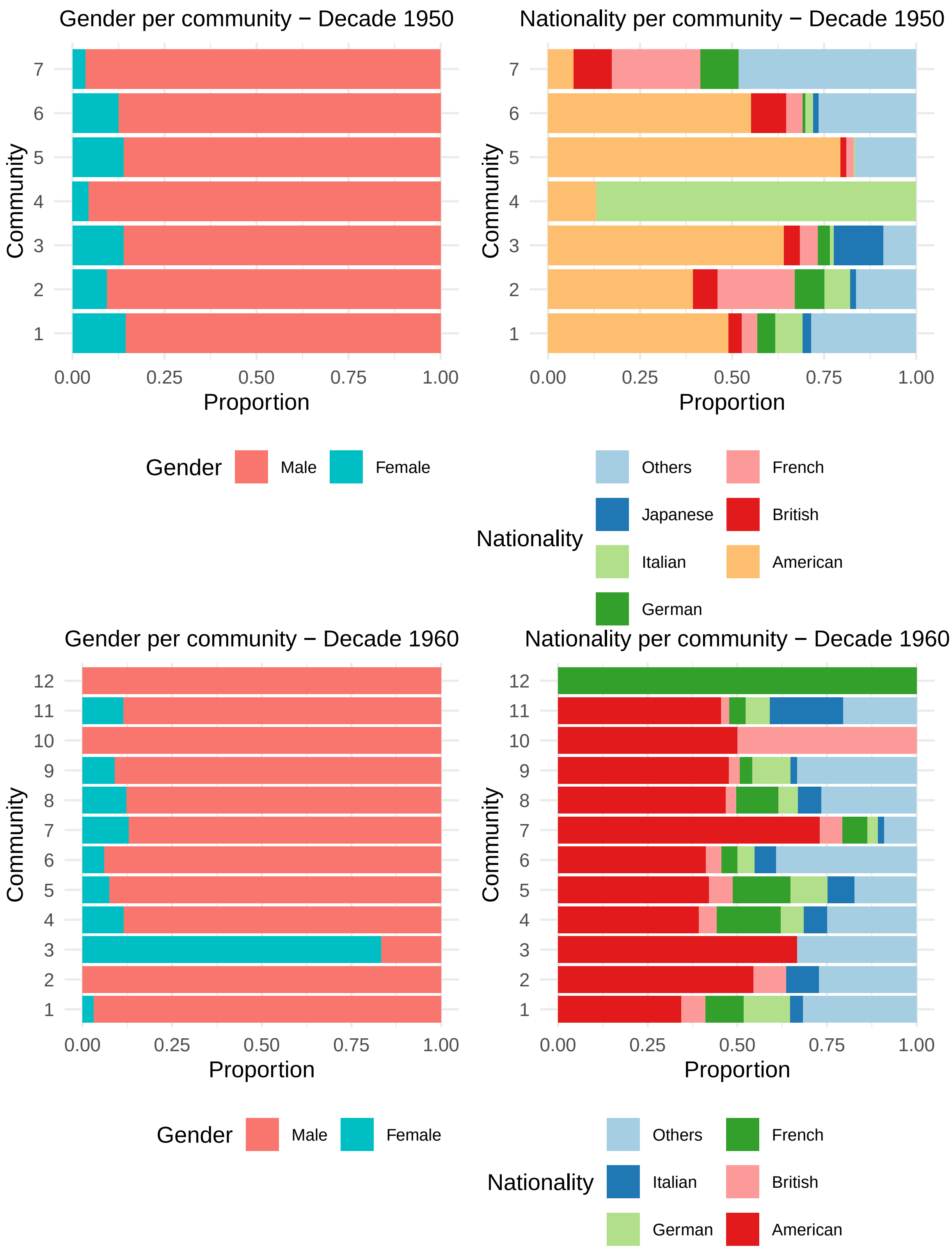


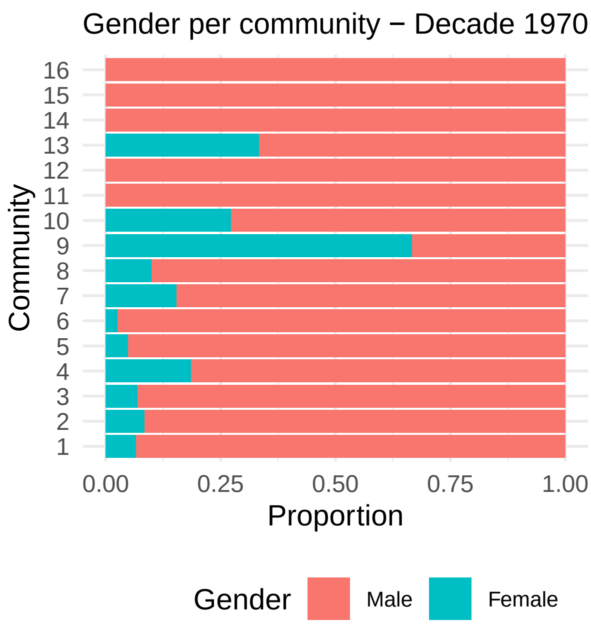

Gender per community − Decade 1930
Community
Proportion
0.00 0.25 0.50 0.75 1.00
1 2 3 4 5 6 7 8 9 10
Gender
Male
Female

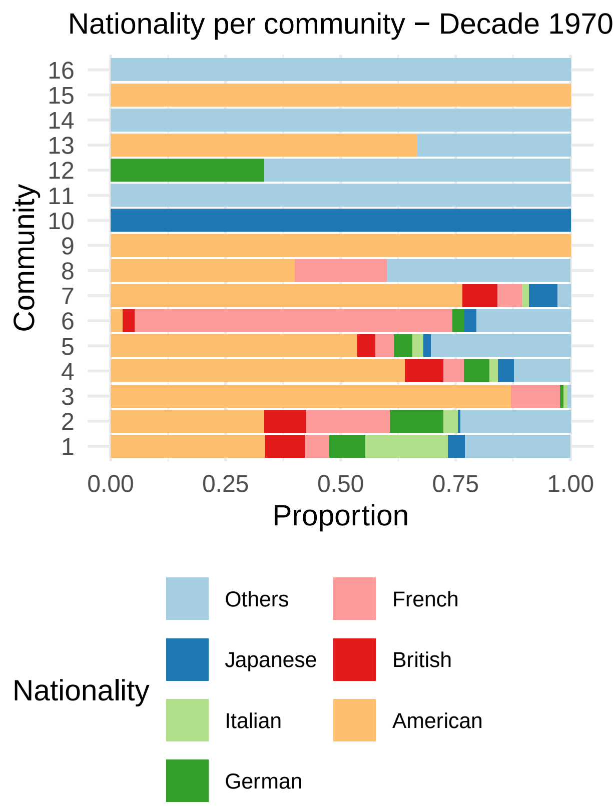

Nationality per community − Decade 1930
Community
Proportion
0.00 0.25 0.50 0.75 1.00
1 2 3 4 5 6 7 8 9 10
Nationality
Others
British
German
American
French

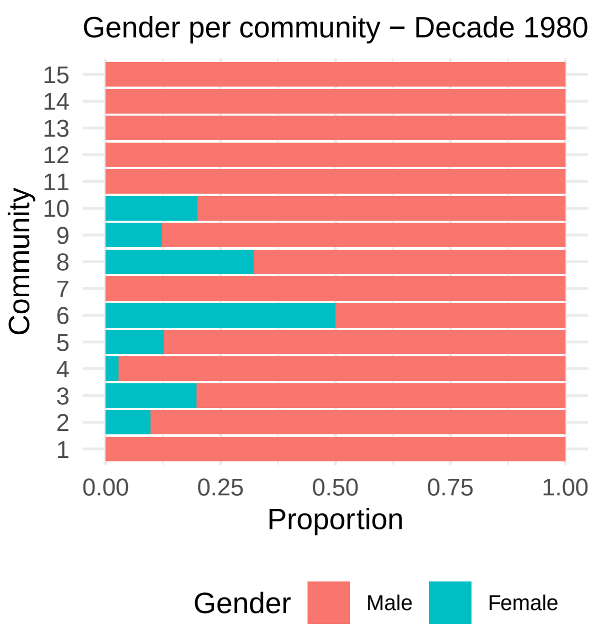

Gender per community − Decade 1940
Community
Proportion
0.00 0.25 0.50 0.75 1.00
1 2 3 4 5 6 7 8 9 10 11 12
Gender
Male
Female

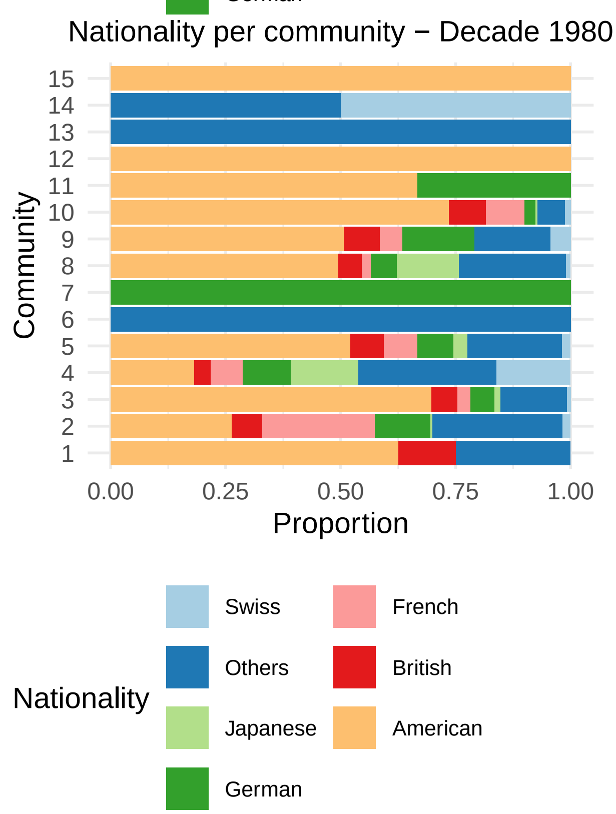

Nationality per community − Decade 1940
Community
Proportion
0.00 0.25 0.50 0.75 1.00
1 2 3 4 5 6 7 8 9 10 11 12
Nationality
Others
French
Mexican
British
Italian
American
German


Gender per community − Decade 1950
Nationality per community − Decade 1950
Community
Proportion
0.00 0.25 0.50 0.75 1.00
Gender Male Female
Nationality Others French Japanese British Italian American German
Gender per community − Decade 1960
Nationality per community − Decade 1960
Community
Proportion
Gender Male Female
Nationality Others French Italian British German American


Gender per community − Decade 1970
Community
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0.00
0.25
0.50
0.75
1.00
Proportion
Gender
Male
Female


Nationality per community − Decade 1970
Community
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0.00
0.25
0.50
0.75
1.00
Proportion
Nationality
Others
French
Japanese
British
Italian
American
German


Gender per community − Decade 1980
Community
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0.00
0.25
0.50
0.75
1.00
Proportion
Gender
Male
Female


Nationality per community − Decade 1980
Community
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0.00
0.25
0.50
0.75
1.00
Proportion
Nationality
Swiss
French
Others
British
Japanese
American
German

Cluster Visualization by Decade

Running the visualization with R was very expensive, and the visualizations were not very clear. I decided to do it with Gephi, and the results are in the GitHub repository I've included in the references.

Plot Analysis

1920s

Very few communities are observed, composed almost entirely of male artists. The predominant nationality is American, with a minority presence of others such as French and Dutch. This suggests a smaller beginning of exhibitions, concentrated on a limited group of artists.

1930s

The number of communities increases notably, although male predominance remains evident. The greater diversity of nationalities (American, British, German, etc.) is notable, indicating a slight shift toward more plural exhibitions, although still with a high percentage of men.

1940s

A wider range of nationalities is evident (for example, Mexican and Italian), while the communities continue to fragment. The majority of artists remains male, but some female contributions begin to be seen. This period reflects a more evident internationalization in the selection of artists.

1950s

The number of communities is not as high as in the previous decade, but the proportion of female artists increases slightly. The emergence of nationalities such as Japanese and a broader "Others" category are notable, signaling a growing interest in artists from more diverse backgrounds.

1960s

The network features numerous communities, with a marked male predominance. However, the nationalities become even more diverse, particularly the presence of French, German, and Italian artists. This suggests that MoMA continued to expand its reach and exhibition programming.

1970s

The number of communities increases, maintaining a male majority, although more women are observed in certain groups. The variety of nationalities increases with the inclusion of Japanese and Italian artists, reflecting an increasingly international trend in exhibitions.

1980s

This period sees the largest number of communities, still predominantly male, but the distribution of nationalities is more heterogeneous (e.g., Swiss, Japanese, British), which points to a much more global curatorial strategy and greater openness to diverse backgrounds.

General Conclusions

In this study, the research questions were addressed by constructing and analyzing a bipartite network connecting artists and exhibitions at MoMA. The results show that, in the early decades, the network exhibited low connectivity, with few co-occurrences, indicating that exhibitions were limited and concentrated in a small group of artists, mostly male, with a strong representation of American artists.

Beginning in the 1930s, a notable increase in the number of artists (nodes) and connections (edges) was observed, evidencing an expansion in exhibition activity and greater diversity in the representation of nationalities. Temporal analysis revealed that the network fragmented into distinct periods, reflected in an increase in components and a larger diameter in some decades, suggesting variations in the overall cohesion of the network.

The approach of segmenting the network by decade before applying clustering techniques allowed for a more precise identification of communities, avoiding the mixing of artists from different eras and highlighting key nodes that act as bridges between groups. Centrality metrics (degree and betweenness) confirmed the presence of influential artists who maintain cohesion in the network, despite fragmentation in other areas.

In conclusion, the objectives set by constructing and analyzing the network, identifying influential artists, and assessing the structural evolution of the network over time have been met. These findings provide a detailed view of the evolution of curatorial practices at MoMA, highlighting the process of internationalization and changes in artists' connectivity over the decades.

References

- Newman, M. E. J. (2018). *Networks*. Oxford University Press.
- Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing Social Networks*. Sage Publications.
- Scott, J. (2017). *Social Network Analysis*. Sage.
- De Nooy, W., Mrvar, A., & Batagelj, V. (2018). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.

Fuente de los datos:
Museum of Modern Art Exhibitions Dataset, disponible en https://github.com/MuseumofModernArt/exhibitions.

GitHub Repository: https://github.com/1637521albert/MoMA_analysis.git

Declaration: Use of AI-Tools I acknowledge the use of generative AI tools in the development of this thesis, specifically [AI tools with links]. Their use was strictly limited to spelling and grammar corrections, and for suggestions on phrasing and word choice. At no point were these tools utilized for generating original content. After utilizing the tool, I thoroughly reviewed and edited the changes and assume full responsibility to the thesis' content. I entered the following prompts, which were slightly varied as needed:

- Tradúceme el texto al inglés.