Danmarks
Tekniske
Universitet

DTU

Social Graphs and Interactions
Final Project

AUTHORS

Davide Venuto - s220331
Pere Ginebra - s223407
Sergi Doce - s211978

Work distribution:
(everyone contributed to all sections to an extent, but X's mark where one has worked more extensively)

Name	Scraping	Inbreeding	Drawing	Network analysis	nlp	Report
Davide Venuto		X		X	X	X
Pere Ginebra	X	X		X		X
Sergi Doce		X	X	X		X

10/12/24

Abstract

For centuries marriages between members of different royal families have been used to secure power, titles, and thrones. This practice often meant marriages between close relatives, which is something that can lead to birth defects and harmful mutations. Even though some studies have been put in place regarding specific cases, there still lacks investigation in a broader context about inbreeding and how it affects rulers and if it has been transmitted until our days. In this project we model the main european royal families into a network graph and perform an analysis of it using techniques such as inbreeding coefficient computation and text analysis of publicly available information to see if we can draw a correlation between royal inbreeding and health issues and the capacity of ruling. We expect our results to give more insight into this topic and fuel further analysis using modern techniques and computation.

Significance Statement

We study the social network composed of royal families within Europe making use of network analysis techniques and computation to investigate the relationship between family inbreeding and other characteristics in monarchs throughout history. There are not a lot of studies that use these techniques to make insightful investigations within this context so we hope that this work will fuel more work involving network analysis and monarchy study.

1 Introduction

Inbreeding within royal families has long been a controversial and intricate facet of their history, entwined with the complexities of lineage, power, and tradition. This practice, often employed to consolidate wealth and maintain a pure dynastic line, has yielded both political advantages and grave genetic consequences. From ancient civilizations to modern monarchies, the repercussions of intermarriage among closely related individuals within royal bloodlines have sparked debates surrounding genetic disorders, the stability of power, and the intricate balance between tradition and progress within these esteemed lineages.

Social network analysis is the process of investigating social structures using networks and graph theory [1]. They are very useful for processing large amounts of relational data and describe the overall relational network structure.

In this work, our objective is to model the main European families into social networks and perform an analysis using network analysis techniques and natural language processing to draw a correlation between royal family inbreeding and genetic disorders or power stability. Our work is based on the data provided by the Wikipedia API [2]

During the course of the project, we perform a generic analysis on the network to investigate about its structure, we compute context specific metrics like the inbreeding coefficient based on the provided data, and we perform NLP (Natural Language Processing) text and sentiment analysis to draw conclusions from publicly available biographies.

2 Results

2.1 Network Creation

For each royal, our data source provided us with the following data: predecessor, parents, sons, daughters, spouses, birth and death dates, start and end of the reign dates, and royal house. Thanks to all of this data, we are able to construct a graph modeling not only the family relationships between royals, but also the power relationships. We also introduce a cutoff year to avoid fetching data endlessly. The cutoff year we have chosen is the year 1450.

The result of this network construction is a network having royal family members as nodes and the relationships between them as edges. As attributes, the nodes have all the data mentioned previously and also some computed attributes such as the inbreeding score, which we explain later in this report. The edges have as attributes the type of relationship that they represent, which can be blood, marriage, or throne (succession)

2.2 Network Analysis and Structure

In this subsection we provide a general analysis of the network to have an understanding of its structure. The results of this analysis are that we have a network with 7365 nodes, 13259 edges and an average degree of 3.6. This average degree makes a lot of sense, as the degree is conditioned by the amount of close relatives one has, which in our case are both parents and the issue one might have.

We have also performed a connectivity analysis, and we have found that the network has one connected component. This means that we can draw a path between any given pair of nodes. In Figure 1 we can see a partial visualization of the network. We can't plot the whole network as it is too hard to understand due to its complexity. We filter the network by choosing which royal houses to plot.

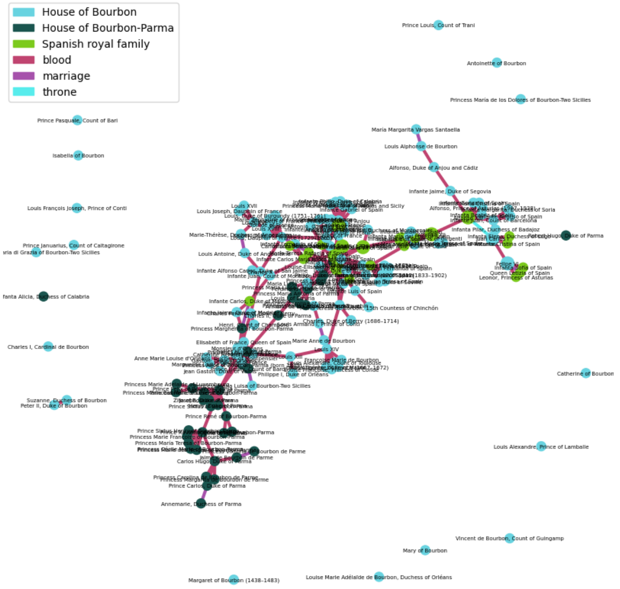
House of Bourbon
House of Bourbon-Parma
Spanish royal family
blood
marriage
throne
Prince Louis, Count of Trani
Antoinette of Bourbon
Princess María de los Dolores of Bourbon-Two Sicilies
Prince Pasquale, Count of Bari
Isabella of Bourbon
María Margarita Vargas Santaella
Louis Alphonse de Bourbon
Alfonso, Duke of Anjou and Cádiz
Infante Jaime, Duke of Segovia
Louis François Joseph, Prince of Conti
Louis XVII
Louis Joseph, Dauphin of France
Infante Gabriel of Spain
Alfonso, Prince of Asturias (1907–1938)
Infante Margarita, Duchess of Soria
Infante Juan, Count of Barcelona
Infanta Pilar, Duchess of Badajoz
Marie-Thérèse, Duchess of Angoulême
Prince Januarius, Count of Caltagirone
Louis Antoine, Duke of Angoulême
Juan Carlos I
Infanta Cristina of Spain
Infante Alfonso Carlos, Duke of San Jaime
Infante Juan, Count of Montizón
Felipe VI
Infanta Sofía of Spain
Queen Letizia of Spain
Leonor, Princess of Asturias
fanta Alicia, Duchess of Calabria
Infante Carlos, Duke of Madrid
Princess Maria Antonia of Parma
Infante Luis of Spain
15th Countess of Chinchón
Charles II, Duke of Parma
Charles, Duke of Berry (1686–1714)
Louis Armand I, Prince of Conti
Henri, Count of Chambord
Princess Margherita of Bourbon-Parma
Charles I, Cardinal de Bourbon
Elisabeth of France, Queen of Spain
Marie Anne de Bourbon
Louis XIV
Monsieur d'Orléans
Louis XIII
Françoise Marie de Bourbon
Louis Alexandre, Count of Toulouse
Anne Marie Louise d'Orléans, Duchess of Montpensier
Jean Gaston, Duke of Valois
Philippe I, Duke of Orléans
Catherine of Bourbon
Princess Marie Adélaïde of Luxembourg
Luisa of Bourbon-Two Sicilies
Suzanne, Duchess of Bourbon
Peter II, Duke of Bourbon
Prince René of Bourbon-Parma
Princess Marie Françoise of Bourbon-Parma
Princess Maria Teresa of Bourbon-Parma
Jaime de Bourbon de Parme
Carlos Hugo, Duke of Parma
Louis Alexandre, Prince of Lamballe
Princess Margarita de Bourbon de Parme
Prince Carlos, Duke of Parma
Annemarie, Duchess of Parma
Vincent de Bourbon, Count of Guingamp
Mary of Bourbon
Margaret of Bourbon (1438–1483)
Louise Marie Adélaïde de Bourbon, Duchess of Orléans

Figure 1: Network Visualization of the Bourbon Royal Houses

2.3 Inbreeding Coefficient

The central piece of our work is the inbreeding coefficient. The coefficient of inbreeding of an individual is the probability that two alleles at any locus in an individual are identical by descent from the common ancestor(s) of the two parents [3]. For each node in the graph, we calculate it using the following formula:

$$fX = \sum 0.5^{n-1} * (1 + f_A)$$

where n is the number of individuals in the loop created by both parents and the common ancestor, and f_A is the coefficient of inbreeding of the common ancestor of X's parents. In Figure 2 we show the distribution we get through the network for the inbreeding coefficient.

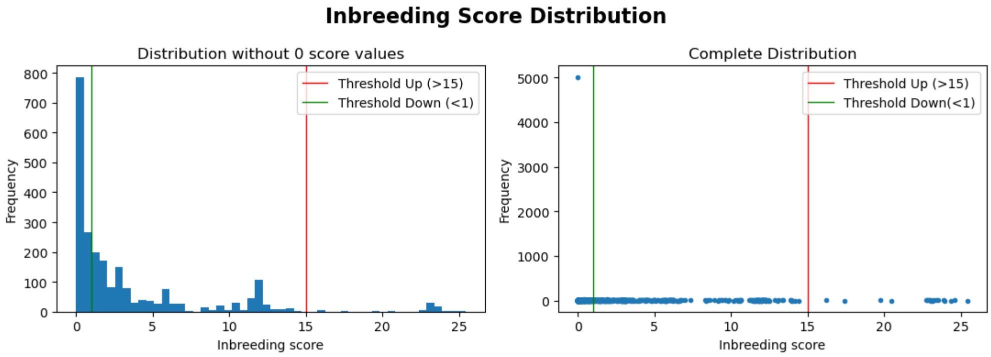

Figure 2: Inbreeding Coefficient Distribution Through the Network

2.4 Royal houses and Inbreeding

As discussed earlier, inbreeding was used to conserve power within the family, it is even said that some houses used it to keep "blood purity". Like is the case in Spanish and Portuguese societies with the coined term "Limpieza de Sangre" [4] (blood cleansing in Spanish). This is the reason we analyzed if there are royal families that were more prone to inbreeding than others, finding very interesting results as seen in the table below.

Position	House	Size	Inbreeding Score	Community Modularity
1	House of Bourbon-Braganza	2	0.103726	0.00008
2	House of Habsburg	113	0.082765	0.00921
3	Spanish royal family	22	0.070572	0.00155
4	Saxe-Gotha-Altenburg	14	0.067366	0.00105
5	House of Bourbon-Two Sicilies	56	0.062819	0.00456
6	OrlĂŠans	2	0.060059	-0.00000
7	House of Habsburg-Lorraine	108	0.048109	0.00622
...	...	...	...	...

To put these values into perspective, the average inbreeding score of the graph is around 0.01, which is close to that of the average score in society according to a study [5]. So there are at least 6 royal families with more than 5 times the expected score, of which 4 have more than 10 individuals. We can see from this that there definitely were families with a bigger inbreeding culture than others, two clear examples are the Bourbon and Habsburg families along with their branches.

2.5 How royal inbreeding has changed over time

In this subsection, it going to be studied the change of the inbreeding habit among royals. This was performed through the analysis of the Inbreeding Coefficient of the people who lived after 1920 compared to the Coefficient of people who lived from 1450 (date of Network start) to 1701.

These two thresholds were chosen to extract two sub-groups of similar size (after removing 0 inbreeding score nodes) to create a visually comparable distribution. In addition, the two dates were chosen as being two historical thresholds, well known for being breaking points of scientific and social changes (the start of the $18th$ century, the ”$RoaringTwenties$”).

Analyzing the figure 3 it could be noticed that differently as one could falsely expect, as today inbreeding is very uncommon even illegal in certain cases, the two distributions would show completely different behavior. They show a similar shape and there is even a slight rise of the mean and median inbreeding scores for the "After 1920" group.

If this may appear wondering at first glance, however, it is well explained inside the inbreeding score calculation 2.3 itself, which, for each new inbreeding case (even among far relatives), takes into consideration not only the new inbreeding occurrence but also how much was inbred the common ancestor itself (f_A). The inbreeding score for a population is a factor that grows in time if inbreeding behavior continues, which well explain the general slight shift of the lower end of the "after 1920" distribution compared to the "Before" group. At the same time, due to social changes and scientific discoveries about the genetic risk of inbreeding, the extremely high values (>15) have significantly dropped. It has also to be stated that the only two outlier values with inbreeding score (>15) in the "After 1920" distribution belong to Princess Maria Antonietta of Bourbon-Two Sicilies and Princess Maria Carolina of Bourbon-Two Sicilies which are both parts of the same House of Bourbon-Two Sicilies; well known for their high inbreeding history and discovered to be one of the most inbred families.

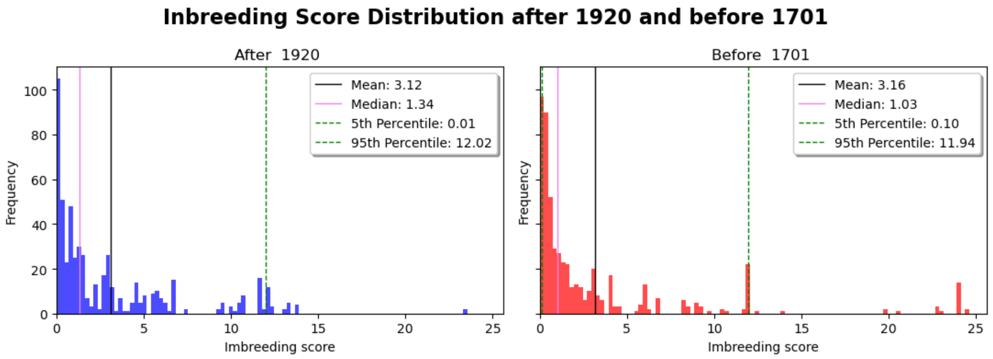

Figure 3: Comparison of the Inbreeding Coefficient Distribution of people in the Network lived before 1701 and after 1920.

2.6 Sentiment Analysis

This section is going to analyze the result of a sentiment score attribute calculated for each node Wiki text and compare the score distribution between an extremely high health risk inbreeding score group (inbreeding score >15) to a low-risk group (inbreeding score <1). As it could be seen in 4 the used Nltk SentimentIntensityAnalyzer shows a similarly shaped bimodal distribution. However, it could be noticed that in the high inbreeding score group one of the two modes is predominant. This causes the shift of the mean and the median value, compared to the distribution for low inbreeding risk, as a result, the extracted sentiment score has some characteristics in common with a (not very precise)classifier.

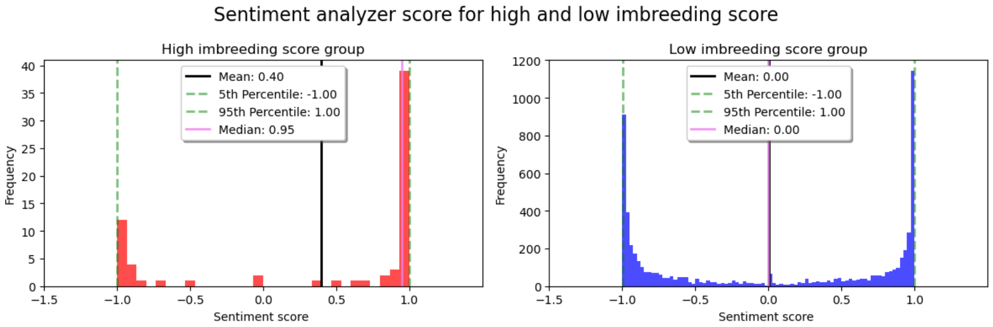

Figure 4: Sentiment score distribution extracted using nltk SentimentIntensityAnalyzer

3 Discussion

Our objective regarding the inbreeding coefficient was to analyse if we could find any relation between high levels of inbreeding and problems stated in the public biographies of the monarchs. We haven't found enough evidence to state that levels of inbreeding correlate with registered issues in public biographies. This could be due to a series of reasons.

One of the reasons is that even though there is inbreeding, the amount of it is not enough to actually cause any relevant issue with the person. Another reason could be that throughout history maybe there have been issues, but they haven't been properly registered due to lack of scientific knowledge or to preserve the image of monarchs.

This project has many limitations, from the data sources, lack of background in certain fields like history or genetics, but also from faults in our own work.

When it comes to the data source, Wikipedia does try to follow some standards in their articles, but these are not always kept. This means that data extraction is not as straightforward as it seems. The more connections you explore within articles, the more irregularities you will find. There is no simple way of checking if an article is a person, and many times, incorrect or unexpected articles are linked in certain fields. But wikipedia isn't the only part of the dataset that is hard to work with. Geopolitics are a complex topic, and there have been many exceptions and irregularities in lines of succession and government types throughout history, as well as many different titles with meanings behind them which makes it understandable for Wikipedia's confusion on what to put in what field. This makes it hard to reach conclusions related to many interesting aspects tied to monarchies.

In conclusion, although we've found interesting results and have learned a lot from working on this project (about real-world data behaviour, networks, text analysis, history...), we feel like there could still be more to be found within the data. With a bigger and more curated dataset, which would require more research and probably collaboration with experts, together with more advanced techniques and tests, some new findings could be obtained.

4 Methods

The Dataset: For the construction of the network, we used the Wikipedia API [2] which provides all the information contained in a monarchs infobox. The infobox of a Wikipedia page is the small box that appears in the top right of the page and provides the main information of the person.

References

[1] "Social Network Analysis Wikipedia." `https://en.wikipedia.org/wiki/Social_network_analysis`. Accessed: 2023-12-05.

[2] "Wikipedia API Documentation." `https://www.mediawiki.org/wiki/API:Main_page/en`. Accessed: 2023-12-05.

[3] "Coefficient of Inbreeding Wikipedia." `https://en.wikipedia.org/wiki/Coefficient_of_inbreeding`. Accessed: 2023-12-06.

[4] "Estatutos de Limpieza de Sangre [Spanish]." `https://pachami.com/Inquisicion/LimpiezaSangre.html`. Accessed: 2023-12-06.

[5] "Inbreeding estimates in human populations." `journals.plos.org/plosone/article?id=10.1371/journal.pone.0196360`. Accessed: 2023-12-06.