# Women in Early Modern Handwritten News: Random Walks and Semantic Wanderings in the Medici Archive

### Gabor Mihaly Toth [orcid](https://orcid.org/0000-0002-4301-1581)

University of Luxembourg

from IPython.display import Image  
display(Image("cover.png"))



[cc-by-nc-nd](https://creativecommons.org/licenses/by-nc-nd/4.0/) © Gabor Mihaly Toth. Published by De Gruyter in cooperation with the University of Luxembourg Centre for Contemporary and Digital History. This is an Open Access article distributed under the terms of the [Creative Commons Attribution License CC-BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Abbreviations:

* ASF MdP: Archivio di Stato di Firenze, Fondo Mediceo del Principato

early modern history, media history, simulation, gender, manuscript news, semantic modelling, Bayesian inference

Early modern Europe witnessed an unprecedented revolution in information culture. This gave rise to new means of communication that included handwritten news sheets. Throughout the 16th and 17th centuries, information circulated in the form of regular manuscript newsletters. Even though these newsletters had a significant impact on European history, their study has remained challenging. Most of the previously existing news sheets did not survive. Those that survived are dispersed in different collections around the world. These collections are somewhat random and incomplete samples. How can we unlock and study the lost early modern corpus of handwritten news? To tackle this problem, this study draws on various methods applied in dynamic system modeling and AI (Bootstrapping, Markovian Simulation, and Random Walks), as well as on Bayesian statistics and semantic modeling. Precisely, the paper investigates a collection of one thousand Italian manuscript newsletters from the Medici Archive (today part of the Florentine State Archive) with a focus on the representation of women in the news. First, it discusses how loss and incompleteness contribute to the randomness of archival collections. Second, it offers an accessible introduction to an array of methods that natural sciences and AI apply to tackle randomness and incompleteness. Third, by analyzing the representation and presence of women in early modern news, the paper presents these methods in action.

## Introduction

### Background

In February 1600, seemingly peculiar news reached the Medici court in Florence: a woman ambassador from Persia arrived at the Ottoman court in Constantinople. The short report about her arrival came as part of a handwritten news sheet compiled in Vienna (ASF MdP, 3087). Were contemporary news consumers surprised reading this news? Or was it common for early modern news consumers to read about women who acted decisively and boldly?

Answering these questions is not easy. Early modern Europe witnessed an information revolution, characterized in part by the continual circulation of handwritten news sheets, known as avvisi in Italian (, , ). Later, from the early 1600s, handwritten news sheets were complemented with printed news, though handwritten news continued to play an important role until the mid-18th century (). News was collected in large urban centers (henceforth, news hubs) of the continent by semi-professional newsagents; they assembled single news items into handwritten news sheets (). Therefore, each news sheet is comprised of a consecutive sequence of news items, organized into separate paragraphs. News agents, who could be postmasters, merchants, or even diplomats and ambassadors residing abroad, sold their news sheets to a diverse range of customers. Subscribers to news sheets could include ducal courts and merchant houses such as for instance the grand ducal court of the Medici in Florence and the Fuggers in Augsburg (,). These news sheets were usually compiled in the native language of the customers. Hence, the Medici received most of the news sheets in Italian and the Fuggers mainly read news in German. The types of information that handwritten news sheets disseminated about women have not been studied, even though this is a relevant question for historians. The broader relevance is highlighted by scholars' efforts to discover and present women's representation in historical sources ().

However, scholarship on early modern women's history cannot answer the question, to which extent the story of the woman ambassador was a genuine novelty for contemporary readers. Recent scholarship on early modern women's history has studied the representation of women in various sources, including letters, literary works, and court records. For instance, various topics, social activities, and roles associated with early modern women have been studied (; ). Recent scholarship in women's history has also challenged the traditional perception of early modern women as passive actors in history (). Through the systematic study of historical records, historians have uncovered various areas of women's agency and pointed out that it was not altogether exceptional. Women's agency manifested in specific domains of knowledge and skills, such as medicine, household management, and information gathering (; ; ; ). Women did engage with political and material culture, as well as with knowledge production, and they did shape the course of history. Throughout the early modern period there were a number of powerful women, such as for instance the English queen Elisabeth I (1558 - 1603), who ruled entire countries. However, scholars of women’s history have not surveyed handwritten news; unlike news pamphlets, avvisi have not been studied from the perspective of women's history (). Therefore, we do not know what representations of womanhood were conveyed in this important group of historical sources. This exploratory study aims to contribute to women's history by addressing this gap.

The lack of studies on women’s representation in early modern handwritten news is due to the difficulty of researching these sources until recently. Most of the surviving news sheets remained unpublished and untranscribed in various archives and libraries worldwide. Recently, thanks to the Euronews Project (<https://euronewsproject.org>) funded by the Irish Research Council and the Medici Archive Project (<https://www.medici.org/>) funded by the Mellon Foundation, a large selection of transcribed Italian-language handwritten news sheets from the Florentine State Archive became publicly available on the MIA platform and database (<https://mia.medici.org/>). This collection includes approximately 1,250 news sheets and 10,800 individual news items. In this essay, I will investigate this selection and I will refer to it as the MIA-Euronews Corpus. Specifically, my main focus will be on women and their representation in early modern handwritten news sheets. This focus on representation aligns with recent developments in media history that explore women's representation in various modern media formats ().

However, this paper also has another key focus: methodology and historiography of early modern handwritten news sheets. The MIA-Euronews Corpus is simply an immense and not representative sample of handwritten news sheets that once circulated in early modern Europe. Given that no explicit principles were followed by the team that aggregated it, the MIA-Euronews Corpus is a random and somewhat ad hoc snapshot of the vivid early modern news culture. In this sense, the MIA-Euronews Corpus represents very well the heart of the methodological and historical problem that the study of the early modern handwritten news sheet market involves: the totality of early modern handwritten news sheets perished; today there is a lost whole of these documents; as a result historians can only work with incomplete snapshots. Since most of the time factors determining which news sheets survived and which ones perished are unknown, the existing news sheet collections seem to be random samples from the lost whole. This leads to an important methodological and historiographical question I will address as the secondary focus of my paper. How can we explore a random sample of historical documents and say something about the original and lost document collection from which the random sample comes from? As part of the second focus of this paper, I will show that we can still discover how the randomness of incomplete samples (from a lost whole) works, and based on this, we can make inferences about the lost whole itself.

The surveying of lost historical documents and collections has recently become an important area of historical research. Historians and digital humanists have adopted methods from ecology and Bayesian statistics to gather information about the size of once existing document collections (; ; ). For instance, Mike Kestemont et al. have applied the so-called unseen species model to study the loss of medieval narratives (). This paper aims to contribute to this novel area of study, though it does not aim to gauge the volume of once circulating early modern handwritten news sheets; instead, it aims to gather insights into the content of these lost documents.

Given the importance of the methodology underlying this paper, a significant amount of space will be devoted to the methodology, specifically to the discussion of randomness and sampling. As part of this discussion, in the first part of the paper, I will outline the methodology (Random Walks, Markovian simulation, Bayesian inference, and semantic modeling) I applied to address randomness; I will keep my explanation accessible to non-specialist audiences and draw on a simplified thought experiment. The goal of this thought experiment is to enable readers with limited or no experience in statistics and computational modeling to form a basic idea of the methodology applied. An in-depth and technical discussion of the methodology will be presented in the section called Implementation, which follows the general conclusions of this paper.

### Research questions

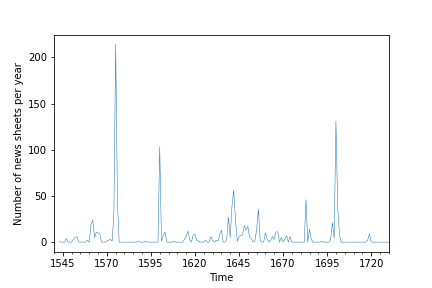
My research questions come from previous research that has studied the representation and agency of early modern women (; ). Previous studies addressed broader questions in the context of agency and representation. Here, I address the same broader questions by focusing on handwritten news. Did handwritten news perpetuate the essentially male order and dominance of early modern times? To what extent did news give space to women and women's actions aiming to challenge the male order? However, these questions need to be translated into more specific queries that a data-driven exploration can answer. In other words, the main question of how women were represented in early modern handwritten news is to be operationalized through a number of more descriptive questions.

First, how often did contemporary readers hear about women? This question leads to further inquiries. Generally, who were the main characters in the news world of the early modern period? What was the social status of those women who appeared in the news? Second, if women appear in the news, in what circumstances are they mentioned? In other words, what are those topics that are associated with women? Before answering this, we need to consider the main themes in the early modern news world. Collectively, addressing these questions will help us understand not only the presence and representation of women in early modern news, but also offer fresh insights into the culture of early modern handwritten news

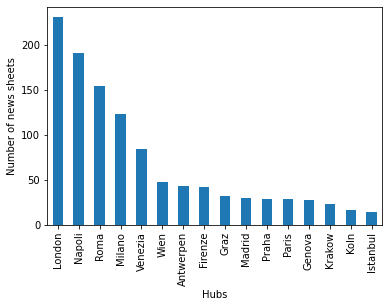
### Limitations of this study

The news sheets I worked with cover a period of 200 years (see Figure 1) and they arrived from a great variety of source locations (see Figure 2). This 200-year period marks a particularly tumultuous time in Europe. The early modern period witnessed not only an information revolution but also other profound social and economic transformations and historical events: the discovery of the new world, the Reformation and subsequent religious divides, witch hunts and long-lasting wars (Thirty Years' War, the war against the Ottoman Empire, etc). Despite the urgent need to study how the presence and the representation of women changed in space and time, the diachronic and synchronic analysis of the MIA-Euronews Corpus will be limited in this study. This is due to the fact that the corpus is not balanced; yet, as Figure 1 and Figure 2 demonstrate, news sheets available in the corpus are quite unevenly distributed in space and time. I will therefore limit the diachronic and synchronic approaches to the study of women’s presence or absence. In particular, I will compare the presence or absence of women in two periods (before and after 1600) and in the news sheets that came from the seven most important news hubs in the MIA-Euronews Corpus: London, Naples, Rome, Milan, Venice, Vienna, Antwerp. The arbitrary temporal boundaries are explained by the unevenness of my dataset. By contrast, when studying the representation of women, I will treat the corpus as one monolithic whole without detailing regional differences; the diachronic and synchronic analysis of the women’s representation will be accomplished once a more balanced data is available.

from IPython.display import Image  
metadata = {  
 "jdh": {  
 "module": "object",  
 "object": {  
 "source": [  
 "The temporal distribution of news sheets in the MIA-Euronews Corpus"  
 ],  
 "type":  
 "image"  
 }  
 }  
}  
display(Image('media/image19.png'), metadata=metadata)



metadata = {  
 "jdh": {  
 "module": "object",  
 "object": {  
 "source": [  
 "The fifteen most frequent news hubs in the transcribed news sheet corpus in the MIA-Euronews Corpus"  
 ],  
 "type":  
 "image"  
 }  
 }  
}  
display(Image('media/image6.png'),metadata=metadata)



## Methodology

### The random walker

The story of the boldly acting woman ambassador was a random finding by my colleague. This story is just one of those millions of stories that the so-called Mediceo del Principato preserves. The Mediceo del Principato is a gigantic archival division in the Florentine State Archive (); it incorporates the Medici family’s major archive from the early modern period. Today, some 6500 archival units, which occupy approximately one and a half kilometres of shelf space, form the core of the Mediceo del Principato. News sheets are dispersed through these 6500 archival units. To understand the problem of randomness and sampling, which are at the heart of this paper, consider the following thought experiment.

There is a person who is literally a random walker; he has two goals. First, he aims to discover the presence or absence of women in the lost corpus of early modern handwritten news sheets. Second, he would like to map those contexts in which women were mentioned and understand the representation of women in early modern handwritten news.

The random walker therefore decides to wander around in the physical spaces where the Mediceo del Principato is today housed. He keeps opening boxes and folders containing handwritten news sheets. He finds that sometimes women are mentioned in the news; sometimes there is silence on them. As the random walker explores the news sheets, he finds that the information about women constantly changes. Sometimes women act boldly; sometimes they are passive agents of history; sometimes the random walker reads about noble women, sometimes he reads about peasant women. The random walker realizes that his effort to develop a firm understanding of women’s presence and representation in handwritten news is hampered by an uncertainty rising from many competing possibilities. With this uncertainty in mind, he concludes that the presence and the absence of women in news sheets, as well as the information about women, feature a high degree of randomness. At the same time, he understands that this randomness is not necessarily true randomness; his experience of randomness can be also due to the fact that he has incomplete information about the history of the Medici Archive, as well as about early modern women and news in general.

The ‘experience’ of randomness and uncertainty that the thought experiment demonstrates is common among historians working with vast amounts of documents in archives and libraries. Archival research often depends on chance, and in the presence of many possibilities, it is difficult to build a firm historical knowledge. Chris Wickham, an eminent historian of early medieval Europe, summarized the randomness of archives and libraries in an interview:

Even in terms of going into a library – you see a whole series of books, shelved in a particular way, and that combination is random. The library as a whole is a random collection. The library, the archive in that respect is a bit like the physical world. (<https://www.ox.ac.uk/research/research-in-conversation/randomness-and-order/chris-wickham>, accessed 1 February, 2024)

It is an open question to what extent the randomness often experienced by historians working in archives is true or simply due to the fact that no one has an overview of the complexity underlying millions of loosely organized documents. This paper views randomness as perceived, as a phenomenon that features unpredictability and uncertainty from the perspective of the posterity. How the true randomness of archives could be tested and distinguished from perceived randomness - i.e. from posterity’s inability to understand the manifold causes and patterns underlying millions of documents - is the subject of further research.

We will continue the thought experiment and see how our random walker copes with uncertainties related to women’s presence and representation in early modern news. As the heart of the matter is the presence of many competing possibilities unearthed throughout the archival research, we will deal with computational methods capable of producing firm and tangible information in the face of many possibilities. Specifically, we will focus on three areas: presence and absence, narratives, and semantics, and see how the behavior of randomness underlying many competing possibilities can be investigated and harnessed to build a stable knowledge.

### Presence and absence of women

Next, the random worker focuses on women’s presence and absence in handwritten news. As a first step, he decides to discover the entire archive and continues to walk around until he has seen all the news sheets. He carefully notes which ones mention women and which ones are silent about them. Although he now has a lot of information about women in the news, he realizes the problem I have already discussed: the collection of news sheets in the former Medici Archive is just an incomplete sample from a lost corpus. More importantly, this sample is not-representative; it is merely one possible snapshot of the original lost corpus, and in this sense it is biased.

As a second step, the random walker modifies his strategy to discover the presence of women. He goes to the archive and randomly selects a sample of 100 news sheets at a time. He notes the number of news sheets that mention at least one woman; then he continues to select further random samples of 100 news sheets. After each round he notes the percentage of news sheets mentioning women in the sub-sample and calculates the average percentage in all previously seen sub-samples. (Notice that the percentage actually represents the probability of finding women in a hundred randomly selected news sheets). The random walker continues to wander around and repeat this task until he has seen all possible combinations of the news sheets in the former Medici Archive. Since the number of possible combinations is extraordinarily large, his task will take an extraordinarily long time.

However, by the end, he has gained important insights into the presence of women in the news sheets preserved in the Medici Archive. First, by determining the probability of finding women in a given random selection of news sheets, the random walker has an idea of ‘what would have happened’ if only this combination had survived; in other words, he knows the probability of women’s presence if only a given combination of news sheets was available today. Second, once his quest for all possible combinations is exhausted, he knows about all possibilities: for instance, women mentioned in 24 out of 100 randomly selected news sheets, women mentioned in 50 out of 100 randomly selected news sheets, and so on. He will also know about the relative likelihood of each possibility. This is an understanding of how the randomness of news sheets - preserved in the Medici Archive - in regard to women’s presence and absence works. In practice, by randomly selecting sub-samples of the news sheet collection in the Medici Archive, the random walker discovers the collection from all possible angles and systematically maps all possibilities. By synthetically creating thousands of small - possible - subsamples, he goes beyond the original dataset, which is just one possible snapshot of the lost corpus of early-modern handwritten news, thus addressing the problem of bias.

The random walker is only a thought experiment; wandering around in the physical spaces where the former Medici Archive is today preserved for thousands of hours and randomly opening boxes and folders are, of course, not viable. Therefore, I implemented the thought experiment algorithmically by using the MIA-Euronews Corpus to represent the Medici Archive. My algorithmic implementation is based on the statistical procedure of bootstrapping, also known as sampling with replacement (). The algorithm randomly fetched a set of 100 news sheets and a set of 100 individual news items from the corpus. First, it reported the number of news sheets in which women are present (henceforth, I will refer to this as random sampling of news sheets). Second, it reported the number of news items (again, a news sheet is a consecutive sequence of news items) that mention women (henceforth, I will refer to this as random sampling of news items). Third, it turned its findings into probabilities, i.e., the probability of finding women in a 100 randomly selected news sheets and news items. The algorithm repeated these tasks until it had seen enough combinations to reach convergence; since the algorithm started to converge relatively quickly, it was not necessary to see all possible combinations as no new information would be gained beyond this point. As a result, the algorithm produced an overview of all possibilities (e.g., women mentioned in 50 out of 100 randomly selected news sheets, women mentioned in 24 out of 100 randomly selected news sheets, and so on) and their respective probabilities. In mathematical terms, the algorithm gave a probability distribution that rendered how the randomness of women’s presence and absence behaved in the MIA-Euronews Corpus.

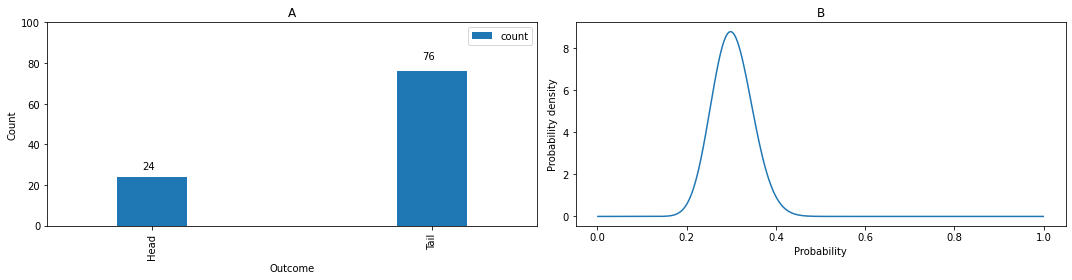
Additionally, I used this algorithmic procedure to explore the presence of certain themes related to women in the news (such as marriage, for instance), as well as to compare women’s presence in news coming from different locations and periods (before and after 1600). Hence, the algorithmic procedure helped me investigate how external factors or variables (such as place and date of compilation) might have influenced the presence or absence of women (see below in the next Results section). Overall, the algorithmic implementation of the random walker thought experiment addressed the randomness of women’s presence and the biased nature of the MIA-Euronews Corpus. But how can we explore how the randomness of women’s presence might have worked in the lost corpus of early modern news sheets?

To answer this question, I worked with two core assumptions deriving from Bayesian statistics (). First, each random sample of 100 news sheets is related to the once existing but now lost whole. Second, the ‘features’ of the lost whole determine the features of each random sample. Of course, here we do not work with exact features but with a range of possible outcomes and their respective probabilities using probability distributions. The range of possible outcomes that we discover by randomly sampling sets of 100 news sheets in the MIA-Euronews Corpus are conditioned by women’s presence in the lost whole. More technically speaking, women’s presence in the once existing totality of circulating news sheets is a prior; this prior conditions the evidence we find when randomly sampling sets of 100 news sheets in the MIA-Euronews Corpus. Bayesian statistics offers a framework for making inferences about the prior based on the evidence we have. The resulting knowledge about the prior is not exact; it is instead an estimate or an informed guess. The following illustrative example makes this line of reasoning more intuitive.

We have an unfair coin; the probability of heads is 0.3 and the probability of tails is 0.7. We are not aware of how biased the coin is; this prior probability setting is hidden. We toss the coin 100 times. The evidence we get by tossing the coin is determined by the prior probability setting. This means that the probability of getting heads will be around 0.3 and cannot be, for instance, 0.8. Figure 3A shows the evidence we get throughout the tossing; Figure 3B shows our respective guess about the prior probability setting. The guess about the prior probability setting (Figure 3B) was calculated from the evidence (Figure 3A) with the help of Bayesian inference (). As Figure 3B shows, the information we have about the prior probability setting is not a fixed value; it is an estimate represented by a range of possible outcomes (or a probability distribution).

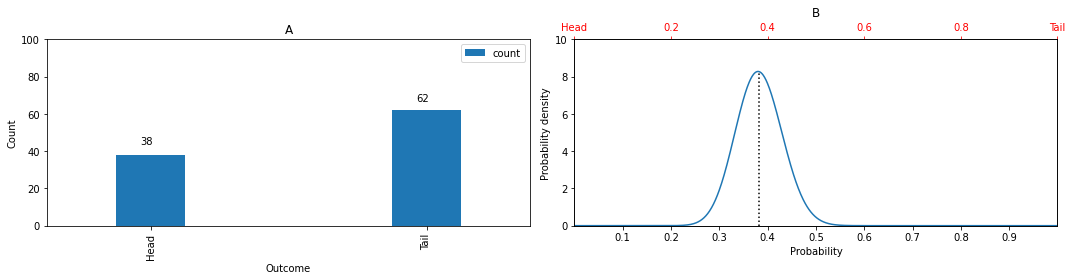
Similarly, we can apply Bayesian inference to uncover the prior probability setting (in the lost corpus) that must have determined women’s presence in thousands of possible sub-samples created from the MIA-Euronews Corpus. Figure 7 in the following section illustrates this prior probability setting. This represents an educated guess or an estimate of women’s presence in the previously existing total population of news sheets (further technical details and the proof are discussed in the Implementation section). This educated guess is not a firm value; rather, it is a range of possible outcomes (or a probability distribution).

display(Image("media/image28.png"))



from scipy.stats import beta  
import numpy as np  
from matplotlib import pyplot as plt  
import os  
from scipy.stats import bernoulli, binom  
import pandas as pd  
figure, axis = plt.subplots(1,2,figsize=(15, 4))  
X = bernoulli(p=0.3)  
X\_samples = X.rvs(100)  
entries = []  
entry = {'outcome': 'Head', 'count':np.where(X\_samples==1)[0].shape[0]}  
entries.append(entry)  
entry = {'outcome': 'Tail', 'count':np.where(X\_samples==0)[0].shape[0]}  
entries.append(entry)  
dfTmp = pd.DataFrame(entries)  
dfTmp.plot.bar(x='outcome',y='count',ax=axis[0],width=0.25)  
axis[0].set\_ylim(ymin=0,ymax = 100)  
axis[0].set\_xlabel('Outcome')  
axis[0].set\_ylabel('Count')  
axis[0].set(title='A')  
for f,p in enumerate(axis[0].patches):  
 if f == 0:  
 axis[0].annotate(str(p.get\_height()), (p.get\_x() \* 0.3, p.get\_height() \* 1.15))  
 else:  
 axis[0].annotate(str(p.get\_height()), (p.get\_x() \* 1.1, p.get\_height() \* 1.08))  
  
x = np.linspace(0, 1, 1002)[1:-1]  
a = dfTmp.iloc()[0]['count']  
b = dfTmp.iloc()[1]['count']  
dist = beta(a+1, b+1)  
mean = beta.mean(a+1, b+1)  
meanY = dist.pdf(mean)  
axis[1].plot(x, dist.pdf(x))  
axis[1].set(title='D: Beta distribution of Coin flipping')  
axis[1].vlines(x = mean, ymin = 0, ymax = meanY, colors = 'black', label = 'Mean value' ,linestyle = 'dotted')  
axis[1].set\_xlabel('Probability')  
axis[1].set\_ylabel('Probability density')  
axis[1].set(title='B')  
axis[1].set\_xlim(0,1)  
axis[1].set\_ylim(0,10)  
ax2 = axis[1].twiny()  
  
labels = [item.get\_text() for item in ax2.get\_xticklabels()]  
labels[0]="Head"  
labels[-1]="Tail"  
ax2.tick\_params(axis='x', colors='red')  
ax2.set\_xticklabels(labels)  
axis[1].set\_xticks([0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9])  
figure.tight\_layout()

/var/folders/f2/skrbd8xn0mdf1wfmzx2l26pc0000gn/T/ipykernel\_21455/2298967200.py:47: UserWarning: FixedFormatter should only be used together with FixedLocator  
 ax2.set\_xticklabels(labels)



### Narratives

Next, our random walker aims to discover those narrative contexts in which women are discussed in the news. As discussed above, this also features uncertainty from the perspective of the random walker. During his discovery of the archive, he faces competing possibilities. His goal is therefore to make the content of news sheets predictable to some extent and build stable knowledge. If a news item mentions a woman, what are the most likely words co-occurring with it? Similarly, what words are likely to precede or follow the mention of a woman? What do women do or what actions are attributed to them?

To answer these questions, he keeps reading the news sheets and “learns” the probabilities of how words follow each other. Based on this knowledge, he starts to generate texts synthetically and observes how the randomness of word sequences functions. During the text generation process, our random walker will eventually exhaust all possibilities and learn the probability of each. The synthetic generation of texts is called simulation; it is a convenient method to explore all possibilities of how information in a large textual dataset can be articulated. If certain conditions are met, the text simulation reaches a stable state (also known as an equilibrium state), resulting in a firm knowledge of all possible ways words can unfold, including their respective probabilities. As discussed below, the stable state reached during the simulation provides a highly informative viewpoint where sample bias is mitigated.

I implemented the procedure above on the MIA-Euronews Corpus by means of a mathematical model named Markov Chain (). First, the model “learned” the probabilities of how words follow each other in the MIA-Euronews Corpus. Second, I ran the Markov Chain model, which is a form of simulation sometimes referred to as a random walk in the literature (). In practice, when we run a Markov Chain to model the possible sequences of words in a large textual collection, we synthetically generate texts (or possible subsamples) consisting of millions of words. Again, the result reveals all possible ways words in a large textual collection may unfold, thereby enabling us to observe how the randomness of word sequences functions in a given textual collection.

The Markovian simulation brought about three pieces of information. First, it revealed the overall or stable state probability of a given word in the MIA-Euronews Corpus. This overall probability also expresses its significance in the corpus (for further details, see the Implementation section). By calculating the overall probability of certain words referring to women and other historical actors, I quantified the significance and importance of these words (see Figures 4 and 8 in Section 3). Then, I manually clustered the most significant words into units of distinct topics (see Figure 7). Second, the Markovian simulation revealed the probability that a given word follows or precedes another word (see Figure 10 and Figures 11 to 14 in Section 4). However, this estimate can be inaccurate for very infrequent terms. Therefore, I focused on salient and frequent associations. In summary, the Markovian approach helped explore words that follow or precede terms denoting women, offering insights into women’s representation and agency in the corpus. Finally, the Markovian simulation provided information about the strength of association between words (see Figure 9 and Figures 15 to 18 in Section 4).

With the approach described above, I could investigate how information surrounding women in early modern news is likely to unfold in narratives. In contrast to other traditional — static— approaches (for instance, Pointwise Mutual Information) that uncover associative relationships between words, the Markovian approach has the advantage of modeling texts as dynamic systems. It allows us to discover all possible ways information is discussed in a large textual collection. Most importantly, the Markovian simulation quantifies the probability of these possible ways. The Markovian simulation also helps address the problem of bias by enabling us to go beyond an original non-representative sample. It does this by synthetically producing additional possible sub-samples. Hence, it helped us come closer to the lost corpus of early modern handwritten news. Finally, the Markovian simulation of a textual collection presents the collection from a highly informative and less biased viewpoint: the stable state (also known as equilibrium state).

However, the Markovian approach did not provide exhaustive information about the possible meanings attached to womanhood in early modern news. The core issue is that meaning often remains tacit, and the Markovian approach cannot address this tacitness. Additionally, the Markovian approach struggles with cases of polysemy and homonymy. Therefore, I modeled the meaning of womanhood by reconstructing the semantic space underlying the MIA-Euronews Corpus.

### Semantics

The semantic space approach () is based on an old assumption: the meaning of a word lies in its relationships to other terms (; ; ). The set of all possible relationships between words in a given textual corpus gives rise to a system of meaning underlying that corpus. We can envisage this system of meaning as a high dimensional space populated by words as “points.” In this high dimensional space connections between terms do not imply bonds (as in a network); connections between terms instead imply spatial proximity and positional similarity; if two terms are closely related semantically, they are also close to each other in the high dimensional space representing the semantic space and the entire system of meaning. Further details about the process of constructing a semantic space out of a textual corpus are discussed in the Implementation section; here I limit the discussion to three broader points.

First, even though a semantic space has some similarities with the two or three dimensional geometric spaces we are accustomed to, it is fundamentally different. A semantic space is a vector space; words that populate the semantic space are not points, but vectors (often named word vectors in the literature). Each word represented as a word vector has certain mathematical features such as angle and magnitude. These mathematical features are used to study how a given word relates to other words, ultimately providing information about the meaning of that word in a textual corpus.

Second, we cannot visualize a high dimensional semantic space; instead, we project it into a lower - two or three - dimensional space. Through these projections, we can observe how words are semantically related to each other (see Figure 19 in Section 4). A key advantage of the semantic space approach is that it can show words not as isolated entities but as parts of groups or clusters; members of a given cluster are semantically related and often form topics. You can observe clusters of semantically related words in Figure 19 in Section 4.

Third, the semantic space approach is an effective method to visualize the set of competing possibilities associated with a given set of words (e.g., terms related to women in our case). Through the visualization of the semantic space projected into a lower dimensional space, we can observe these possibilities and construct a solid understanding.

To sum up, the semantic space approach reveals the possible connections a word can have in a textual corpus. The possible connections include synonymy, associative bonds, and relations to larger clusters of semantically related words, or topics. In the context of this research, the semantic space approach has helped address the following questions: What are the groups of words related to words denoting women in the MIA-Euronews Corpus? What do these groups tell us about the representation of women in early modern news? What are the topics that underlie groups of words semantically related to women in the MIA-Euronews Corpus? Similar questions can be raised about men as well; the semantic analysis of men is the subject of my current and future research; its discussion is therefore limited here. At the same time, the drawback of the semantic space approach is that it does not give easy access to the lost corpus of early modern handwritten news, a problem that remains unsolved in this paper.

## Results A: Presence of Women in the News

In 1618, there was an interesting exchange of letters between officials of the Medici Court. The letters offer a rare glimpse into the news consumption in the Medici court.

Today after dining, the Most Serene Arch Duchess our Ladyship wants to hear the reading of the newsletter from Venice. And Her Most Serene Highness would be pleased if Your Lordship would secure their delivery every week, writing to this effect to that friend so that he might say how much he is to be given annually. (ASF MdP 4634, fo. 242; cited and translated by , pp. 726–727 )

The letter refers to Maria Magdalena von Habsburg (1587-1631), the consort of Cosimo II (1609-1621). Apparently, Maria Magdalena was an eager reader of handwritten news. In this section, I will explore how often she might have read about fellow women.

The early modern handwritten news world was populated by a great variety of actors such as kings, ambassadors, and consorts. However, these actors were not equally important protagonists. Figure 4 renders those personages who are the most important on the news sheets of the Medici Archive. There is a striking thing in Figure 6: women are significantly less important than men (observe words such as donna (woman), moglie (wife), signora (lady)). This difference is even more striking when comparing word pairs with feminine and masculine forms. For instance, according to my estimation, the probability of reading about sons was three times higher than the probability of hearing about daughters. I found a similar pattern with other word pairs: brother (fratello) - sister (sorella), duke (duca) - duchess (duchessa), man (uomo) - woman (donna), she (lei) - he (lui). All this highlights the dominance of male actors in the early modern handwritten news world.

import json  
from msmtools.estimation import transition\_matrix,count\_matrix, is\_connected  
from pyrandwalk import \*  
import pandas as pd  
  
# Load the data  
  
f = open('data/input/newsSheetsEncoded.json')  
finalDocs = json.load(f)  
finalDocs = [np.array(element) for element in finalDocs]  
  
f = open('data/input/dictionaryMarkovianSimulation.json')  
dictionary = json.load(f)  
  
f = open('data/input/actorsTerms.json')  
actorsTerms = json.load(f)  
  
# Create a count matrix  
  
def createCountMatrix(trajectories,dictionary,window=1):  
 countMatrix = np.zeros((len(dictionary),len(dictionary)))  
 for i in range(1,window+1):  
 countMatrix = countMatrix + count\_matrix(trajectories,lag=i)  
 countMatrix = np.array(countMatrix)  
 assert len(dictionary) == countMatrix.shape[0]  
 return countMatrix  
  
cm = createCountMatrix(finalDocs,dictionary,5)  
assert len(np.where(cm.sum(axis=1)==0)[0]) == 0  
  
# Create the transition matrix  
  
def createTransitionMatrix(trajectories,dictionary,window=1):  
 countMatrix = np.zeros((len(dictionary),len(dictionary)))  
 for i in range(1,window+1):  
 countMatrix = countMatrix + count\_matrix(trajectories,lag=i)  
 countMatrix = np.array(countMatrix)  
 P = transition\_matrix(countMatrix)  
 assert len(dictionary) == P.shape[0]  
 return P  
  
P = createTransitionMatrix(finalDocs,dictionary,window=5)  
  
# Train the Markov chain and get the statitionary probabilityof each word  
  
def printStableProbabilities(P,dictionary):  
 assert len(P) == len(dictionary)  
 vocabNumerical = [i for i,element in enumerate(dictionary)]  
 rw = RandomWalk(vocabNumerical, P)  
 stationaryProbs = rw.final\_dist()  
 dfStationaryProb = pd.DataFrame(stationaryProbs,columns=['prob'])  
 dfStationaryProb['term'] = dictionary  
 dfStationaryProb['scaledToMedian'] = dfStationaryProb['prob'] / dfStationaryProb['prob'].median()  
 dfStationaryProb.sort\_values('prob',ascending=False,inplace=True)  
 return dfStationaryProb  
  
stationaryProb = printStableProbabilities(P,dictionary)  
stationaryProb = stationaryProb.sort\_values('prob',ascending=False)  
  
# Print the stationary probability of key actors, as well as whether they are part of the upper quartile  
  
dfTemp = stationaryProb.reset\_index()  
quantile = []  
for element in dfTemp.prob.to\_list():  
 if element > dfTemp.prob.quantile(0.75):  
 quantile.append(True)  
 else:  
 quantile.append(False)  
dfTemp['upperQuantile'] = quantile  
dfTemp[dfTemp.term.isin(actorsTerms)]

index prob term scaledToMedian upperQuantile  
1 20 0.038665 signore 59.498249 True  
2 79 0.030361 re 46.719378 True  
4 2 0.011632 lui 17.899817 True  
5 85 0.010445 duca 16.073625 True  
12 142 0.007740 principe 11.909641 True  
13 18 0.007502 cardinale 11.543861 True  
15 45 0.007446 papa 11.457722 True  
25 41 0.005529 ambasciatore 8.507827 True  
44 194 0.003981 generale 6.126149 True  
50 40 0.003665 monsignore 5.640325 True  
57 372 0.003281 marchese 5.049596 True  
58 185 0.003261 soldato 5.018316 True  
63 165 0.003023 regina 4.652124 True  
66 731 0.002941 uomo 4.525110 True  
68 462 0.002874 imperatore 4.422910 True  
80 141 0.002432 figlio 3.742086 True  
109 269 0.001927 governatore 2.965480 True  
112 478 0.001890 signora 2.908886 True  
122 98 0.001824 fratello 2.806491 True  
156 5 0.001436 vescovo 2.210089 True  
172 672 0.001347 elettore 2.073071 True  
174 735 0.001331 padre 2.047642 True  
262 100 0.000910 moglie 1.399733 False  
276 97 0.000869 lei 1.337030 False  
291 559 0.000830 duchessa 1.277568 False  
312 489 0.000769 figlia 1.183296 False  
347 99 0.000687 madre 1.056894 False  
411 78 0.000587 donna 0.903468 False  
462 114 0.000525 prete 0.808342 False  
682 192 0.000335 sorella 0.514934 False  
713 736 0.000283 marito 0.434842 False

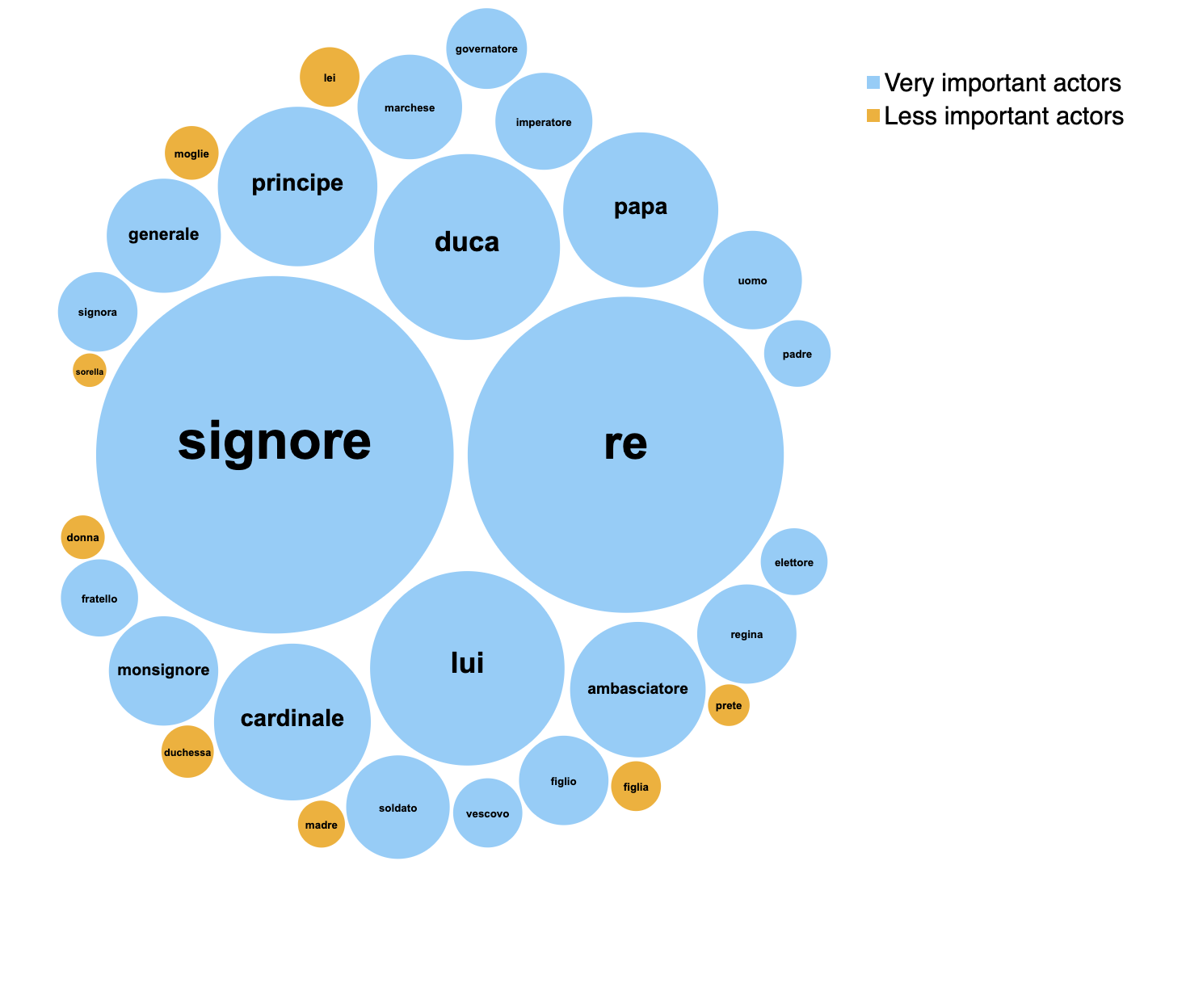
The process of random sampling described above uncovered a general tendency. While news sheets tended to be silent about women, they almost always mention men (see Figure 5 and Figure 6). More specifically, out of a hundred randomly selected news items, approximately fourteen news items mention at least one woman; by contrast, approximately eighty-five news items mention at least one man. At the same time, it is important to note there was no complete silence about women. Early modern news readers did not get much news about women in general, but they often heard about at least one woman when they read an entire news sheet (see Figure 7).

Although there was no complete silence about women as a group, I found that there was silence about non-elite women. My findings suggest that early modern consumers of handwritten news overwhelmingly read about elite women (see Figure 4 and note the small size of the bubble that represents the importance of donna (woman), a word usually referring to women of lower social status).

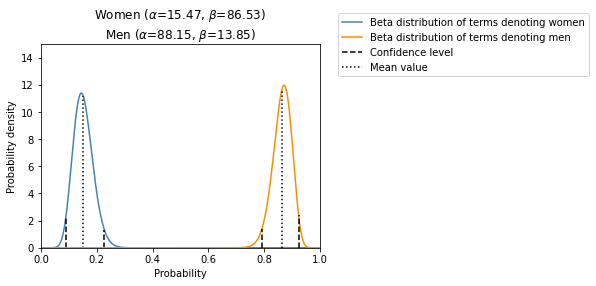
I also investigated whether the presence of women increased over time. This investigation revealed that the presence of women did not change significantly over time (see Figure 6). Similarly, I did not find any significant variation in the presence of women when comparing news sheets from different news hubs.

To summarize, Maria Magdalena did not hear a lot of news related to women, but she still learned something about women when she read the news.

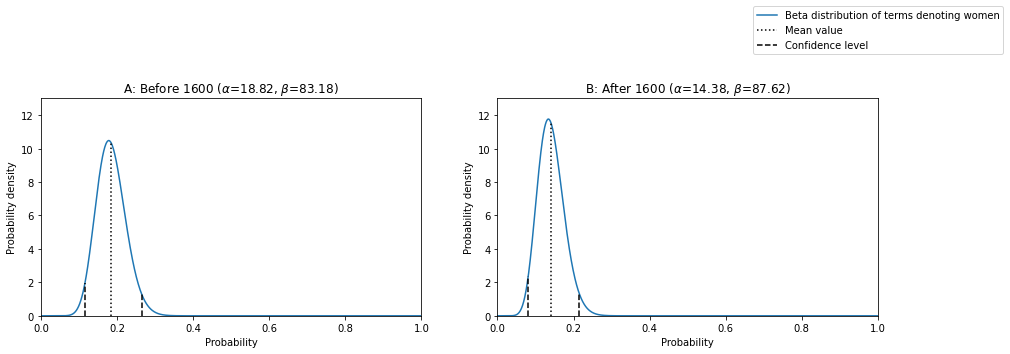
Image('media/image12.png',width=800)



Image('media/image25.png')

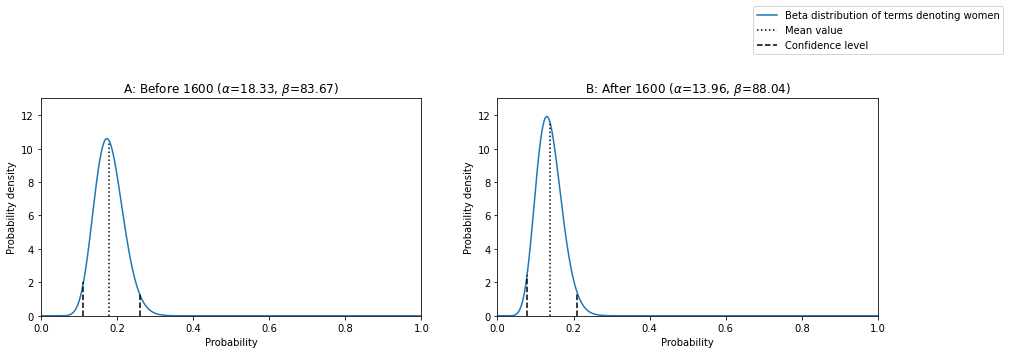


Image('media/image14.png')

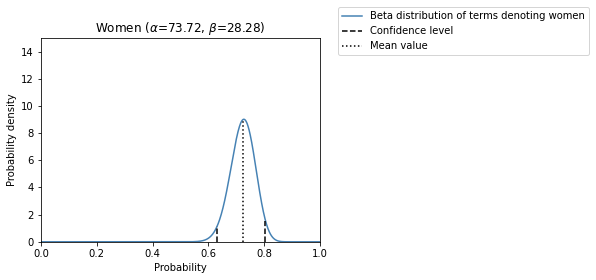


from scipy.stats import beta  
  
def dt\_parse(s):  
 y,m,d = s.split('-')  
 return pd.Period(year=int(y), month=int(m), day=int(d), freq='D')  
  
  
df = pd.read\_csv('data/input/WomenMenAbsencePresence.csv')  
df['date']=df.date.apply(dt\_parse)  
df = df.set\_index('date')  
  
  
n\_trials = 100  
sampleSize = 100  
finalResults = []  
temporalBoundaries = [1600,1599]  
for f,time in enumerate(temporalBoundaries):  
 results = []  
 for i in range(0,n\_trials):  
 if f == 0:  
 #print (df[(df.index.year<time)].shape)  
 sample = df[(df.index.year<time)].sample(n=sampleSize)  
 else:  
 #print (df[(df.index.year>time)].shape)  
 sample = df[(df.index.year>time)].sample(n=sampleSize)  
 success = sample[sample.woman==True].shape[0]  
 failure = sample[sample.woman==False].shape[0]  
 result = np.array([success,failure])  
 results.append(result)  
 results = np.vstack(results)  
 a = results.mean(axis=0)[0]  
 b = results.mean(axis=0)[1]  
 dist = beta(a+1, b+1)  
 X = np.linspace(0, 1, 1002)[1:-1]  
 y = dist.pdf(X)  
 # Confidence levels  
  
 lower, upper = beta.interval(0.95, a+1, b+1, loc=0)  
 meanY = dist.pdf(mean)  
 lowerY = dist.pdf(lower)  
 upperY = dist.pdf(upper)  
 a = np.format\_float\_positional(a+1,precision=2)  
 b = np.format\_float\_positional(b+1,precision=2)  
 if f == 0:  
 titleText = 'A: Before 1600 ('+r'$\alpha$='+str(a)+r', $\beta$='+str(b)+')'  
 else:  
 titleText = 'B: After 1600 ('+r'$\alpha$='+str(a)+r', $\beta$='+str(b)+')'  
  
 finalResult = {'title':titleText,  
 'a':a,'b':b,  
 'x':X,'y':y,  
 'meanX':dist.mean(),  
 'meanY':dist.pdf(dist.mean()),  
 'lowerY':lowerY,  
 'lowerX':lower,  
 'upperY':upperY,  
 "upperX":upper}  
  
 finalResults.append(finalResult)   
  
### Plot it  
  
A = np.array(finalResults)  
B = np.reshape(A, (-1, 2))  
  
figure, axis = plt.subplots(B.shape[0], B.shape[1],figsize=(15, 4))  
  
for i in range(B.shape[0]):  
 for f in range(B.shape[1]):  
 axis[f].plot(B[i][f]['x'], B[i][f]['y'],label='Beta distribution of terms denoting women')  
 axis[f].set\_title(B[i][f]['title'])  
 axis[f].vlines(x = B[i][f]['meanX'],ymax= B[i][f]['meanY'],ymin = 0,linestyle = 'dotted',label='Mean value',colors = 'black')  
 axis[f].vlines(x = B[i][f]['upperX'],ymax= B[i][f]['upperY'],ymin = 0,linestyle = 'dashed',label='Confidence level',colors = 'black')  
 axis[f].vlines(x = B[i][f]['lowerX'],ymax= B[i][f]['lowerY'],ymin = 0,linestyle = 'dashed',colors = 'black')  
 axis[f].set\_ylim(ymin=0,ymax=13)  
 axis[f].set\_xlim(xmin=0,xmax=1)  
  
 axis[f].set\_ylabel(r'Probability density')  
 axis[f].set\_xlabel(r'Probability')  
  
  
  
handles, labels = axis[f].get\_legend\_handles\_labels()  
figure.legend(handles, labels,bbox\_to\_anchor=(0.9, 1.2), loc='upper center', borderaxespad=0)  
  
  
plt.show()

/opt/homebrew/Caskroom/miniconda/base/envs/euronews/lib/python3.10/site-packages/scipy/stats/\_continuous\_distns.py:624: RuntimeWarning: overflow encountered in \_beta\_ppf  
 return \_boost.\_beta\_ppf(q, a, b)



Image('media/image36.png')

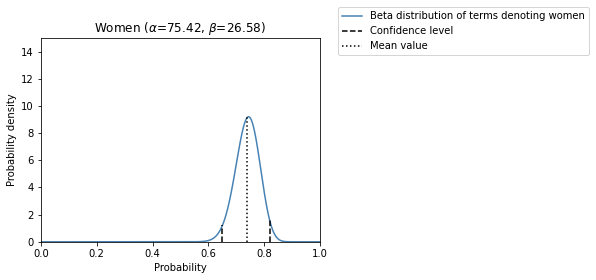


from scipy.stats import beta  
  
dfTemp = df.docid.value\_counts().to\_frame('docid')  
indices = dfTemp[dfTemp.docid>7].index.to\_list()  
df = df[df.docid.isin(indices)]  
  
  
# Sample women  
alpha\_values = []  
beta\_values = []  
n\_trials = 100  
sampleSize = 100  
results = []  
for i in range(0,n\_trials):  
 #sample = df.sample(n=sampleSize)  
  
 sample = df.docid.sample(n=sampleSize)  
  
 resultM = []  
 resultW = []  
 for dId in sample.to\_list():  
 resW = True in df[df.docid==dId].woman.to\_list()  
 resM = True in df[df.docid==dId].man.to\_list()  
 resultW.append(resW)  
 resultM.append(resM)  
  
 # Postprocess sample  
 sample = sample.to\_frame()  
 sample['woman'] = resultW  
  
  
 success = sample[sample.woman==True].shape[0]  
 failure = sample[sample.woman==False].shape[0]  
 result = np.array([success,failure])  
 results.append(result)  
results = np.vstack(results)  
a = results.mean(axis=0)[0]  
b = results.mean(axis=0)[1]  
alpha\_values.append(a)  
beta\_values.append(b)  
  
  
means = []  
linestyles = ['#4682B4']  
x = np.linspace(0, 1, 1002)[1:-1]  
labels = ['Beta distribution of terms denoting women','Beta distribution of terms denoting men']  
  
#------------------------------------------------------------  
# plot the distributions  
fig, ax = plt.subplots(figsize=(5, 3.75))  
num = 0  
for a, b, ls, label in zip(alpha\_values[0:1], beta\_values[0:1], linestyles[0:1],labels[0:1]):  
 lower, upper = beta.interval(0.95, a+1, b+1, loc=0)  
 mean = beta.mean(a+1, b+1)  
  
 dist = beta(a+1, b+1)  
 meanY = dist.pdf(mean)  
 lowerY = dist.pdf(lower)  
 upperY = dist.pdf(upper)  
 plt.plot(x, dist.pdf(x), c=ls,  
 label=label)  
 if num == 1:  
 plt.vlines(x = lower, ymin = 0, ymax = lowerY, colors = 'black' ,linestyle = 'dashed')  
 plt.vlines(x = upper, ymin = 0, ymax = upperY, colors = 'black', label = 'Confidence level' ,linestyle = 'dashed')  
 plt.vlines(x = mean, ymin = 0, ymax = meanY, colors = 'black', label = 'Mean value' ,linestyle = 'dotted')  
 else:  
 plt.vlines(x = lower, ymin = 0, ymax = lowerY, colors = 'black' ,linestyle = 'dashed',label = 'Confidence level')  
 plt.vlines(x = mean, ymin = 0, ymax = meanY, colors = 'black', linestyle = 'dotted',label = 'Mean value')  
 plt.vlines(x = upper, ymin = 0, ymax = upperY, colors = 'black' ,linestyle = 'dashed')  
 print("Mean:"+str(mean))  
 print ("Positive deviation from the mean:"+str(upper-mean))  
 print ("Negative deviation from the mean:"+str(mean-lower))  
 means.append(dist.mean())  
 num +=1  
plt.xlim(0, 1)  
plt.ylim(0, 15)  
  
  
  
plt.xlabel('Probability')  
plt.ylabel(r'Probability density')  
a\_W = np.format\_float\_positional(alpha\_values[0]+1,precision=2)  
b\_W = np.format\_float\_positional(beta\_values[0]+1,precision=2)  
  
titleTextW = 'Women ('+r'$\alpha$='+str(a\_W)+r', $\beta$='+str(b\_W)+')'  
  
titleText = titleTextW  
plt.title(titleText)  
  
plt.legend(bbox\_to\_anchor=(2,1.2), borderaxespad=1)

Mean:0.7394117647058823  
Positive deviation from the mean:0.0800633681593258  
Negative deviation from the mean:0.08895254555358567

/opt/homebrew/Caskroom/miniconda/base/envs/euronews/lib/python3.10/site-packages/scipy/stats/\_continuous\_distns.py:624: RuntimeWarning: overflow encountered in \_beta\_ppf  
 return \_boost.\_beta\_ppf(q, a, b)

<matplotlib.legend.Legend at 0x17e55df90>



## Results B: Central themes of early modern handwritten news

When consuming news, Maria Magdalena must have encountered a variety of information about the world. However, certain themes were particularly recurrent on the news sheets she read. Generally speaking, newsmongers today and in early modern times tend to encounter similar types of information. Previous scholarship has identified various themes that were allegedly prevalent in early modern news ():

* international politics
* war
* social affairs (such as royal weddings, royal births and official celebrations)
* fires, miracles,
* bloody crimes
* catastrophes (unusual weather, plagues)
* court news (such as royal journeys, royal deaths, and triumphal entries, etc)
* church news
* sensational news (monsters, distorted humans, etc)

However, earlier scholarship did not determine the importance of these topics. While some of these topics were particularly recurrent, others were less frequent. Similarly, the most important topics discussed in the context of women's experience have remained unknown. In this section, I will therefore identify the most important topics of news in the MIA-Euronews Corpus, as well as their relationship to women and women’s experience in the early modern period. This discovery process will be supported by the analysis of words strongly associated with women, and by studying the semantic space underlying terms denoting women. Figure 19 illustrates this semantic space and shows the groups of words to which terms related to women are semantically connected.

# Get those words the stationary probability of which is in the upper quartile  
keyTerms = stationaryProb[stationaryProb.prob>stationaryProb.prob.quantile(0.75)]  
  
# Print those that belong to a certain group of topics, including their stationary probability  
dfKeyTerms = pd.read\_csv('data/input/keyTermsTopics.csv',delimiter=';')  
dfKeyTerms = dfKeyTerms.groupby('Topic')['Term'].apply(list).to\_frame()  
for row in dfKeyTerms.iterrows():  
 print('Topic - '+row[0]+':\n')  
 for word in row[1]['Term']:  
 try:  
 print(word)  
 prob = keyTerms[keyTerms.term==word]['prob'].values[0]  
 except:  
 continue  
 print('\n')

Topic - Catastrophe, disease, and death:  
  
male  
morire  
morte  
danno  
  
  
Topic - Economics:  
  
dare  
pagare  
denaro  
danaro  
spesa  
carico  
  
  
Topic - Festivity :  
  
festa  
bello  
  
  
Topic - Governance:  
  
ordine  
servire  
corte  
ministro  
governo  
popolo  
regno  
affare  
impedire  
principale  
ordinare  
  
  
Topic - Information:  
  
dire  
mandare  
scrivere  
lettera  
sentire  
ricevere  
spedire  
avviso  
risposta  
  
  
Topic - Mental activity and intentions:  
  
volere  
credere  
vedere  
intendere  
sapere  
parere  
dubitare  
pensare  
  
  
Topic - Movement:  
  
andare  
venire  
passare  
portare  
partire  
restare  
aspettare  
parto  
ritornare  
condurre  
entrare  
arrivare  
mare  
via  
ritorno  
tornare  
strada  
viaggio  
  
  
Topic - Religion:  
  
cattolico  
religione  
chiesa  
  
  
Topic - War:  
  
guerra  
soldato  
cavallo  
trattare  
armata  
pace  
terra  
galera  
forza  
nemico  
fante  
ritirare  
guardia  
arma  
nave  
esercito  
prigione  
fortezza  
truppa  
campo  
campagna

# Print those words that are the most likely to accompany terms related to women  
  
womenTerms = ['regina',  
 'signora',  
 'contessa',  
 'marchesa',  
 'duchessa',  
 'principessa',  
 'sposa',  
 'moglie',  
 'madre',  
 'consorte',  
 'sorella',  
 'figlia',  
 'donna',  
 'nipote',  
 'gentildonna']  
  
terms = pd.read\_csv('data/input/keyWordsSurroundingWomen.csv',delimiter=';')  
addendum = ['vestire','tagliare','difesa',  
 'prigione','attaccare','assedio','sorella','figlia','figlio','deliberare','entrata','moglie']  
for el in addendum:  
 entry = pd.DataFrame([{'Term':el}])  
 terms = pd.concat([terms, entry], ignore\_index = True, axis = 0)  
terms = terms[~terms.Term.isin(['povero','peste','dire','ducato','condurre','celebrare','visita','robbe'])]  
  
# First run the simulation of the corpus  
  
def simulateText(P,dictionary,length):  
 vocabNumerical = [i for i,element in enumerate(dictionary)]  
 rw = RandomWalk(vocabNumerical, P)  
 states, probs = rw.run(ntimes=length)  
 return states  
  
states = simulateText(P,dictionary,1000000)  
  
# Get those terms that accompany women terms  
  
def postProcessSimulationResultsV2(probabilities,term,stationaryProb):  
 entries = []  
 for key in probabilities:  
 try:  
 entry = {'term':key,'prob':probabilities[key][-1]}  
 except:  
 entry = {'term':key,'prob':np.nan}  
 entries.append(entry)  
  
 df = pd.DataFrame(entries).sort\_values('prob',ascending=False)  
 df['scaledToMedian'] = df['prob']/df.prob.median()  
  
 p\_a = []  
 for ter in term:  
 prob = stationaryProb[stationaryProb['term']==ter]['prob'].values[0]  
 p\_a.append(prob)  
 p\_a = np.array(p\_a).sum()  
 results = []  
 for row in df.iterrows():  
 term = row[1]['term']  
 p\_b\_a = row[1]['prob']  
 if np.isnan(p\_b\_a):  
 p\_a\_b = np.nan  
 else:  
 p\_b = stationaryProb[stationaryProb['term']==term]['prob'].values[0]  
 p\_a\_b = bayes\_theorem(p\_a,p\_b,p\_b\_a)  
 results.append(p\_a\_b)  
 df['bayesRescaled'] = results  
 return df  
  
def getContextualProbabilitiesV2(simulatedStates,term,dictionary,direction=None):  
  
 counter = {element:[] for element in dictionary}  
 probabilities = {element:[] for element in dictionary}  
 means = {element:[] for element in dictionary}  
 positions = {element:[] for element in dictionary}  
  
 if direction == 'preceding':  
 for i,element in enumerate(np.where(np.array(simulatedStates) == dictionary.index(term))[0]):  
 try:  
 neighbour = simulatedStates[element-1]  
 p = len(counter[dictionary[neighbour]])/i  
 probabilities[dictionary[neighbour]].append(p)  
 means[dictionary[neighbour]].append(np.array(probabilities[dictionary[neighbour]]).mean())  
 counter[dictionary[neighbour]].append(1)  
 position = element  
 positions[dictionary[neighbour]].append(position)  
 except Exception as e:  
 #print(e)  
 pass  
 return means, positions  
 elif direction == 'following':  
 for i,element in enumerate(np.where(np.array(simulatedStates) == dictionary.index(term))[0]):  
 try:  
 neighbour = simulatedStates[element+1]  
 p = len(counter[dictionary[neighbour]])/i  
  
 probabilities[dictionary[neighbour]].append(p)  
 means[dictionary[neighbour]].append(np.array(probabilities[dictionary[neighbour]]).mean())  
 counter[dictionary[neighbour]].append(1)  
 position = element  
 positions[dictionary[neighbour]].append(position)  
 except Exception as e:  
 #print(e)  
 pass  
 return means, positions  
  
 elif direction == None:  
 indices = []  
 for ter in term:  
 index = dictionary.index(ter)  
 indices.append(index)  
 locations = []  
  
 for ind in indices:  
 loc = np.where(np.array(simulatedStates) == ind)[0]  
 locations.extend(loc)  
  
 for i,element in enumerate(locations):  
 try:  
 neighbour = simulatedStates[element+1]  
 p = len(counter[dictionary[neighbour]])/i  
  
 probabilities[dictionary[neighbour]].append(p)  
 means[dictionary[neighbour]].append(np.array(probabilities[dictionary[neighbour]]).mean())  
 counter[dictionary[neighbour]].append(1)  
 position = element  
 positions[dictionary[neighbour]].append(position)  
 except Exception as e:  
 #print(e)  
 pass  
 try:  
 neighbour = simulatedStates[element-1]  
 p = len(counter[dictionary[neighbour]])/i  
  
 probabilities[dictionary[neighbour]].append(p)  
 means[dictionary[neighbour]].append(np.array(probabilities[dictionary[neighbour]]).mean())  
 counter[dictionary[neighbour]].append(1)  
 position = element  
 positions[dictionary[neighbour]].append(position)  
 except Exception as e:  
 #print(e)  
 pass  
 return means, positions  
  
 else:  
 raise Exception("Direction should be either preceding or following or None")  
  
  
def bayes\_theorem(p\_a, p\_b, p\_b\_given\_a):  
 # calculate P(A|B) = P(B|A) \* P(A) / P(B)  
 p\_a\_given\_b = (p\_b\_given\_a \* p\_a) / p\_b  
 return p\_a\_given\_b   
  
probabilities,positions = getContextualProbabilitiesV2(states,womenTerms,dictionary)  
dfResult = postProcessSimulationResultsV2(probabilities,womenTerms,stationaryProb)  
dfTemp = dfResult[dfResult.term.isin(terms.Term.to\_list())]  
dfTemp = dfTemp.fillna(0)  
  
# Print those selected terms that are most likely to follow or precede women terms  
display(dfTemp.sort\_values('prob',ascending=False)[['prob','term']])  
  
# Print the selected terms above with the strength of association between them and women terms  
  
dfTemp['bayesRescaled'] = dfTemp.bayesRescaled/dfTemp.bayesRescaled.sum()  
display(dfTemp.sort\_values('bayesRescaled',ascending=False)[['prob','term']])

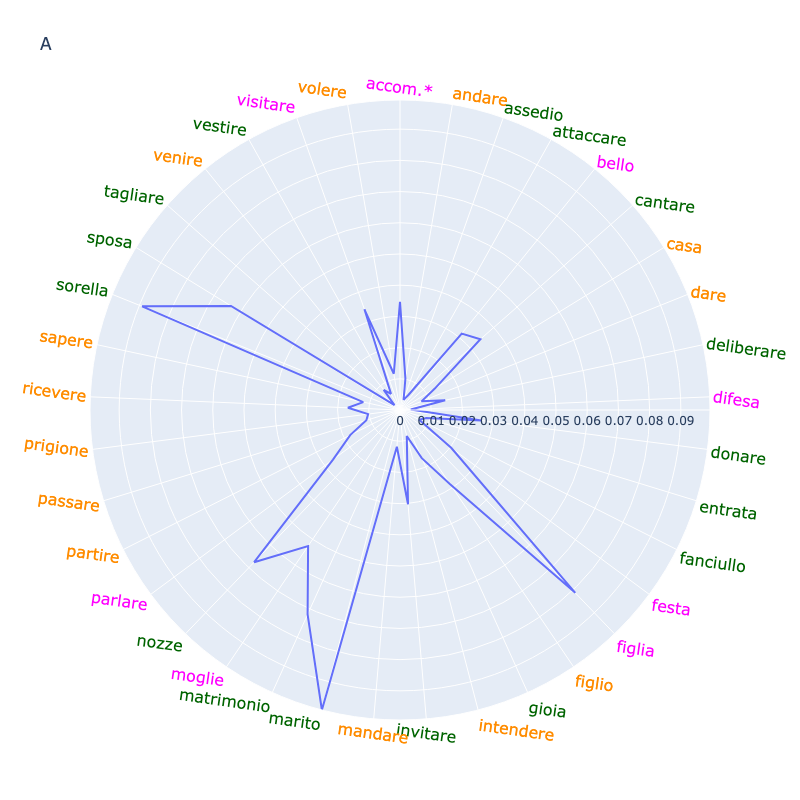
prob term  
27 0.024490 andare  
135 0.019923 passare  
124 0.016042 volere  
17 0.014438 mandare  
6 0.013632 dare  
115 0.013489 partire  
141 0.013351 figlio  
489 0.012582 figlia  
33 0.012275 venire  
11 0.012068 sapere  
100 0.010547 moglie  
23 0.009963 casa  
121 0.009739 matrimonio  
30 0.009117 intendere  
51 0.009112 bello  
93 0.008957 sposa  
446 0.007860 visitare  
96 0.006862 accompagnare  
31 0.006361 ricevere  
736 0.005577 marito  
370 0.005419 festa  
192 0.005174 sorella  
590 0.003930 nozze  
422 0.003904 parlare  
312 0.003568 prigione  
164 0.003035 donare  
87 0.002803 invitare  
416 0.001778 deliberare  
543 0.001752 entrata  
443 0.001257 gioia  
427 0.001086 cantare  
350 0.001046 assedio  
551 0.000967 tagliare  
178 0.000760 attaccare  
690 0.000674 difesa  
54 0.000633 vestire  
733 0.000000 fanciullo

prob term  
736 0.005577 marito  
489 0.012582 figlia  
121 0.009739 matrimonio  
93 0.008957 sposa  
192 0.005174 sorella  
100 0.010547 moglie  
590 0.003930 nozze  
164 0.003035 donare  
51 0.009112 bello  
446 0.007860 visitare  
87 0.002803 invitare  
96 0.006862 accompagnare  
416 0.001778 deliberare  
141 0.013351 figlio  
370 0.005419 festa  
422 0.003904 parlare  
543 0.001752 entrata  
443 0.001257 gioia  
115 0.013489 partire  
11 0.012068 sapere  
427 0.001086 cantare  
135 0.019923 passare  
23 0.009963 casa  
312 0.003568 prigione  
31 0.006361 ricevere  
27 0.024490 andare  
124 0.016042 volere  
30 0.009117 intendere  
17 0.014438 mandare  
551 0.000967 tagliare  
350 0.001046 assedio  
33 0.012275 venire  
6 0.013632 dare  
54 0.000633 vestire  
178 0.000760 attaccare  
690 0.000674 difesa  
733 0.000000 fanciullo

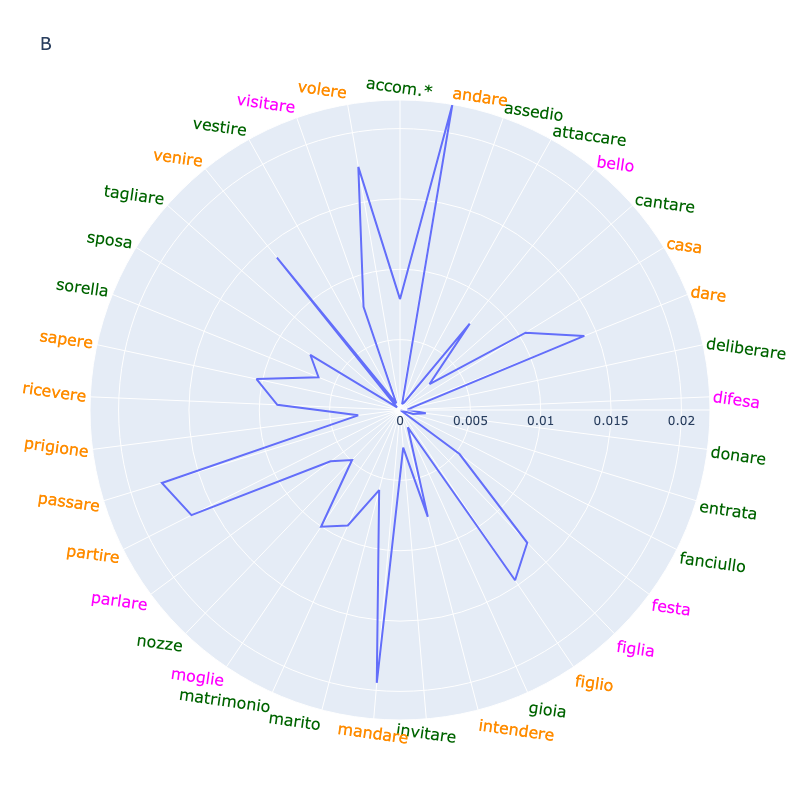
Image('media/image8.png',width=800)



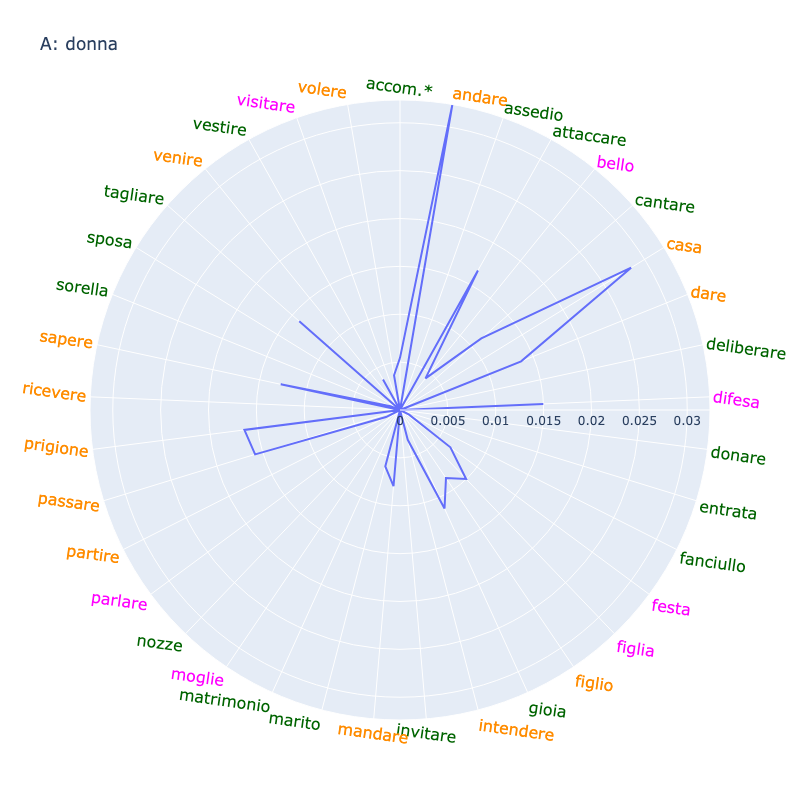
Image('media/image15.png')



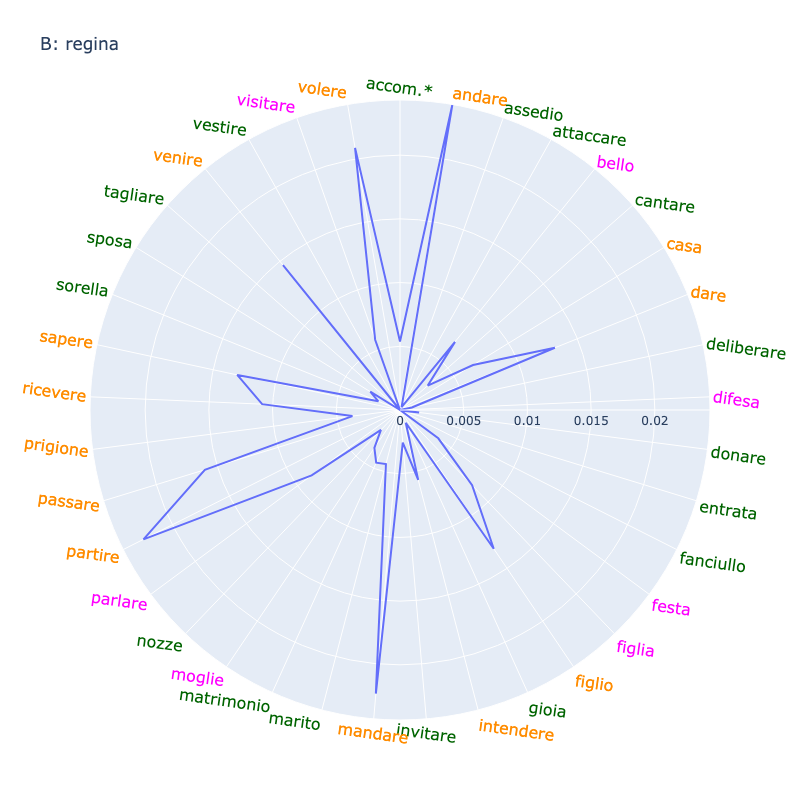
Image('media/image32.png')



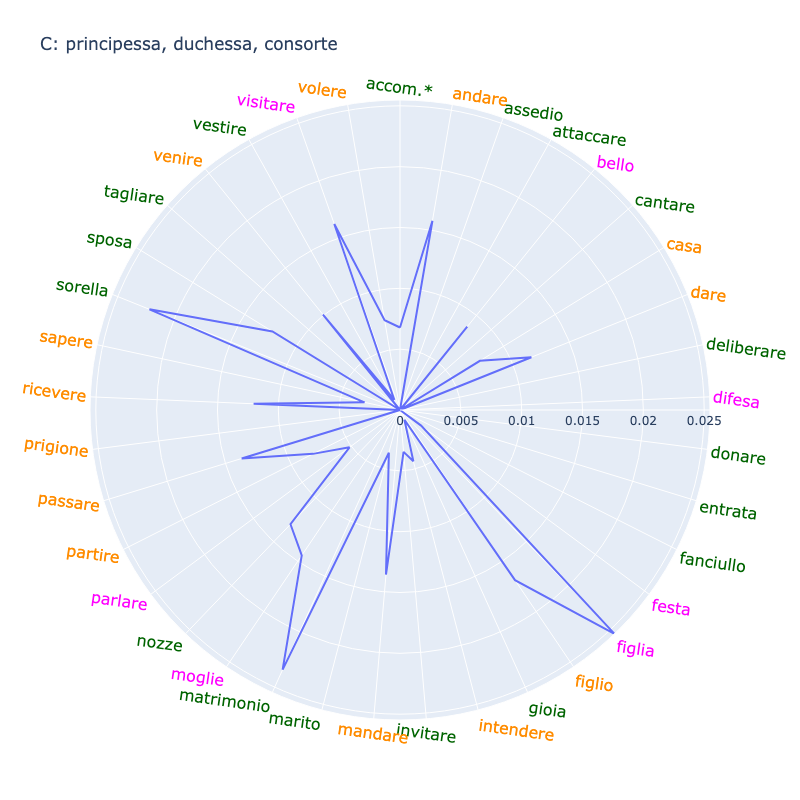
Image('media/image29.png')



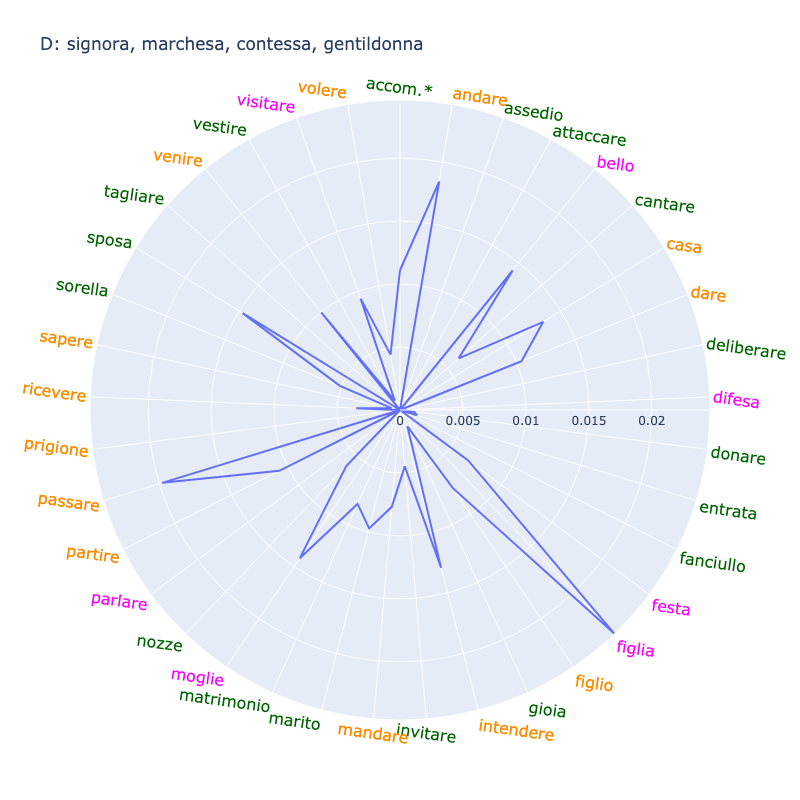
Image('media/image21.png')



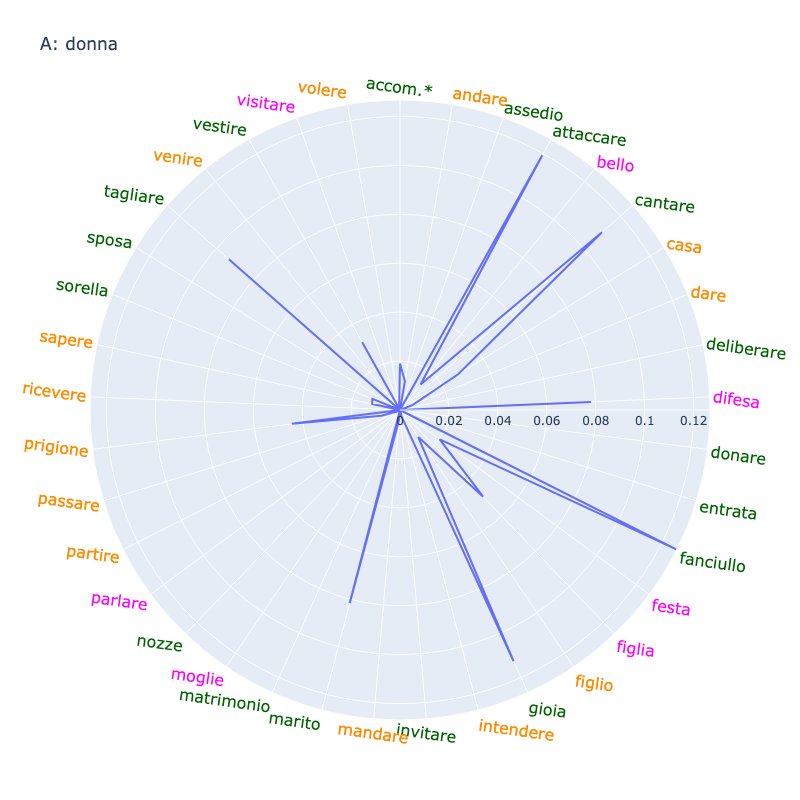
Image('media/image20.png')



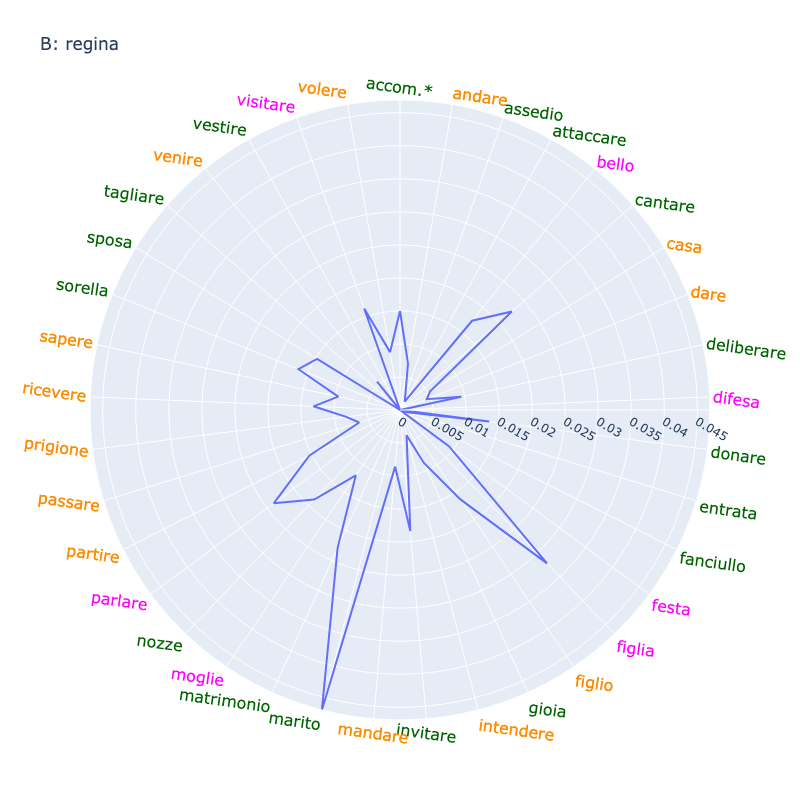
Image('media/image30.png')



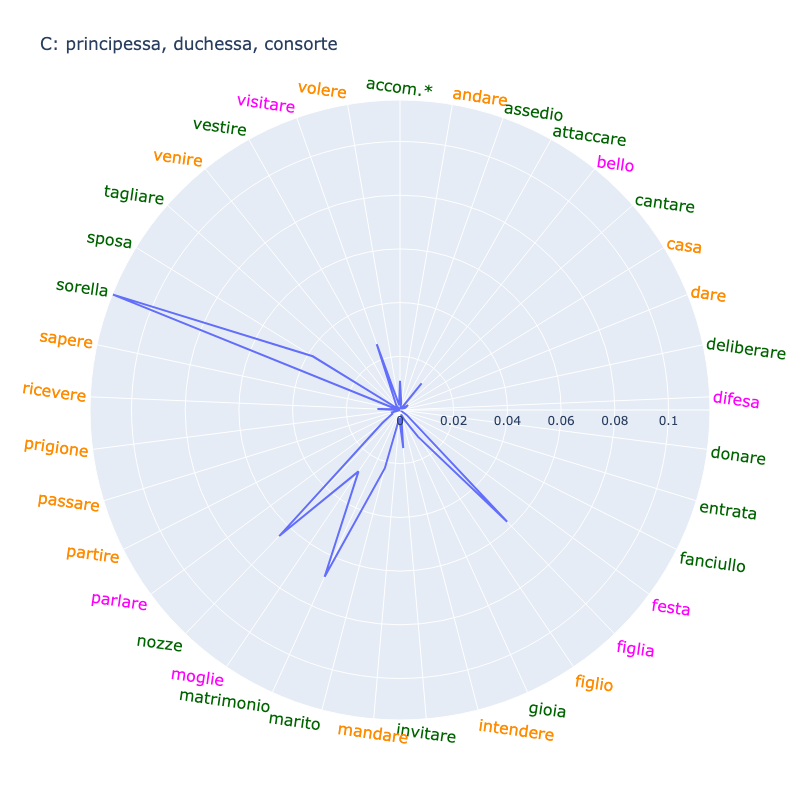
Image('media/image24.png')



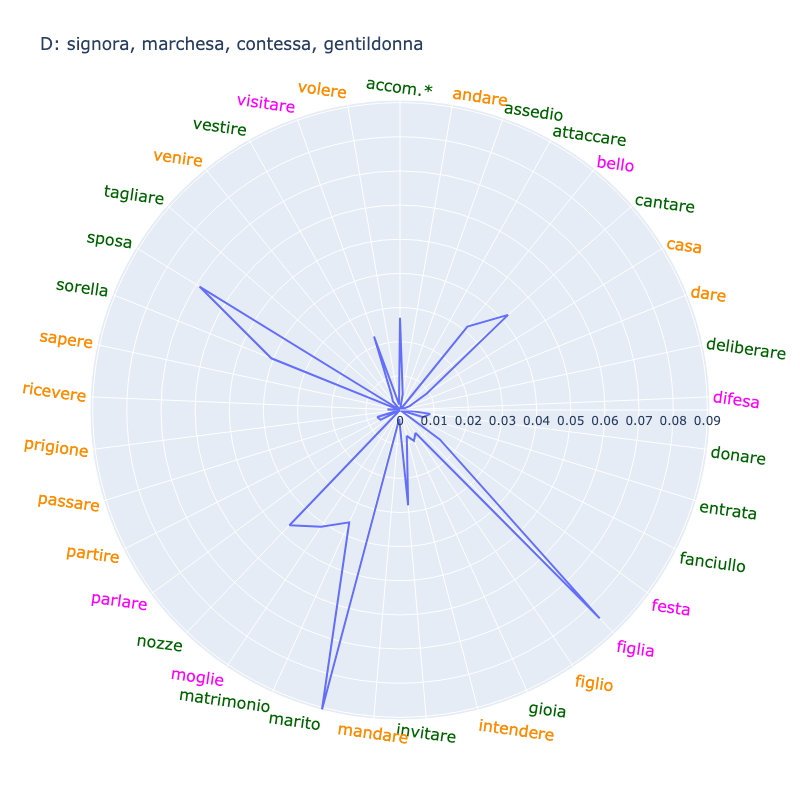
Image('media/image26.png')



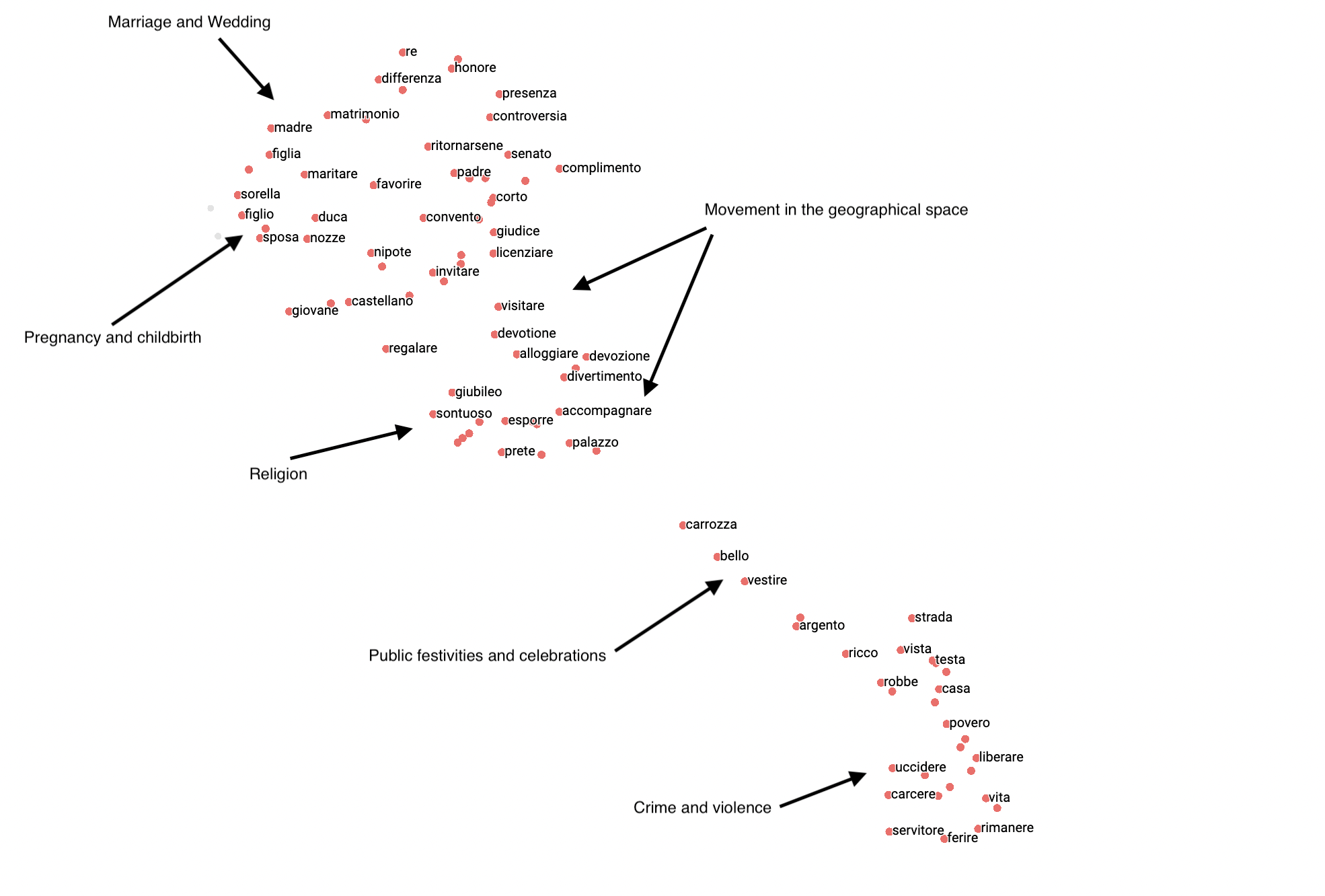
Image('media/image10.png')



Image('media/image9.png')



Image('media/image5.png')



### Movement in the geographical space

In August of 1551, the Medici court received a news sheet from Vienna (ASF MdP, 4572) giving notice about the departure of Duke Ferdinand of Austria (Holy Roman Emperor, 1556-1564) to Bohemia. The same news sheet also informed its readers about the departure of the Hungarian aristocrat, Tamas Nadasdy, to his castles. The news sheet from Vienna typifies the single most important theme in the MIA-Euronews Corpus, and most probably in early modern handwritten news sheets in general. This theme is the movement of historical actors and goods through geographical space. Early modern news extensively reported on the arrivals, departures, and returns of historical actors such as sovereigns, churchmen, and ambassadors. News about the arrival of messengers and ships with precious cargo was also recurrent. As road conditions dramatically improved and courts frequently traveled between summer and winter residences, the early modern period became a world in movement (). Movements of sovereigns were highly visible events, often accompanied by public rituals and processions.

In light of this, it is not surprising that one of the most likely topics discussed in the context of women is movement and traveling (see Figure 10). Terms denoting spatial movements such as andare (to go), venire (to come), passare (to pass by), and partire (to depart) are some of the most likely terms to co-occur with terms denoting women. Of course, if social status is taken into consideration, there are important differences. While the news extensively discusses the movement of women belonging to the elite, the discussion of the movement of women of humble origin, usually mentioned by the general term donna (woman), is less comprehensive (see Figures 11 to 14). At the same time, it is important to underline that movement in the geographical space is not a topic that is strongly associated with women only. In other words, terms denoting movement in space are very likely to occur in the context of women, but the strength of association between these terms and women terms still tends to be low (see Figures 15 to 18). This is also highlighted by the absence of travel words from the semantic space around women terms (see Figure 19).

There are, however, three terms that are loosely related to the topic of movement in geographic space and also strongly connected with terms denoting women. The first one is visitare (to visit). Early modern news often discusses the spatial movement of actors aiming to visit another actor or special locations such as churches, palaces, or events related to devotion (see the proximity of travel words to terms related to religion in Figure 19). In these visits, women of the elite often had prominent roles, which is demonstrated by the relatively strong associative relationship between women of the elite and the term visitare (see Figures 16 to 18 and Figure 19). The second term is accompagnare (to accompany). Women did not travel alone; they were often accompanied by their husbands or their entourage. The news often discusses who accompanied women throughout their travels. The third term is alloggiare (to lodge). News frequently highlight who put up foreign visitors and dignitaries, which explains why alloggiare is positioned close to visitare in the semantic space.

### Intentions and mental states of historical actors

Scholarship has emphasized that news was a tool to monitor geopolitical developments in the world. Newsmongers, such as the high-ranking officials of the Medici court and the grand dukes themselves, read handwritten news sheets to gather insights into the ambitions and goals of other courts and sovereigns. This function of handwritten news explains why the topic that I describe as the intentions and beliefs (or mental states) of historical actors is a particularly important theme in my dataset. Words such as sapere (to know), credere (to believe), volere (to want), pensare (to think), and dubitare (to doubt) belong to the topic of intentions and beliefs of historical actors.

This topic is generally unrelated to women; it is associated with men. News is unlikely to discuss the beliefs and intentions of women (see Figures 11 to 14 and see the absence of terms related to beliefs and intentions in Figure 19). The only exception is queen (most probably referring to Elizabeth I of England (1558 - 1603)). Volere (to want), intendere (to intend), and sapere (to know) are likely to co-occur with the term queen (see Figure 12), but the strength of association between regina (queen) and these terms is still low (see Figure 16).

### Information circulation

In May 1642, it was reported from London that the English king, Charles I (1625 - 1649), received a letter from the governors of Ireland who thanked him for his interest in Irish affairs (ASF MdP, 4201). This story exemplifies how court officials and sovereigns used news to keep track of not only geopolitical developments but also information circulation. News sheets often report the arrival of important letters and messengers; occasionally they inform readers about letters and messengers sent abroad. The importance of the topic I call information circulation is signaled by a number of key terms: mandare (to send), rispondere (to answer), lettera (letter), and scrivere (to write). Furthermore, news often discusses public opinion and information circulating through word of mouth, which is signaled by the term dire (to say). Information circulation is a theme that is strongly associated with men. For instance, mandare (to send), inviare (to send), and spedire (to send) are all associated with men.

### War

Figure 8 highlights another key theme of early modern news: war. I did not find a strong relationship between this topic and women. My findings suggest that news focused on the impacts of war on the people (gente in Italian) in general. There is, however, one aspect of war that is particularly related to women of humble origins (donna). This is violence, which is discussed in the following section.

### Catastrophes, diseases, and epidemics

Natural catastrophes and diseases are another key topic frequently featuring in early modern news. This theme is similar to the topic of war. It severely affected women, but news generally discusses it in the context of people in general (gente) and does not specifically highlight women's perspectives.

### Public festivities and celebrations

Among the most recurrent topics of early modern news, we can find one that is specifically related to women and their experience. This is the topic of festivities and celebrations (see Figure 19). Public festivities were highly visible events that expressed the prestige of the locations where they took place. Women had important roles in these festivities; they, for instance, participated in public processions and parades. There is a specific term closely associated with festivities: bello (beautiful). In this context, beauty does not necessarily imply the physical beauty of women who participated; rather, it implies the splendor of the entire setting, such as the beauty of clothes and various decorative elements. This explains why bello (beautiful) is closely positioned next to vestire (to dress) and carrozza (wagon) in the semantic space of womanhood (see Figure 19).

### Governance and politics

News writers often gave notice about the political developments in a given location. This is, of course, a highly recurrent theme signaled by a number of key terms: ordine (order), impedire (to obstruct), regno (kingdom), affare (affair), governo (government), etc. My findings (see the absence of terms related to governance in Figure 19) suggest that this topic was largely unrelated to women.

### Religion

The last central theme that my analysis revealed is religion. Even though religion and women were deeply connected in the early modern period, I found that this connection in the handwritten news of the MIA-Euronews Corpus remained loose. Words related to religion do not typically occur in close proximity to terms denoting women (references to nuns and cloisters are rare). As Figure 19 shows, religious celebrations and church visits are sub-themes that establish connections between women and religion. One explanation for the loose connection between religion and women can be that religion is mainly discussed in the context of politics and wars-of-religion. Spirituality, which was the main theme connecting women and religion, is rarely discussed in the news, however.

## Results C: Non-central themes of early modern news

The exploratory analysis in the previous section uncovered the themes that formed the core of early modern handwritten news sheets. In addition to these, I also identified other, less frequent themes. Maria Magdalena and her contemporaries read about these themes, but less frequently than the core themes. My findings suggest that women were more often associated with these non-core themes.

### Marriage and wedding

Marriage was viewed as the cornerstone of social order in the early modern period. Along with the family, marriage was the fundamental social framework that shaped the lives of most women and men (;;). It was believed—both by Catholics and Protestants—to be a divine creation aimed at maintaining the harmony of social life. Marriage was particularly important for noble families since it contributed to social status and ensured continuity over generations. Surprisingly, I found that marriage and weddings are non-central themes of the MIA-Euronews Corpus. I provide a possible explanation for this in the Conclusion section.

It is not surprising that marriage and weddings are the themes most strongly associated with women in the MIA-Euronews Corpus. Beyond the topic of movement in physical space, marriage and weddings are the most important topics in which women are mentioned (see matrimonio (marriage) and nozze (wedding) in Figure 9). However, there are differences in terms of social status. While elite women are strongly associated with marriage, the same cannot be said for women of humble social origins: the term donna (woman), which usually denotes women of humble social origins, is not associated with marriage (see Figures 15 to 18).

A close reading of marriage news revealed an unexpected aspect of the strong associative relationship between women and marriage. In news reporting about marriages, women usually remain in the background; sometimes, the news does not even mention the name of the bride. For instance, in July 1574, a news sheet was sent from Milan to Florence. The news sheet (ASF MdP 3254) contains a notice of marriage between a certain Pietro Antonio Grasso, most probably a local nobleman, and an unnamed woman. As with other marriage reports, this one mainly focuses on the dignitaries who were present at the wedding. This explains why the term presenza (presence) is salient in the semantic space underlying women's terms (see Figure 19).

Marriage notices in the MIA-Euronews Corpus tend to be short and concise. They give the names of the husband and wife, as well as the day when the marriage was celebrated or published. For instance, in April 1698, a news sheet was sent from Hamburg to Florence (ASF MdP 4191a) and reported the marriage between Frederick IV (1695–1702), the reigning Duke of Holstein-Gottorp, and Hedvig Sophia Augusta of Sweden. The news is brief and to the point; it simply provides information about the marriage union without offering any details about the ceremony. Notably, marriage gifts and the amount of the dowry remain unmentioned. Some of the marriage news focus on marriage planning. For example, a news sheet from Parma, compiled in January 1576, reported that the Duke of Parma, Ottavio Farnese (1547–1586), traveled to Turin to discuss a marriage plan for his daughter. Again, the focus of the news is Ottavio’s travel and marriage plan; the name of the daughter is not even mentioned. Other news reports marriage contracts and negotiations of dowries. Again, this group of news highlights the material aspect of marriage alliances while future wives remain in the background.

Material aspects and dignitaries were the focal points of news about wedding ceremonies as well. In April 1600, the Austrian city of Graz hosted the wedding ceremony of Maria Anna of Bavaria with the future Emperor of the Holy Roman Empire, Ferdinand II (1619–1637). Through news sheets sent directly from Graz and Vienna, the Medici in Florence could follow the event. These news sheets discuss the arrival of foreign dignitaries, list the names of ambassadors present, and describe the lavishness of wedding banquets and processions. Other wedding news in the MIA-Euronews Corpus focuses on gifts presented at the ceremony. For instance, a news sheet (ASF MdP 4201) that arrived from London in May 1641 describes in great detail the marriage gifts by William II, Prince of Orange (1647–1650), to Maria Enrichetta Stuart, including the price of each gift. Little is actually said about the bride in marriage news; instead, material aspects are salient.

In summary, even though marriage is believed to have been the cornerstone of early modern social order, it is a marginal theme in handwritten news. Furthermore, the news typically discusses it through the lens of prestige and power. As a result, brides are not the focal points of marriage notices; they often remain in the background, treated as just one of many protagonists. Marriage and wedding descriptions in the MIA-Euronews Corpus tend to be silent about the feelings and aspirations of future wives, thus perpetuating the traditional view—questioned by recent scholarship ()—that early modern marriage was an emotionless bond primarily aimed at establishing political and social alliances between families.

### Pregnancy and childbirth

The reproductive role of women was highly important in early modern societies. For the elite, women as mothers assured the continuity of a dynasty or of a noble family. For artisanal and merchant families, children contributed to the completeness of the household that often drew on their labor. Both Protestant and Catholic theologians underlined the sanctity of the household, which would remain incomplete without children (). Despite the apparent importance of childbirth and pregnancy, neither of them is a significant theme in the MIA-Euronews Corpus. However, not surprisingly, both are strongly related to women (see figlio (son) and figlia (daughter) in Figures 10).

Generally speaking, in the MIA-Euronews Corpus, there are only occasional and sporadic mentions of pregnancy. What is also indicative is the fact that pregnancy does not occur as a primary news topic; it is usually discussed in tandem with other news. For instance, in April 1575, a piece of news arrived from Lyon through Rome (ASF MdP 3082); it extensively reported about fights with Huguenots and, as a side note, also mentioned the pregnancy of the Queen of Navarre, Marguerite de Valois-Bourbon. Sometimes, news in the MIA-Euronews Corpus comments on the development of the pregnancy, including the number of months a woman has completed (for instance, ASF MdP 4191a, ASF MdP 3082, ASF MdP 4572). Occasionally, they give an account of a miscarriage, although, as the MIA-Euronews Corpus suggests, this was a particularly rare theme.

Childbirth and baptism are equally unimportant themes in this dataset. When the news discusses them, it focuses on the celebratory aspects. By contrast, biological and emotional aspects of childbirth usually remain unaddressed. In the MIA-Euronews Corpus, we can also find examples of news about childbirth where the name of the mother is not even communicated. For instance, there is news about childbirth in a news sheet compiled in June 1643 in Florence (ASF MdP 384); it reports the birth of the second son of a local nobleman, Count Altoviti, mentioning the birth of the child and the name of the father but not the name of the mother. News about baptism is somewhat similar to news about marriage and weddings. The most salient aspect is the presence of dignitaries and their representatives. Hence, both childbirth and baptism are treated as prestige events, with women and children again remaining in the background.

### Crime and violence

Crime and violence are not particularly important topics in the MIA-Euronews Corpus. However, as Figure 19 shows, they are semantically related to terms denoting women. There is a remarkable dichotomy between women of the elite and women of humble origins. Crime and violence are closely associated with women of humble origins. In fact, if women of humble origins are mentioned in the news, they are quite likely to be discussed in the context of crime, violence, and social control (see the axes *attaccare* (attack) and *difesa* (defence) in Figures 11 and 15). The relationship between women and crime has two facets in the MIA-Euronews Corpus.

On the one hand, there are news articles about crimes committed by women. For instance, a news sheet that arrived from Rome in August 1719 (ASF MdP 100) reported about a woman who killed her husband. News sometimes reports on the punishment of women, including public whipping, execution, and imprisonment. Reports about crimes by women are not surprising. In the early modern belief system, women and transgressive social behavior were closely related (). The connection was in part based on religious ideas. According to the Christian tradition, women were daughters of Eve, which established a direct connection with the Original Sin and the Fall of Man. As a result, in the early modern period, it was widely believed that women were vulnerable to the incursions of the devil and to their own natural passions. This belief was also reinforced by premodern “scientific” ideas. According to the elemental theories of Galen, Aristotle, and other ancient Greek philosophers, the female body is incomplete and vulnerable to uncontrollable emotions and passions. While men were associated with rationality and self-control, women were linked to nature, passion, and savageness. Women were also often subject to social control, which, for instance, gave rise to the regulation of clothes that women were allowed to wear in public. News articles sometimes discuss this theme as well. In short, everyday news reinforced the traditional misogynist beliefs of early modern times by relating women of humble origins to crime and transgressive behavior.

On the other hand, news often also report about violence against women. Again, these reports usually concern women of humble origins. A recurrent theme is violence by soldiers. For instance, a news sheet from Graz compiled in October 1600 discusses the state of affairs in the war against the Ottomans (ASF MdP, 51248), reporting violence by French mercenaries against women. Another news sheet from Cremona, compiled in June 1575, highlights the cruelty of German and Italian soldiers (ASF MdP, 3254). This news sheet pinpoints the heart of the matter. In early modern times, peasants were forced to accommodate soldiers, who often took advantage of this situation and committed violence, crime, and looting. Sometimes, news also mentions petty crimes against women, though this remains sporadic in the MIA-Euronews Corpus.

There are some interesting and unexplained silences in the MIA-Euronews Corpus. First, witchcraft is surprisingly underreported. It is estimated that in the early modern period, between approximately 40,000 and 60,000 people, mainly women, were executed for witchcraft (). However, apart from some exceptions, the news in the MIA-Euronews Corpus is silent about witchcraft. Second, scholarship () has discovered a great variety of crimes that were usually ascribed to women in the early modern period. These include fornication, infanticide, adultery, and premarital sex. News is not particularly specific about the types of crimes women commit. Generally, physical violence is the main crime discussed, both against women and by women.

## Conclusion

The focus of scholarship on early modern women's history has been agency. Specifically, historians have studied how women “skirted or even reshaped the patriarchal structure of the day” (, p. 21). Numerous examples of women challenging the traditional male order have been unearthed and analyzed with the purpose of understanding how gender differences in the early modern period were constructed (; , ; ). Hence, I am evaluating the findings of my investigation in light of this question.

My findings suggest that handwritten news reinforced existing ideas about gender and gendered differences. First, handwritten news highlighted the male dominance of the period. We have seen that even though women were not completely absent from the news, their presence was rather sporadic. By contrast, men were almost always present and were the main protagonists of the news. As has been pointed out, no significant difference was found between different countries and regions. Second, when women were mentioned in the news, they were usually featured in traditional settings such as celebrations, marriages, travel, and childbirth. Similarly, women tended to assume traditional roles. As a result, I conclude that in handwritten news, women did not challenge the patriarchal order of their day; rather, they acted in accordance with it. Therefore, the news did not challenge the traditional gender differences of the period.

The reason news sheets conveyed traditional gender roles and did not fundamentally alter the existing construction of gendered differences likely lies in the medium itself. It is probable that news sheets were compiled by men, which significantly impacted the types of information that were included or ignored. Furthermore, the key topics of handwritten news uncovered in this essay indicate that these news sheets were primarily a political medium. They were intended to help court officials and sovereigns follow political events, including wars and interstate conflicts. This explains why themes related to women, such as childbirth and marriage, were relatively rare. In light of all this, I can conclude that the story of the woman ambassador presented at the beginning of this paper must have been an anomaly.

## Implementation

The methodology behind this research comprises three components and includes a pre-processing stage. Here, I will present these components with a special emphasis on how they address the issues of randomness and archival research discussed earlier. This section requires advanced knowledge in mathematics and computer science and is intended for specialists.

### Preprocessing

The MIA-Euronews Corpus, preserved in the MIA Database of the Medici Archive Project, consists of transcriptions of news sheets alongside basic metadata (place and date of compilation, source location of news items). As a first step, the transcriptions (in JSON format) were tokenized, lemmatized, and POS tagged using spaCy’s tools for modern Italian (<https://spacy.io/>). Since the news sheets reaching the Medici court were predominantly written in the Tuscan dialect, a precursor to modern Italian, the lemmatizer performed effectively and yielded satisfactory results. During preprocessing, names of persons were identified with a Named Entity Recognizer; their gender and social rank were inferred computationally under human expert supervision. The inference of social rank was facilitated by the presence of titles (e.g., marchese, princess, duke) preceding names in the MIA-Euronews Corpus. As these titles typically correspond to those in modern Italian, the Named Entity Recognizer functioned well, achieving 85% accuracy in tests on a small sample of data.

### Repeated random sampling of news sheets and news items

To accomplish repeated random sampling, the standard procedure of bootstrapping, or sampling with replacement, was applied (). Bootstrapping involves the repeated and randomized selection of a small subset from a large original sample. In this research, at each step of randomized selection, I picked one hundred complete news sheets and one hundred news items, then counted the number of news items and news sheets that mention at least one woman. Thus, I treated women's presence in the news as a binomial random variable with two possible outcomes: present or absent. By averaging the outcomes of the randomized selection steps, I calculated the mean number of news items and news sheets in which women are present or absent when one hundred news items and news sheets are randomly selected. In short, the process of repeated random sampling of news sheets and news items resulted in a binomial probability distribution representing women's presence or absence in the MIA-Euronews Corpus.

To discover the probability distribution that represents the presence of women in the entirety of news sheets that once circulated in the early modern period, I applied Bayes’ Theorem. This mathematical framework helps identify the most likely prior distribution that gives rise to the collected evidence from a series of independent trials. In this research, the evidence is the binomial probability distribution resulting from the bootstrapping procedure, and the prior is the probability distribution representing women’s presence in the original corpus of news sheets. We can formalize this as follows:

* Let X be the prior, represented by a continuous distribution with support [0, 1], effectively denoting the probability of presence and absence. For instance, if X = 0.2, then the probability of women’s presence is 0.2 and the probability of their absence is 0.8.
* Let H and T be the evidence: H is a random variable representing the number of news sheets (or news items) in which women are present, and T is a random variable representing the number of news sheets (or news items) in which women are absent.

We can formalize the relationship between the prior (X) and the evidence (H and T) as the probability (P) of observing presence *h* number of times and absence *t* number of times given that the prior probability distribution of presence is *x*.

We can also reverse the statement above:

This term expresses the probability of the prior given the evidence. It indicates how the prior probability distribution adjusts in light of the evidence (more on this later).

For convenience, we can express the prior probability distribution with its probability density function (PDF):

Similarly, we can express the second term as a PDF:

Bayes’ theorem connects the two terms in the following way ():

Given that the conjugate prior of the binomial distribution is the beta distribution, the PDF required is the beta PDF:

where

and

is the Gamma function.

In the context of this research, the two parameters of the beta PDF, and , represent the average number of times women are present and absent, i.e., and . To compute the distribution of women’s presence and absence in the totality of news sheets that once circulated, we need to accomplish two steps. First, we set the prior distribution so that the probabilities of absence and presence are equally likely:

When and both equal one, the beta distribution is actually the uniform distribution.

Next, we need to update this prior belief based on the evidence collected during the bootstrapping process as follows:

Notice that the update step in the case of the beta distribution is straightforward and involves only simple additions. Following the update step, we obtain the most likely prior distribution given the evidence:

### Markovian simulation

The Markovian simulation of texts is based on the state space model (). It also assumes that the Markovian property holds: any future state depends only on the current state. The so-called implied time scale test is used to validate the Markov model ().

The textual universe underlying the MIA-Euronews Corpus was therefore represented as a state space with individual words as discrete states. Next, the transition probabilities between words as states were computed from the corpus data and subsequently encoded in a transition matrix.

Let this transition matrix be *P*; as a corollary of the Chapman-Kolmogorov equation (), we can use *P* to simulate the development of the underlying Markovian system after *n* steps and with *m* as the starting step:

In practice, as the equation demonstrates, throughout the simulation, we need to raise P to the power of *n+m*. Given that a Markovian system is ergodic, the simulation will reach a stable probability distribution (π) that satisfies the following equation ():

The stationary probability distribution π is a non-negative row vector that sums to one. It is the left eigenvector of the transition matrix and remains unchanged when multiplied by the transition matrix. The stationary probability has several features important in the context of this research. First, it expresses the non-conditional probability of each state in a finite random system modeled as a Markovian system. The stationary probability of a given state is the maximum probability that the simulated system will be in that state, regardless of the starting distribution. Hence, the stationary probability distribution is a limiting distribution ():

Throughout this paper, I referred to the stationary probability of a given word as the "overall probability" of that word in the news system. Second, the stationary probability distribution expresses the significance of each state in a dynamic system represented with a Markovian framework (), showing how important a given word is in the MIA-Euronews Corpus.

The transition matrix can also be used to run random walks on the underlying Markovian system () and to simulate how states (words, in the context of this research) follow each other. First, one needs to pick a given starting state, ideally at random. Second, the row corresponding to this state in the transition matrix is treated as a probability distribution and used to randomly select the next state. Throughout the random walk, one must repeat the second step for a reasonably large number of times (or until the random walk reaches convergence). With the help of this random walk, I could observe how words are likely to follow each other. Specifically, I could uncover the words that are most likely to occur before or after terms denoting women.

At the same time, as explained earlier, the likelihood that word B follows or precedes word A is not necessarily informative about the strength of the relationship between them. To estimate the strength of association, one needs to consider the independent probabilities of A and B. For instance, the fact that donna (woman) is very likely to be followed by essere (to be) does not indicate a strong associative relationship, given that essere is a very common term. Again, I used Bayes' theorem and its terminology to describe and resolve this problem as follows:

1. Evidence: the probability (P) of B given A (the result of the random walk on the transition matrix)

P (B | A)

1. Priors: the probability of B independently of A (the stationary probability of B) and the probability of A independently of B (the stationary probability of A)

P (B) and P (A)

1. Posterior: probability of A given B (the expression of how much our knowledge about B and A is changing as result of the evidence)

P (A | B)

To calculate the posterior, I again used Bayes' theorem (see above). In short, the higher the posterior probability, the more closely related the two terms are to each other.

### Semantic modeling

To model the semantic space underlying the MIA-Euronews Corpus, I applied a standard word2vec model (). First, I represented the corpus with a TF-IDF model and then trained the word2vec model (). To identify terms semantically related to words denoting women, I used cosine similarity. I projected the semantic space into two dimensions using Uniform Manifold Approximation and Projection ().

## Data Availability

The data used for this research is freely and publicly available in the MIA database of the Medici Archive Project (<http://mia.medici.org/>).

## Funding

This research was funded by the Irish Research Council.

## Acknowledgements

The author is thankful to the researchers who completed the transcriptions of news sheets in the MIA-Euronews Corpus and made them publicly available in the MIA database. The author also thanks Alessio Assonitis, the director of the Medici Archive Project, and Brendan Dooley, the Principal Investigator of the Euronews Project, for facilitating this research. Special gratitude is extended to the research team of the Euronews Project for their contribution to the MIA-Euronews Corpus: Davide Boerio, Wouter Kreuze, Sara Mansutti, Carlotta Paltrinieri, and Lorenzo Allori. The author is especially thankful to Davide Boerio for bringing attention to the story of the woman ambassador.

## Bibliography