jnkqqTTKW8km

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

### anonym

[ ©. 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 include handwritten news sheets. Throughout the 16th and 17th century 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 once existing news sheets did not survive. Those news sheets that survived are dispersed in different collections of the world. These collections are just random and incomplete samples from the once existing corpus of handwritten news. 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 presence and the perception of women in the news. First, it discusses how loss and incompleteness generate the randomness of archival assets. Second, it offers an accessible introduction into 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 a 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 mongers surprised reading this news? Or was reading about a woman who acted decisively and boldly common for early modern news consumers?

Answering these questions is not easy. Early modern Europe witnessed an information revolution featured - in part - by the continual circulation of handwritten news sheets or avvisi in Italian (*Droste, Salmi-Niklander 2019*, *Davies, Fletcher 2014*,*De Vivo 2019*). 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*Briggs, Burke 2010*. 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 *Droste, Salmi-Niklander 2019*. Hence each news sheet is a consecutive sequence of news items that are organized into separate paragraphs. News agents, who could be postmasters, merchants or even diplomats and ambassadors residing abroad, sold their news sheets to a great variety of customers including ducal courts and merchant houses such as for instance the grand ducal court of the Medici in Florence and the Fuggers in Augsburg (*Infelise 2002*,*Bauer 2011*). 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. What types of information handwritten news sheets spread about women have not been studied, even though it is a relevant question for historians. The broader relevance is given by scholarship’s efforts to discover and present women’s representation in historical sources *Ross 2013*.

However, scholarship on early modern women history cannot answer the question, to which extent the story of the woman ambassador was a genuine novelty for contemporary readers. Recent scholarship focusing on early modern women history studied the representation of women in a great variety of sources, including letters, literary works, and court records. For instance, topics, social activities and roles associated with early modern women have been studied (*Filosa 2022*; *Brizio 2022*). Recent scholarship of women’s history has also challenged the traditional perception of early modern women as passive protagonists of history (*Wiesner-Hanks 2019*). Through the systematic study of historical records, historians have uncovered different areas of woman agency and pointed out that woman agency was not altogether exceptional. Women agency manifested in specific domains of knowledge and skills such as medicine and household management, and information gathering(*Benson 2010*;*Davies, Fletcher 2014*; *Davis, Davis 1995*;*Daybell 2011*). 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, students of women’s history have not surveyed handwritten news; unlike news pamphlets, avvisi have not been studied from the perspective of women’s history*Warfield 2013*. We do not therefore 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 lacuna.

The lack of studies investigating women’s representation in early modern handwritten news is due to the fact that these sources were not easy to research until not long ago. Most of the surviving news sheets remained unpublished and untranscribed in various archives and libraries of the world. Recently, thanks to the Euronews Project (https://euronewsproject.org) funded by the Irish Research Council and to the Medici Archive Project (https://www.medici.org/) funded by the Mellon Foundation, a large selection of transcribed Italian language handwritten news sheets (a corpus of approximately 1250 news sheets and 10800 individual news items, see the description of data set below) from the Florentine State Archive became publicly available on the MIA platform and database (https://mia.medici.org/, last accessed 1 February, 2024). In this essay I will investigate this selection and I will refer to it as the MIA-Euronews Corpus. Specifically, the main focus of my investigation will be women and their representation in early modern handwritten news sheets. The focus on representation is in line with most recent developments in media history discovering women representation in different modern media formats *Ross 2013*.

However, this paper also has another key focus: methodology and historiography of early modern handwritten news sheets. The MIA-Euronews Corpus is just 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 who 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 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 (*Tiihonen, Tolonen, Lahti 2021*;*Riddell, Betancourt 2021*;*Kestemont, Karsdorp, de Bruijn, Driscoll, Kapitan, Ó Macháin, Sawyer, Sleiderink, Chao 2022*). For instance, Mike Kestemont et al. have applied the so-called unseen species model to study the loss of medieval narratives (*Kestemont, Karsdorp, de Bruijn, Driscoll, Kapitan, Ó Macháin, Sawyer, Sleiderink, Chao 2022*). This paper aims to contribute to this novel area of study, though its aim is not to gauge the volume of once circulating early modern handwritten news sheets; it rather aims to gather insights into the content of these lost documents.

Due to the importance of the methodology underlying this paper, a significant amount of space will be devoted to 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 indepth 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 (*Howell 2019*; *Wiesner-Hanks 2020*). Previous studies addressed broader questions in the context of these themes. Here I address the same broader questions focusing on handwritten news. Did handwritten news perpetuate the essentially male order and dominance of early modern times? To which extent did news give space to women and women’s action aiming to challenge the male order? However, these questions need to be translated into more specific questions 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 implies further questions. Generally, who were the main characters in the news world of the early modern period? What was the social status of those women who occurred in the news? Second, if women appear in the news, in what context are they mentioned? In other words, what are those topics that are associated with women? Before answering this question, I will address the question, what are the main themes in the early modern news world? As a whole, answering these questions will help understand not only the presence and representation of women in early modern news; it will also offer fresh insights into the early modern handwritten news culture.

### Limitations of this study

The new 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). These 200 years mark a particularly tumultuous period 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 (30 year war, the war against the Ottoman Empire, etc). Despite the urgent need to study how the presence and the perception 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 new sheets that came from the seven most important news hubs in the MIA-Euronews Corpus: London, Naples, Rome, Milan, Venice, Vienna, Antwerp (these arbitrary temporal boundaries are explained by the unevennes of my dataset). By contrast, when studying the perception of women, I will treat the corpus as one monolithic whole without detailing regional differences; the diachronic and synchronic analysis of the women’s perception will be accomplished once a more balanced data will be available.

## Methodology

### The random walker

The story of the boldly acting woman ambassador is 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 (*Pavone 1983*); it incorporates the Medici family’s major archive from the early modern period. Today, some 6500 archival units, which occupy approximately one and half kilometers 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 in 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 and 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 continues to explore news sheets, he also finds that the type of information conveyed about women is in constant change. 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 Wicham, 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)

To which extent the randomness often experienced by historians working in archives is true or it is only due to the fact that no one has an overview of the complexity underlying millions of loosely organized documents is an open question. This paper views randomness as perceived and 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 posteriority’s inability to understand the manifold causes and patterns underlying millions of documents, is 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 an 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, 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 keeps walking around until he has seen all the news sheets. He carefully notes which ones mention women and which ones are silent on them. Even though 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. He notes how many of them mention at least one woman and 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 indicates the probability of finding women in a hundred randomly selected news sheets). The random walker keeps wandering around and repeating this task until he has seen all possible combinations of news sheets in the former Medici Archive. Since the number of possible combinations is extraordinarily large, his wandering will take almost forever.

However, by the end he has important insights into the presence of women in new sheets preserved in the Medici Archive. First, by uncovering the probability of finding women in a given random combination 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 were available today. Second, once his quest for all possible combinations exhausts, he knows about all possibilities: for instance, women mentioned in 24 news sheets out of 100 randomly selected news sheets, women mentioned in 50 news sheets out of 100 randomly selected news sheets, and so on. He will also know about the relative likeliness 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, and addresses 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. I therefore implemented the thought experiment algorithmically by letting the MIA-Euronews Corpus represent the Medici Archive. My algorithmic implementation is based on the statistical procedure of bootstrapping, also known as sampling with replacement (*Mooney, Mooney, Mooney, Duval, Duvall 1993*). 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 new 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. into the probability of finding women in the 100 randomly selected news sheets and news items. The algorithm repeated these tasks until it had seen all possible combinations of news sheets and news items (since the algorithm started to converge relatively quickly, there was actually no need to see all possible combinations; once the point of convergence is achieved, the algorithm did not give new information anymore). As a result, the algorithm produced an overview of all possibilities (again, for instance, women mentioned in 50 news sheets out of 100 randomly selected news sheets) and their respective probabilities. In mathematical terms, it resulted in a probability distribution that rendered how the randomness of women’s presence and absence behaved in the MIA-Euronews Corpus; the resulting probability distribution gave rise to a firm knowledge of all possibilities.

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

To answer this question, I worked with two core assumptions deriving from Bayesian statistics (*Lo 1986*). 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 - again with 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 hundred news sheets in the MIA-Euronews Corpus. Bayesian statistics offers an analytical framework to make inferences about the prior based on the evidence we have. The resulting knowledge about the prior is not an exact knowledge; it is instead an estimate or an informed guess. The following illustrative example makes this train of thought 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 a hundred times. The evidence we get by tossing the coin is determined by the prior probability setting. Determination means that the probability of getting heads will be around 0.3 and it cannot be 0.8 for instance. 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 (*Lo 1986*). As Figure 3B shows, the information we have about the prior probability setting is not a firm 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 renders this prior probability setting. This is an educated guess or an estimate of women’s presence in the once 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 but rather a range of possible outcomes (or a probability distribution).

### 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 somehow predictable and build a stable knowledge. If a news item mentions a woman, what are the most likely words co-occurring with it? Similarly, what are those words that are likely to precede or follow the mentioning of a woman?

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 words following each other works. Throughout the generation of texts, our random walker will at some point exhaust all possibilities and he will learn the probability of each possibility. 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, simulation of texts reaches a stable state (aka 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 throughout the simulation is a highly informative view point where sample bias is mitigated.

I implemented the procedure above on the MIA-Euronews Corpus by means of a mathematical model named Markov Chain (*Bini, Latouche, Meini 2005*). First, the model “learned” the probabilities of how words follow each other in the MIA-Euronews Corpus. Second, I run the Markov Chain model, which is a form of simulation sometimes named random walk in the literature (*Blum 2020*). In practice, when we run a Markov Chain modeling 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 exposes all possible ways words in a large textual collection may unfold; it thus enables us to observe how the randomness of words following each other in a given textual collection works.

The Markovian simulation brought about three pieces of information. First, it revealed the overall or the stable state probability of a given word in the MIA-Euronews Corpus; the overall probability of a word also expresses its significance in the corpus (for further details see the Implementation section). Hence, by calculating the overall probability of certain words referring to women and other historical actors, I could quantify the significance and importance of these words (see Figure 4 and Figure 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), though this estimate can be inaccurate in case of very infrequent terms. Hence I focused on salient and frequent associations. In short, the Markovian approach helped explore the possible words following or preceding terms denoting women, which in turn offered insights into women’s representation in the corpus. Finally, the Markovian simulation gave 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. As opposed to other traditional - static - approaches (for instance Pointwise Mutual Information) uncovering associative relationships between words, the advantage of the Markovian approach is its ability to model texts as dynamic systems, i.e. it lets us discover all possible ways information is discussed in a large textual collection. Most importantly, the Markovian simulation helps quantify the probability of these possible ways. The Markovian simulation also helps tackle the problem of bias; it enables us to go beyond an original non-representative sample by synthetically producing further possible sub-samples. Hence it helped get closer to the lost corpus of early modern handwritten news. Finally, the Markovian simulation of a textual collection shows the collection from a highly informative and less biased viewpoint: stable state (aka equilibrium state).

However, the Markovian approach did not produce an exhaustive set of information about the possible meanings attached to womanhood in early modern news. The heart of the matter is that meaning often remains tacit and the Markovian approach cannot handle tacitness. Furthermore, it is difficult to handle cases of polysemy and homonymy with the Markovian approach. I therefore modeled the meaning of womanhood by reconstructing the semantic space underlying the MIA-Euronews Corpus.

### Semantics

The semantic space approach (*Baroni, Lenci 2010*) is based on an old assumption: the meaning of a word lies in its relationship to other terms (*Wittgenstein 2010*; *Harris 1954*; *Firth 1949*). 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 is related to other words, which ultimately gives 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 isolated 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 on Figure 19 in Section 4.

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

To sum up, the semantic space approach reveals those possible connections that a word can have in a textual corpus. The possible connections include synonymity, 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 those groups of words that are 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 underline 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 letter offers 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 *Barker 2016*, pp. 726–727 )

The letter is referring 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 on 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 if we compare word pairs that have feminine and masculine forms. For instance, according to my estimation, the probability that contemporary news mongers read about sons was three times higher than the probability that they heard 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.

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 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 emphasize that 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).

Even though 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 women belonging to the elite (see Figure 4 and observe the small size of the bubble representing 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 uncovered that the presence of women did not change significantly over time (see Figure 6). Similarly, I could not find any significant variation when I compared the presence of women in news sheets that arrived from different news hubs

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

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

When consuming news, Maria Magdalena von Habsburg must have encountered all sorts of information about the world. However, there must have been certain themes that were particularly recurrent on the news sheets she read. Generally speaking, news mongers today or back in the early modern times tend to encounter similar types of information. Previous scholarship came up with a variety of themes that were allegedly prevalent in early modern news (*Ettinghausen 2016*):

* 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 produce information about how important these topics were. While some of them must have been particularly recurrent, others were less frequent themes. Similarly, the most important topics discussed in the context of women’s experience have also remained unknown. In this section, I will therefore reveal 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 those words that are strongly associated with women, as well as by the study of the semantic space underlying terms denoting women. Figure 19 renders this semantic space and shows those groups of words to which women terms are semantically related.

### 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 Boemia. The same news sheet also informed its readers about the leaving 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 new sheets in general. This is the movement of historical actors and goods in the geographical space. Early modern news extensively reported about the arrivals, the departures, and the returns of historical actors such as sovereigns, churchmen, and ambassadors. News about the arrival of messengers and ships with precious cargo were also recurrent. With road conditions dramatically improving and with courts frequently traveling between summer and winter residences, the early modern period was a world in movement (*Mączak 1995*). 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 ones of the most likely terms to co-occur with terms denoting women (see Figure 10). Of course, if social status is taken into consideration, there are important differences. While the news extensively discuss 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 movement in the geographical space and also strongly connected with terms denoting women. The first one is visitare (to visit). Early modern news often discuss the spatial movement of actors aiming to visit another actor or to visit special locations such as churches, palaces or events related to devotion (see the proximity of travel words to terms related to religion on Figure 19). In these visits women of the elite often had prominent roles, which is demonstrated (see Figures 16 to 18 and Figure 19) by the relatively strong associative relationship between women of the elite and the term visitare. The second one is accompagnare (to accompany). Women did not travel alone; they were often accompanied by their husbands or by their entourage. The news often discuss who accompanied women throughout their travels. The third one 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 were tools to monitor the geopolitical and political developments in the world (*Baron, Dooley 2005*). News mongers, such as the high ranking officials of the Medici court, as well as 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 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), dubitare (to doubt) belong to the topic intentions and beliefs of historical actors.

This topic is generally unrelated to women; it is associated with men. News are 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 on Figure 19). The only exception is queen (most probably referring to Elisabeth I. of England (1558 - 1603)). Volere (to want), intendere (to intent), 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 the king’s interest in affairs in Ireland (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 about the arrival of important letters and messengers; occasionally, they inform readers about letters and messengers sent abroad. The importance of this topic that I name information circulation is signaled by a number of key terms: mandare (to send), rispondere (to answer), lettera (letter), 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), spedire (to send) are all associated with men and unrelated to women (see Figures 9 & 10).

### War

Figure 8 brings forward 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 rather focused on the impacts of war on the people (gente in Italian) in general. There is however one aspect of wars 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 is 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 discuss it in the context of people in general (*gente*) and do 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 women experience. This is the topic of festivities and celebrations signaled by two key terms in the semantic space underlying womanhood (see Figure 19): bello (beautiful) and festa (feast). Both are associated with women and likely to occur in the context of women (see Figure 9 and 10 and Figure 19). Public festivities were highly visible events and they expressed the prestige of locations where they took place. Women had important roles in these festivities; they for instance participated in public processions and parades. Beauty does not necessarily imply the physical beauty of women who participated; it rather implies the splendor of the entire setting, i.e. beauty of clothes and of various decorative elements. This explains why bello (beautiful) is closely positioned to vestire (to dress) and carrozza (wagon) in the semantic space (see Figure 19). Not surprisingly, the theme festivities and celebrations feature stronger relationships to women of the elite than to women of humble origins (see Figures 10 to 14 and 15 to 18).

### Governance and politics

News writers often gave notice about the political development 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 on Figure 19) suggest that this topic was very much 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 thousands of handwritten news in the MIA-Euronews Corpus remained loose. Words related to religion do not typically occur in the close proximity of terms denoting women (references to nuns and cloisters are rare). As Figure 19 shows, religious celebrations and church visits are the 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 however rarely discussed in the news.

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

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

### Marriage and wedding

Marriage was viewed as the cornerstone of social order in the early modern period. In tandem with family, marriage was the most rudimentary social framework that shaped the life of most (but not all) women and men (*Dolan 1998*;*Flandrin 1979*;*Froide 2005*). It was believed - both by catholics and protestants - to be a divine creation that aimed to maintain the harmony of social life. Marriage was particularly important for noble families since it contributed to social status and assured continuity over generations. Surprisingly, I found that marriage and wedding are non-central themes of the MIA–Euronews Corpus. I give a possible explanation for this in the Conclusion section.

What is not surprising is the fact that marriage, including wedding, is the theme that is most strongly associated with women in my newsheet corpus. Beyond the topic of movement in physical space, marriage and wedding are the most important topics in the context in which women occur (see Figure 9 and Figure 12). There are however differences in terms of social status. While women of the elite are strongly associated with marriage, the same cannot be said about women of humble social origins. The term donna (woman), which usually denotes women of humble social origins, is in fact not associated with marriage (see Figures 15 to 18).

The close reading of news reporting about marriages 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 do 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. Just like other marriage news, this report mainly focuses on those dignitaries who were present at the wedding. This explains why the term presenza (presence) is salient in the semantic space underlying women terms (see Figure 19).

Generally, marriage notices in the MIA-Euronews Corpus tend to be short and concise. They give the name of the husband and the wife, as well as the day when the marriage was celebrated, consumed or published. For instance, in April 1698, a news sheet arrived 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 gives 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 instance, 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 report about marriage contracts and negotiations of dowries. Again, this group of news highlights the material aspect of marriage alliances and future wives remain in the background.

Material aspects and the presence of dignitaries were the focal points of news about wedding ceremonies as well. In April 1600 the Austrian city, 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, give the names of ambassadors who were present, and describe the lavishness of wedding banquets and processions. Other wedding news in MIA-Euronews Corpus focus on gifts presented at the ceremony. For instance, a news sheet (ASF MdP 4201) that arrived from London in May, 1641 describes the marriage gifts by William II, Prince of Orange (1647 - 1650) to Maria Enrichetta Stuart in great detail, 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 keystone of early modern social order, it is a marginal theme of handwritten news. Furthermore, the news typically discuss it through the lens of prestige and power. As a result, brides are not the focal points of marriage notices, yet they often remain in the background and treated as only one of the many protagonists. Marriage and wedding descriptions in the MIA-Euronews Corpus tend to be silent about the feelings and aspirations of future wives; they thus perpetuate the traditional view, questioned by recent scholarship (Davis & Davis, 1995), that early modern marriage was an emotionless bond that primarily aimed to establish 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 and children as heirs 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 that would remain incomplete without children (*Hardwick 2010*). Despite the apparent importance of childbirth and pregnancy, neither of them are significant themes in the MIA-Euronews Corpus. On the other hand, they are both strongly associated with women (see Figures 9 & 10, 11 to 14, and 15 to 18).

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 news alone; 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 Hugenottes and as a side note it also mentioned the pregnancy of Queen of Navarra, Marguerite de Valois-Bourbon. Sometimes news in the MIA-Euronews Corpus comment 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 account of a miscarriage, though as the MIA-Euronews Corpus suggests this was a particularly rare theme.

Childbirth and baptism are equally unimportant themes in my dataset. When the news discuss them, they focus on the celebrative aspects. By contrast, biological and emotional aspects of childbirth usually remain unaddressed. In the MIA-Euronews Corpus we can also find examples of childbirth news where the name of the mother is not even communicated. For instance, there is a childbirth news in a news sheet compiled in June 1643 in Florence (ASF MdP 384); it reports about the birth of the second son of a local nobleman (Count Altoviti); it mentions the birth of the child and name of the father but not the name of the mother. News about baptism are 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. But 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 (defense) on Figure 11 and 15). The relationship between women and crime has two facets in the MIA-Euronews Corpus.

On the one hand, there are pieces of news 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 report about 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 (*Rublack 2001*). 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 physics of Gallen, 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; this for instance gave rise to the regulation of clothes that women were allowed to wear in public. News sometimes discuss this theme as well. In short, everyday news reinforced the traditional misogynist beliefs of the early modern times by relating women of humble origins to crime.

On the other hand, news also often 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 extensively discusses the actual state of affairs in the war against the Ottomans (ASF MdP, 51248). It reports about violence of French mercenaries against women. Another news from Cremona, compiled in June 1575, highlights the cruelties of German and Italian soldiers (ASF MdP, 3254). This piece of news pinpoints the heart of the matter. In early modern times peasants were forced to give accommodation to soldiers who often took advantage of this and committed violence, crime, and looting. Sometimes, news also mention 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 approximately 40.000 and 60.000 people, mainly women, were executed for witchcraft *Barry, Hester, Roberts 1998*. However, apart from some exceptions, the news in the MIA-Euronews Corpusare are silent about witchcraft. Second, scholarship *Walker 2003* has discovered a great variety of crimes that were usually ascribed to women in the early modern period. This includes for instance fornication, infanticide, adultery, and premarital sex. News are not particularly specific about the types of crimes women commit. Generally, physical violence is the main crime against and by women that is discussed.

## Conclusion

The focus of scholarship on early modern women history has been agency. Precisely, historians have studied the way women “skirted or even reshaped the patriarchal structure of the day. (*Howell 2019*, p. 21)” Numerous examples of women challenging the traditional male order have been unearthed and studied with the purpose of understanding how gendered differences in the early modern period were constructed (*Whyte 2015*; *Harline 2013*,*Montenach 2013*; *Müller 2013*). 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 news, their presence was rather sporadic. By contrast, men were almost always present; yet they were the main protagonists of the news. As it has been pointed out, no significant difference was found between different countries and regions. Second, if women occur in news, they are usually mentioned in traditional settings: celebration, marriage, travel, childbirth, etc. Similarly, women tend to assume traditional roles. As a result, I conclude that in handwritten news women do not challenge the patriarchal order of their day; yet they act according to this order. News therefore does not challenge the traditional gender differences of the period.

The reason why news sheets conveyed the traditional gender roles and did not fundamentally alter the existing construction of gendered differences must be sought in the medium itself. Most probably, news sheets were compiled by men, which must have had a significant impact on the types of information that were included or ignored. Furthermore, the key topics of handwritten news uncovered in this essay pinpoint the fact that handwritten news sheets were mainly a political medium. News sheets were meant to help court officials and sovereigns follow political events, including wars and interstate conflicts. This explains why themes related to women, such as for instance child birth 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 consists of three different components and includes a preprocessing stage. Here I will present these three components with a special emphasis on how they resolve the problems outlined in the section addressing randomness and archival research. This section requires advanced knowledge in mathematics and computer science, and it is for specialists.

### Preprocessing

The MIA-Euronews Corpus is preserved in the MIA Database of the Medici Archive Project; it consists of transcriptions of news sheets and basic metadata (place and date of compilation, source location of news items). As a first step, the transcriptions (represented in JSON) were tokenized, lemmatized, and POS tagged. For this task a lemmatizer and POS tagger for modern Italian by spaCy (<https://spacy.io/>) was used. Given that the news sheets that reached the Medici court were mainly written in the Tuscan dialect, which is the predecessor of modern Italian, the lemmatizer performed reasonably well and delivered acceptable results. Throughout the preprocessing names of persons were identified with a Named Entity Recognizer; their gender and social rank were inferred computationally with the supervision of a human expert. What made the computational inference of social rank possible is that names are usually preceded by titles (such as marchese, princess, duke, etc.) in the MIA-Euronews Corpus. Given that titles in the corpus usually correspond to titles in modern Italian, the Named Entity Recognizer worked well. It was tested on a small sample of data and achieved good results (85% accuracy).

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

To accomplish repeated random sampling, the standard procedure of bootstrapping, also known as sampling with replacement, was applied *Mooney, Mooney, Mooney, Duval, Duvall 1993*. Bootstrapping involves the repeated and randomized selection of a small subset of a large original sample. In this research, throughout each step of randomized selection, I picked one hundred complete news sheets and one hundred news items; then I counted the number of news items and news sheets that mention at least one woman. Hence I treated women’s presence in news as a binomial random variable with two possible outcomes: present and absent. Next, 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 if a hundred news items and new sheets are randomly selected. In short, the process of repeated random sampling of news sheets and news items brought about a binomial probability distribution that represents women’s presence or absence in the MIA-Euronews Corpus.

To discover the probability distribution that represents the presence of women in the lost totality of news sheets that once circulated in the early modern period, I applied Bayes’ Theorem *Puga, Krzywinski, Altman 2015*. This is a mathematical framework that facilitates the discovery of the most likely prior distribution giving rise to a set of evidence gathered throughout a series of independent trials. In the context of this research, the evidence is the binomial probability distribution that the bootstrapping procedure brought about and the prior is the probability distribution that represents women’s presence in the lost totality of once existing news sheets. We can formalize all this in the following way:

* Let X be the prior and let X be a continuous distribution with support [0,1] so that it can effectively represent 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; let H be a random variable for the number of news sheets (or news items) in which women are present, and let T be a random variable for 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. Qualitatively, it tells how the prior probability distribution is changing in light of the evidence we have (more about 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 (Puga et al., 2015):

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

where

and

is the Gamma function.

In the context of this research, the two parameters of the beta PDF, and are the average number of times women are present and absent, i.e. *H* and *T*. Precisely, to compute the distribution of women’s presence and absence in the lost whole of news sheets that once circularted, we need to accomplish two steps. First, we set the prior distribution so that the probability of absence and presence are equally likely:

Next, we need to update this prior belief in light of the evidence we collect throughout the bootstrapping as follows:

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

### Markovian simulation

The Markovian simulation of texts is based on the state space model *Hinrichsen, Pritchard 2011*. It is also based on the assumption that the Markovian property holds: any future state depends on the current state only. The so-called implied time scale test is used to validate the Markov model*Bowman, Pande, Noé 2013*. 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 *Encyclopedia of Statistical Sciences 2006*, 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 on 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 *Papoulis 1965*:

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 it remains unchanged if multiplied by the transition matrix. The stationary probability has a number of features that are important in the context of this research. First, it expresses the non-conditional probability of each state of 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 irrespectively of the starting distribution. Hence, the stationary probability distribution is a limiting distribution as follows *Papoulis 1965*:

Throughout this paper I therefore described it as the ‘overall probability’ of a given 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 *Meyn, Tweedie 2009*; it thus shows how important a given word is in the MIA-Euronews Corpus.

The transition matrix can be also used to run random walks on the underlying Markovian system *Chung 2012* and to simulate how states (words in the context of this research) are following each other. First, one needs to pick (ideally randomly) a given starting state. Second, the row corresponding to this state in the transition matrix is treated as a probability distribution and it is used to randomly select the next state. Throughout the random walk one has to 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. Precisely, 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 likeliness that word B follows or precedes word A is not necessarily informative about the strength of relationship between them. To estimate the strength of association, one needs to take into consideration the independent probability 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 drew on Bayes theorem (see above). In short, the higher the posterior probability, the more the two terms are related to each other.

### Semantic modeling

To model the semantic space underlying the MIA-Euronews Corpus, I applied a standard word2vector model *Mikolov, Karafiát, Burget, Černocký, Khudanpur 2010*. First, I represented the corpus with a TF-IDF model and then trained the word2vector model (*Manning, Raghavan, Schütze 2008*). To identify terms semantically related to words denoting women, I used cosine similarity. I projected the semantic space into two dimensions with the help of Uniform Manifold Approximation and Projection *McInnes, Healy, Saul, Großberger 2018*.

## 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 accomplished the transcriptions of news sheets in the MIA-Euronews Corpus and made them publicly available in the MIA database. The author is also thankful to Alessio Assionitis, the director of the Medici Archive Project, and to Brendan Dooley, the Principal Investigator of the Euronews Project, for facilitating this research. The author is especially grateful to the research team of the Euronews Project for their contribution to the MIA-Euronews Corpus: Davie Boerio, Wouter Kreuze, Sara Mansutti, Carlotta Paltrinieri, and Lorenzo Allori. The author is specially thankful to Davide Boerio for calling his attention on the story of woman ambassador.

## Bibliography

Barker, S.. “ ‘secret and Uncertain’: A History of Avvisi at the Court of the Medici Grand Dukes”. News Networks in Early Modern Europe, edited by Joed Raymond and Noah Moxham, BRILL, pp. 716–38, http://dx.doi.org/10.1163/9789004277199\_032.

Baroni, M.and A. Lenci. “Distributional Memory: A General Framework for Corpus-based Semantics”. Computational Linguistics, vol. 36, no. 4, pp. 673–721, https://doi.org/10.1162/coli\_a\_00016.

Baron, S. A.and B. Dooley. The Politics of Information in Early Modern Europe. Routledge.

Barry, J., et al.. Witchcraft in Early Modern Europe: Studies in Culture and Belief. Cambridge University Press.

Bauer, O.. Zeitungen Vor Der Zeitung: Die Fuggerzeitungen (1568-1605) Und Das Frühmoderne Nachrichtensystem. Oldenbourg Verlag.

Benson, P. J.. Invention of the Renaissance Woman: The Challenge of Female Independence in the Literature and Thought of Italy and England. Penn State Press.

Bini, D. A., et al.. “INTRODUCTION TO MARKOV CHAINS”. Numerical Methods for Structured Markov Chains, Oxford University PressOxford, pp. 3–2, http://dx.doi.org/10.1093/acprof:oso/9780198527688.003.0001.

Blum, A.. “Random Walks and Markov Chains”. Foundations of Data Science, Cambridge University Press, pp. 62–108, http://dx.doi.org/10.1017/9781108755528.004.

Bowman, G. R., et al.. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Springer Science & Business Media.

Briggs, A.and P. Burke. A Social History of the Media: From Gutenberg to the Internet. 3. ed. reprinted, Polity Press.

Brizio, E.. “Family Affairs: How a Woman Cared for Her Casa and Its Honour in Cinquecento Siena”. Women’s Agency and Self-fashioning in Early Modern Tuscany: (1300-1600), edited by Simona Lorenzini and Deborah Pellegrino, Viella Libreria Editrice.

“Chapman–kolmogorov Equation”. “Chapman–kolmogorov Equation”. Encyclopedia of Statistical Sciences, John Wiley & Sons, Ltd, https://onlinelibrary.wiley.com/doi/abs/10.1002/0471667196.ess0225.pub2.

Chung, K. L.. Markov Chains: With Stationary Transition Probabilities. Springer Science & Business Media.

Davies, S.and P. Fletcher. News in Early Modern Europe: Currents and Connections. BRILL.

Davis, H. C. L. P. of H. E. N. Z.and N. Z. Davis. Women on the Margins: Three Seventeenth-century Lives. Harvard University Press.

Daybell, J.. “Gender, Politics and Diplomacy: Women, News and Intelligence Networks in Elizabethan England”. Diplomacy and Early Modern Culture, edited by Robyn Adams and Rosanna Cox, Palgrave Macmillan UK, pp. 101–19, http://link.springer.com/10.1057/9780230298125\_7.

De Vivo, F.. “Microhistories of Long-distance Information: Space, Movement and Agency in the Early Modern News\*”. Past & Present, vol. 242, no. Supplement\_14, pp. 179–214, https://doi.org/10.1093/pastj/gtz042.

Dolan, C.. Le Notaire, La Famille Et La Ville: Aix-en-provence À La Fin Du Xvie Siècle. Presses Univ. du Mirail.

Droste, H.and K. Salmi-Niklander. Handwritten Newspapers: An Alternative Medium During the Early Modern and Modern Periods. BoD - Books on Demand.

Ettinghausen, H.. “International Relations: Spanish, Italian, French, English and German Printed Single Event Newsletters Prior to Renaudot’s Gazette”. News Networks in Early Modern Europe, BRILL, pp. 261–79, http://dx.doi.org/10.1163/9789004277199\_012.

Filosa, E.. “Lapa Acciaiuoli: A Businesswoman”. Women’s Agency and Self-fashioning in Early Modern Tuscany: (1300-1600), edited by Simona Lorenzini and Deborah Pellegrino, Viella Libreria Editrice.

Firenze, A. di S. di .. Archivio Mediceo Del Principato: Inventario Sommario.

Firth, J. R.. “The Semantics of Linguistic Science”. Lingua, vol. 1, pp. 393–404, https://doi.org/10.1016/0024-3841(49)90085-6.

Flandrin, J. L.. Families in Former Times. Cambridge University Press.

Froide, A. M.. Never Married: Singlewomen in Early Modern England. Oxford University Press.

Hardwick, J.. Practice of Patriarchy: Gender and the Politics of Household Authority in Early Modern France. Penn State Press.

Harline, C.. “Big Sister as Intermediary: How Maria Rolandus Tried to Win Back Her Way- Ward Brother”. Attending to Early Modern Women: Conflict and Concord, University of Delaware, pp. 3–3.

Harris, Z. S.. Distributional Structure.

Hinrichsen, D.and A. J. Pritchard. Mathematical Systems Theory I: Modelling, State Space Analysis, Stability and Robustness. Springer Science & Business Media.

Howell, M.. “The Problem of Women’s Agency in Late Medieval and Early Modern Europe”. Women and Gender in the Early Modern Low Countries, 1500 - 1750, edited by Sarah Joan Moran and Amanda C. Pipkin, Brill, pp. 21–30.

Infelise, M.. Prima Dei Giornali: Alle Origini Della Pubblica Informazione, Secoli XVI E XVII.

Kestemont, M., et al.. “Forgotten Books: The Application of Unseen Species Models to the Survival of Culture”. Science, vol. 375, no. 6582, pp. 765–69, https://doi.org/10.1126/science.abl7655.

Kleinberg, S. J.. Retrieving Women’s History: Changing Perceptions of the Role of Women in Politics and Society.

Lo, A. Y.. “Bayesian Statistical Inference for Sampling a Finite Population”. The Annals of Statistics, vol. 14, no. 3, https://doi.org/10.1214/aos/1176350061.

Manning, C. D., et al.. “Scoring, Term Weighting, and the Vector Space Model”. Introduction to Information Retrieval, Cambridge University Press, pp. 100–23.

McInnes, L., et al.. “UMAP: Uniform Manifold Approximation and Projection”. Journal of Open Source Software, vol. 3, no. 29, p. 861, https://doi.org/10.21105/joss.00861.

Mączak, A.. Travel in Early Modern Europe. John Wiley & Sons.

Meyn, S.and R. L. Tweedie. Markov Chains and Stochastic Stability. Cambridge University Press.

Mikolov, T., et al.. “Recurrent Neural Network Based Language Model”. Interspeech 2010, ISCA, http://dx.doi.org/10.21437/interspeech.2010-343.

Müller, M.. “Peasant Women, Agency and Status in Mid-thirteenth to Late Fourteenth- Century England: Some Reconsiderations”. Married Women and the Law: Coverture in England and the Common Law World, edited by Tim Stretton and Krista J. Kesselring, McGill-Queen’s Press - MQUP, pp. 91–113.

Montenach, A.. “Legal Trade and Black Markets: Food Trades in Lyon in the Late Seven- Teenth and Early Eighteenth Century”. Female Agency in the Urban Economy: Gender in European Towns, 1640-1830, edited by Deborah Simonton and Anne Montenach, Routledge, pp. 17–35.

Mooney, C. Z., et al.. Bootstrapping: A Nonparametric Approach to Statistical Inference. SAGE.

Papoulis, A.. Probability, Random Variables, and Stochastic Processes: Solutions to the Problems in Probability, Random Variables and Stochastic Processes.

Pavone, C., editor. Guida Generale Degli Archivi Di Stato Italiani. Edited by C. Pavone, Ministero per i beni culturali e ambientali, Ufficio centrale per i beni archivistici.

Puga, J. L., et al.. “Bayes’ Theorem”. Nature Methods, vol. 12, no. 4, pp. 277–78, https://doi.org/10.1038/nmeth.3335.

Riddell, A.and M. Betancourt. “Reassembling the English Novel, 1789–1919”. Journal of Cultural Analytics, vol. 6, no. 1, https://doi.org/10.22148/001c.19102.

Ross, K.. “Gender and Media A Very Short History”. The Handbook of Communication History, Routledge.

Rublack, U.. The Crimes of Women in Early Modern Germany. Oxford University Press on Demand.

Tiihonen, I., et al.. “Probabilistic Analysis of Early Modern British Book Prices”. Proceedings of the Conference on Computational Humanities Research, https://ceur-ws.org/Vol-2989/short\_paper9.pdf.

Walker, G.. Crime, Gender and Social Order in Early Modern England. Cambridge University Press.

Warfield, A.. The Media Representation Fo the Crime of Witchcraft in Early Modern Germany: An Examination of Non-periodical News-sheets and Pamphlets, 1533-1669. National University of Ireland, Maynooth, https://books.google.lu/books?id=8rLknQAACAAJ.

Weduwen, A. der .and A. Pettegree. The Library: A Fragile History. Profile Books.

Whyte, N.. ““ ‘with a Sword Drawne in Her Hande’: Defending the Boundaries of Household Space in Seventeenth-century Wales””. Women, Agency and the Law, 1300–1700, edited by Bronach Kane and Fiona Williamson, Routledge, pp. 141–57.

Wiesner-Hanks, M. E.. Women and Gender in Early Modern Europe. Cambridge University Press.

Wiesner-Hanks, M.. Challenging Women’s Agency and Activism in Early Modernity. Amsterdam University Press.

Wittgenstein, L.. Philosophical Investigations. John Wiley & Sons.