**Lausanne 2024 Abstract Submission**

**Title**  
Evolution in the Wine Industry: A Patent Mining Approach

**I want to submit an abstract for:**  
Conference Presentation

**Corresponding Author**  
Francesca Chiaradia

**E-Mail Corresponding Author**  
fc73@st-andrews.ac.uk

**Affiliation**  
University of St Andrews

**Keywords**  
Wine Patents, Patent Mining, Innovation, Evolution

**Research Question**  
How has patenting in the wine industry evolved between 1893 and 2023?  
Why many of these patents are currently invalid?

**Methods**  
I use a generative probabilistic model coupled with a Graph Neural Network algorithm to extract relevant information from patent titles and abstracts and perform a classification task.

**Results**  
Preliminary evidence indicates that patent filings can offer powerful tools to analyze past trends and read current practices as a byproduct of technological evolution.

**Abstract**  
Patents in the wine industry have enabled the diffusion and uptake of new technologies, with persisting long-run effects on costs, profits, and competition (WIPO 2023). In the short-run, however, bureaucratic and administrative inefficiencies in the patenting system, coupled with questions around the feasibility and usefulness of innovations, may have important welfare implications for both right owners and potential patentees (Williams 2017). In this light, understanding historical patterns of patenting activity can go a long way towards explaining how the wine industry has evolved and predicting how it may change in the immediate future.

In this paper, I use the universe of wine patents filed between the end of the nineteenth century and the present day to assess how viniculture and oenological practices have developed over the years. This exercise is policy-relevant for two reasons. First, examining the evolution of patenting provides evidence of how winemakers have responded to regulatory pressures from national and international governments. As the size of wine patents registered worldwide continues to rise (EPO, EU IPO 2022), this research reveals how patent policy can foster or hamper the emergence of new subject areas. Second, it shows how inventions and innovations shape and are shaped by changing consumers' tastes.

The research strategy proceeds in two steps. First, I retrieve detailed, public data from PatentLens, a comprehensive database provided by Lens.org. As the initial dataset consists of structured and unstructured data, I exploit recent advances in patent mining techniques to categorize relevant keywords and synthesize past
technological trends. In particular, I use a generative probabilistic model to extract relevant semantics and obtain a topic-probability matrix. Next, I employ a Graph Neural Network (GNN) algorithm to provide a visual illustration of node embeddings, formulating predictions for future technological advances.

Since nearly 63% of wine patents are no longer enforceable, I then turn to analyze the legal status of existing filings. There is concrete evidence from the literature that incentives for patenting highly depend on two policy mechanisms, the quality of the examination process and clear rules for patentability (Hall 2007; Picard and van Pottelsbergh de la Poterie 2013). In the spirit of Torrisi et al. (2016), I provide empirical evidence that invalid wine patents differ from valid ones due to a general lack of precision in the description of novelty. This perspective underscores the importance of strategic patenting among applicants and, more generally, a need for expert legal support at the filing stage.

Preliminary evidence indicates that patents have become an increasingly valuable mean of protection for the wine industry. To induce wine operators to protect innovations with technical and economic potential, however, it is important for patent policy to develop a robust system of legal statistics that discourages the registration of invalid patents and enhances consumer protection. These aspects are crucial for the harmonization of examination procedures and enhancement of the international patent system.

References


Evolution in the Wine Industry: A Patent Mining Approach

Francesca Chiaradia
University of St Andrews

1 Introduction

Patents in the wine industry have enabled the diffusion and uptake of new technologies, with persisting long-run effects on costs, profits, and competition (WIPO 2023b). In the short-run, however, bureaucratic and administrative inefficiencies in the patenting system, coupled with questions around the feasibility and usefulness of innovations, may have important welfare implications for both right owners and potential patentees (Williams 2017). In this light, understanding historical patterns of patenting activity can go a long way towards explaining how the wine industry has evolved and predicting how it may change in the immediate future.

In this paper, I use the universe of wine patents filed between the end of the nineteenth century and the present day to assess how viniculture and oenological practices have developed over the years. This exercise is policy-relevant for two reasons. First, examining the evolution of patenting provides evidence of how winemakers have responded to regulatory pressures from national and international governments. As the size of wine patents registered worldwide continues to rise (EPO, EUIPO 2022), this research reveals how patent policy can foster or hamper the emergence of new subject areas. Second, it shows how inventions and innovations shape and are shaped by changing consumers’ tastes.

The research strategy proceeds in two steps. First, I retrieve detailed, public data from PatentLens, a comprehensive database provided by Lens.org. As the initial dataset consists of structured and unstructured data, I exploit recent advances in patent mining techniques to categorise relevant keywords and synthesise past technological trends. In particular, I use a generative probabilistic model to extract relevant semantics and obtain a topic-probability matrix. Next, I employ a Graph Neural Network (GNN) algorithm to provide a visual illustration of node embeddings, formulating predictions for future technological advances. Since nearly 63% of wine patents are no longer enforceable, I then turn to analyse the legal status of existing filings. There is concrete evidence from the literature that incentives for patenting highly depend on two policy mechanisms, the quality of the examination process and clear rules for patentability (Hall 2007; Picard and van Pottelsberghe de la Potterie 2013). In the spirit of Torrisi et al. (2016), I provide empirical evidence that invalid wine patents are no longer enforceable.

---

1 From the early days of patenting, the concepts of invention and innovation were categorised separately. In this study, I will use them interchangeably. See Adams (2019) for more evidence on this difference.

2 Unenforceable rights might result from discontinued, expired, or inactive patents. See the Appendix section for an overview.
patents differ from valid ones due to a general lack of precision in the description of novelty. This perspective underscores the importance of strategic patenting among applicants and, more generally, a need for expert legal support at the filing stage.

Preliminary evidence indicates that patents have become an increasingly valuable mean of protection for the wine industry. To induce wine operators to protect innovations with technical and economic potential, however, it is important for patent policy to develop a robust system of legal statistics that discourages the registration of invalid patents and enhances consumer protection. These aspects are crucial for the harmonization of examination procedures and enhancement of the international patent system.

2 Background

Patents grant owners monopoly rights over the patented product, service, or process for a twenty-year period. Although a patenting system was introduced in the late fifteenth century, it was not until the US Patent Act of 1836 was signed into law that the modern history of wine patents began. This legislation was the first to establish a formal system of patent classification similar to the one in use today (USPTO 2020). In what follows, I provide an overview of key legal developments for wine patenting and a review of the relevant literature.

2.1 A Brief Primer

The history of wine patents is tightly linked to the development of a modern system of intellectual property (IP) protection. Until the beginning of the twentieth century, indeed, wine patents registered at a national office need not meet the criteria of patentability currently in force. Offices were purely deposit systems and registrars had little to no expertise to assess the validity of claims. During this period, wine patents were exclusively filed in three jurisdictions, namely the United Kingdom, the United States, and Canada. Contrary to the majority of states, these countries were all signatories to the 1883 Paris Convention and could boast modern systems of IP protection.

At the turn of the twentieth century, legal improvements to existing patent regulations favoured the filing of applications in other nations, including Australia, Spain, and Ireland. In sharp contrast to the surge of the first decade, however, the Interwar period considerably slowed the patent granting process and, in turn, the growth rate of applications in most countries, with the exception of the United Kingdom and Spain. Whether motivated by patent secrecy programs or the need to redeploy resources from the agricultural to the industrial sector, nations that initially contributed to the patent race remained silent until the early 1970s.

---

3 The conditions for patentability include novelty, inventiveness, and industrial application (Hart et al. 2019).
4 The 1883 Paris Convention is recognized as the first policy aimed at establishing and harmonizing patent systems worldwide. See WIPO (2023a) for a complete overview.
5 Examples include Australia, which recorded its last wine patent in 1911 and the subsequent one in 1973. Similarly, Canada registered the last patent in 1917 and the subsequent one in 1975.
While nowadays IP law remains tied to national jurisdictions, most patent systems have subscribed to the 1970 Patent Cooperation Treaty. Under this legislation, winemakers or their lawyers have to file a patent application consisting of a title, abstract, claims, a description and, if applicable, figures and citations of prior art. Coupled with structured data, this material forms the basis for the statistical output released by national offices, intergovernmental organizations, and NGOs. Although this information is a first step towards patent data standardisation, the sophistication of filings and absence of an international quality benchmark are major hurdles in the use and interpretation of the output. In turn, this legal complexity creates a problem of information disclosure, which can only be dealt with deep learning techniques.

This brief historical background sketches two relevant aspects for this study. First, differential adoption timings of formal specifications and examination systems suggest that this analysis may lose important information about wine innovations in laggard countries, that is regions in which patents laws were initially lax or nonexistent. Second, assessments of inventiveness are subject specific and inevitably tied to national jurisdictions. Both limitations call for a deeper exploration of the problem using text mining techniques and serve to motivate this research.

2.2 Literature Review

A large body of literature in economics has analysed the role of intellectual property in the wine industry (Meloni and Swinnen 2013; Saïdi et al. 2020). While most of the academic focus has been on geographical indications, other forms of IP protection have been largely ignored. To the best of my knowledge, this is the first paper to study the evolution of patenting in the wine industry.

From a methodological perspective, this study complements a recent strand of the literature that uses machine learning techniques to extract relevant economic information from textual data (see Gentzkow et al. (2019), Ash and Hansen (2023), and Korinek (2023) for excellent reviews). In wine economics, Yang et al. (2022) use a sample of the Wine Spectator’s reviews to assess whether the latter can accurately predict wine quality. After converting reviews into numeric vectors, they compare the performance of three feature extraction methods, finding that Doc2Vec models offer the highest accuracy level for the sample considered. While, similar to this literature, I use natural language processing for classification purposes, my contribution is to employ knowledge graph embeddings to facilitate information retrieval and association among findings.

Thematically, the paper closest to mine is Baiano et al. (2013). They offer the first qualitative survey of wine patents, with a specific focus on production and vinification processes. While we both leverage patent data to trace the evolution of winemaking over the years, I neatly distance myself by using advanced deep learning techniques. These methods not only enhance but also sharpen our comprehension of significant industry developments, providing a more nuanced perspective on the wine landscape.

Ultimately, while most studies on wine innovation are country-specific, looking at the universe of filed patents enables me to offer a broader outlook on novel practices across time and space.
3 Methodology

3.1 Sample and Preprocessing

This analysis uses detailed data queried from the Lens.org, an open-access repository consisting of three database packages, *Patents*, *PatCite*, and *PatSeq*. Unlike other registers, the Lens.org provides information about patents filed in 95 jurisdictions, complementing this with the scholarly and non-scholarly literature citing the filing. My sample ranges from 1893 to 2023, covering a period of 130 years. The final dataset yields information on 68424 wine patents. This data is unique because, until the time of writing, there is no other quantitative analysis on wine patents for such an extensive time period.

Table 1 summarises all the fields recorded. Variables of interest are titles, abstracts, jurisdictions and legal status, priority and publication dates, inventors’ and owners’ names, and classification codes. I provide further information about the sample in the Appendix.

<table>
<thead>
<tr>
<th>Patent Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Jurisdiction</td>
<td>Document Type</td>
</tr>
<tr>
<td>2 Kind</td>
<td>Has Full Text</td>
</tr>
<tr>
<td>3 Display Key</td>
<td>Cites Patent Count</td>
</tr>
<tr>
<td>4 Lens ID</td>
<td>Cited by Patent Count</td>
</tr>
<tr>
<td>5 Publication Date</td>
<td>Simple Family Size</td>
</tr>
<tr>
<td>6 Publication Year</td>
<td>Extended Family Size</td>
</tr>
<tr>
<td>7 Application Number</td>
<td>Sequence Count</td>
</tr>
<tr>
<td>8 Application Date</td>
<td>CPC Classifications</td>
</tr>
<tr>
<td>9 Priority Numbers</td>
<td>IPCR Classifications</td>
</tr>
<tr>
<td>10 Earliest Priority Date</td>
<td>US Classifications</td>
</tr>
<tr>
<td>11 Title</td>
<td>NPL Citation Count</td>
</tr>
<tr>
<td>12 Abstract</td>
<td>NPL Resolved Citation Count</td>
</tr>
<tr>
<td>13 Applicants</td>
<td>NPL Resolved Lens ID(s)</td>
</tr>
<tr>
<td>14 Inventors</td>
<td>NPL Resolved External ID(s)</td>
</tr>
<tr>
<td>15 Owners</td>
<td>NPL Citations</td>
</tr>
<tr>
<td>16 URL</td>
<td>Legal Status</td>
</tr>
</tbody>
</table>

Here, I restrict attention to the title and abstract of each case file. As customary in textual analysis (Ash and Hansen 2023), I preprocess data by lowering upper case words, removing trailing whitespaces, non-ascii characters, punctuation, numbers, and stopwords, splitting contractions, and tokenising the text. Next, I employ a latent Dirichlet allocation (LDA) method to reduce the complexity of filings and extract key labels from the patent.

---

6See https://about.lens.org/ for further reference.

7In this literature, relevant codes are the International Patenting Classification (IPC), established by the World Intellectual Property Organization (WIPO), and the Cooperative Patent Classification system (CPC), jointly run by the United States Trademark Office (USPTO) and the European Patent Office (EPO). Noteworthy is also the United States Classification system (USPC), which, however, will not be used in this analysis.
corpora. Figure 1 shows a wordcloud of the most frequent terms in the abstracts (a) and titles (b). Noteworthy is that, unlike patent classification codes, text algorithms provide a richer sense of context to model the evolution of the wine industry over the years.

Figure 1: Summary of Top Words.

(a) Top 50 Words (Abstracts). (b) Top 50 Words (Titles).

Source: Author’s Calculations.

3.2 Graph Neural Network

A Graph Neural Network (GNN) is a representation learning technique used to model the relationship between words in large textual corpora and produce labels for summarization and data visualization. The basic structure is a graph \( G = (V, E) \), which consists of vertices (e.g. key sectoral innovations \([V]\)) and edges (e.g. the interaction between patents \([E]\)). The graph is commonly defined as an adjacency matrix \( A \in \mathbb{R}^{V \times V} \), whose entries represent the edges between two vertices. Due to its positive performance, this method is being increasingly employed for node clustering and classification (Zhou et al. 2020).

In this analysis, I proceed in three steps. I begin by converting patent abstracts and titles into graph-of-words, whose vertices provide sectoral information and edges identify term co-occurrences. Next, I use graph convolution operators to convolve the word-graph and aggregate information from subgraphs in the neighbourhood of our vertices, \( V \). This, in turn, enables me to extract feature vectors from the GNN output for each element in the corpora. Subsequently, I use the elbow method to determine the optimal number of centroids \( k \) and apply a k-means algorithm to the aggregate feature set to group similar terms in the same cluster. Based on this initial classification layer, I then assign each element to one of the ten clusters based on its feature vector. Lastly, I compare the performance of the LDA and k-means algorithms to build evolutionary maps for patent landscaping. In Figure 2, I provide a schematic illustration of the pipeline just described.
4 Future Work

This preliminary setup opens new avenues for future work. First, while the global adoption of patent standards has led to increased patent applications even in countries where viticulture plays a lesser role in the economy, I ask: which sectors of the industry have benefited most from patenting? Which innovation waves have lasted the longest? Which nations have led the patenting race over the years? Additionally, there remain important questions about the legal status of invalid patents. In particular, is there evidence that patent legislation encourages the filing of invalid applications? Where is this phenomenon most common? Does the right continue to afford the owner the needed protection in all patent classes? These questions will guide my research in the coming months.

Overall, as the art of winemaking continues to evolve, preliminary evidence indicates that patent filings can offer powerful tools to analyze past trends and read current practices as a byproduct of technological evolution.
References


7


Appendix

General Overview

Figure A.1: Countries by Number of Patent Filed.

Notes: The figure lists countries by the number of patent filings published. China tops the list with 54,026 published filings, followed by the US (24,544), Korea (19,566), Australia (9,449), and France (9,061). Notice that the top five patent offices have registered over 88% of all publications.

Source: Author’s Calculations.
Figure A.2: Number of Published Patents per Year.

Notes: The figure shows the total count of published filings over the period 1893-2023. It should be noted that years 2018 (5299), 2021 (5296), 2022 (4971), 2017 (4877), and 2020 (4586) recorded the largest number of published patents.

Source: Author’s Calculations.
Legal Status

Figure B.1: Legal Status Information.

Notes: The figure contains information about the legal status of patents in the wine sample ($n = 68423$). More than 33% of all patents filed were discontinued. In other words, the initial application was either rejected, withdrawn, or refused on absolute or relative grounds. Nearly 28% of sampled patents is active. For around 16% of all filings, the patent has expired. This means that the right has reached the term date and is no longer enforceable. Nearly 14% of patents is inactive, which means that the right has not been renewed before the term expiry date. To 9% of the sample, the patent has not been attributed yet and the application is in either the filing, examination, or pre-grant stage. Only 53 patents have been registered at a regional office. Data on the legal status of the remaining observations is unknown.

Figure B.2: Legal Status by Top Jurisdictions.
Figure B.3: Analysis of IPCR Codes for Discontinued, Active, Expired, and Inactive Patents.

Notes: The figures shows that IPCR code C12G has the largest share of discontinued (61.4%), active (29.3%), and inactive (39.4%) patent applications. Code B65D is the most common among expired filings (28%).

Source: Author’s Calculations.
Notes: The figures shows that IPCR code C12G has the largest share of pending (49%) and patented (28.3%) patent applications. Code B65D is the most common among unknown filings (40%).

Source: Author’s Calculations.
Figure B.5: Analysis of CPC Codes for Discontinued, Active, Expired, and Inactive Patents.

Notes: The figures shows that CPC code C12G has the largest share of discontinued (37.5%), active (29.6%), expired (22.3%), and inactive (31.8%) patent applications.

Source: Author’s Calculations.
Notes: The figures shows that CPC code C12G has the largest share of pending (44.2%) and patented (24.5%) patent applications. Code C12H is the most common among unknown filings (100%).

Source: Author’s Calculations.