International Journal of Wine Business Research

BIG DATA ANALYSIS OF SPANISH WINE CONSUMERS REVIEWS

Journal:	*International Journal of Wine Business Research*
Manuscript ID	IJWBR-12-2023-0084.R2
Manuscript Type:	Research Article
Methods:	review, Natural Language Processing (NPL), Text Mining, website, Biterm Topic Model
Topics:	Consumers, Marketing strategy, Wines, Experience, Quality assessment

BIG DATA ANALYSIS OF SPANISH WINE-~~LOVING~~ CONSUMERS REVIEWS

Abstract

Purpose: Wine is a complicated and difficult product to know, which makes it extremely difficult for people with little knowledge to choose the wine they want. The aim of this work is to analyse whether the vocabulary used in the reviews on wine written by experts and amateurs on the specialized website Vivino is useful for those consumers who wish to search for information on this website to choose a wine.~~the terms used by wine-loving consumers when they write reviews about different types of wine on a specialised website.~~

Design/methodology/approach: The analysis combines Text mining, Natural Language Processing and the ~~Latent Dirichlet Allocation~~ Biterm Topic model applied to ~~49,765~~25847 reviews, evaluating a total of ~~12,028~~13263 Spanish wines, made by ~~28~~ 17 selected users of a specialised wine website.

Finding: ~~The results show that the wine-loving consumers-users of the selected website who write wine reviews, describing the organoleptic qualities and their reflections on the tasting focus on aromas, taste and appearance with appropriate terminology. The information conveyed by the reviews is useful because it comes from a consumer with knowledge of wine, which is a reliable source to inform novice consumers about the characteristics of a given wine, facilitating decision-making when deciding on one wine or another in the purchasing process~~. The results show that wine consumers and users of the specialized wine website who write reviews can be divided into expert users and amateur users. Both experts and amateurs use a specific vocabulary related to the wines they review. Unlike amateurs, experts have a broader and more precise vocabulary, and greater consistency in the use of words with the aspects of the wine. revised wines; they address fewer and more specific aspects of wine (such as vintages), but they do so with more depth and rigor.

Originality/Value: The originality and value of this research work lies in addressing two aspects that have hardly been analysed: 1) the reviews of ~~wine-loving~~experts- consumers, and amateurs-consumers~~who are not professional experts~~, and 2) the textual information referring to the Spanish language, which distinguishes this analysis from other similar analyses carried out on the English language.

Keywords: wine; Website~~review~~; ~~wine-loving consumer~~; Natural Language Processing (NPL); ~~Latent Dirichlet Allocation (LDA)~~,Biterm Topic Model, Text Mining

1. Introduction

Wine is a complicated and difficult product to understand. This complexity makes it difficult for people with little knowledge to choose the desired wine. experts (professionals from the word of wine) and wine enthusiasts know the words and their meaning in a unique language to communicate about the sensory characteristics of wines, such as aroma, flavor, appearance and mouthfeel (Katumullage et al., 2022). However, for those consumers with little knowledge, choosing a wine from the large number of different existing wines can be overwhelming. For the consumer, information on grape type, the region of the wine or the vintage can be useful. However, information about taste and aroma is not enough with the numerical scores given by critics~~reputable experts~~ like Parker. Consumers need more detailed qualitative information and seek it from professional wine magazines and wine websites where wine reviews are published. However, currently, there is no systematic way of using the large number of available reviews on the sites, which helps, on the one hand, consumers to choose the wine most suitable to their preferences and, on the other hand, producers and distributors to design the most appropriate marketing strategy for their target audience.

There is a~~n~~ debate about whether wine reviews provide meaningful information about wine properties and quality (Yang et al., 2022). In other words, does the text of reviews add significant and different information for the ~~consumer beyond~~consumer beyond the quality score, and objective and observable characteristics such as vintage, winemarket, variety and region of the wine? Wine reviews and wine ratings contain latent sensory information about wines, which consumers cannot obtain from the objective characteristics measured in wines (Katumullage et al., 2022). Therofore, wine reviews, tasting notes and wine ratings provide more information for consumers to more easly select wines they like (Katumullage et al., 2022). The accurate description of a wine includes not only flavours and aromas, but also characteristics such as acidity, tannin~~,~~ and structure. Furthemore, within each of these categories, there are a multitude of attributes or forms that each can take. The above described generates the possibility that two people simultaneously see the same wine differently and, at the same time, be able to share and detect all the same attributes.

However, wine reviews can sometimes include opaque and abstract descriptions because of the terminology used (e.g. complex, aggressive~~attractive~~, balance etc.), which is difficult to understand for consumers who are less informed in the context of wine (Gawel, 2007). For this reason, wine reviews have been labelled as useless and uninformative regarding the sensory properties of wines (Shesgreen, 2003; Silverstein, 2006; Quandt, 2007; Levinson and Majid, 2014).

Research on wine reviews has mainly focused on wine experts (professionals from the world of wine), leaving a gap in wine reviews written by consumers (Katumullage et al., 2022). For this reason and in accordance with the above, the aim of this research is to analy~~z~~se whether the vocabulary used in the reviews on wine written by experts and amateurs on the specialized website Vivino is useful for those consumers who wish to search for information on this website to choose a wine.~~the information from the reviews written by wine-loving consumers on a specialised wine website. In order to achieve this objective, we asked the following questions: RQ1 Do wine reviews written by wine consumers contain useful information? If they contain useful information, RQ2.1 What topics do they mention when commenting on the websites? RQ2.2 Do they follow a set script to describe wines, starting with appearance, then smell, taste and ending with mouthfeel (Paradis and Eeg-Olofsson, 2013)? and RQ3. What kind of terms do wine consumers use to write their reviews on websites?~~

To achieve the proposed objective and answer the above questions, techniques are applied that can be included within artificial intelligence, specifically, Text Mining, Natural Language Processing (NLP) and Biterm Topic Modelling (BTM) (Yan et al. 2013).~~Latent Dirichlet Allocation (LDA) mathematical model (Jelodar, H. et al., 2019)~~ Therefore, an attempt is made to retrieve useful information in a systematic and traceable way and analyz~~s~~e it in order to help (1) consumers to make informed purchasing decisions, and (2) winemarkers and distributors to design their marketing strategy towards their target audience. The results obtained show that wine consumers and users of the specialized wine website who write reviews can be divided into expert users and amateur users. Both experts and amateurs use a specific vocabulary related to the wines they review. Unlike amateurs, experts have a broader and more precise vocabulary, and greater consistency in the use of words with the aspects of the wine. revised wines; they address fewer and more specific aspects of wine (such as vintages), but they do so with more depth and rigor.~~ the consumer-users of the selected specialised wine website write wine reviews with precise wine language terms combining the description of the organoleptic qualities of the tasted wine and their reflections on the tasting, focusing on aromas, taste and appearance~~. ~~The information in the reviews is useful to the uninformed consumer because it comes from another wine-loving consumer who has knowledge of wine without being a wine professional or expert. It is concluded that the information from wine-loving consumers with some knowledge of wine is a reliable source to inform those other less informed consumers who want to choose a wine.~~

This research paper is organiz~~s~~ed as follows: After the introduction in Section 1, the literature on the topic under study is reviewed in Section 2. Subsequently, the methodology is described in Section 3 to discuss the results in Section 4. Finally, in Section 5 conclusions are drawn and Section 6 limitations and future lines of research as well as implications are pointed out.

2. Literature Review

Wine is an example of a product offering an experience where consumer information is limited. It is therefore common for third parties to act in this market to provide information. Specifically, one means of providing information about wine is currently online reviews by experts~~ (wine professionals)~~ or~~ wine-loving~~ consumers (amateurs). It is common for consumers to want specific characteristics about a wine without having any knowledge about grape varieties or regions where the desired characteristics would be generated. This situation of consumers will make them resort to description or review of a wine in order to obtain information about the bottle of wine they are holding in their hands. Unfortunately, these reviews may not have meaning for all those consumers who do not have a knowledge of the precise language of wine because they do not understand the language used with technical descriptors (McCannon, 20 20).

Wine reviews have become a common way of publiciz~~s~~ing a particular wine on the Internet. Consumers, if they want information other than that provided by the wine label and the winery in order to choose a wine, can search and read the ratings that wines receive from the experts, as well as read the reviews that~~ wine professionals and~~ wine~~-loving~~ consumer (amateurs) have written on specialized wine websites~~ the internet~~.

Numerous authors have explored the relationship between wine price and wine quality as measured by numerical ratings provided by expert tasters (McCannon, 2020).

Research on numerical wine ratings and wine reviews has been increasing over time (Mutkoski, 2011; Xu and Wang, 2013; Lemionet et al., 2015; Huang, 2018). As for numerical ratings, there is an extensive literature that has investigated the correlation between numerical ratings and price through demand, with inconclusive results. Some authors such as Ashenfelter and Jones (2003) found that numerical ratings of wines by wine experts~~professionals~~ are not significantly associated with price. Other authors such as Dubois and Nauges (2010) found correlation among~~between~~ ratings of the wine professionals or experts and wine scores. In fact, Ali et al. 2008 exploit an exogenous change in the timing of the publication of the numerical rating of a highly respected wine critic (Robert Parker) to document its influence on wine price. In addition to the relationship between numerical score and wine price, the prediction of the score a wine can achieve based on characteristics such as grape variety and winery location (Xu and Wang, 2017) or physicochemical attributes such as acidity, pH, sulphate (Lemionet, et al. 2015) has also been frequently analyz~~s~~ed. In relation to ~~All of~~ the above, McCammon (2020) suggests that the demand of a wine~~-loving~~ consumer is not only affected by the differentiating characteristics of the wine and price, but also by information from the textual descriptions (reviews) of the wine. ~~and the price paid McCannon (2020). .~~ Therefore, wine reviews can provide more valuable information to the consumer than a single numerical rating (Katumullage et al., 2022).

In the case of wine reviews, research is still limited (Yang et al., (2022~~)~~). As for wine reviews by wine professional, Storchmann (2012) provided a review of work studying the role of played by the opinions of these professionals. Chen et al. (2014) collected keywords (i.e. attributes that explain wine properties, such as "aroma", "full-bodied", "blackberries" and "tannins") from Wine Spectator and named this domain knowledge as "computational wine wheel". Katumullage et al. (2022) analyz~~s~~ed the performance of neural network algorithms for predicting wine rating class based on latent sensory information contained in wine text reviews. These authors found that wine reviews provide useful information about wine ratings because the accuracy of wine ratings based on wine reviews is quite high. In another study, Yang et al. (2022) found that wine review descriptors are more accurate in predicting wine quality ratings than numerical information. McCannon (2020) found that text review appears to be a significant predictor of wine price. However, the effect often disappears when the type of wine and its rating are included in the model. Quandt (2007) presented an example found in wine reviews by comparing legitimate professional wine reviews with random artificial reviews generated from a wine lexicon. Klimmek (2013) provided a new metric to distinguish meaningful wine reviews from redundant wine reviews, citing that reviews with a higher level of specificity tend to be more informative.

Finally, research on wine reviews by ~~wine-loving~~ consumers is scarce compared to research on wine reviews by professionals. Authors such as Weil (2007) conducted an experiment to demonstrate that wine consumers cannot match the descriptions of wines from professionals. Similarly, Salomon, (1997) found that reviews written by wine professionals are more accurately matched to wines than those of less informed consumers.

3. Methodology

This research work aims to fill the gap identified by Katumullage et al. (2020) in analy~~s~~zing information from reviews written by ~~wine-loving~~ consumers on a speciali~~s~~zed wine website. To achieve this goal, an appropriate methodology is required because text data, compared to numerical data, create a challenge in data analysis

~~Therefore, in this study, text mining techniques have been used, as procedures commonly employed in NLP for the pre-processing and preparation of texts and, subsequently, use the Latent Dirichlet Allocation (LDA) model. NLP procedures are used for correct textual processing and analysis and subsequently obtain reliable results. The LDA model is used to extract the most relevant terms and topics from the set of texts analysed. Term Frequency (TF) and Inverse Document Frequency (IDF) metrics were also used to assess the importance of a word within a corpus (Cai-Zhi et al., 2018) and other data processing and visualisation software tools were used to work with multiple types of data sources, process them and present the results obtained in a clear and understandable way to the end consumer.~~

The methodology used in the first stage is natural language processing (NLP) techniques to accurately process and analyze textual data and obtain reliable results. After pre-processing the selected text data, a topic modeling analysis has been performed on reviews from a website specialized in wines. These reviews are a type of text with some distinctive characteristic such as being relatively short in length. This motivates the need for a method that can effectively handle short texts from social media sources, such as the set of reviews under study. Qiang et al. (2022) propose three distinct categories of models for the application of topic modeling in short texts. The first category includes models based on Dirichlet multinomial mixture (DMM), the second category includes models based on global word co-occurrences (Biterm Topic Modelling (BTM)) and the third category includes models based on auto-aggregation. They conclude that methods based on global word co-occurrence (BTM) achieve better performance without any additional information than methods based on auto-aggregation and LDA. Specifically, BTM is the best method for topic classification and consistency. The best results for evaluation metrics of topic consistency, purity and accuracy, as well as number of iterations, are obtained by applying BTM (Qiang et al., 2022).

3.1. Selection of textual data sources

This research focuses on Spain in order to select the users who wrote the wine reviews. According to the Spanish Wine Federation (2022), Spain is the first vineyard in the world (approx. 13% of the world total) and in 2020 the second largest exporter in the world in volume (more than 2,012 million litres) and the third largest exporters in the world in value, (almost 2,616 million euros exported), and the third largest producer in the world (37.3 million hectolitres). Finally, it has a wide variety of wines as it is produced in 17 different regions and a wide variety of recogni~~s~~zed quality figures: 70 Denominations of Origin, 42 Protected Geographical Indications and 26 single vineyard wines.

The textual data analy~~s~~zed come from a speciali~~s~~zed wine website[i]. This website offers 38,000 wines from all over the world and which is used by thousands of consumers both to write and consult reviews of wine consumed and to choose the wine that best suits their preferences. This website was chosen among other sources of information because it is the only one focused on wines. Other websites provide information that is either do not have a strong relationship ~~with wine~~with wine to be included in a subsequent analysis of relevant terms and topics, or are very focused on the touristic experience, in general or ecotourism, in particular.

The selected website has 12,027 Spanish wines, distributed in 6 typologies (red, white, rosé, sparkling, fortified and dessert) and a wide range of regions. Given the wide variety, it is necessary to focus the work on those typologies that generate the greatest interest. To do this, an exploration is carried out with a data ~~visualisation~~visualization tool for *business intelligence* and the types of red, white and rosé wines are selected. This selection is made taking into account that these 3 types of wines 1) have a similar production process; 2) cover 91.5% of the volume of wine on the website analy~~s~~zed (Red 68%, White 21% and Rosé 2.5%); 3) and cover 90% of the total volume of reviews (4,127,192 out of the 4,593,308 available) with a similar distribution between the three; and 4) cover a wide price range (from 1.00 € to 490.00 €).

~~.~~ All reviews and textual information used in this analysis are confined to the Spanish language, which also distinguishes this analysis from other similar analyses conducted on the English language. There are works such as that of Antonio et al. (2018) applied to hotels, who demonstrate that both the ratings and the textual components of the reviews differ depending on the language in which they are written and that the textual component of the reviews can reveal even more about the influence of users' cultural origins on their preferences, likes and dislikes. However, this paper does not aim to carry out a comparative analysis of wine reviews written in various languages.

3.2. User selection and reviews

The basis of this research is the textual reviews made by wine consumers on Vivino~~the selected web~~. In order to choose the most consistent reviews possible, those made by the users with the highest popularity and reputation within the community of users of the web shop analysed were selected. This decision has been made on the assumptions that 1) the same user will hold similar criteria when commenting on different wines, giving stability to the study and 2) by basing the analysis on popular users, we assume that it will increase the chances of capturing relevant and useful content for the study. However, user ratings are not directed at the same set of wines, although many of the wines rated by users may overlap. The stages followed for the selection of users and reviews can be seen in Figure 1.

[Insert Figure 1]

The initial selection of users has been made according to (1) the ranking of the selected website and (2) the users recommended by the website. Therefore, priority has been given to the users with the best ranking and some other users from the community that the selected website highlights in the "recommended featured users" section.

The final selection of participating users was based on (1) the ranking provided by the website and, subsequently, an own ranking was established on this initial subset of users, using the aforementioned indicators to establish the ordinal value of the user within the ranking. This own ranking or *user engagement* aims to include information on the content generated by users already identified as "relevant" by the virtual *marketplace* itself. To calculate this value, the weighted average of the average number of *likes* per comment, the total number of reviews and the total number of followers of the users has been taken into account.

The number of *likes* per review is a metric of the interest generated by the comment, increasing the likelihood of it providing relevant information. The number of reviews, although it may not be directly related to the quality of the likes, has a certain weight in terms of the volume of reviews contributed to the study. Finally, as for the total number of followers, it is assumed that a user has followers because their reviews and ratings arouse interest, so it is also a metric to be taken into account.

For the reason mentioned above, more weight was given to the number of *likes* per review (60%) and the same weight to the other two variables (20% to the total number of reviews and 20% to the total number of followers). The criterion for selecting the end users to be part of the research was the median weighting value. It was decided to use the median instead of the mean because extreme values are observed in the weights, and therefore the median will give better information about the central point of the data. All users with a weighting greater than or equal to the limit set at >= 3.0 are selected. The median weighting for red wines was 3.45, white wines 3.6 and rosé wines 3.0. In total ~~28~~ 17 users were selected as shown in Table 1. The identifying name of the user has been hidden for privacy of personal information.

[Insert Table 1]

As illustrated in Table 1, the majority of reviews refer to red wines, followed by white wines and then rosés. The objective of this study is to differentiate between wine experts and amateurs, who possess a certain degree of knowledge about the world of wine. Vivino's privacy policy prevents having personal information about the identity of Vivino users, not being able to find out which users are experts and what others are amateurs among all the selected users. To this end, a threshold value of reviews has been established for each type of wine, which classifies the user as either an expert or an amateur. In the case of red wine, the threshold is 702 reviews. It is acknowledged that the threshold is relatively high; however, it is important to note that not all comments are wine reviews. The values selected for the purpose of dividing the data set were based on the total number of reviews for each type of wine. It could be argued that the figure is relatively high, but it is also thought that the level of knowledge of an individual who has tasted or reviewed 702 wines is likely to be lower than someone who has tasted 1,464. For rosés, the threshold was set at 38, while for whites, it was set at 474.

Once the users had been selected, the reviews published by them on the website were retrieved. The selection of user review data includes the following information: textual comment, winery, name of the wine and vintage, web rating, number of reviews made by the user, number of likes received, region, user rating, user name, web ranking for Spain and number of followers. In order to have a high volume of reviews, all the reviews available for each user up to the date of the analysis, carried out on 25 April 2022, have been collected.

3.3. Mathematical definition

~~For the automatic analysis of the texts obtained, statistical models such as term relevance metrics and topic extraction and clustering techniques have been used.~~

Firstly, for the automatic analysis of the texts obtained, natural language processing techniques will be used to process and cleanse the texts, and then BTM will be used to discover the underlying topics. In terms of implementation, the R libraries udpipe (Wijffels et al., 2018) and BTM (Wijffels et al., 2013) have been used.

3.3.1. ~~Relevant terms~~ Natural Language Processing

~~The TF-IDF metric (Aizawa, 2003) is used to identify relevant terms in the corpus, i.e. the set of wine reviews.~~

$$\cancel{TF\text{-}IDF = TF * IDFF}$$

~~TF = Number of times the term "X" appears in Document / Total number of terms in Document~~

~~$IDF = log$ (Number of documents in $Corpus$ / Number or documents in which the $t$erm "X" appears)~~

~~Where:~~

- ~~**Term:** individual words (tokens)~~
- ~~**Document**: the set of all the reviews of the same user as if it were a single text.~~
- ~~**Corpus**: the set of all documents.~~

~~This metric identifies terms that allow usto assign the most relevant terms to a text. The relevant terms are those that appear most frequently in the document (TF) and, furthermore,t are not excessively common terms in the subject matter (IDF).~~

It is essential for good analysis to have as clean a text as possible, so in this study each review was processed in the following way. All characters were converted to lowercase, then hyperlinks and mentions of other users were removed. Non-alphanumeric characters and extra spaces were removed. Also, punctuation marks and blank reviews were removed.

Once the texts have been cleaned, the next stage is to perform the tokenisation tasks, which involve separating the words that make up the text, eliminating empty words or stopwords, and carrying out a syntactic analysis to determine the type of word, such as a verb, noun, adjective, or so on. The reviews are divided into three categories according to whether they are red, rosé or white, and the reviews on each type of wine are analysed independently.

The analysis process consists of several steps: first, the nouns and adjectives that occur most frequently in all reviews are calculated, then the co-occurrences at sentence level are calculated, which allows a visual analysis of the frequency of words that occur together in the same sentence. Correlations are then calculated, showing how often terms occur together in all reviews. While co-occurrences focus on frequency, correlation measures between two terms can be high even if two terms occur only a few times, but always together. Once these processes have been applied, topic modelling can be performed.

3.3.2. Topic modeling

The extraction of topics present in the set of reviews is done by applying the BTM model ~~LDA model~~. ~~This statistical model assumes that each document is a mixture of a small number of topics, and that each topic is a probability distribution over the set of words contained in a *corpus* or subset of words in the language relating to an area. Thus, Bayesian inference is used to estimate the probability distribution of topics and words in the corpus. In this particular model, the Dirichlet distribution is used to model the probability distribution over topics and words in each topic (Jelodar et al., 2019).~~

~~More specifically, the initial assumptions used by the model are as follows:~~

1. ~~We assume that there are: D documents in our text set, K different topics in the document set, and W words in the corpus under analysis.~~
2. ~~For document d = 1...D, the vector of probabilities of topics of length K,~~ $\cancel{\theta_d}$ ~~, can be considered as drawn from a Dirichlet distribution (α), where~~ $\cancel{\alpha_k > 0}$ ~~for topics k = 1...K.~~
3. ~~For subject k = 1...K, the vector of probabilities of terms of length W,~~ $\cancel{\phi_k}$ ~~, can be considered as drawn from a Dirichlet distribution (β), where~~ $\cancel{\beta_w > 0}$ ~~for terms w = 1...W.~~
4. ~~The probabilistic model states that for the jth token or word in document d, a latent topic,~~ $\cancel{z_{dj}}$ ~~, is extracted, where~~ $\cancel{P(z_{dj} = k) = \theta_{dk}}$ ~~for document d = 1...D, token j = 1...n_d, and topic k = 1...K.~~
5. ~~Then, the j-th token of the d-th document,~~ $\cancel{Y_{dj}}$ ~~, is extracted from the vocabulary of terms according to~~ $\cancel{P(Y_{dj} = w|z_{dj}) = \phi(z_{dj}, w)}$~~, for document d = 1...D, token j = 1...n_d, and term w = 1...W.~~

~~The model is able to indicate as a result which texts have distributions of terms that characterise them as belonging to the same group. To achieve this result, the co-occurrences of the words contained in the texts are taken into account and the "weight" that each term has relative to each set of texts in particular is indicated, based on the probability of occurrence of the term in the different sets of texts in the collection. It is the analyst's responsibility to identify the theme that encompasses the different sets of texts by examining the most representative terms.~~

~~An LDA algorithm is used to perform the probability fitting calculations on the basis of complete texts. It works by iteratively updating probability distributions on topics and words, based on the texts received as input, until it reaches convergence. It starts with an initial guess for the probability distributions α and β, which it then updates using Bayesian inference. It continues to update the distributions until they converge to a stable solution (Zhao et al., 2020).~~

~~In this analysis, the implementation of the LDAModel algorithm provided by the Gensim library (Řehůřek & Sojka, 2010) has been used.~~

In particular, we will be utilizing the BTM implementation postulated by Yan et al. (2013). A detailed and comprehensive explanation of BTM and its variants can be found in the article by Cheng et al. (2014). This section presents the operational aspects of the algorithm, as extracted from the aforementioned article.

BTM performs the topic modelling by the generation of biterms. The fundamental premise is that if two words co-occur with greater frequency, they are more likely to be associated with the same topic. Based on this premise, it is assumed that the two words in a biterm are drawn without dependence from a topic, where a topic is sampled from a topic mixture over the entire corpus.

Given a corpus with N_D documents, suppose it contains N_B biterms $\mathbf{B} = \{\boldsymbol{b_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{N_B}}$ with $b_i = (\omega_{i,1} \mid \omega_{i,2})$, and K topics expressed over W unique words in the vocabulary. Let $z \in [1,K]$ be a topic indicator variable, we can represent the prevalence of topics in the corpus (i.e., $P(z)$) by a K-dimensional multinomial distribution $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^{K}$ with $\theta_k = P(z = k)$ and $\sum_{k=1}^{K} \theta_k = 1$. The word distribution for topics (i.e., $P(\omega \mid z)$) can be represented by a $K \times W$ matrix $\mathbf{\Phi}$ where the kth row ϕ_k is a W-dimensional multinomial distribution with entry $\phi_{k,\omega} = P(\omega \mid z = k)$ and $\sum_{\omega=1}^{W} \phi_{k,\omega}? = 1.$

In accordance with the established convention of LDA, symmetric Dirichlet priors are employed for $\boldsymbol{\theta}$ and ϕ_k with single-valued hyper-parameters α and β, respectively. The formal description of the generative process of BTM is as follows:

1. Draw $\boldsymbol{\theta} \sim$ Dirichlet(α).
2. For each topic $k \in [1,K]$
    1. draw $\phi_k \sim$ Dirichlet(β).
3. For each biterm $b_i \in \mathbf{B}$
    1. draw $z_i \sim$ Multinomial($\boldsymbol{\theta}$), and
    2. draw $\omega_{i,1}, \omega_{i,2} \sim$ Multinomial(ϕ_{zi}).

Following the above procedure, we can write the probability of biterm b_i conditioned on the model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$:

$$P(b_i \mid \theta, \Phi) = \sum_{k=1}^{K} P(\omega_{i,1}, \omega_{i,2}, z_i = k \mid \theta, \Phi).$$

$$= \sum_{k=1}^{K} P(z_i = k|\theta_k) P(\omega_{i,1} \mid z_i = k, \phi_{k,\omega_{i,1}}) P(\omega_{i,2} \mid z_i = k, \phi_{k,\omega_{i,2}})$$

$$= \sum_{k=1}^{K} \theta_k \phi_{k,\omega_{i,1}} \phi_{k,\omega_{i,2}}$$

Given the hyperparameters α and β, we can obtain the probability of b_i by integrating over $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$:

$$P(B \mid \alpha,\beta) = \prod_{i=1}^{N_B} \iint \sum_{k=1}^{K} \theta_k \Phi_{k,\omega_{i,1}} \Phi_{k,\omega_{i,2}} \, d\theta \, d\Phi$$

4. Results and dicussion

~~Data processing has been carried out, first, by applying NLP pre-processing techniques and then by applyingthe LDA model that allow obtaining the relevant terms and themes.~~

~~NLP techniques were applied to carry out tokenisation and lemmatisation. Tokenisation is the extraction of tokens or words contained in the comment, including the removal of empty words. Lemmatisation consists of transforming each token into its lemma or term representing all the forms derived from the word. In the end, 4,098 unique lemmatised words (tokens) were obtained from a total of 49,765 reviews. Table 2 shows a sample of the NLP pre-processing that has been carried out with some written user reviews of each of the types of wines analyzed.~~

~~[Insert Table 2]~~

~~Once the texts had been pre-processed, an initial exploratory analysis of the content was carried out. For this analysis, the TF-IDF (Aizawa, 2003) was calculated using as a "document" the concatenation of all the texts written by the same author and referring to the same type of wine (red, white or rosé). In this way, it is possible to identify (1) those relevant terms within the vocabulary and written expression of the same web user and (2) the possible differences with other authors or characteristic patterns of each user.~~

~~Tables 3, 4 and 5 show the results obtained (10 relevant terms per user for each type of wine) visually represented on a heat map, where each row corresponds to a user, and the deep blue colour indicates a higher TF IDF value and the yellow green tone a lower value.~~

~~[Insert Table 3]~~

~~[Insert Table 4]~~

~~[Insert Table 5]~~

~~In the results in Tables 3 and 4, it is showed that the relevant terms are related to the specific characteristics of each type of wine, which ensures the consistency of the results. In addition, the three wines share terms in highly relevant positions, such as "fruit", "nose", "colour", "mouth", etc. This shows that technical terms emerge in the evaluation of wines made by users or wine-loving consumers , which shows the knowledge that users have about the appropriate language that should be used when assessing a wine.~~

~~The extraction of topics present in the set of reviews is performed by applying the LDA model. It is interesting to note that the LDA algorithm asks the analyst to provide as input the number of topics (or groups of documents) that he/she expects the set of documents to contain. In order to determine this parameter, a *parmeter tuning* process was followed, which consisted of considering a range of number of topics from 2 to 20, adjusting the model for each number and checking the parameters of *Coherence Score* as the main parameter and *Perplexity* as the differentiating parameter, if necessary, obtained for each model adjustment.~~

~~Based on the results provided by the algorithm on the texts analysed, it can be concluded that there are mainly 2 themes, identified by the authors as: "tasting and sampling" (denoted with value 1 in Table 6) and "the product and its processing" (denoted with value 2 in Table 6). The theme of tasting and tasting covers 57% of the terms in the corpus and the theme of the product and its processing covers 43% of the terms in the corpus. In the first theme, looking at Table 6, it is possible to distinguish terms closely related to the organoleptic qualities of the wine (colour, taste, aroma, textures, sensations, etc.) and to more technical aspects of the tasting process (tasting, edge, tear, mouth, note, etc.). In the topic "product and elaboration" we find terms related to the general characteristics as a product (type, label, price, etc.) and terms related to the field of winemaking (winery, type of grape, barrel, etc.). There are also neutral terms that could fit in both options.~~

~~[Insert Table 6]~~

~~Given the interest in the relationship between the relevant terms and the themes found, it can be said that the themes obtained are generic, although they seem to hide other latent sub-themes in them. In the case of tasting and sampling, since this theme is closely related to the organoleptic qualities of wine (what we can perceive through the senses), categories or sub-themes related to flavor and aroma, and appearance (colour of the wine and its effects in the light) are identified. An important part of the tasting is the final reflection on the tasting, in particular, the set of flavours, aromas, colours and effects. These categories include terms that are often used to describe these reflections and sensations together with the terms used to describe technical aspects of the wine such as "persistence", "length", "intensity", etc.~~

~~In the topic "product and processing", the terms used to comment on the wine as a final product are related to the bottle, the type of wine, the vintage and the price. While the terms used to comment on the winemaking process are usually related to the type of grape, the type of wood used in the barrel or aspects of the winemaking process.~~

~~Table 7, 8 and 9 shows a list of relevant terms for each type of wine, ordered by the number of users who have used them within the 50 most relevant terms. The topic with the most relevant terms is "tasting and sampling" and the subtopics with the most relevant terms are, in this order: tasting, flavour and aroma, production, appearance and product. When a sub-topic has a greater number of relevant terms, it is assumed that it is of greater interest to users when evaluating a wine.~~

~~[Insert Table 7]~~

~~[Insert Table 8]~~

~~[Insert Table 9]~~

~~Finally, if we compare the most prominent aromas in the above tables with those contained in the wine aroma wheel designed by Ann C. Noble (1984, 1987), we can deduce the degree of knowledge that consumers and selected users have about this organoleptic property. For all three types of wines, users mention some first level aromas such as "fruit", "spice" or "floral" and many more third level aromas such as "cherry", "blackberry", "liquorice", "strawberry", "raspberry" leaving unnamed second level aromas such as "tropical fruit", "citrus", "caramel". Therefore, it can be observed that~~

~~consumers move from the more generic aromas to the more specific ones.~~

As mentioned above, BTM is used to extract the relevant topics and the words that carry the most weight in each of them. We have chosen to extract eight topics for each type of wine and to represent the five words with the highest weight.

4.1. Red wine experts vs. red wine amateurs

The number of expert reviews on red wine is 16.540 and the number of amateurs is 2.797.

Comparing Graphic 1, which corresponds to the most frequent words in the reviews of the experts, with those in Graphic 2, which belong to the amateurs, seven of the ten most frequent words are common to both groups, although the frequency of each is different. The most frequently used words by the experts are fruit (fruta), mouth (boca) and oak (roble), while the amateurs use them in positions 3 to 5, giving more importance to terms such as note (nota) and color, which are less frequently used by the experts.

[Insert Graphic 1]

[Insert Graphic 2]

The analysis of the most frequent adjectives presented in Graphic 3 and 4 yields a similar interpretation. In this case, the initial terms are identical, albeit in a different order. It is noteworthy that experts and amateurs employ distinct adjectives to describe wine. For instance, experts use terms such as tall (alto) French (francés) or mature (maduro), which pertain to the wine's origin and taste characteristics. In contrast, amateurs often use terms like natural cork (corcho natural) or waxed cork (parafinado), which relate to the presentation and sealing of the bottle. Additionally, fruit is frequently cited as an adjective, particularly in reference to the wine's aroma or taste, and is often perceived as a quality associated with a fruity wine.

[Insert Graphic 3]

[Insert Graphic 4]

In the Graphic 5 and Graphic 6," it can be observed that in the upper part of the amateurs (image 6), terms referring to the sealing of the bottle usually appear together, while in the lower part, the word "red" (rojo) centralizes the rest, being the one that is most related to the others. In the case of the experts, the resulting network is more complex. While the word "red" (rojo) appears to centralize the network, other words, including fruit (fruta), mouth (boca), nose (nariz) and good (bueno) also have significant weight. This suggests that experts use a greater range of word relationships than amateurs.

[Insert Graphic 5]

[Insert Graphic 6]

The correlations are demonstrated in Graphic 7 and 8. Upon examination of Graphic 7 on the right, a cluster of terms pertaining to the winemaking process can be discerned, including oak (roble) and barrel (barrica) In the primary network, it is evident that there is a robust interconnection between the terms fruit (fruta), mouth (boca), nose (nariz) aroma (aroma) and red (rojo) To provide further detail, the words nose (nariz) and aroma (aroma) appear together with others, such as intensity (intensidad) or fine (fino) which pertains to the aromatic properties of the wine. Similarly, the word fruit (fruta) appears together with touch (toque), tannin (tanino), acidity (acidez) or entrance (entrada), and mouth (boca), which pertains to the taste properties. Additionally, the words fruit (fruta), red (rojo), color, black (negro), and cherry (cereza), appear with a strong relationship, which pertains to the visual characteristics. In Graphic 8, two distinct networks are evident among the amateurs. In the lower part of the network, words referring to the finish and bottling are observed, while in the central part, words related to the red color (rojo), nose (nariz) and fruit (fruta) are present, though the relationships between these elements are less clear than in the experts. A minor subnetwork pertaining to the winemaking process is also visible on the right. Additionally, there are a few peripheral relationships. This suggests that experts tend to form more intricate word associations when describing the characteristics of red wines, in comparison to amateurs.

[Insert Graphic 7]

[Insert Graphic 8]

Graphic 9 and Table 2 show the topics identified by BTM for expert users, while Graphic 10 and Table 3 show the topics identified for amateur users. The thickness of the line in the images is indicative of the strength of the relationship between the terms. The tables illustrate the probability of the term occurring in the reviews associated with a given topic.

[Insert Graphic 9]

[Insert Table 2]

[Insert Graphic 10]

[Insert Table 3]

For experts, topic one pertains to the visual characteristics of the wine, encompassing its red-black hue and intensity. The second topic concerns the production process, with a particular focus on the barrels used. Topic three is centered on the olfactory characteristics of red wine. The fourth topic once again refers to color, but the appearance of the word "capa" indicates that it is a tasting experience. Topic five addresses the provenance of the wine and the vineyards from which it is sourced. It is noteworthy that the sixth topic indicates that experts differentiate based on the various vintages of the wine. The seventh topic addresses the sensations experienced in the mouth when tasting the wine, with a particular focus on its taste characteristics, including acidity. Finally, topic 8 is concerned with the characteristics of fruity wines, including their color and mouthfeel. It is noteworthy that in several of the topics, the adjectives "good" and "average" are employed, which suggests that experts use a range of adjectives to evaluate the characteristics of the wine.

Regarding the amateur sector, the initial topic is that of bottle closure. The second topic pertains to the winemaking process. The third topic addresses the designation of origin, and the grape varietal used in the winemaking process. The fourth topic concerns wine tasting experiences. The fifth and sixth topics appear to be somewhat more expansive in scope, though both appear to pertain to the visual and aromatic characteristics of the wine. The seventh topic is a combination of the characteristics of the bottles, the designation of origin and the type of grape used. Finally, the eighth topic addresses the gustatory characteristics of the wine. In this instance, there is no extensive use of adjectives of degree in the topics.

A comparative analysis of experts and amateurs reveals that experts tend to address topics in a more focused manner. They use reviews to discuss a particular aspect of wine, whereas amateurs often cover multiple topics in a single review. Regarding the nature of the topics, both address winemaking, the origin and the visual, olfactory and gustatory characteristics of the wine. In terms of differential topics, experts tend to focus on the specific vintages of the wine, whereas amateurs tend to concentrate more on the sealing and presentation of the bottle. Furthermore, it is evident that experts possess a more refined vocabulary, using precise terminology to delineate the various aspects of wine, whereas amateurs tend to employ a more generic lexicon, applying the same terms to multiple topics, as evidenced in topic five, six and seven.

4.2. Rosé wine experts vs. rosé wine amateurs

The number of experts on rosé wine is 359 and the number of amateurs is 211. It is worthy of note that the number of reviews referring to rosé wine is less than that of the other two types of wine under analysis.

Regarding the vocabulary employed by experts and amateurs in the context of rosé wines, as illustrated in Graphic 11 and 12, respectively, it is evident that the initial three nouns are identical. The nouns color and mouth (boca) are used by experts to define rosé wines with a similar frequency. In a second level of frequency, the nouns fruit (fruta), acidity (acidez), winery (bodega), and note (nota) are used. The nouns once (vez), sensation (sensación), and nose (nariz) are used less frequently. Among the rosé wine amateurs, the term fruit (fruta) is the most frequently used, followed by the term mouth (boca). The remaining terms are used with a similar frequency. The terms color, aroma (aroma), acidity (acidez), sparkle (destello), note (nota), intensity (intensidad), touch (toque), and brilliance (brillantez) are also used with some regularity. The experts employ a greater number of terms to differentiate between the two types of wine, utilising the terms winery (bodega), once (vez), sensation (sensación) and nose (nariz). The term winery (bodega) is used to denote the vineyard or winemaking facility responsible for producing the wine, while once (vez), sensation (sensación) and nose (nariz) are used to describe the sensory experiences associated with tasting, the texture or mouthfeel of the wine and its aromatic qualities, respectively.

[Insert Graphic 11]

[Insert Graphic 12]

In contrast, the terms employed by amateurs to differentiate between the two types of wine are sparkle (destello), intensity (intensidad), touch (toque) and brilliance (brillantez). These terms are used to describe the visual appearance, intensity, texture and overall quality of the wine, respectively.

Adjectives that qualify nouns can be found in Graphic 13 and 14, respectively, for experts and amateurs. It can be observed that the first four names exhibit a similar frequency of occurrence, except for the adjective "red" (rojo), which is more prevalent among experts. The remaining adjective employed by both groups is the middle grade adjective. The experts use the adjectives fruity (afrutado), light (ligero), intense, (intense) and fresh (fresco), which pertain to the gustatory and olfactory attributes of the wine. The adjective "new" (nuevo) may be indicative of the fact that expert users place a higher value on wines that have recently been launched on the market. In contrast, the amateurs employ the adjectives "long" (largo), "mouth-filling" (boca) and "dry" (seco), which pertain to the gustatory characteristics, and fruity (afrutado), which pertains to both gustatory and olfactory characteristics

[Insert Graphic 13]

[Insert Graphic 14]

The co-occurrences of experts in rosé wine are illustrated in Graphic 15, while those of amateurs are presented in Graphic 16. Both sets of data demonstrate a degree of similarity. In the experts, the terms red (rojo), nose (nariz), mouth (boca), good (bueno), fruit (fruta), note (nota), and acidity (acidez) demonstrate a comparable degree of interrelationship. Other terms manifest in a peripheral manner with comparatively weaker relationships with the word fruit (fruta). In the amateur cohort, a robust correlation is evident between the terms fruit (fruta) red (rojo), mouth (boca), good (bueno), and nose (nariz) with a comparatively weaker association observed between aroma (aroma), pink (rosa) and color. This suggests that these three terms are used to a lesser extent within this group. Similarly, as with the experts, terms related to fruit (fruta) also emerge peripherally. From this image, it can be inferred that the experts tend to use the terms in a more balanced manner, whereas the amateurs employ the same terms with greater frequency in their descriptions of the wines.

[Insert Graphic 15]

[Insert Graphic 16]

The correlations of the experts in Graphic 17 and that of the amateurs in Graphic 18 serve to reinforce this idea, with a greater number of words appearing together in the experts. The strongest relationship is observed in both groups, namely the terms red (rojo) and fruit (fruta). Nevertheless, experts evince a greater number of word associations, thereby indicating that they deploy a more diverse linguistic repertoire when discussing rosé wines than do amateurs.

[Insert Graphic 17]

[Insert Graphic 18]

The themes that emerge from the BTM analysis for experts are presented in Graphic 19 and Table 4. The initial theme that emerges pertains to the characteristics that the wine acquires during the storage process in barrels. The second theme encompasses the tasting experiences that occur within the context of the winery. The third theme concerns comparisons with red and white wines. The fourth theme is primarily concerned with the color of the wine. The fifth theme deals with the taste characteristics. The sixth theme is an amalgamation of the color, mouthfeel and smell of the wine. The seventh theme refers to new wine releases, comparing them with older ones, and to the type of soil where the vine is grown. Finally, the eighth theme concerns the type of winery where the wine is produced.

[Insert Graphic 19]

[Insert Table 4]

The outcomes of the BTM analysis in the amateur cohort are illustrated in Graphic 20 and Table 5. The initial theme that emerges appears to be focused on the visual characteristics of the wine. The second theme is evidently associated with the maturation process in the barrel. The third theme is primarily concerned with the taste and visual characteristics. The fourth theme is focused on the visual and olfactory characteristics. The fifth theme is related to the type of sealing of the bottle. The sixth theme does not appear to be a coherent theme, as all words have an equal probability of occurrence. The seventh and eighth themes address the taste characteristics. The seventh theme is more focused on the finish, while the eighth theme is more focused on the initial sensation in the mouth.

[Insert Graphic 20]

[Insert Table 5]

Both experts and amateurs engage in discourse on common topics pertaining to rosé wine, including the processes involved in its production in barrels and the visual, gustatory, and olfactory characteristics of the wine. However, experts delve more profoundly into comparisons with other types of wine, the cultivation process, and the winery of origin. In contrast, the topics discussed by amateurs appear to be more diffuse, with several topics merged into a single discussion. This often involves talking about various characteristics or topics that are very similar but have slight differences, as illustrated by the last two examples. It is also noteworthy that both experts and amateurs focus on the sealing of the bottle.

Based on these observations, it can be concluded that both experts and amateurs employ a significant number of similar terms to discuss rosé wines, with the most frequent being the most used by both groups. However, experts demonstrate a more expansive vocabulary, as evidenced by the co-occurrences and concurrences. With regard to the topics discussed, experts exhibit a broader range of topics than amateurs.

4.3. White wine experts vs. white wine amateurs

The number of expert reviews on white wine is 3.638 and the number of amateurs is 2.390.

The most frequently used vocabulary by experts in their reviews on white wines is showed in Figure 21, while the most frequently used vocabulary by amateurs is shows in Figure 22. In both cases, the most frequently occurring term is mouth (boca), which pertains to the gustatory attributes of the wine. Other frequently occurring terms are color and note (nota), as well as fruit (fruta), which are the most commonly used by both groups, indicating that the reviews pertain to visual and gustatory characteristics. The experts employ vocabulary such as persistence (persistencia) and intensity (intensidad) which are evident examples of specialized vocabulary pertaining to the sensory experience of the wine. In contrast, amateurs employ more general terms, such as touch (toque) or flash (destello) to describe the sensation in the mouth. It is noteworthy that the experts use the word “final”, which is more prevalent in their lexicon than in that of the general public. This indicates that experts tend to prioritize the aftertaste of the wine, whereas amateurs tend to focus on the initial sensation. Regarding olfactory characteristics, the term aroma (aroma) is frequently employed by amateurs, whereas experts tend to use the term nose (nariz) with greater frequency.

[Insert Graphic 21]

[Insert Graphic 22]

The adjectives used are shown in Graphic 23 for the experts and Graphic 24 for the amateurs. As with nouns, the most frequently used adjective in both cases is good (buen) to define the quality of a certain characteristic, followed by yellow (amarillo), which undoubtedly refers to the color of the wine. The experts also use the adjective medium (medio) to define the qualities, which is not used by the amateurs. In both cases we find nose (nariz), which comes from nasal and refers to the olfactory qualities, while the other adjectives refer to the color or flavour. We find citrus (citríco) in fifth place for the experts and eighth place for the amateurs. This adjective is a specialized term used to define the characteristics of the flavour or aroma. Therefore, according to the results, experts use it more frequently, indicating a greater familiarity with the vocabulary.

[Insert Graphic 23]

[Insert Graphic 24]

The co-occurrences of the experts ~~is~~are showed in Graphic 25, while those of the amateurs are presented in Graphic 26. In the network of words displayed in the lower left quadrant, there is a cluster of terms with a strong relationship between them. Additionally, a series of characteristics that are related to the adjective good (bueno) in a weaker manner can also be discerned. In other words, a group of words that appear with high frequency in the same sentence can be identified, including "good color," "good ending," "mouth," and "fruit." This indicates that the same lexical items are employed with considerable frequency. In contrast, the network of terms among the amateurs is characterized by a strong relationship between good fruit, mouth and yellow. However, the relationships between these terms and the others are less intense than those observed among the experts. In this instance, the amateur cohort demonstrates a greater diversity of vocabulary.

[Insert Graphic 25]

[Insert Graphic 26]

Upon initial observation, it is evident that the correlations of the experts in Graphic 27 and those of the amateurs in Graphic 28 exhibit notable discrepancies. The number of terms employed is greater in amateurs, thereby confirming that they use a more extensive vocabulary. The strongest relationship observed in the ~~experts~~expert's network is that between color and yellow (Amarillo). In the case of the amateurs, there is a notable correlation between the color, yellow and brightness, which subsequently gives rise to a greater number of adjectives. Another robust correlation is observed between the terms nose (nariz), aroma (aroma) fruit (fruta), and brightness (brillantez), which pertain to olfactory characteristics. Additionally, a significant relationship is evident between mouth (boca) and entrance (entrada) which pertains to taste characteristics.

[Insert Graphic 27]

[Insert Graphic 28]

The results of the BTM for experts are illustrated in Graphic 29 and Table 6, while Graphic 30 and Table 7 present the corresponding results for amateurs.

A review of the vocabulary used in the initial topic for experts suggests that it pertains to the maturation process of wine. The second topic is clearly concerned with the origin of the wine, with the use of terminology such as 'vineyard', 'winery' and 'designation of origin'. The third most frequently occurring theme is the taste characteristics of the wine, while the fourth is a comparison of different vintages. The fifth topic reiterates the discussion of taste characteristics, with a particular emphasis on the aftertaste. The sixth topic appears to concentrate on the color and visual aspect of the wine, although it also addresses the olfactory characteristics. The seventh topic addresses wines with fruity and citrus notes on the palate, while the eighth discusses the grape varieties used.

In the case of the amateurs, the initial topic is concerned with the gustatory characteristics of the wine, with particular attention paid to its aftertaste. The second topic is related to the closing of the bottle. The third topic concerns the geographical origin and grape variety of the wine. The fourth topic is challenging to discern with very low probabilities; in this case, it is not possible to identify a specific topic to which it refers. The fifth topic concerns the initial sensory experience of the wine upon entering the mouth. The sixth topic is concerned with the color and olfactory characteristics of the wine. The seventh topic addresses the maturation process in barrels, while the eighth and final topic is concerned with the aromas and fruity notes on the nose and palate.

[Insert Graphic 29]

[Insert Table 6]

[Insert Graphic 30]

[Insert Table 7]

In the case of white wine, the two profiles address the same topics, with the only notable difference being that experts address the topic of different vintages, while amateurs focus on the closing of the wine.

The three analyses collectively indicate that expert users employ a more sophisticated and technical vocabulary when describing red and rosé wines, whereas amateurs tend to use a more general vocabulary when discussing white wines. The nouns and adjectives employed to characterize the wines are consistent across both user profiles. General terms are used to describe the geographic characteristics, olfactory characteristics and appearance. These include terms such as mouth (boca), nose (nariz) and color. The most notable divergences in vocabulary and specialized terminology emerge in the less frequently used words. The two profiles address common topics that can be classified into four categories: taste, smell, appearance, and winemaking. Those with expert knowledge tend to engage with more complex topics, such as the different vintages, whereas those with less expertise or amateurs focus on other aspects, such as encapsulation. It is noteworthy that, in general, the topics addressed by experts are more structured and concrete, whereas those of amateurs are vaguer and more imprecise.

5. Conclusions and implications

In markets where the consumption of the product is an experience, consumers need information. Wine is an experiential product where there are consumers who have little knowledge of wine and do not know to accurately use ~~-~~the technical terms of the wine world. The aim of this work is to analyze whether the vocabulary used in the reviews on wine written by experts and amateurs on the specialized website Vivino is useful for those consumers who wish to search for information on this website to choose a wine.~~analyse the terms used by consumers when they write reviews about different types of wine on a specialised website~~, by applying Natural Language Processing techniques and more specifically, the ~~LDA~~ BTM model for topic modelling. ~~This objective is specified in the research questions formulated in the Introduction section, which, in the light of the results obtained, are answered in this section.~~

Both experts and amateurs use a specific vocabulary related to the wines they review. Unlike amateurs, experts have a broader and more precise vocabulary, and greater consistency in the use of words with the aspects of the wine. revised wines; they address fewer and more specific aspects of wine (such as vintages), but they do so with more depth and rigor. This conclusion is in line with Kontoya et al. (2018) who concluded that Vivino users had a level of knowledge of wine similar to professional experts

~~As to whether wine reviews contain useful information (QR1), it can be answered that, c~~Contrary to Weil (2007) and Salomon (1997), wine consumers provide useful information about the wines they have consumed in their reviews, using a specific and common vocabulary or terminology within the world of wine. However, there are differences between reviews from different users regarding the breadth of vocabulary, aspects of the wine analyzed and internal consistency. Therefore, users who consult Vivino are advised to use more sources of information when they need to be informed when choosing a wine. ~~As consumers write reviews with useful information on the selected website, it is possible to answer the second research question (QR2.1) related to the mentioned topics. When wine-loving consumers or users of the website write their reviews, they, firstly refer to the tasting and sampling, i.e. the sensory aspect, and secondly mention the product and its elaboration or technical aspect. Regarding research question Q2.2 related to the script followed by consumers-web users in their reviews, it can be concluded that, contrary to what Paradis and Eeg-Olofsson, (2013) advocate, consumers write their reviews by first mentioning the smell, then the taste and mouthfeel and finish with the appearance. Finally, among the terms used by consumers of different types of wine to express their reviews (RQ3), the term "fruit" stands out for the three types of wine analysed (red, rosé and white), "intense" and "body" when referring to the mouthfeel, the term "red" when describing the appearance of the wine, the terms "mouth" and "nose" when tasting it. As for the product and its processing, the most used terms are "crianza" and "rosé" when they are commenting on the product, and "barrel" and "cellar" when consumers are commenting on the production.~~

Therefore, in general the users of the wine website or ~~wine-loving~~ consumers write useful reviews mainly about the tasting and sampling, and then about the product and its production. However, these topics are further broken down into sub-themes such as taste and aroma, appearance, production and product. From the above it follows that the wine consumer who writes reviews about the wines consumed has informed knowledge about wine, in general, and about the wine consumed, in particular. These consumers show wine profiles and inform others consumers with less knowledge about those aspects of wine that are more difficult for users to know when evaluating a wine. ~~The consumer profile of the selected users (the most influential of the selected web in Spain) is a wine-loving consumer profile with knowledge of the product due to the technical language used. The above enables the wine-loving consumer to write constructive reviewes when evaluating the tasting and the product.~~

The main implication of the results obtained from this research concerns less informed consumers. These consumers may have a credible source of information in the reviews of the consumer-user of a speciali~~s~~zed wine website to decide the wine they want to buy. This implication is relevant since wine is a product with an important social value and, to reduce the risk of making a mistake with the wine chosen for a social event is important for less informed consumers. From the previous implication for consumers, others are derived for wineries. First, winemakers should contrast the organoleptic characteristics of their wine with those described in the reviews made by wine~~-loving~~ consumers and, consequently, making~~e~~ a decision about their future wines and redesign the marketing strategy, if necessary. Second, wineries that are not present with their wines on websites should make an effort to be~~ on them~~ since they could be recommended by these wine-~~loving~~ consumers who write reviews with knowledge about the characteristics of the reviewed wines and, in addition, have numerous followers.

6. Limitations and future lines research

This study has~~ some~~ limitations. One of them is that the analysis carried out does not analyz~~s~~e additional information on the utility that the consumer will receive from the wine beyond the level of quality and the objective and observable characteristics. Wine is a product that conveys an experience and consumers are looking for an experience in that consumption. It becomes necessary to analyz~~s~~e the emotional aspect that the consumer has conveyed in the comment made along with the sensory and technical characteristics. This need becomes evident when we ask ourselves what would be the consumer's feeling in relation to each relevant term? To solve this limitation, it is proposed to carry out a sentiment analysis of the reviews analyz~~s~~ed.

A second limitation is having used reviews of the Spanish~~ wine-loving~~ consumers, which means that the conclusions cannot be generalized to the non-Spanish ~~wine-loving~~ consumers. It would be interesting to carry out a similar study with reviews of non-Spanish wine~~-loving~~ consumers and analyz~~s~~e whether the conclusions obtained would be similar or different.

~~Finally, another future line of research would be to analyse the reviews of wine professionals and compare them with the reviews of wine-loving consumers.~~

References

Aiken, T., and Meister, C. (2018), "Applying natural language processing to the world of wine", available at http://cs230.stanford.edu/projects_spring_2018/ reports/8290440.pdf.

Akiko A. (2003), "An information-theoretic perspective of TF-IDF measures", *Information Processing & Management*, Vol. 39, No. 1, pp. 45-65, https://doi.org/10.1016/S0306-4573(02)00021-3.

Ali, H. H., Lecocq, S. and Visser, M. (2008), "The impact of gurus: Parker grades and en primeur wine prices, *Economic Journal,* Vol. 118, No.529, pp.158-173

Antonio, N., de Almeida, A., Nunes, L., Batista, F., & Ribeiro, R. (2018). Hotel online reviews: different languages, different opinions. *Information Technology & Tourism*, Vol. 18, pp. 157-185.

Ashenfelter, O and Jones, G.V. (2003), The demand for expert opinion: Bordeaux Wine, *Journal of Wine Economic,* Vol. *8,* No. 3, pp. 285-293.

~~Cai-Zhi, L.; Yan-Xiu, S.; Zhi-Quiang, W. and Yong-Quan, Y. (2018). "Research of text classification based on improved TF-IDF Algorithm", available at: https://d1wqtxts1xzle7.cloudfront.net/69479694/liu2018-with-cover-pagev2.pdf~~

Chen, B., Rhodes, C., Crawford, A., and Hambuchen, L. (2014). "Wineinformatics: Applying data mining on wine sensory reviews processed by the computational wine wheel". *IEEE International Conference on Data Mining Workshop*, pp. 142-149. doi:10.1109/ ICDMW.2014.149.

Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26 No.12, pp. 2928-2941.

Dubois, P. and Nauges, C. (2010), "Identifying the effect of unobserved quality and expert reviews in the pricing of experience goods: Empirical application on Bordeaux wine, *International Journal of Industrial Organization* Vol. 28 No.3, pp. 205-212.

Gawel, R. (2007), "The use of language by trained and untrained experienced wine tasters", *Journal of Sensory Studies*, Vol. 12 No. 4, pp. 267-284.

Hendrickx, l., Lefever, E., Croijmans, l., Majid, A., and van den Bosch, A. (2016), "Very quaffable and great fun: Applying NLP to wine reviews", In Proceedings of the 54th Annual Meeting of the Association far Computational Linguistics (Volume 2: Short Papers) (Vol. 2, pp. 306-312.

Jelodar, H., Wang, Y., Yuan, C. et al. (2019), "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey", *Multimed Tools Appl* Vol. 78, pp. 15169-15211. https://doi.org/10.1007/s11042-018-6894-4.

Katumullage, D., Yang, Ch.,Barth, J. and Cao, J. (2022), "Using neural network models for wine review classification", *Journal of Wine Economics*, Vol. 17 No. 1, pp. 27-41.

Klimmek, M. (2013), "On the information content of wine notes: Some new algorithms?" *Journal of Wine Economics*, Vol. No. 3, pp. 318-334.

Kontoya,N., de Cristofaro; P. and de Cristofaro, E. (2018), "Of wines and reviews. Measuring and modelling the Vivino wine social network, *IEEE/ACM*

International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 387-392, https://doi.org/10.1109/ASONAM.20188508776

Lemionet, A., Liu, Y., and Zhou, Z. (2015), "Predicting quality of wine based on chemical attributes" Available from http://cs229.stanford.edu/proj2015/245_report.pdf

Levinson, S. C., and Majid, A. (2014), "Differential ineffability and the senses", *Mind & Language,* Vol. 29 No 4, pp. 407-427. doi:10.1111/mila.12057.

McCannon, B. C. (2020), "Wine descriptions provide information: A text analysis". *Journal of Wine Economics,* Vol. 15 No.1, pp. 71-94.

~~Noble, A. C., Arnold, R. A., Buechsenstein, J., Leach, E. J., Schmidt, J. O., and Stern, P. M. (1987), "Modification of a standardized system of wine aroma terminology", *American Journal of Enology and Viticulture*, Vol. 38, pp. 143-146.~~

~~Noble, A. C., Arnold, R. A., Masuda, B. M., Pecore, S. D., Schmidt, J. O., and Stern, P. M. (1984), "Progress towards a standardized system of wine aroma terminology" *American Journal of Enology and Viticulture,* Vol. 35, pp. 107-109.~~

Yang, Ch., Barth, J., Katumullage D. and Cao, J. (2022), "Wine review descriptors as quality predictors: Evidence from Language Processing Techniques", *Journal of Wine Economics,* Vol. 17 No. 1, pp. 64-80.

Paradis, C., and Eeg-Olofsson, M. (2013), "Describing sensory experience: The genre of wine reviews" *Metaphor and Symbol*, Vol. 28 No.1, pp. 22-40. doi:10.1080/10926488.2013.742838

Pietsch, A. S., and Lessmann, S. (2018). Topic modeling for analyzing open-ended survey responses. *Journal of Business Analytics*, *1*(2), 93-116.

Qiang, J., Qian, Z., Li, Y., Yuan, Y., and Wu, X. (2022). "Short Text Topic Modeling Techniques, Applications, and Performance: A Survey," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp. 1427-1445, 1 March 2022, doi: 10.1109/TKDE.2020.2992485

Quandt, R. E. (2007), "On wine bullshit: Some new software?" *Journal of Wine Economics,* Vol. 2 No. 2, pp. 129-135. doi:10.1017/S1931436100000389.

Sievert, C. and Shirley, K. (2014), "LDAvis: A method for visualizing and interpreting topics". DOI: 10.13140/2.1.1394.3043.

Řehůřek, Radim and Sojka, Petr. (2010), "Software framework for topic modelling with large corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks At: Malta, pp. 45-50. DOI: 10.13140/2.1.2393.1847.

Solomon, G. (1997), "Conceptual changes and wine expertise". *Journal of the Learning Sciences,* Vol. 6 No. 1, pp. 41-60.

Shesgreen, S. (2003), "Wet dogs and gushing oranges: Winespeak for a new millennium". *Chronicle of Higher Education*, Vol. 49 No. 26, pp. 15-16.

Silverstein, M. (2006), "Old wine, new ethnographic lexicography". *Annual Review of Anthropology*, Vol. 35 No.1, pp. 481-496.

Storchmann, K. (2012), "Wine economics". *Journal of Wine Economics,* Vol. 7 No.1, pp. 1-33.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1445-1456).

Xu, K., and Wang, X. (2017), "Wine rating prediction", available at http://cs229.stanford.edu/proj2017/final-reports/5217737.pdf

Weil, R L. (2007), "Debunking critics' wine words: Can amateurs distinguish the smell of asphalt from the taste of cherries?" *Journal of Wine Economics* Vol. 2 No.2 pp. 136-144.

Wijffels, Jan, Milan Straka, and Jana Straková. "Package 'udpipe'." *Also available at https://cran. microsoft. com/snapshot/2018-06-12/web/packages/udpipe/udpipe. pdf* (2018).

Wijffels, Jan, BNOSAC and Xiaohui Yan "Package 'BTM'." *Also available at http:// https://cran.r-project.org/web/packages/BTM/BTM.pdf* (2023).

Zhao, F., Ren, X., Yang, S., Han, Q., Zhao, P., and Yang, X. (2020), "Latent Dirichlet Allocation model training with differential privacy", *IEEE Transactions on Information Forensics and Security*, Vol.16, pp. 1290-1305.

[i] The name of the websites analysed as well as the names of the selected users have not been disclosed for data protection reasons.

Figure 1: Stages of user selection and user reviews

1. Initial selection of users: Based on the ranking position of each user in the Spanish version of the website and other aspects that could be relevant for our study.

2. Selection of reviews: Selection of the reviews made by the selected users. In addition to the content of the review, information on the rating, number of likes and complete user information (name, ranking, number of followers) is collected.

3. Final selection of reviews by type of wine: With the data from step 2, and for each type of wine, a final selection will be made of the users who will participate in the study and the reviews published by them, according to the aforementioned parameters.

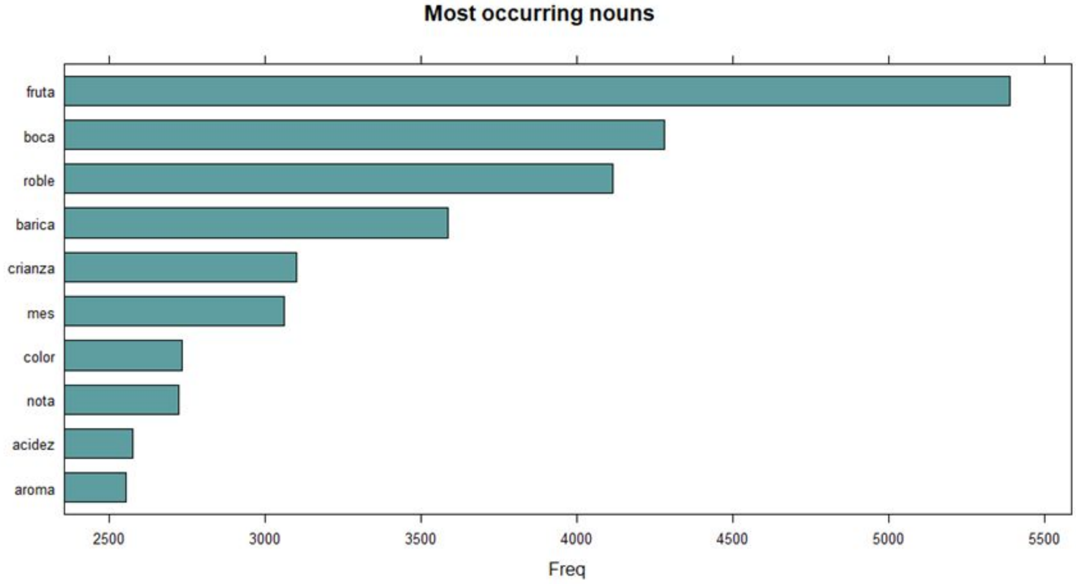
Source: Own elaboration

Table 1. List of selected users and their main characteristics

			Red wine		Rosé wine		White wine	
User	Web rank	No. of followers	No of reviews	Profile	No of reviews	Profile	No of reviews	Profile
1	-	28.981	1232	Expert	22	Amateur	110	Amateur
2	1	2.329	1464	Expert	87	Expert	528	Expert
3	2	3.227	2013	Expert	22	Amateur	532	Expert
4	5	4.021	560	Amateur	36	Amateur	360	Amateur
5	6	1.188	1193	Expert	-		-	
6	7	890	1752	Expert	-		-	
7	9	1.370	642	Amateur	15	Amateur	92	Amateur
8	14	2.413	1619	Expert	33	Amateur	803	Expert
9	15	1.635	1185	Expert	22	Amateur	565	Expert
10	16	1.184	702	Amateur	41	Expert	291	Amateur
11	17	1.190	302	Amateur	9	Amateur	91	Amateur
12	18	2.057	2315	Expert	38	Amateur	322	Amateur
13	19	2.554	1714	Expert	47	Expert	661	Expert
14	22	473	438	Amateur	8	Amateur	132	Amateur
15	24	1.334	1619	Expert	84	Expert	461	Amateur
16	31	1.669	939	Expert	56	Expert	474	Amateur
17	133.836	10.877	153	Amateur	6	Amateur	57	Amateur

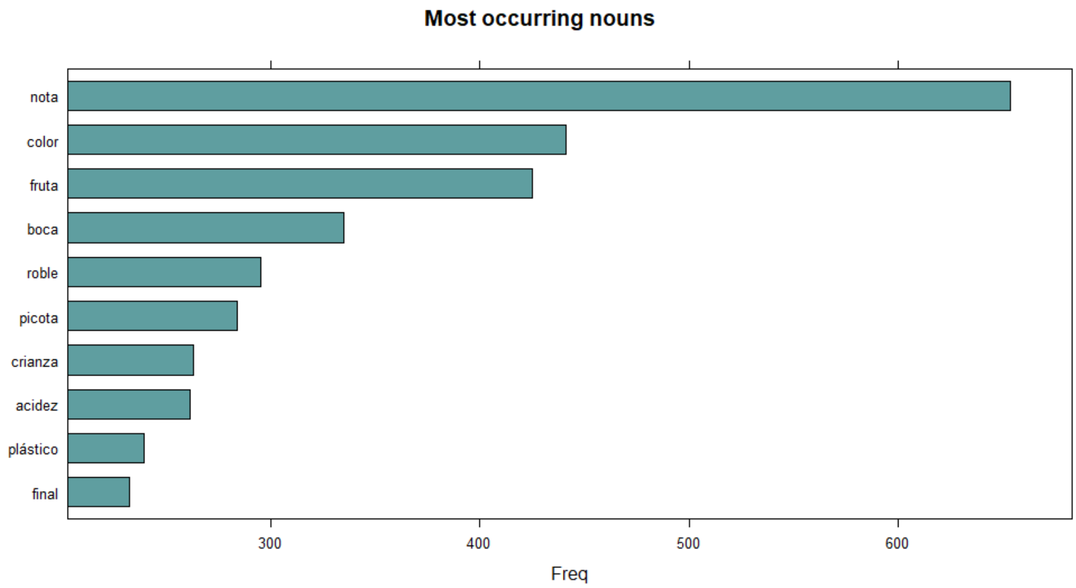
Source: Own elaboration

Graphic 1. Most occurring nouns red wine experts



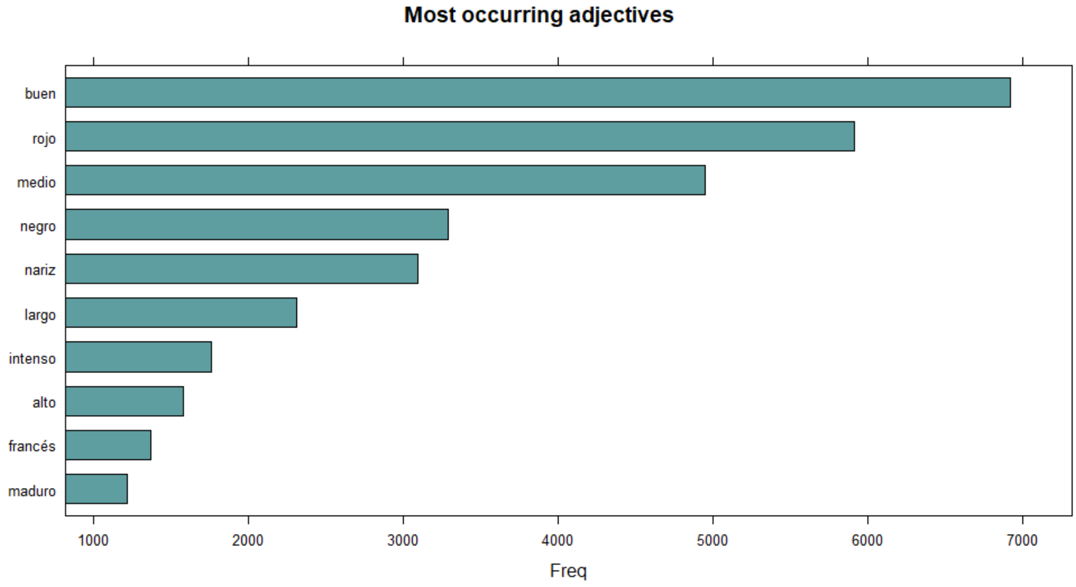
Source: Own elaboration

Graphic 2. Most occurring nouns red wine amateurs



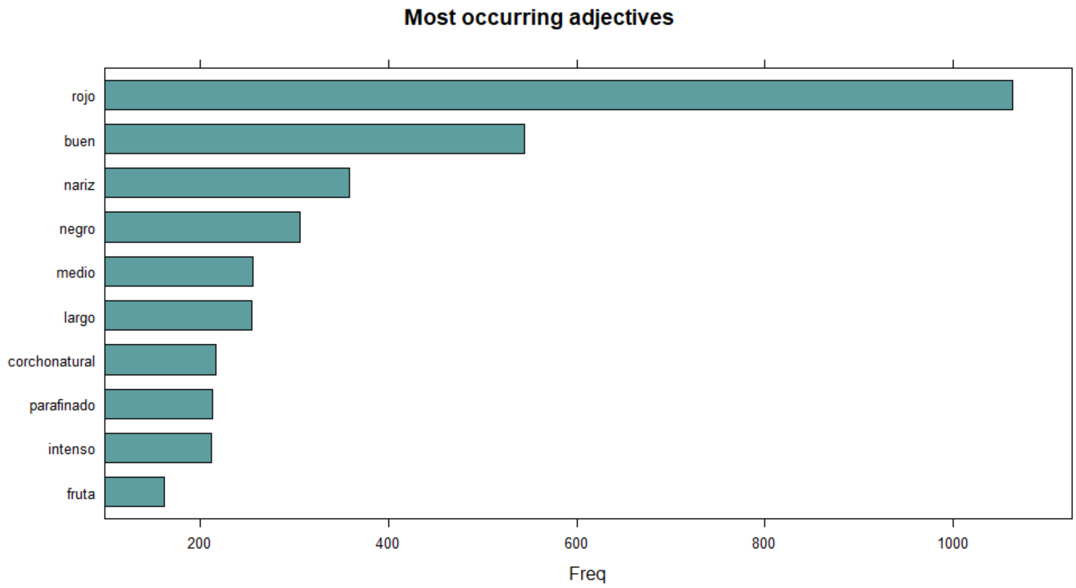
Source: Own elaboration

Graphic 3. Most occurring adjectives red wine experts



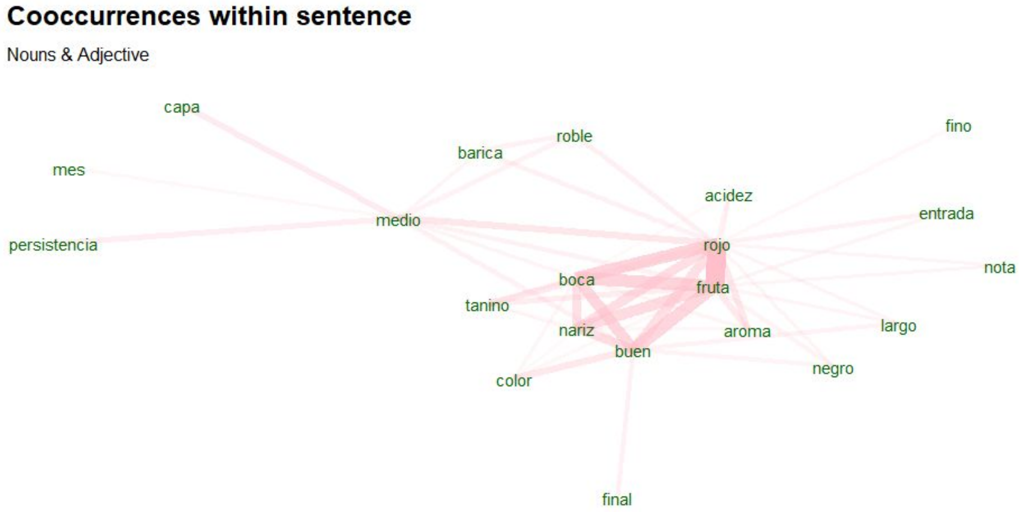
Source: Own elaboration

Graphic 4. Most occurring adjetives red wine amateurs



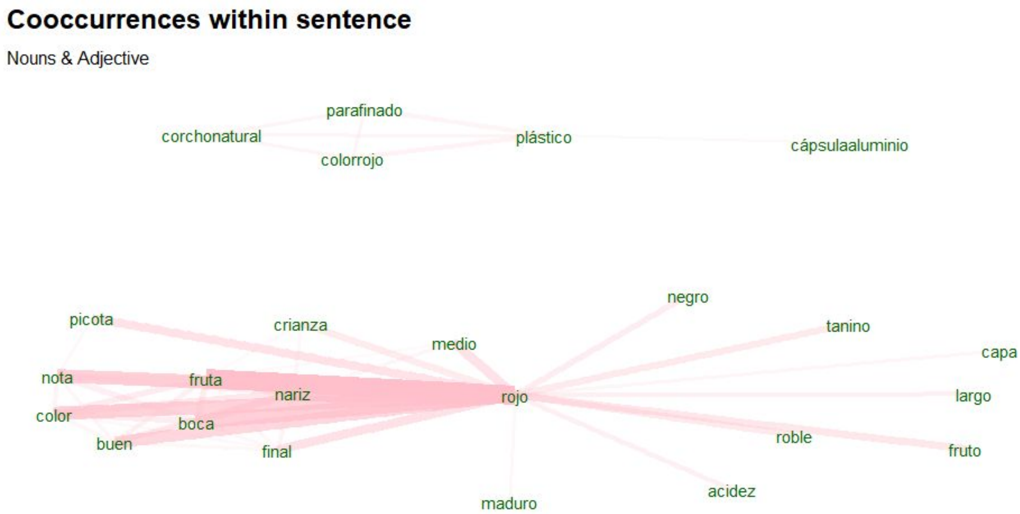
Source: Own elaboration

Graphic 5. Cooccurrences within sentence red wine experts



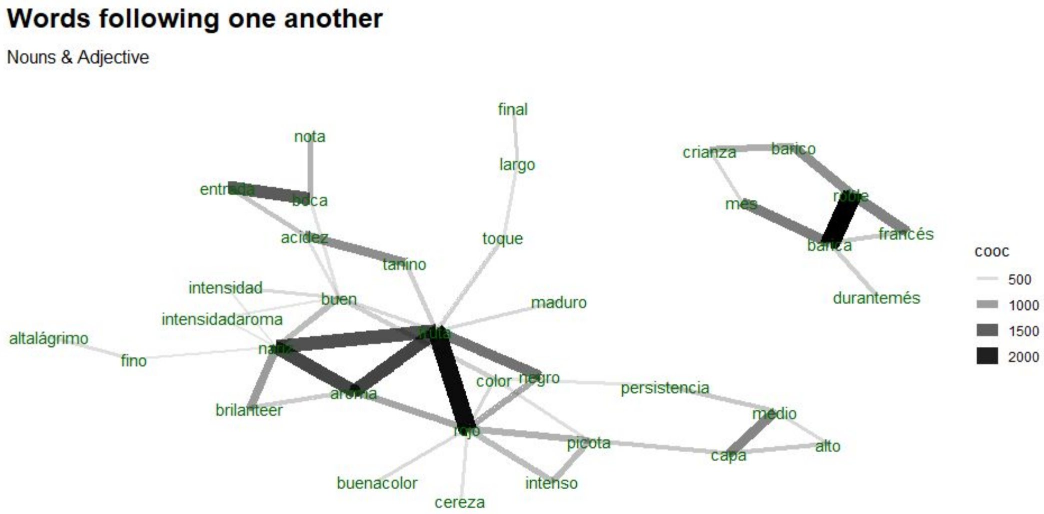
Source: Own elaboration

Graphic 6. Cooccurrences within sentence red wine amateurs



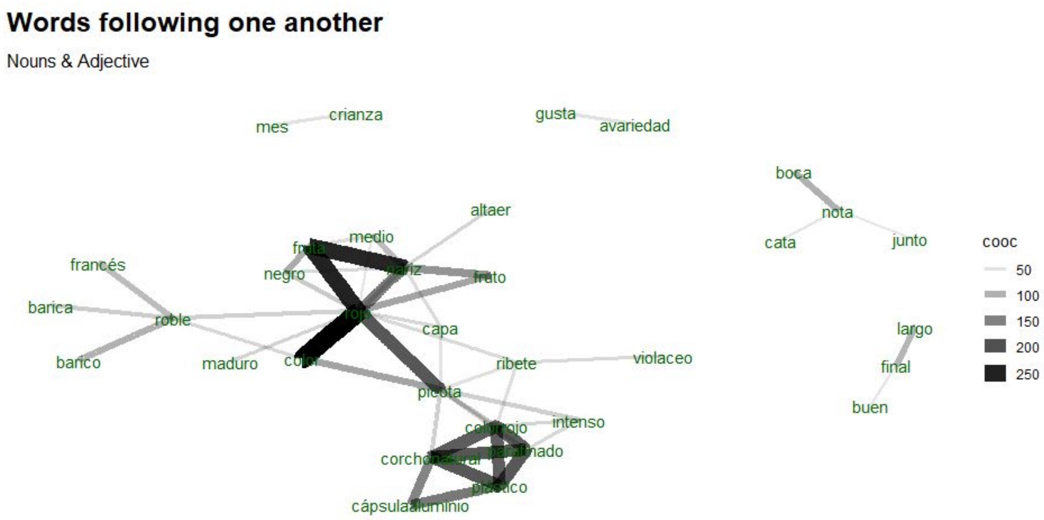
Source: Own elaboration

Graphic 7. Correlations red wine experts



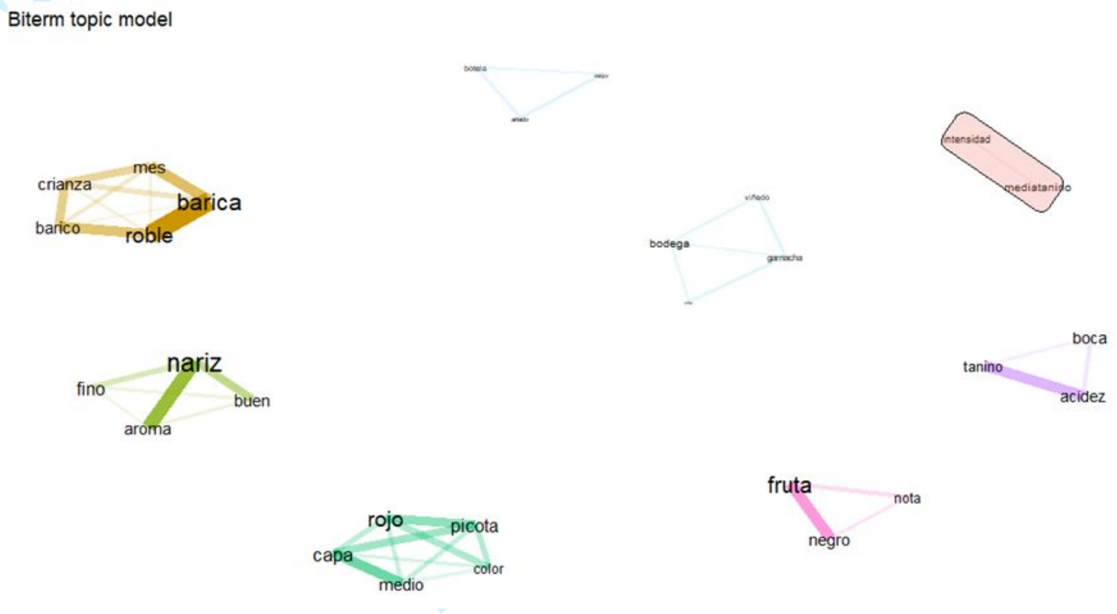
Source: Own elaboration

Graphic 8. Correlations within sentence red wine amateurs



Source: Own elaboration

Graphic 9. BTM red wine experts



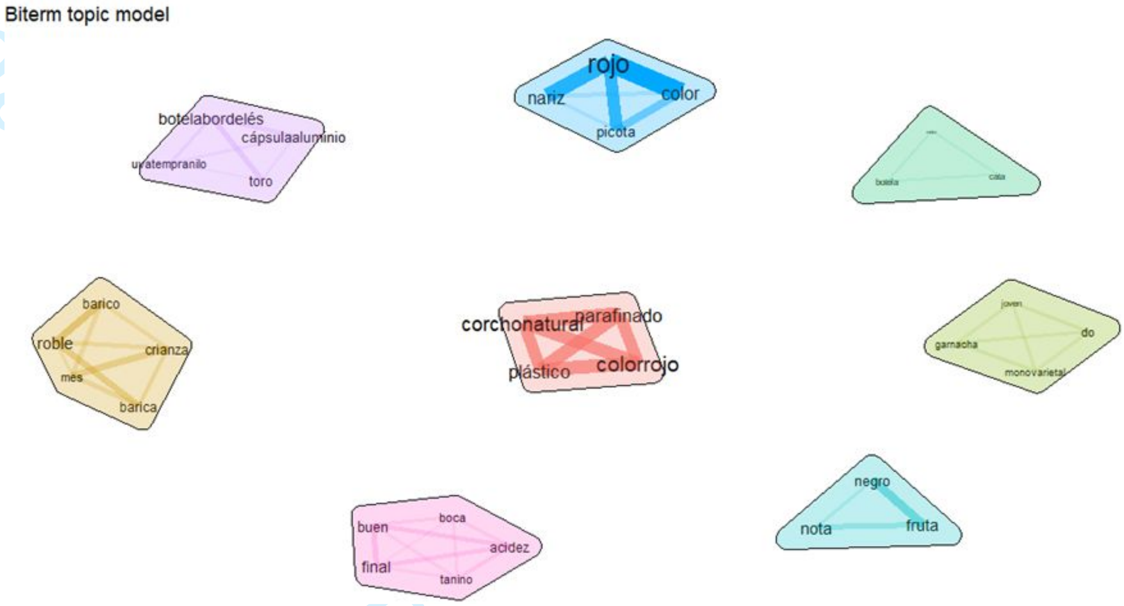
Source: Own elaboration

Table 2. BTM and terms red wine experts

	Topic 1	Prob	Topic 2	Prob	Topic 3	Prob	Topic 4	Prob	Topic 5	Prob	Topic 6	Prob	Topic 7	Prob	Topic 8	Prob
1	Rojo	0.028	Barrica	0.107	Nariz	0.129	Rojo	0.088	Bodega	0.022	Buen	0.031	Buen	0.049	Fruta	0.094
2	Media-tanino	0.023	Roble	0.091	Fino	0.584	Capa	0.069	Garnacha	0.016	Botella	0.012	Medio	0.048	Rojo	0.061
3	Medio	0.022	Barrico	0.053	Buen	0.512	Picota	0.057	Mes	0.016	Añada	0.012	Boca	0.046	Negro	0.054
4	Negro	0.020	mes	0.053	Aroma	0.048	Medio	0.052	Viñedo	0.013	Bodega	0.011	Acidez	0.044	Boca	0038
5	intensidad	0.017	Crianza	0.050	medio	0043	color	0.034	viña	0.010	mejor	0.010	tanino	0.038	nota	0.035

Source: Own elaboration

Graphic 10. BTM red wine amateurs



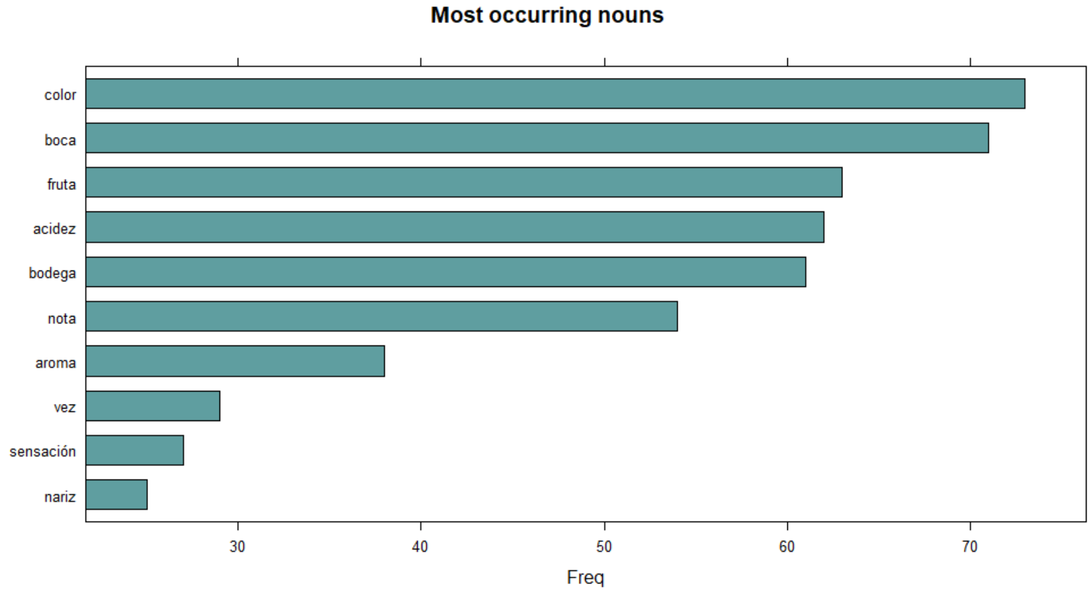
Source: Own elaboration

Table 3. BTM and terms red wine amateurs

	Topic 1	Prob	Topic 2	Prob	Topic 3	Prob	Topic 4	Prob	Topic 5	Prob	Topic 6	Prob	Topic 7	Prob	Topic 8	Prob
1	Color_rojo	0.089	Roble	0.059	D.O.	0.030	Nota	0.021	Rojo	0.063	Rojo	0.140	Botella bordelesa	0.048	Final	0.049
2	Corcho_natural	0.081	Crianza	0.037	Buen	0.022	Botella	0.018	Nota	0.060	Nariz	0.066	Cápsula aluminio	0.039	Buen	0.043
3	Parafinado	0.078	Barrica	0.036	Garnacha	0.020	Cata	0.018	Fruta	0.046	Color	0.063	Toro	0.034	Acidez	0.036
4	Plástico	0.075	Barrico	0.035	Monovarietal	0.020	Mejor	0.011	Negro	0.040	Picota	0.044	Uva_tempranillo	0.029	Boca	0.032
5	picota	0.041	mes	0.030	joven	0.018	buen	0.011	nariz	0.027	fruta	0.042	plástico	0.023	tanino	0.031

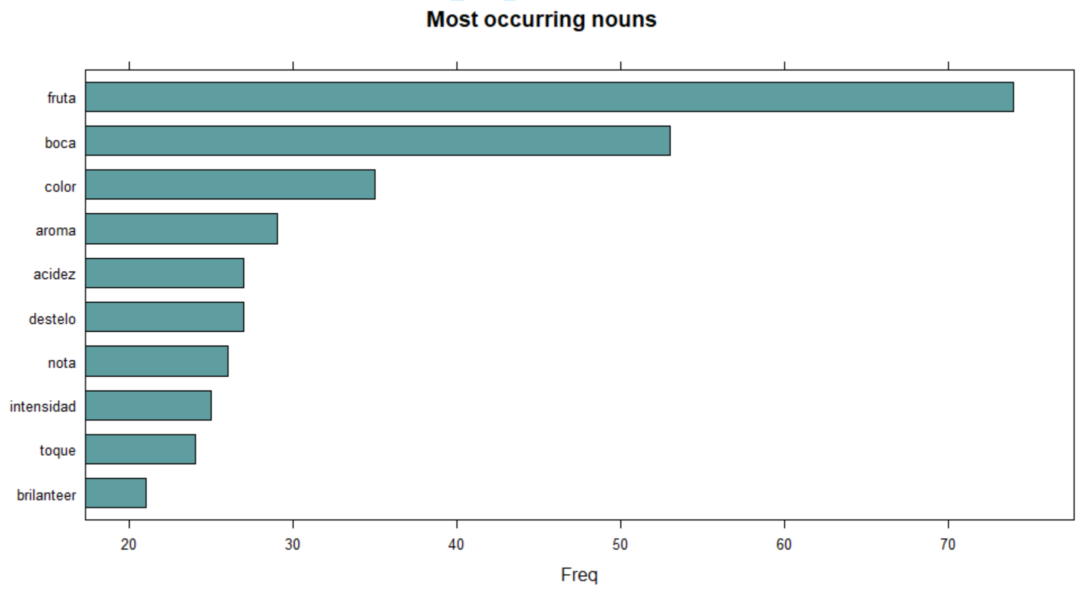
Source: Own elaboration

Graphic 11. Most occurring nouns rosé wine experts



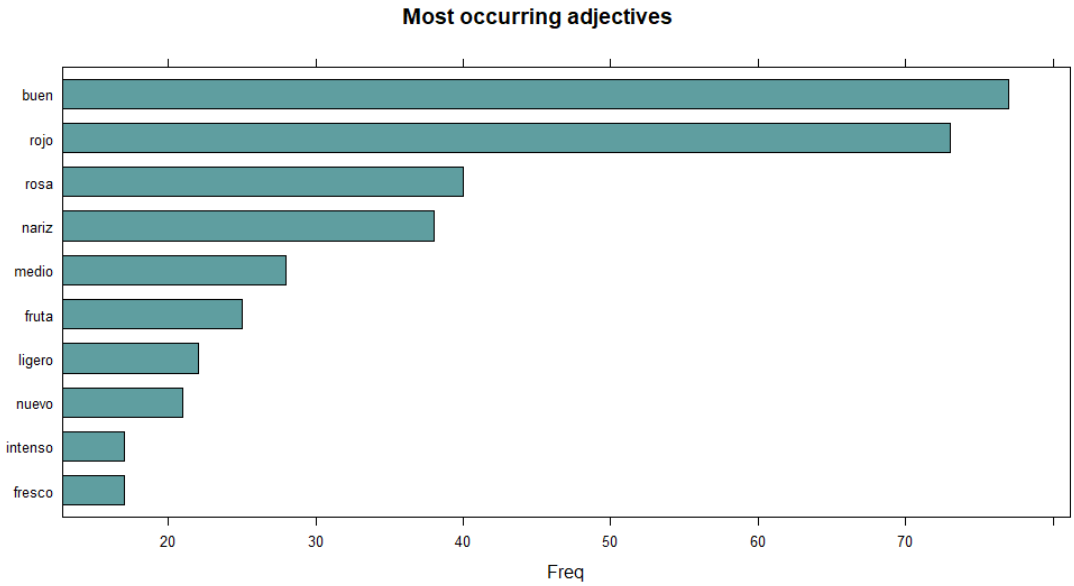
Source: Own elaboration

Graphic 12. Most occurring nouns rosé wine amateurs



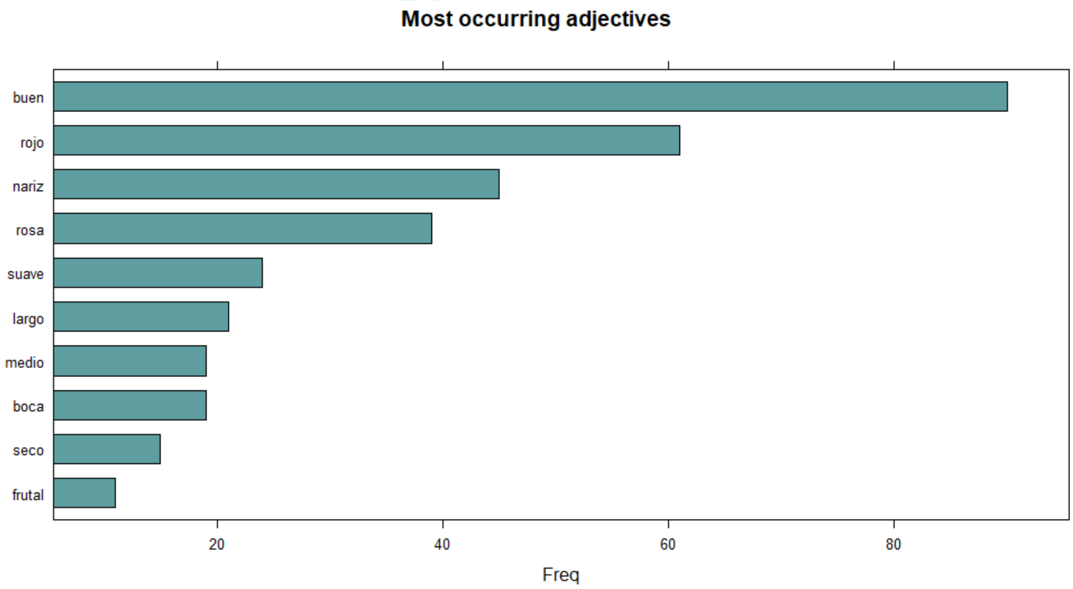
Source: Own elaboration

Graphic 13. Most occurring adjetives rosé wine experts



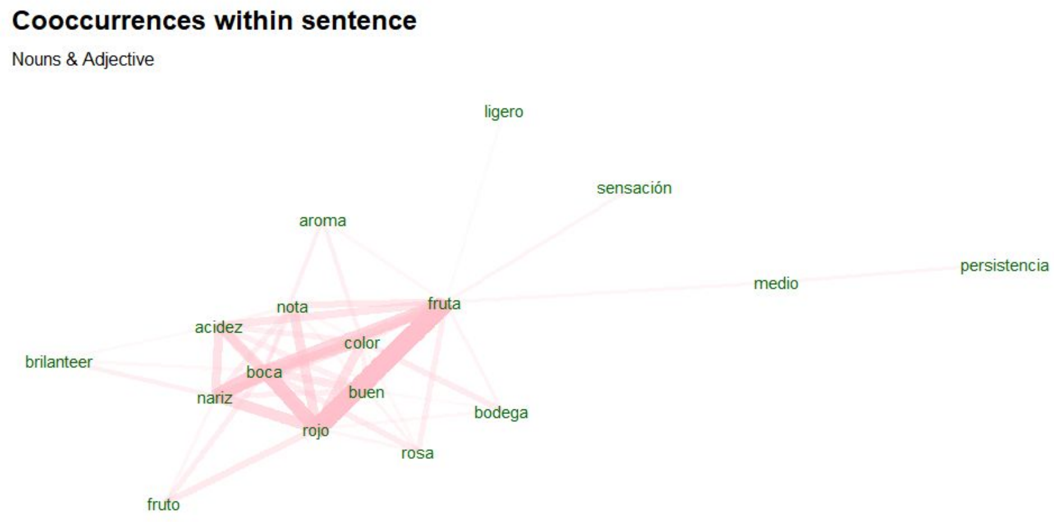
Source: Own elaboration

Graphic 14. Most occurring adjectives rosé wine amateurs



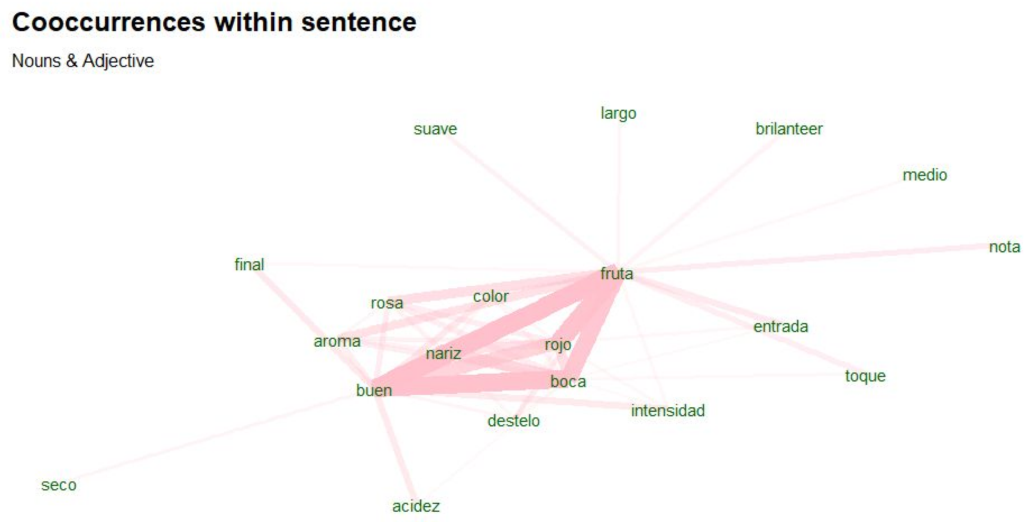
Source: Own elaboration

Graphic 15. Cooccurrences within sentence rosé wine experts



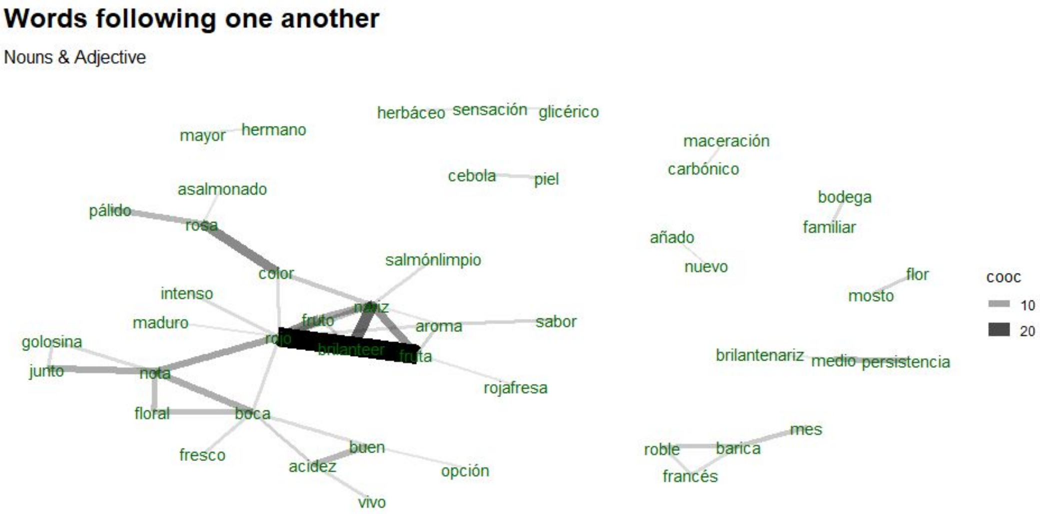
Source: Own elaboration

Graphic 16. Cooccurrences within sentence rosé wine amateurs



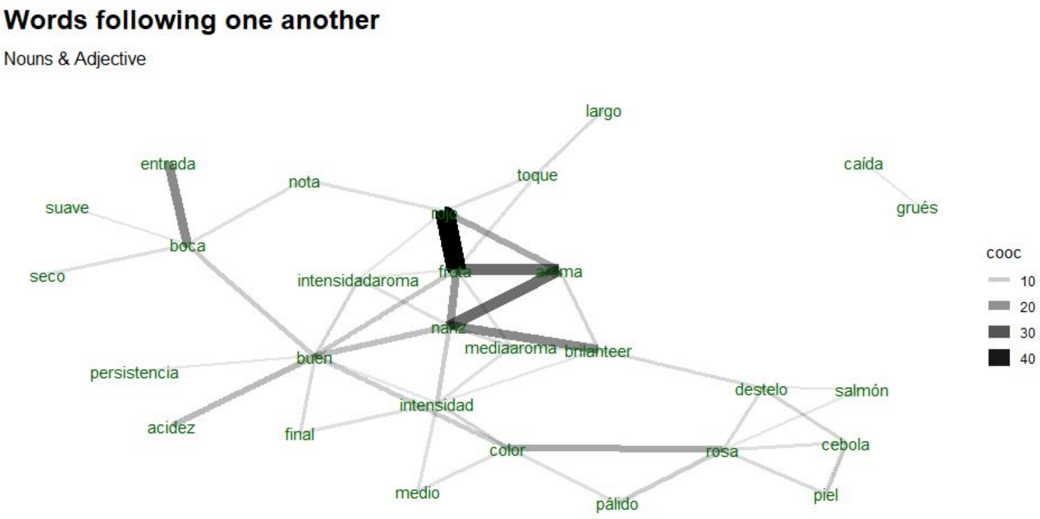
Source: Own elaboration

Graphic 17. Correlations rosé wine experts



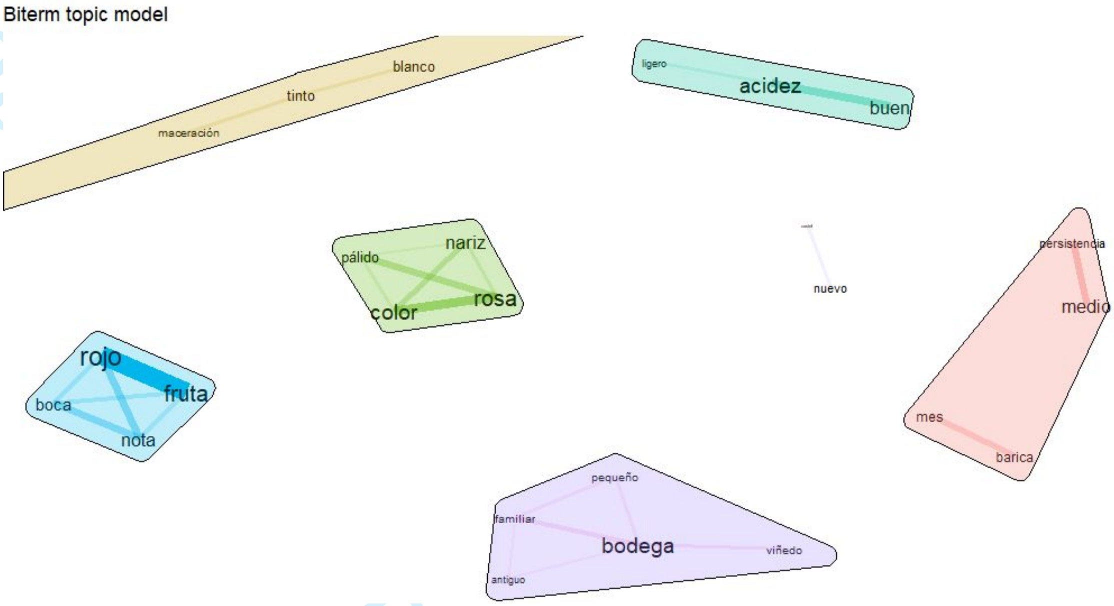
Source: Own elaboration

Graphic 18. Correlations within sentence rosé wine amateurs



Source: Own elaboration

Graphic 19. BTM rosé wine experts



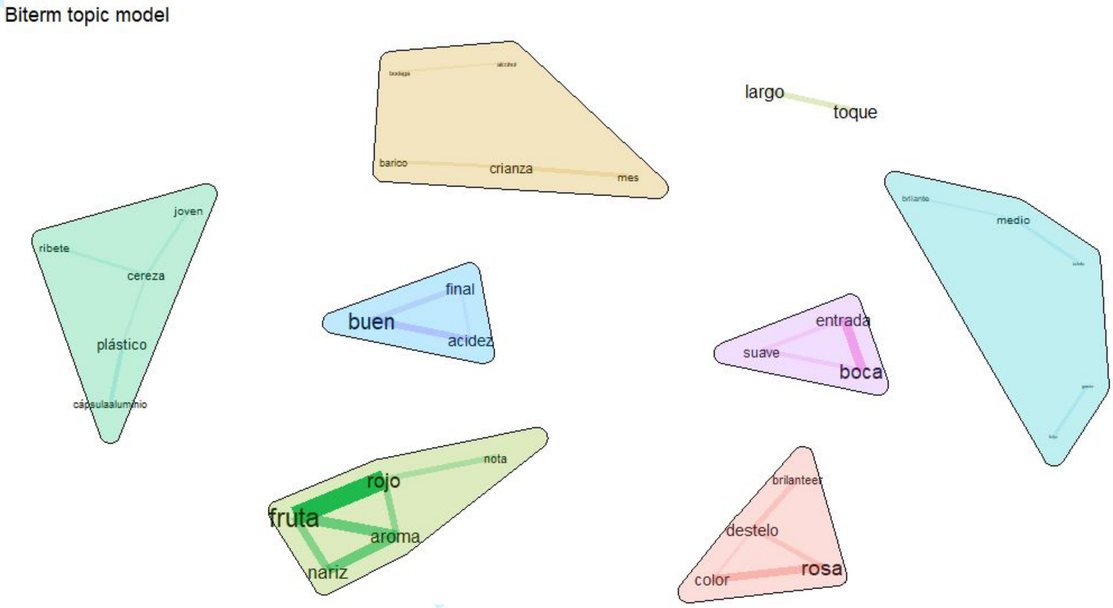
Source: Own elaboration

Table 4. BTM and terms rosé wine experts

	Topic 1	Prob	Topic 2	Prob	Topic 3	Prob	Topic 4	Prob	Topic 5	Prob	Topic 6	Prob	Topic 7	Prob	Topic 8	Prob
1	Medio	0.036	Buen	0.035	Tinto	0.027	Rosa	0.058	Acidez	0.051	Rojo	0.078	Nuevo	0.018	Bodega	0.051
2	barrica	0.023	Bodega	0.026	Blanco	0.026	Color	0.053	Buen	0.042	Fruta	0.062	Suelo	0.014	familiar	0.017
3	mes	0.023	Cata	0.017	Garnacha	0.021	Nariz	0.042	Boca	0.029	Nota	0.042	Viejo	0.012	Viñedo	0.017
4	color	0.021	Nota	0.015	Variedad	0.017	Rojo	0.031	Fruta	0.016	Boca	0.036	Buen	0.011	Pequeño	0.016
5	persistencia	0.019	punto	0.015	maceración	0.016	pálido	0.026	ligero	0.015	nariz	0.034	castel	0.009	antiguo	0.014

Source: Own elaboration

Graphic 20. BTM rosé wine amateurs



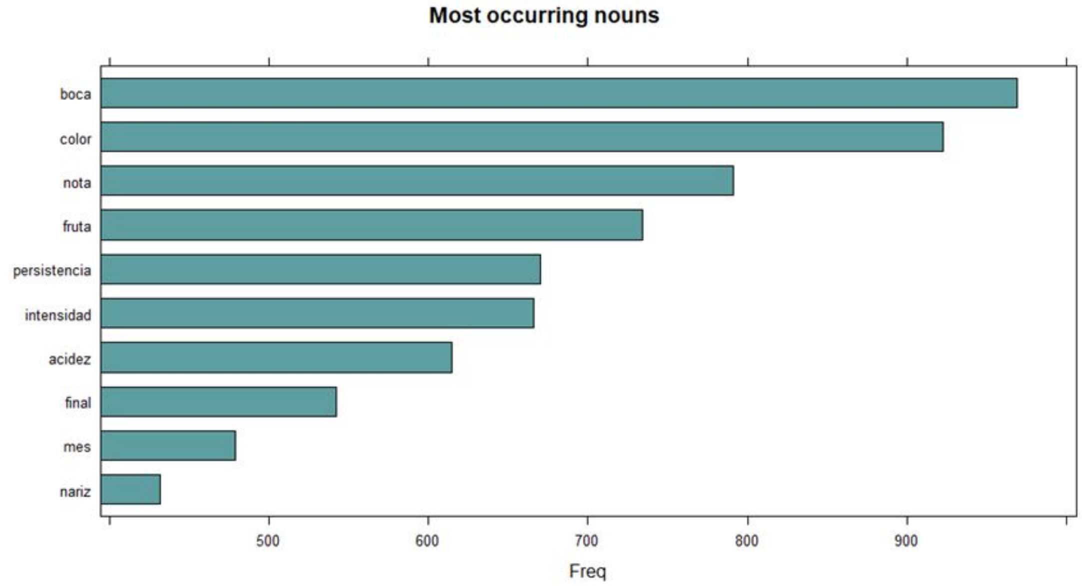
Source: Own elaboration

Table 5. BTM and terms rosé wine amateurs

	Topic 1	Prob	Topic 2	Prob	Topic 3	Prob	Topic 4	Prob	Topic 5	Prob	Topic 6	Prob	Topic 7	Prob	Topic 8	Prob
1	Rosa	0.094	Crianza	0.042	Largo	0.070	Fruta	0.141	Plástico	0.047	Medio	0.035	Buen	0.114	Boca	0.086
2	Destello	0.066	Mes	0.032	Toque	0.069	rojo	0.090	Cereza	0.040	Brillante	0.024	Acidez	0.059	Entrada	0.062
3	Nariz	0.061	Barrica	0.030	Buen	0.059	Nariz	0.082	Joven	0.032	Botella	0.021	Final	0.058	Suave	0.048
4	Color	0.053	Alcohol	0.022	Fruta	0.050	Aroma	0.072	Ribete	0.032	Gracia	0.021	Boca	0.046	Fruta	0.045
5	Brillante en	0.039	bodega	0.022	rojo	0.036	nota	0.039	Capsula-aluminio	0.031	lonja	0.021	medio	0.024	nota	0.033

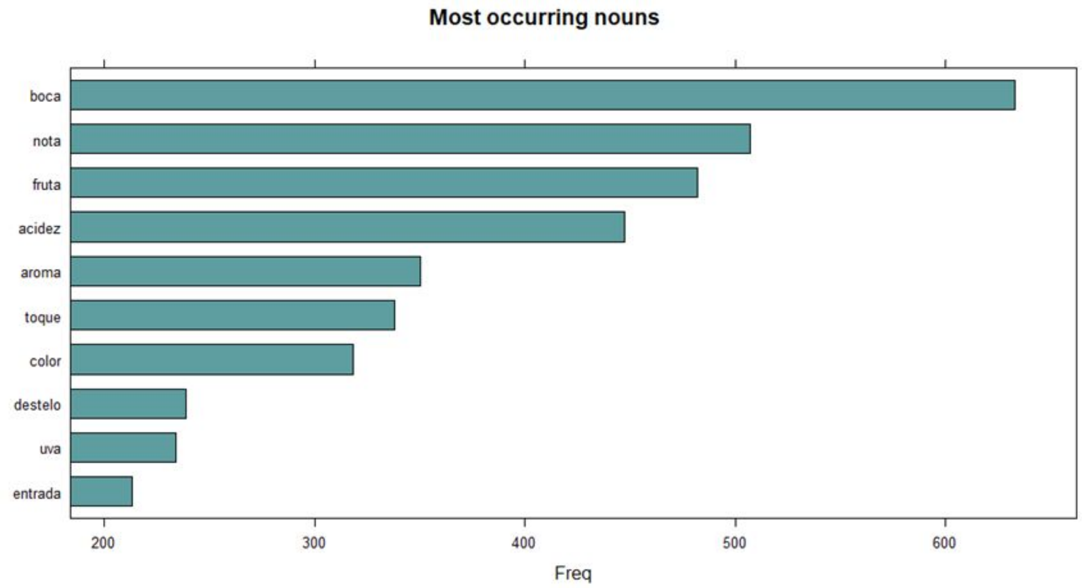
Source: Own elaboration

Graphic 21. Most occurring nouns white wine experts



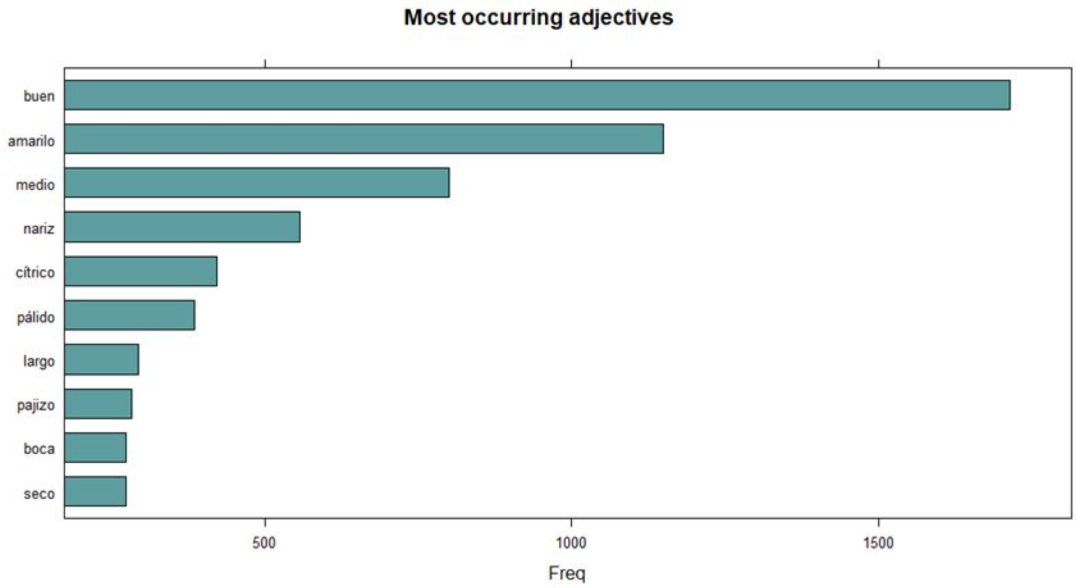
Source: Own elaboration

Graphic 22. Most occurring nouns white wine amateurs



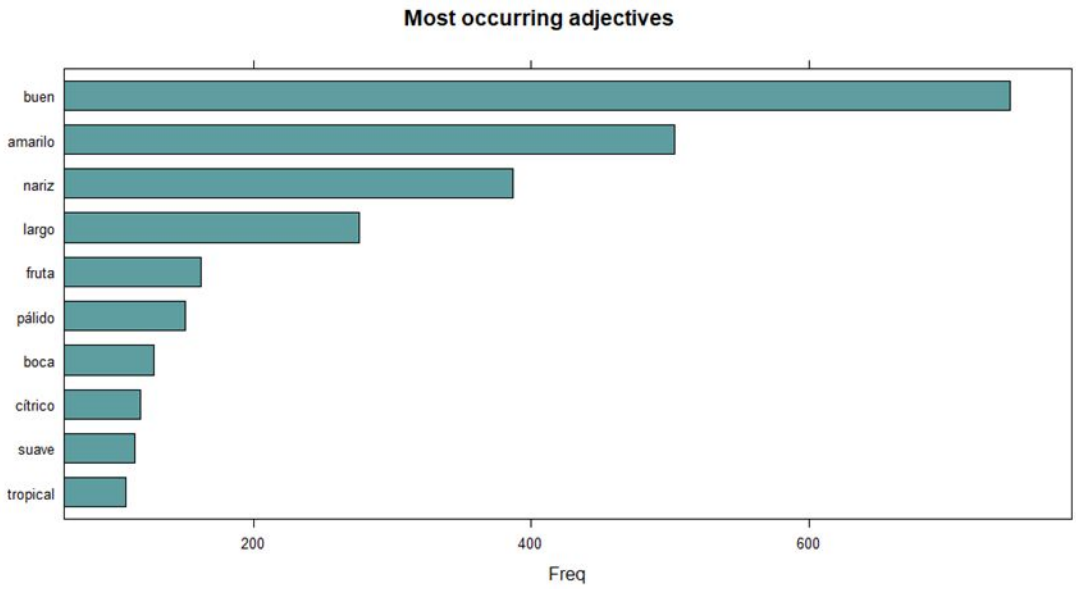
Source: Own elaboration

Grphic 23. Most occurring adjectives white wine experts



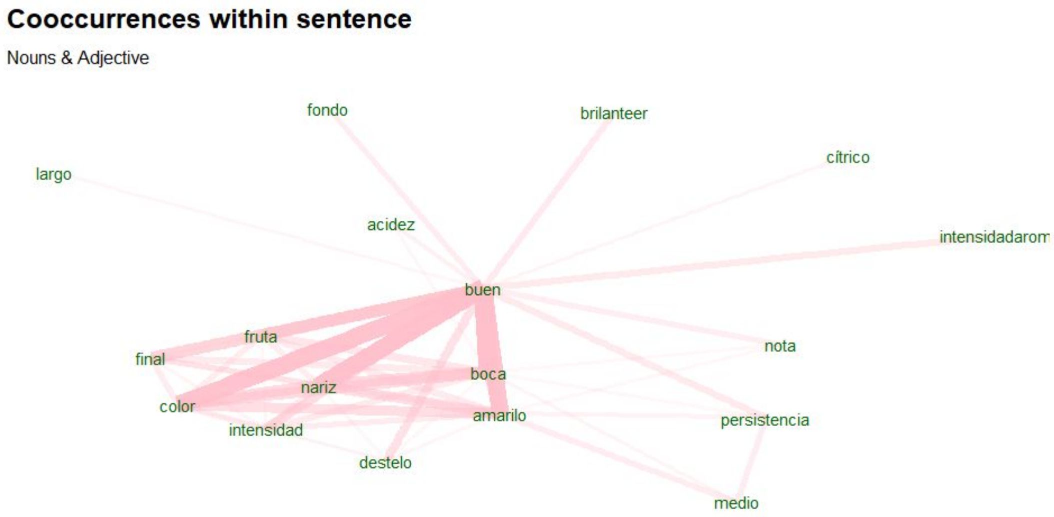
Source: Own elaboration

Graphic 24. Most occurring adjectives white wine amateurs



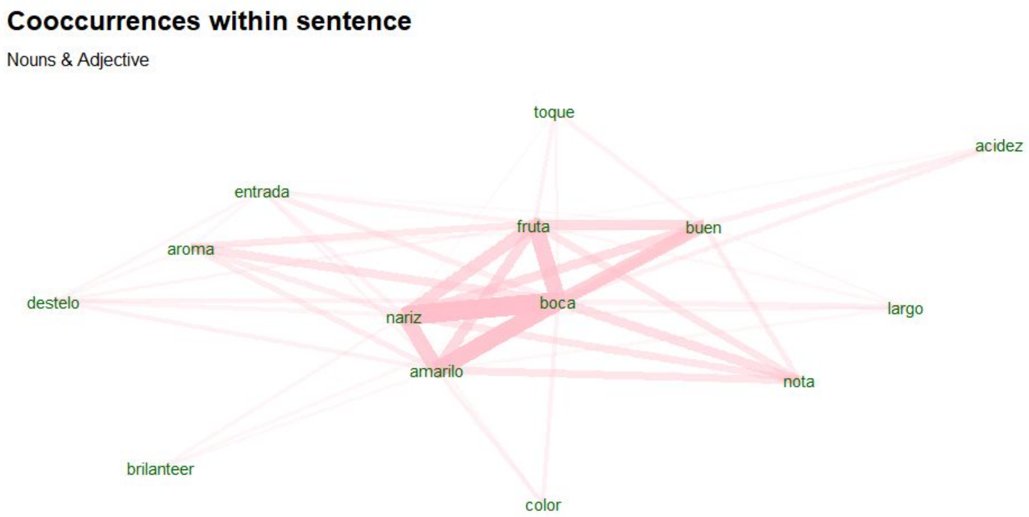
Source: Own elaboration

Graphic 25. Cooccurrences within sentence white wine experts



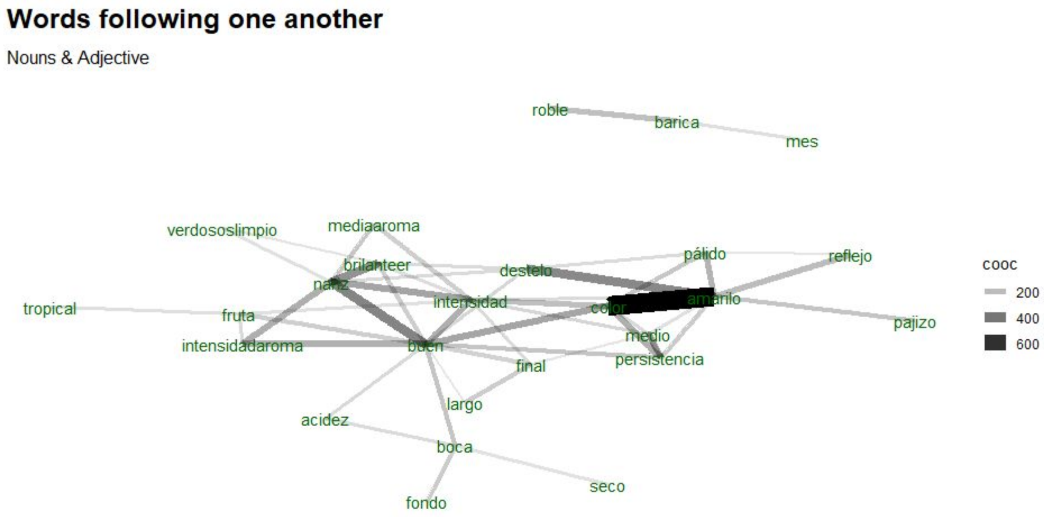
Source: Own elaboration

Graphic 26. Cooccurrences within sentence white wine amateurs



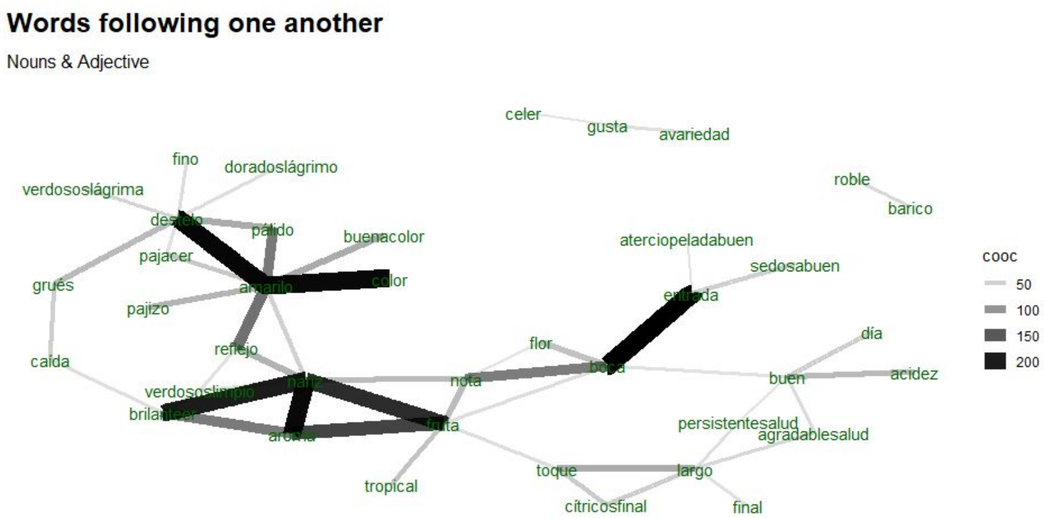
Source: Own elaboration

Graphic 27. Correlations white wine experts



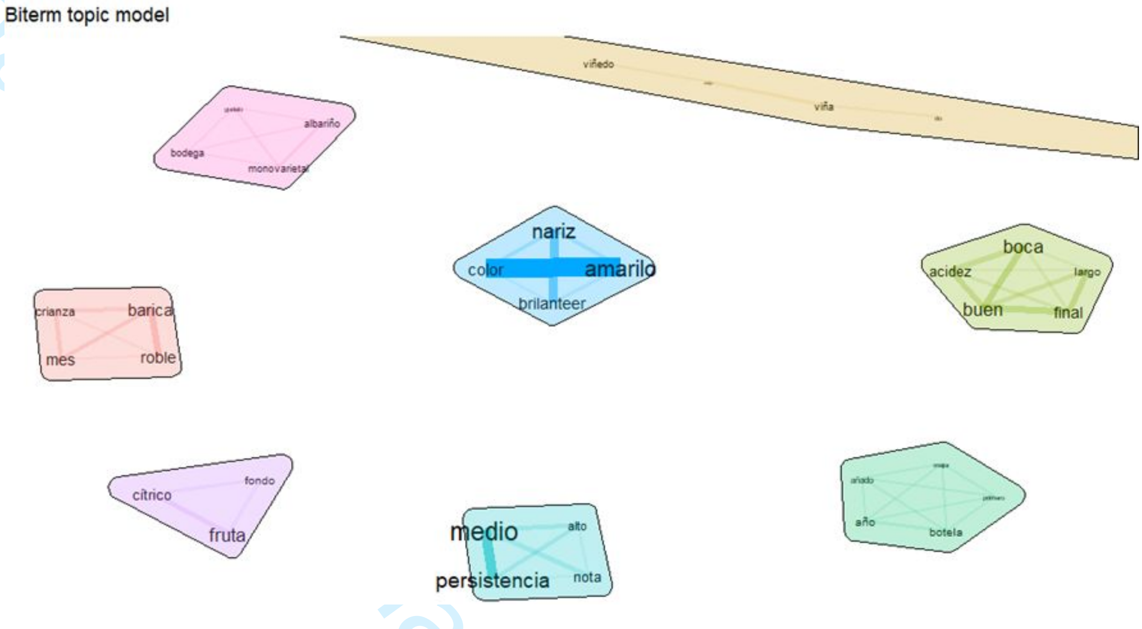
Source: Own elaboration

Graphic 28. Correlations within sentence white wine amateurs



Source: Own elaboration

Graphic 19. BTM white wine experts



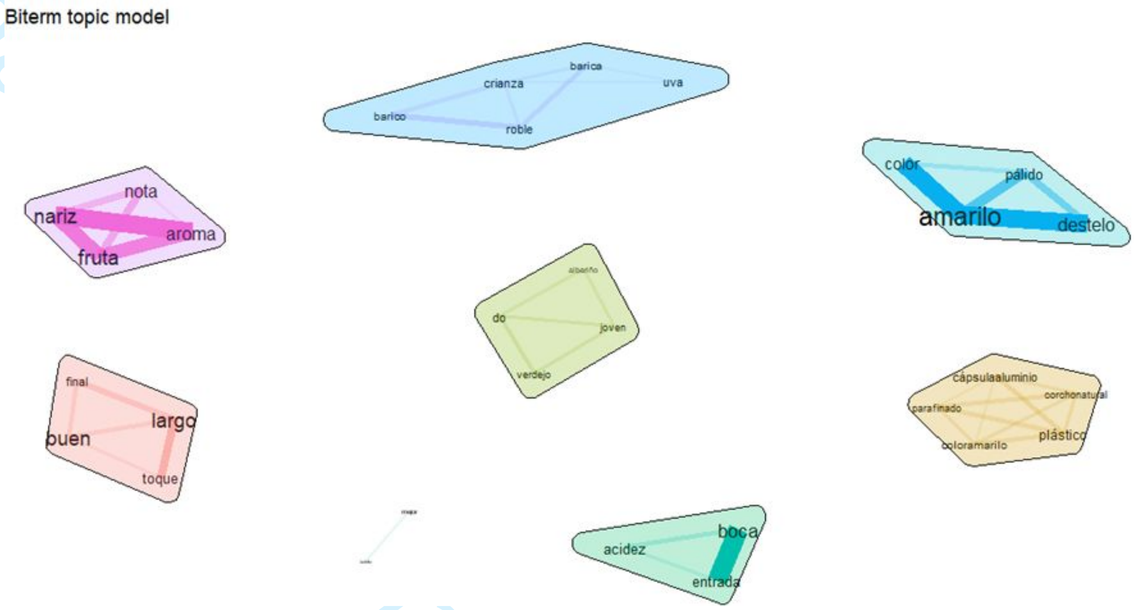
Source: Own elaboration

Table 6. BTM and terms white wine experts

	Topic 1	Prob	Topic 2	Prob	Topic 3	Prob	Topic 4	Prob	Topic 5	Prob	Topic 6	Prob	Topic 7	Prob	Topic 8	Prob
1	Barrica	0.049	Viña	0.016	Boca	0.066	Año	0.021	Medio	0.129	Amarillo	0.093	Fruta	0.063	Buen	0.018
2	Mes	0.046	Viñedo	0.016	Buen	0.066	Botella	0.020	Persistencia	0.083	Nariz	0.086	Cítrico	0.035	Albariño	0.018
3	Roble	0.045	Bodega	0.015	final	0.042	Añada	0.013	Final	0.040	Buen	0.058	Nota	0.033	Monovarietal	0.018
4	Amarillo	0.031	D.O.	0.010	Acidez	0.037	Mejor	0.010	nota	0.038	Co.or	0.050	Boca	0.030	Bodega	0.017
5	crianza	0.028	viejo	0.008	largo	0.021	primero	0.010	alto	0.024	Brillante en	0.046	fondo	0.021	godelo	0.010

Source: Own elaboration

Graphic 20. BTM white wine amateurs



Source: Own elaboration

Table 7. BTM and terms white wine amateurs

	Topic 1	Prob	Topic 2	Prob	Topic 3	Prob	Topic 4	Prob	Topic 5	Prob	Topic 6	Prob	Topic 7	Prob	Topic 8	Prob
1	Buen	0.080	Plástico	0.032	D.O.	0.028	Buen	0.021	Boca	0.070	Amarillo	0.135	Uva	0.029	Fruta	0.095
2	Largo	0.078	Cápsula aluminio	0.027	Buen	0.026	mejor	0.013	Acidez	0.042	Destello	0.062	Roble	0.028	Nariz	0.080
3	Toque	0.04	Color amarillo	0.023	Verdejo	0.021	Nota	0.011	Entrada	0.038	Color	0.049	Crianza	0.027	Nota	0.066
4	Acidez	0.031	Parafinado	0.022	Joven	0.019	Botella	0.011	Fruta	0.024	Nariz	0.047	Barrica	0.023	Aroma	0.064
5	Final	0.028	Corcho natural	0.020	albariño	0.015	gusta	0.009	buen	0.023	pálido	0.041	barrico	0.022	boca	0.037

Source: Own elaboration