Additive trees for the categorization of a large number of objects, with bootstrapping strategy for stability assessment. Application to the free sorting of wine odor terms

L. Koenig, V. Cariou, Ronan Symoneaux, C. Coulon-Leroy, Évelyne Vigneau

Additive trees for the categorization of a large number of objects, with bootstrapping strategy for stability assessment. Application to the free sorting of wine odor terms.

L.Koenig[a,b], V.Cariou[b], R.Symoneaux[a], C.Coulon-Leroy[a], E.Vigneau[b]

[a] USC 1422 GRAPPE, INRAE, Ecole Supérieure d'Agricultures, SFR 4207 QUASAV, 55 rue Rabelais 49100 Angers, France
[b] StatSC, ONIRIS, INRAE, 44322 Nantes, France

Keywords: additive tree, bootstrap, stability of categories, free sorting, wine, odor terms

Abstract

In the field of clustering techniques, little attention has been paid to the recovery of a set of clusters from the structure of an additive tree. To bridge this gap, this work presents an original partitioning technique which aims to reveal clusters from an additive tree that represents a large set of objects. Specifically, an algorithm that splits a tree into successive subtrees was developed, based on a ratio of the lengths of edges. The stability of the clusters obtained with this technique was then evaluated using measurements of cohesion and isolation that were generated using a bootstrapping strategy. Finally, the degree of association of each object to clusters was analyzed to gain insight into their internal structure. This analysis was performed on the results of a sorting task conducted by 156 subjects, who were asked to sort 96 terms associated with the odor of wine. The methodology developed in this paper represents an innovative way to highlight groups of terms within a large set of wine odor attributes, with the ultimate goal being to improve the structure of the lexicon.

1. Introduction

In sensory analysis, holistic techniques involving untrained subjects for product characterization are becoming increasingly popular due to their simplicity, speed, and cost effectiveness (Delarue et al., 2015; Varela & Ares, 2014). Moreover, the results obtained from these approaches have been shown to be largely similar to those obtained with a conventional profiling approach. Data collected from holistic techniques, such as free sorting or projective mapping, are usually recorded as a proximity matrix between the objects evaluated. Subsequently, a global representation is set up to highlight how the products are perceived by the subjects. Generally speaking, a set of objects can be visualized by means of either spatial or network representations on the basis of their proximity matrix. In spatial models, each object is represented as a point in a geometric space issued from a factorial approach; instead, in network models each object is represented as a node in a graph.

Network representations are commonly advocated when it becomes difficult to interpret a spatial configuration in a low-dimensional space. Besides, it has been shown that factorial approaches seem to be more appropriate for perceptual stimuli while network approaches, in particular additives trees, are more suited for conceptual objects, with a large number of features (Beller, 1990; Navarro & Lee, 2002; Pruzansky et al., 1982; Vanpaemel et al., 2010). As presented by Lahne (2020), Euclidean spatial approximations of non-Euclidean data (as binary sorting data) may produce apparent artifacts leading to a false interpretation. This is one of the reasons which explains the interest to alternative methods as network models. Furthermore, when a representation must include a large set of objects, the number of dimensions required in spatial methods is likely to make data exploration cumbersome (Shepard, 1974; Shepard & Arabie, 1979). Thus, in sociology or in cognition fields, network approaches are preferred as it yields better insights into the structure of the data (Abdi, 1990; Abdi et al., 1984; Chollet et al., 2011; Waterman et al., 1977).

With regard to free sorting, most data analyses involve spatial models derived from multi-dimensional scaling (Carroll & Chang, 1970; Faye et al., 2004; King et al., 1998; Lawless et al., 1995; Van der Kloot & Van Herk, 1991), correspondence and multiple correspondence analyses (Cadoret et al., 2009; Cariou & Qannari, 2018; Takane, 1981, 1982) and multiblock approaches with DISTATIS or sortCC (Abdi et al., 2007; Qannari et al., 2010). Interestingly, new approaches derived from connected graph analyses were recently proposed in the scope of holistic methods such as free sorting (Lahne, 2020) or projective mapping (Orden et al., 2019; Orden et al., 2021). In these works, a connected graph is determined so as to represent the proximity perceived among the objects by means of a strength of connections between the associated nodes of the graph. In addition, to enhance the readability of such a graph, Lahne et al. (2020) proposed a backbone extraction to represent only significant connections, i.e. those with a high strength value. Among network-based approaches, we focus on additive trees, which have been widely applied in different domains from biology to psychology. Additive trees were introduced by Buneman (1971) for filiation of manuscripts. Several algorithms were further proposed among which we can cite Sattah & Tversky (1977), Carrol & Pruzanski (1975) and De Soete (1980). For a detailed presentation of additive tree, including historical aspects, criteria, algorithms and kinds of applications, refer to Abdi (1990). Briefly, additive tree is a connected graph where every pair of objects is connected by a unique path (De Soete & Carroll, 1996 ; Sattath & Tversky, 1977). In particular, objects are represented as leaves and the path from one leave to another one reflects the dissimilarity between the two objects. This path is supported by edges connecting nodes, which can be terminal (i.e. leaves) or alternatively internal. A tree representation has the advantage to propose a parsimonious description and a convenient graphical display (Pruzansky et al., 1982). This method has been used to analyse free sorting data for semantic categorization tasks (Dubois, 2000; Koenig et al., 2020), acoustic categorization tasks (Berland et al., 2015; Guastavino, 2007), tasks related to odor categorization, both perceptually and conceptually (Chrea et al., 2004). It was also carried out in sensory studies for product characterization (Blancher et al., 2007; Chollet et al., 2011).

Along with representation techniques, clustering analysis offers a convenient way to reduce the complexity of the relationships among large sets of objects by grouping objects into homogeneous clusters. Regardless of the method used to collect information, a clustering analysis aims to group objects in such a way that similarities of objects in the same group are maximized while pairwise similarities between objects of different groups are minimized. The two techniques—similarity representation and clustering—complement each other to visualize the similarities between objects with the ultimate goal of grouping them into a small number of disjoint clusters. While the clustering of objects is currently performed using Hierarchical Ascendant Clustering (HAC) or k-means method on the initial dissimilarity matrix or dissimilarities obtained from the low-dimensional configuration, the determination of a partition from an additive tree is not straightforward. Indeed, in HAC, the splitting procedures are commonly based on the variation of an optimization criterion (or height) providing guidelines for identifying a partition. To the best of our knowledge, no clear computational strategy has yet been published with respect to the determination of clusters from an additive tree structure. To fill this gap, we propose a recursive strategy of tree rebuilding, which selects an edge to split at each step. The algorithm takes into account both the topology of the tree structure and the length of the edges between nodes.

Once the partition of the objects is obtained, its stability can be evaluated using resampling techniques. In practice, a partition is assumed to be stable if small changes in the dataset do not have any significant effect on cluster membership. Among the various resampling techniques available, bootstrapping has been widely applied in sensory profiling to assess the stability of a configuration through the generation of confidence ellipses (Blancher et al., 2012; Cadoret & Husson, 2013; Courcoux et al., 2012; Rossini et al., 2012). In the present work, a bootstrapping procedure is applied to evaluate the quality of the partition obtained from an additive tree, through the use of cohesion and isolation indices (Bel Mufti et al., 2012).

To illustrate our approach, we consider a free sorting task of 96 wine odor terms, with the ultimate goal of structuring these terms into a lexicon, based on a partition of these terms. Sorting tasks are easy to perform and involves a relatively intuitive method for assessing the similarities among a large set of objects. This approach has already been successfully used to establish relationships among terms (Gawel et al., 2000; Spencer et al., 2016). Several variations of the free sorting task have also been proposed to provide more insight into the similarities among objects, e.g., taxonomic free sorting (Courcoux et al., 2012; Withers et al., 2014), hierarchical free sorting (Cadoret et al., 2011;

Honoré-Chedozeau et al., 2017; Santosa et al., 2010) and multiple free sorting (Dehlholm, 2015). In our case study, a panel of 156 subjects performed a variation of a hierarchical sorting task that include 96 terms related to wine odors. The data collected from subjects were aggregated and represented as a dissimilarity matrix of 96 wine-odor terms. An additive tree representation was determined, which was subsequently partitioned into clusters of wine-odor terms.

This study aims to present (a) an original method for the creation of a set of disjoint clusters from an additive tree representation, (b) a way to assess the stability of the clusters, by introducing criteria of cohesion and isolation computed with a bootstrapping strategy, and (c) guidelines for the construction of a wine-odor lexicon from a free sorting task. In the first section, the case study is introduced. Then, the methodology for the determination of a partition from an additive tree is detailed together with the stability of the partition thus obtained. Subsequently, the quality of object' assignment to clusters is discussed by examining the degree of association of each object to its assigned cluster as well as to all other ones. Finally, the results obtained from the free sorting task dataset are discussed and compared to other existing representations of a wine-odor lexicon.

2. Case study

The case study pertained to the olfactory characterization of wine, and had the ultimate goal of creating a hierarchically structured sensory tool using a defined lexicon of aroma terms. A sorting task was designed to investigate the semantic relationships among the various terms. As such, this task corresponded to a word sorting procedure, as proposed by Steinberg (1967), in which subjects are asked to group together terms according to their semantic similarity. After consultation of 35 scientific papers (e.g. Coulon-Leroy et al., 2017; Esti et al., 2010; Lawrence et al., 2013; Noble et al., 1984), 96 terms were selected and constituted the lexicon submitted to the sorting task (the detailed list of terms is available in Koenig et al., 2020).

The sorting task was performed by 156 subjects, with different levels of expertise. Among them, 72% did not belong to the viticulture sector while 28% were professionals from this sector (for more details, see Koenig et al., 2020).

An ascendant hierarchical free sorting procedure (Courcoux et al., 2012; Withers et al., 2014) was adopted and designed with regard to the large number of objects considered. During the task, each subject was presented the 96 terms in the format of paper labels with "Odor of…". First, subjects were instructed to form as many groups as they wanted. Thereafter, they were invited to merge the groups they have previously identified as many times as they wanted. This additional step aimed to reveal the semantic structure of the groups that subjects had formed in the initial step. It should be noted that in this hierarchical free sorting procedure, subjects were not instructed to iterate the procedure until there were only two groups left. Besides, they were also allowed to merge more than two groups at each step. In practice, most of the subjects performed two or three aggregation steps.

For each subject, an individual distance matrix among the 96 odor terms was computed from the hierarchical free sorting data, using cophenetic distance as in Courcoux et al. (2012). The distance between two terms is equal to the first level at which they are grouped together divided by the number of levels achieved by the subject. Thus, the first grouping associated with the initial sorting procedure corresponds to level zero. At the extremes, two objects have a cophenetic distance equals to zero if they are already grouped together at level zero, while their pairwise distance is equal to one if they are never grouped together. All in all, the later the terms are grouped together, the greater their cophenetic distance. Fig. 1 depicts the way an individual cophenetic distance matrix is computed on the basis of the sorting of five objects by one subject, with a two-steps sorting procedure. At the first step, that is to say at level zero, let us suppose that the subject performs three groups: {Apple, Pear}, {Strawberry, Raspberry} and {Cherry}. This leads to a first matrix with a pairwise distance equals to zero for two objects belonging to the same group and equals to one otherwise. Then, suppose that at level one, the subject chooses to aggregate the two latest groups into a single one: {Strawberry, Raspberry, Cherry}. The resulting distance matrix is derived following the same rationale as previously. Finally, the cophenetic distance matrix is obtained by summing up the matrices corresponding to the two levels, and by dividing each value by the number of levels set by the subject (here two). Thus, pairwise distances corresponding to objects sorted together at level zero remain to zero, the cophenetic distance between 'Cherry' and 'Strawberry', grouped at level one, is equal to ½, while the cophenetic distance between 'Apple' and 'Strawberry' is equal to one since they were never grouped together.

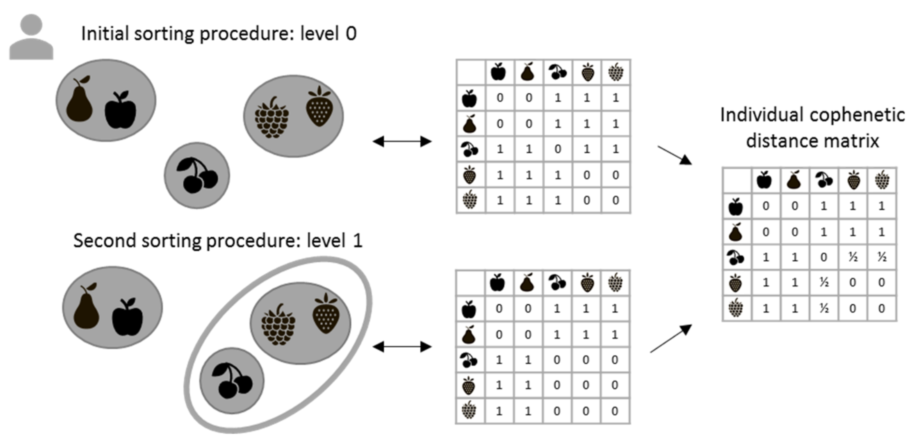

Fig. 1: Example of free sorting of five objects, for one subject. The first step of sorting task corresponds to the top of the figure with the associated distance matrix. The second step of the sorting task is illustrated on the bottom of the figure with the associated distance matrix. On the

right, the combination of the two steps into the cophenetic distance matrix for this subject is figured out.

3. Method

3.1. Notations

Suppose that S subjects have sorted N objects leading to S distance matrices as illustrated in Fig. 2(a). These latter ones are summed together to obtain an overall distance matrix between the N objects. Let us denote as $\mathbf{D}$ this $N \times N$ matrix (Fig. 2(b)), with $\mathbf{D}(i,j)$ being the distance between the pair of objects i and j (with $i,j = 1, \dots, N$). Thereafter, a representation of the N objects is carried out on the basis of distance matrix $\mathbf{D}$ by an additive tree procedure (Fig. 2(c)). From this additive tree, a partition $\mathcal{P}$ of the N objects is determined (see section 3.2 for details). Let us denote C, one cluster of the partition $\mathcal{P}$.

In order to evaluate the stability of the partition $\mathcal{P}$, a bootstrap procedure is performed. This consists in generating B virtual panels, also called bootstrap panels and by choosing S subjects at random and with replacement from the initial panel of subjects. Thus, for the b^{th} virtual panel, a distance matrix $\mathbf{D}_b$ ($b = 1, \dots, B$) can be derived. $\mathcal{P}_b$ refers to the partition of the N objects which is obtained on the basis of the additive tree set up from $\mathbf{D}_b$.

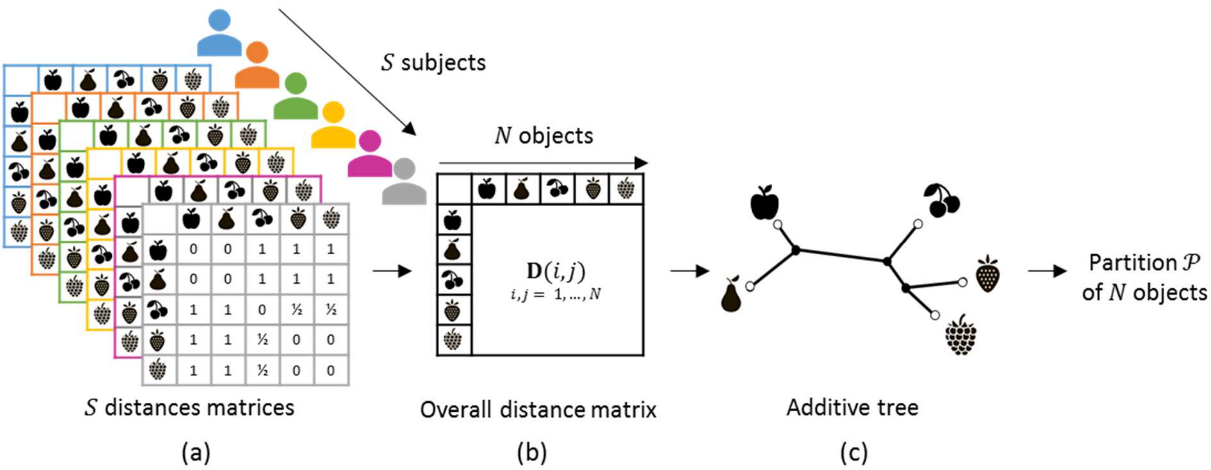

Fig. 2: (a) distances matrices from the S subjects; (b) overall distance matrix from the S subjects, and (c) the additive tree built from the overall distance matrix. The open points correspond to the objects, positioned as terminal nodes or leaves while solid points correspond to the internal nodes.

3.2. Tree and subtree construction

As described above, the overall distance matrix $\mathbf{D}$ is the sum of the individual matrices of the S subjects. The representation of the N objects is therefore carried out by approximating $\mathbf{D}$ by an additive tree. In such a representation, objects correspond to the leaves or terminal/external nodes of the tree (Fig. 2(c)). The distance in the tree between two objects is equal to the length of the path that joins their associated nodes (Sattath & Tversky, 1977). The configuration of an additive tree should approximate as closely as possible the observed distance between each pair (leaves or terminal nodes) with regard to the observed value in $\mathbf{D}$ (Abdi, 1990). Here, the Neighbor-Joining (NJ) algorithm proposed by Saitou & Nei (1987) is used to determine the additive tree, which is also called phylogenetic tree in evolutionary biology.

Once the additive tree obtained, the aim is to investigate how to cut several edges in order to get a partition of the objects. In the case study considered herein, such a partition leads to retrieve groups of odor terms that synthesized all the sorting results performed by the panel of subjects. Unlike a clustering scheme, an additive tree is not rooted, which may lead to different hierarchies of partitions or clusters. To the best of our knowledge, only a few studies have explored the determination of a partition from an additive tree and authors generally adopted an empirical approach that relies on prior knowledge (Guénoche, 2011; Sattath & Tversky, 1977). Nevertheless, these approaches share a common strategy which consists in splitting several edges of the tree to create subtrees, which ultimately provides clusters. Following the same rationale, the original divisive algorithm, proposed herein, is based on recursive steps of splitting and rebuilding. Besides, a stopping criterion is defined so that the number of clusters can be automatically defined.

The selection of the edge to be split is made on the basis of an internal criterion. In the context of phylogenetic trees, the choice is generally based on the length of edges (Sattath & Tversky, 1977), because in this context the length of the edges between two nodes can be interpreted as the evolutionary distance between species, either ancestral or extant. However, as underlined by Gambette et al. (2012), this choice may be irrelevant for other kind of trees such as those based on co-occurrence distances between words, and other criteria should be investigated. Among the criteria that were compared by Guénoche & Garreta (2002) and Gambette et al. (2012), two criteria seem to outperform the others. The first, denoted *LengthRatio*, is defined as the mean distance between all pairs in which the objects are on opposite sides of a given edge, divided by the mean distance of all pairs in which the objects are on the same side of the edge. The second, *Rtrip*, is the proportion of well-designed triples of objects. More precisely, three objects – say i, j, and z with i and j on one side of the edge and z on the other side – are said to be well-designed if i and j are closer to each other than i and z or j and z. Hereafter, we discuss only the *LengthRatio* criterion, as the results were similar with both criteria while the computational time required was much lower when using the *LengthRatio* criterion. It should be noted that this criterion can be set as an input parameter of our partitioning method which can also be applied with alternative criteria.

Using the *LengthRatio* splitting criterion, it is possible to select the edge that, once split, has the highest potential to reveal consistent subtrees. This principle is applied recursively leading to an algorithm which consists of the repetition of three main operations, described as follows:

(a) The creation of an additive tree that represents a set of objects from their associated dissimilarity matrix. In the first step, this encompasses the entire set of terms, and only a subset of terms thereafter;

(b) The division of the tree at hand into two subtrees by splitting the edge with the highest value of *LengthRatio*, thus creating two disjoint subsets of objects;

(c) The computation of the homogeneity of each subset obtained. According to Guénoche and Garetta (2002), the homogeneity of a set of objects may be evaluated by its diameter, i.e. the maximum pairwise distance within the set. These authors also advocated the use of the diameter ratio as a useful measure of homogeneity. It corresponds to the diameter of any subset of objects divided by the diameter of the whole set of N objects of the tree. The lower the diameter ratio, the more homogeneous the subset. This latter criterion is integrated in the algorithm as a stopping rule. More precisely, a threshold value equal to 80% has been set to decide if the subset under consideration should be split further or not. As in the case of the edge to be split criterion, such a threshold is an input parameter of the algorithm which can be adapted depending on the context under study or the dataset to be analyzed.

The recursive algorithm stops when there are no more candidate groups of objects for splitting, that is to say when all subsets already retrieved have a diameter ratio value lower than 80%. The partition is finally composed of the subsets of objects thus obtained.

An example of the recursive algorithm is presented in Fig. 3 with the partition of five objects. A first step leads to a five-terminal nodes additive tree that represents the global dissimilarity matrix. The edge with the highest value of *LengthRatio* corresponds to the splitting (dashed segment) of the set of objects into two groups: {Apple,Pear} and {Cherry, Strawberry, Rasberry}. Dissimilarity matrices of each subset and respective diameter ratios are thereafter computed. The stopping criterion of homogeneity makes the cluster {Apple,Pear} to be no longer a candidate group of splitting. Conversely, the subset {Cherry, Strawberry, Rasberry} is further split into two groups: {Cherry} and {Strawberry, Raspberry}. Finally, a three clusters partition: {Apple, Pear}, {Cherry}, {Strawberry, Raspberry} is obtained from this toy example.

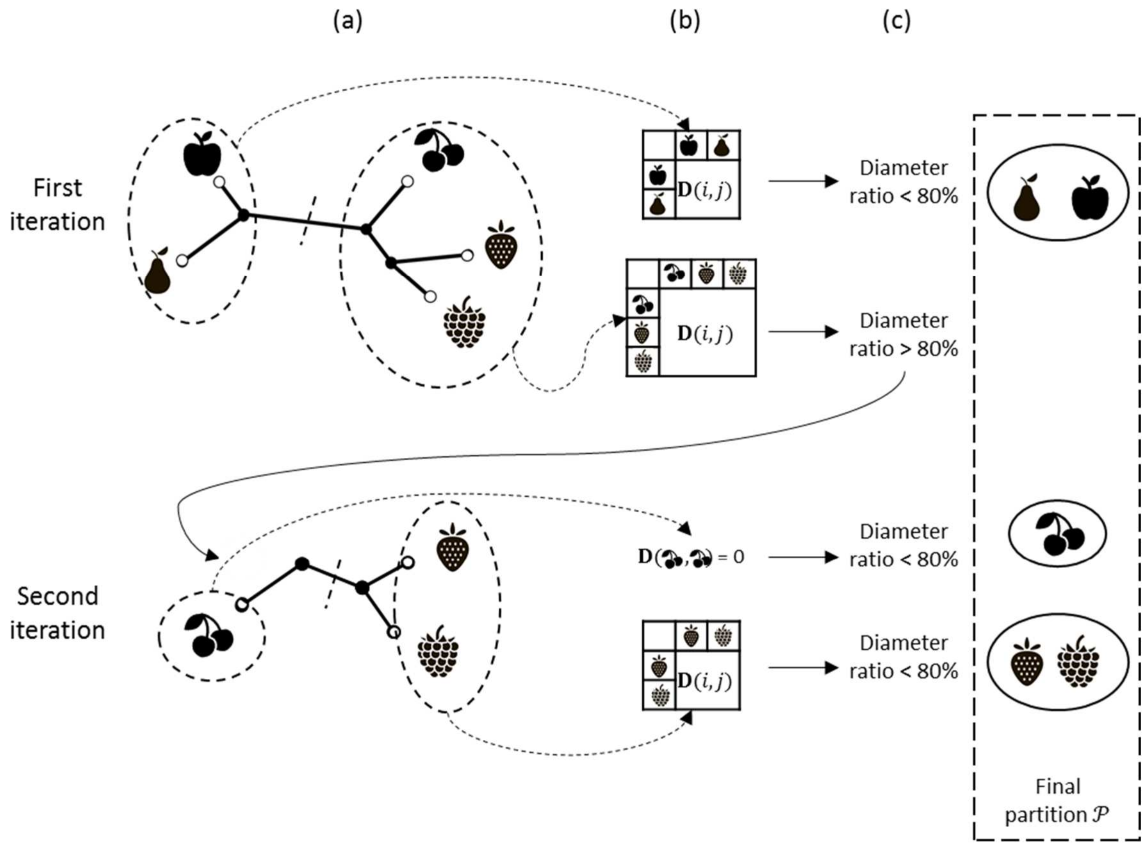

Fig. 3: Recursive algorithm with an example of five objects (Apple, Pear, Cherry, Raspberry and Strawberry) with two iterations, leading to a final partition into three clusters {Apple, Pear}, {Cherry}, {Strawberry, Raspberry}.

The algorithm is written in R software 3.4.3 and the NJ algorithm is carried out using the *nj* function in the ape package (Paradis & Schliep, 2019). The algorithm is available upon request to the corresponding author.

3.3. Assessing the stability of a partition

The partitioning algorithm is carried out on the basis of the distance matrix $\mathbf{D}$ summing up the distance matrices provided by the S subjects of the panel; this results in the determination of a partition $\mathcal{P}$. In order to assess its stability, we submit it to a bootstrapping procedure on subjects. Each bootstrap sample results from the random selection, with replacement, of S subjects from the panel. As defined in section 3.1, a distance matrix $\mathbf{D}_b$ is computed from each virtual panel of subjects ($b = 1, \dots . B$). The corresponding additive tree is submitted to the partitioning procedure, leading to the creation of a partition $\mathcal{P}_b$ ($b = 1, \dots B$). This collection of B (say 1000) "bootstrapped" partitions are then used to assess the stability of the reference partition $\mathcal{P}$. Fig. 4 presents schematically how the $\mathcal{P}_b$ partitions are obtained from a bootstrapping procedure on six different subjects.

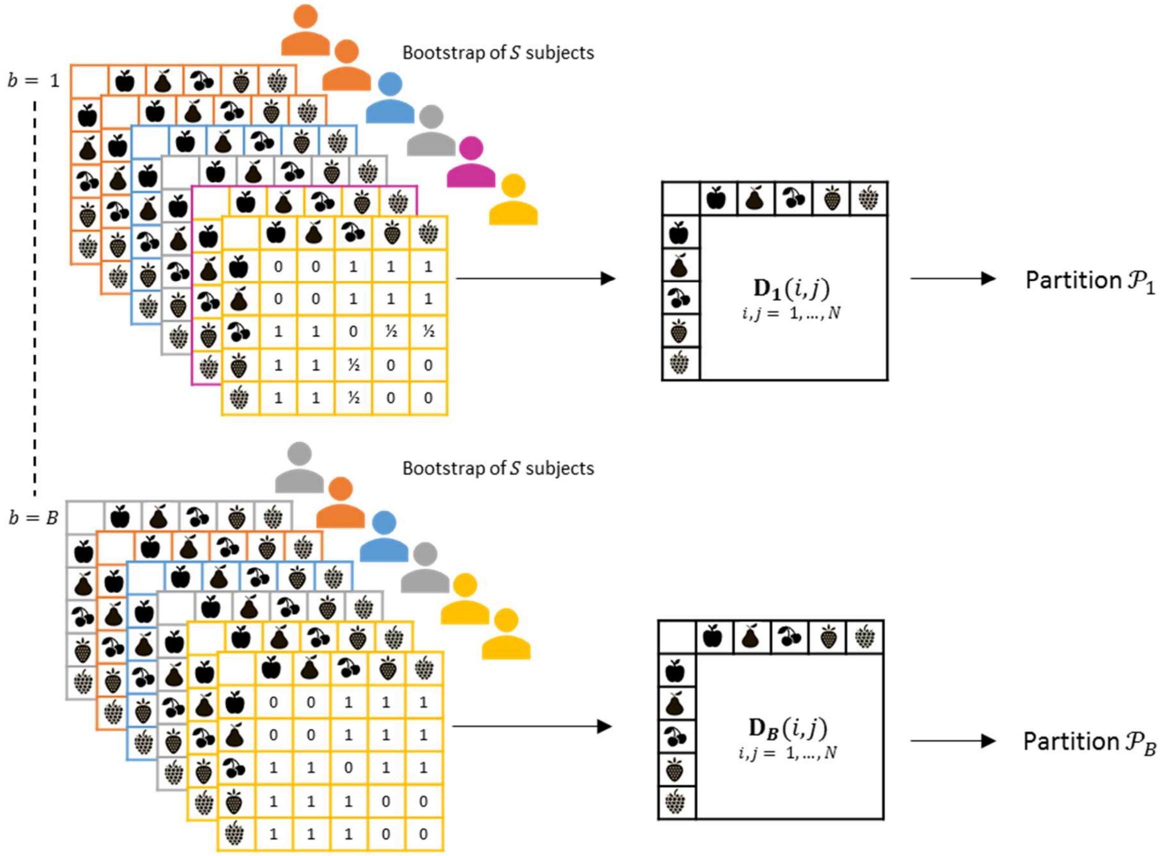

Fig. 4 : B bootstrapping within a set of six subjects (for illustration) to obtain B distance matrices and B partitions used to assess the stability of the partition

3.3.1. Cohesion and isolation of the clusters

To evaluate the stability of a partition, it is necessary to measure the stability of its clusters. For this, several authors advocated the use of cohesion and isolation measures (Bertrand & Bel Mufti, 2006; El Moubarki, 2009; Lenca et al., 2008). A cluster is considered cohesive if the objects which belong to it, remain together after a perturbation of the dataset. Here, the perturbation is generated by the bootstrap resampling. Likewise, a cluster is considered isolated if two objects separated regarding this cluster remain separated after a perturbation. According to Bertrand & Bel Mufti (2006) and El Moubarki (2009), each stability measure is defined as the probabilistic measure of a quality rule. The rules for cohesion and isolation are defined as follows:

- Rule for the cohesion of a cluster C: If two objects are grouped together in cluster C of partition $\mathcal{P}$, then they must be together in partition $\mathcal{P}_b$ obtained after perturbation.

- Rule for the isolation of a cluster C: If two objects are separated according to the partition $\{C, \bar{C}\}$ of $\mathcal{P}$ ($\bar{C}$ being the complement of C), then they must also be separated in partition $\mathcal{P}_b$ obtained after perturbation.

The probabilistic measures for both rules are respectively evaluated by:

$$Co_C = \frac{1}{B}\sum_{b=1}^{B}\frac{n_{11,b}^{C}}{n_c(n_c-1)/2} \tag{1}$$

$$Is_C = \frac{1}{B}\sum_{b=1}^{B}\frac{n_{00,b}^{C}}{n_c(N-n_C)} \tag{2}$$

with $n_{11,b}^{C}$ the number of pairs of objects of cluster C of $\mathcal{P}$ that are also grouped together in partition $\mathcal{P}_b$ $(b = 1, \ldots, B)$; $n_{00,b}^{C}$ the number of pairs of objects separated into $\{C, \bar{C}\}$ and also separated in partition $\mathcal{P}_b$; n_C the size of cluster C and N the total number of objects.

Cohesion and isolation measures (Eq. (1) and (2)) range in value between 0 to 1; the lower the value, the lower the cohesion (isolation) of the cluster. It is worth noting that these indices rely on the principle of conditional probability (Agrawal et al., 1993).

Cohesion and isolation measures are computed for each cluster C of the partition $\mathcal{P}$. Subsequently, weighted sum of cluster's cohesion and isolation indices are used to estimate respectively the cohesion and the isolation of the global partition $\mathcal{P}$. Weights are introduced in order to take into account of the size of the different clusters, as reported in Appendix (Eq. (4) and Eq. (5)). Further details are given in El Moubarki (2009).

As demonstrated by Bel Mufti et al. (2012), there is a tight connection between these indices and the Rand index. Indeed, the Rand index can be expressed on the basis of both cohesion and isolation measures. For more details, the equations defined by El Moubarki (2009) are reported in Appendix (Eq. (6)). In the context of free sorting data analysis, the Rand index is commonly used either to evaluate the agreement between two subjects' partitions or as the key criterion to determine a consensus partition of the objects from the various subjects' partitions (Courcoux et al., 2014 ; Qannari et al., 2014). Undeniably, the Rand index could also be used to compare the observed partition $\mathcal{P}$ with the "bootstrapped" ones, $\mathcal{P}_b$. However, while the Rand index only operates at a partition level, cohesion and isolation measures provide a characterization at both levels, clusters and partition, giving better insights into the structure, as shown in Eq. (6).

As we seek herein a fine analysis, criteria of cohesion and isolation are retained to characterize the stability of each cluster. Indeed, cohesion yields information on the stability of the core of each cluster: a highly cohesive cluster lumps together a strong core of objects, similar to the concept of a "strong pattern" (Diday & Simon, 1976). In addition, the isolation measure evaluates how separate one cluster is from others: in a poorly isolated cluster, the object(s) can often be moved out and assigned to other clusters.

At this stage, the analysis of cohesion and isolation values provides information on the stability of each cluster but it does not enlighten on the objects responsible for such a level of stability.

3.3.2. Investigating the degree of association of objects to clusters

To obtain better insight into the partition $\mathcal{P}$, a deeper evaluation was made to analyze how strongly an object i is related to the cluster it belongs to.

For this purpose, the empirical degree of association of an object i to a cluster C has been computed, using the same rationale of the B boostrapped partitions $\mathcal{P}_b$ that reflect the perturbations of the dataset. Let us denote as $f_{i,j}^{B}$ the estimation of the degree of association between the objects i and j ($i \neq$ j), i.e. the frequency of merging together i and j across the $\mathcal{P}_b$ partitions ($b = 1, \ldots . B$). These values are aggregated in reference to each cluster of the initial partition $\mathcal{P}$ in order to compute the $D_{i,k}$ criterion, defined as follows, for the k^{th} cluster C of $\mathcal{P}$:

$$D_{i,k} = \frac{1}{n_k} \sum\nolimits_{j \in C, j \neq i} f_{i,j}^{B} \qquad \text{if } i \notin C \qquad (3a)$$

$$D_{i,k} = \frac{1}{n_k - 1} \sum\nolimits_{j \in C, j \neq i} f_{i,j}^{B} \qquad \text{if } i \in C \qquad (3b)$$

where n_k is the size of the k^{th} cluster of $\mathcal{P}$.

For an object i, the more concentrated its distribution of $D_{i,k}$ (for k = 1, …, K) is, the higher its specificity (or typicality) to the cluster it belongs to.

4. Results

4.1. Tree representation and clustering of odor terms

The term-by-term matrix $\mathbf{D}$, computed from the set of the 96 wine-odor terms on the basis of the 156 subjects' sorting data (section 3.1), was subjected to the additive tree (NJ) algorithm. The resulting unrooted tree is presented in Fig. 5.

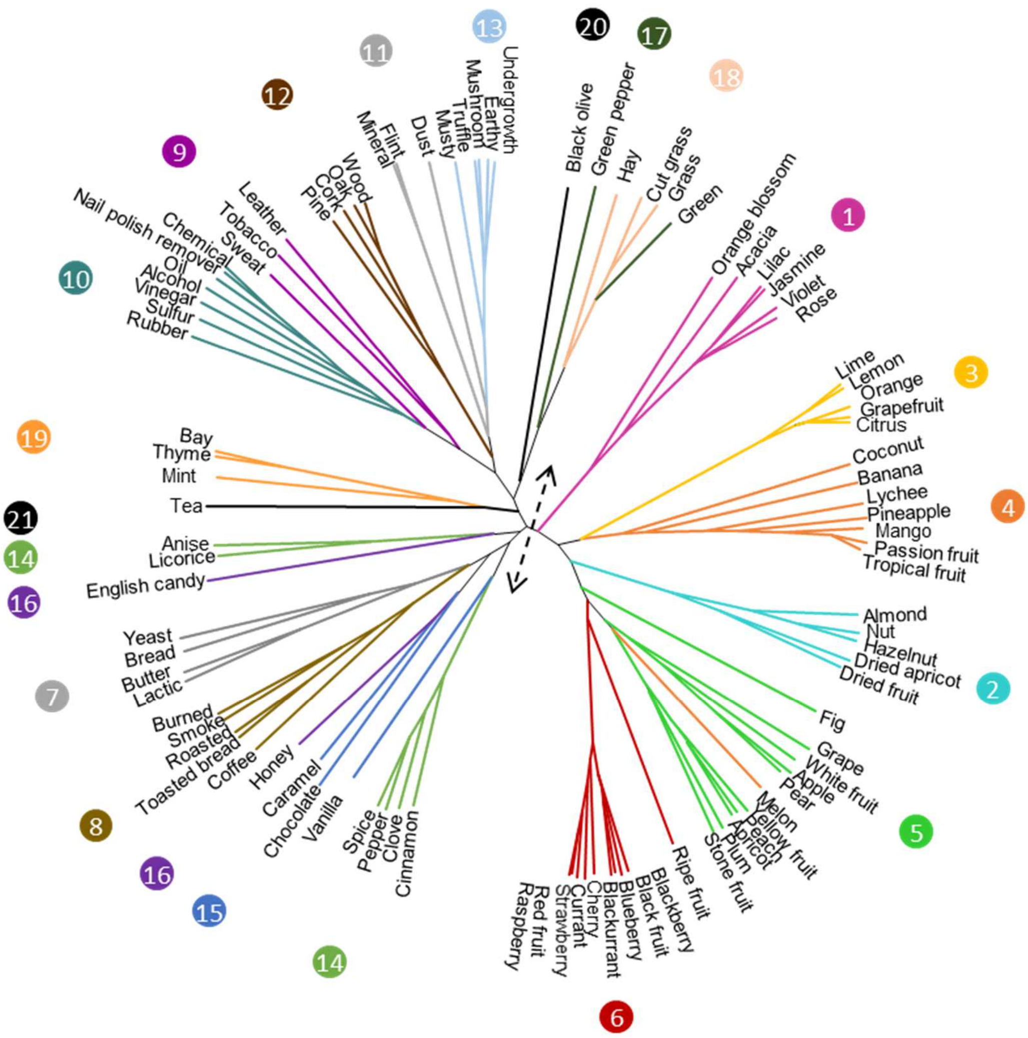

Fig. 5: Additive tree carried out from distance matrix $\boldsymbol{D}$ *of the sorting data associated with wine-odor terms. The first edge split by the recursive partitioning algorithm is indicated with a dotted double arrow. Colors represent the 21 clusters recovered by the sequential partitioning algorithm along with their clusters' number as listed in Fig. 6.*

The *LengthRatio* criterion was calculated for all edges of the additive tree. Then, the edge with the highest value of *LengthRatio* was split, leading to the creation of two subtrees. From this split (see the dotted double arrow in Fig. 5), a group of 44 odor terms was distinguished from another group of 52 odor terms. The splitting strategy described in section 3.2 was recursively applied. The sequence of the clusters thus obtained is schematically represented in Fig. 6. With a threshold of 80% for the stopping criterion (diameter ratio), the partition of the odor terms was ultimately composed of 21 clusters (Fig. 6). Each cluster varied from 1 to 10 odor terms.

Fig. 6: Schematic representation of the successive dichotomous splitting of the 96 odor terms, leading to a partition into 21 clusters (stopping criterion threshold value equal to 80%).

First of all, it appears that some clusters that were highlighted by the algorithm corresponded to well-identified branches of the global additive tree (e.g., clusters 1, 2, and 7 in Fig.5). Nevertheless, some situations did not follow the same rationale and turned out to be more complex. For instance, the term 'Vanilla' (dark blue, southwest position in Fig. 5) was linked in the additive tree to the branch that represented the terms 'Spice', 'Pepper', 'Clove', and 'Cinnamon' (dark green, south-southwest position in Fig. 5). However, the recursive splitting algorithm grouped 'Vanilla' with 'Caramel' and 'Chocolate' in cluster 15 (dark blue, southwest position in Fig. 5). These differences can be explained by the fact that the proposed algorithm determines the tree again at each step.

4.2. Cohesion and isolation of the partition

In order to assess the stability of the 21 clusters corresponding to partition $\mathcal{P}$, B bootstrapped samples of subjects were generated and B partitions, $\mathcal{P}_b$, (b =1…, B, with B = 1000) were computed. Each $\mathcal{P}_b$ partition contained between 17 and 27 clusters.

Values of cohesion and isolation indices for each of the 21 clusters of $\mathcal{P}$ are depicted in Fig. 7, along with the cohesion and isolation measures for partition $\mathcal{P}$ as a whole. Clusters 1, 2, 7, 11, and 8, as well as clusters 3 and 10, showed high values of cohesion and isolation, while others appeared to be less stable, as indicated by one or both of their cohesion and isolation values. For instance, cluster 12 (brown, northwest position in Fig. 5) was associated with a very high measure of cohesion, which indicates that the objects belonging to it, namely ‘Wood’, ‘Oak’, ‘Cork’, and ‘Pine’, were always grouped together in the bootstrapped partitions. However, the isolation measure of this cluster was low because other terms were lumped with these odor terms in several $\mathcal{P}_b$ partitions. On the contrary, cluster 14, composed of ‘Cinnamon’, ‘Clove’, ‘Spice’, ‘Pepper’, ‘Anise’, and ‘Licorice’ (dark green, southwest position in Fig. 5), appeared to be isolated but did not show a high cohesion value. This demonstrates that the objects belonging to this cluster did not usually lumped together in the bootstrapped partitions but, at the same time, hardly any terms outside of this group were grouped with them.

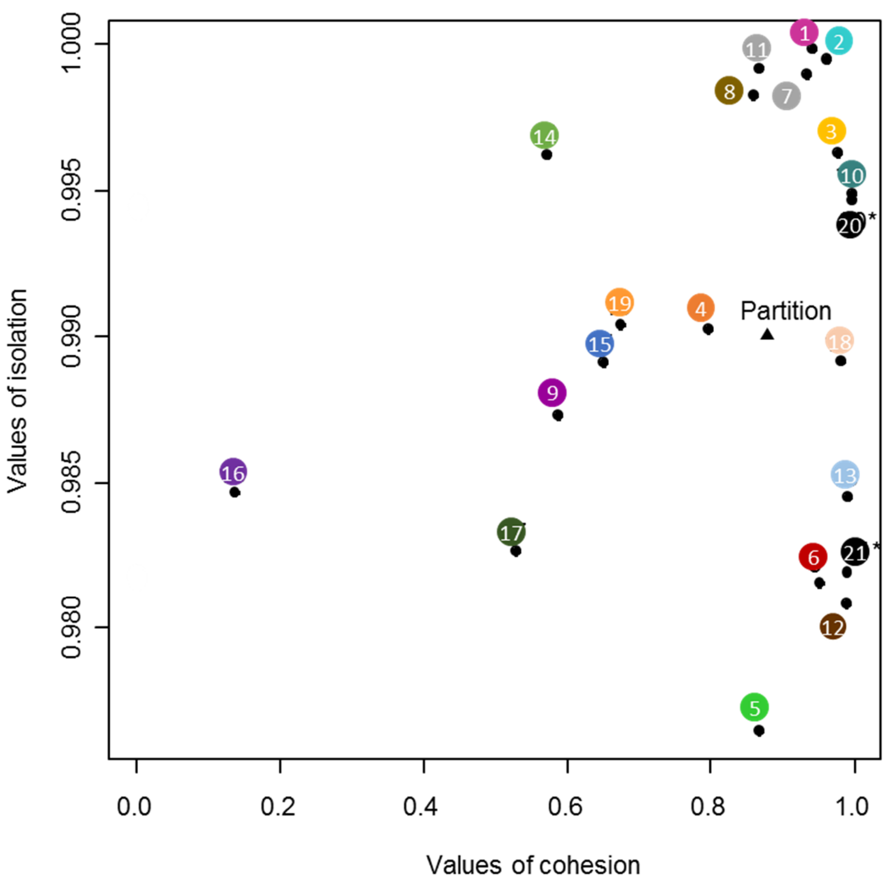

Fig. 7: Cohesion and isolation values for the 21 clusters (circles) and for the global partition (triangle). The clusters singletons (clusters 20 and 21) are indicated by an asterisk.

It should be noted that clusters 20 and 21 corresponded to singletons. These clusters were composed only of one term (‘Black olive’ and ‘Tea’, respectively). In such cases, values of cohesion cannot be computed (see Eq. (1)) and these values were set by default to 1. It is nevertheless apparent that ‘Black olive’ (cluster 20) was more isolated than ‘Tea’ (cluster 21), as ‘Black olive’ formed a cluster by itself more frequently with regard to the $\mathcal{P}_b$ partitions than the 'Tea' term did. Regarding the other small-sized clusters, those that consisted of two terms (e.g. clusters 16 and 17) were among the less cohesive ones and also showed relatively low values of isolation.

The analysis of the clusters' cohesion and isolation provided a broad perspective on their stability and made it possible to identify some strong patterns, as well as to distinguish, the most instable groupings.

4.3. Degree of association of objects to clusters

In order to obtain better insight into the clustering structure, we computed the degree of association of each object to its cluster. From this index, it was possible to determine whether a term highly belonged to one single cluster or whether it matched with several clusters. Fig. 8 depicts the values of $D_{i,k}$ (Eq. (3)) corresponding to the degree of association of the wine-odor terms with the various clusters of partition $\mathcal{P}$. Clusters are arranged in an ascending order according to their number (index k) while each line corresponds to an object, i. The darker the cell is, the higher the degree of association. In addition, distributions of three exemplifying terms ('English candy', 'Vanilla', and 'Melon'; highlighted in Fig. 8) are represented in barplots in Fig. 9. In both figures, a shading from black to white is used to represent the degree of association of objects to their clusters (see the legend of Fig. 8). Regarding the interpretation of the values associated with singleton groups ('Black olive' and 'Tea'), it should be noted that both terms have obviously a degree of association equal to one according to their respective cluster, but they also present values of degree of association rather high with other clusters.

In Fig. 8, dark rectangles represent the clusters whose objects presented high values of degree of association. These clusters matched exactly those with the highest measures of cohesion and isolation (e.g., clusters 1, 2, 3, 7, 8, 10, and 11). For example, 'Citrus', 'Lemon', 'Lime', 'Grapefruit', and 'Orange' showed values of degree of association to cluster 3 which were equal or higher than 0.97.

Regarding other clusters, the analysis revealed lower values of degree of association. The term 'Melon' in cluster 4 had a relatively low degree of association with respect to the other terms of its cluster (represented by a lighter shade). Indeed, this term was almost as frequently associated with clusters 4 and 5 (Fig. 9(c)).

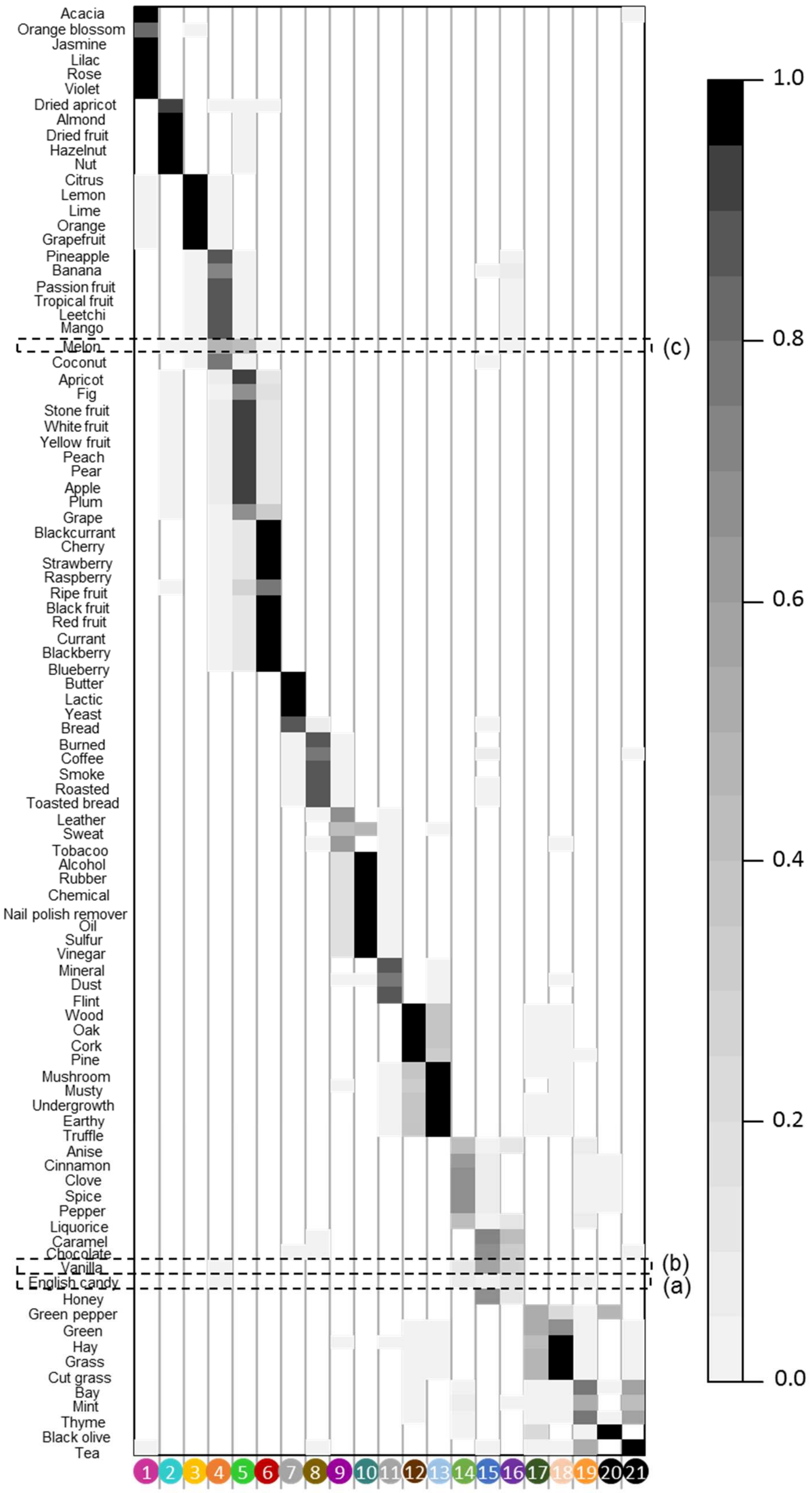

Fig. 8: Degree of association values, $D_{i,k}$, of an object i to the k^{th} cluster, of the reference partition, $\mathcal{P}$ computed from the sorting data of wine-odor terms. The darker the cell is, the higher the degree of association. Dotted rectangles indicate the objects that were highlighted in Fig. 9.

Interestingly, a tight connection can be pointed out between some pairs of clusters, for instance clusters 5 and 6, clusters 9 and 10, or, even more obviously, clusters 12 and 13. Indeed, association degrees of the terms are high for their cluster and also for the cluster paired with it. The link between each pair of clusters may also be noticed in Fig. 5, as terms from these outlined clusters are already closed in the additive tree.

Finally, we can note weaker patterns which were identified during the last recursive stages of the partitioning procedure (clusters 14 to 17, 19 and 21). Regarding these clusters, they did not correspond to strongly shaded rectangles. For each term included, light-shaded rectangles were also observed within several other clusters. For example, cluster 16 was especially weak, as the two terms belonging to the cluster – that is to say ‘English candy’ and ‘Honey’ – had degrees of association values only equal to 0.14 with respect to their own cluster, while ‘Caramel’ and ‘Chocolate’ terms belonging to cluster 15, had degrees of association values equal to 0.43 and 0.35 with respect to cluster 16, respectively. To illustrate this aspect, barplots of ‘English candy’ and ‘Vanilla’ are displayed in Fig. 9(a) and Fig. 9(b). In both cases, there was no clear association between the term and a single cluster.

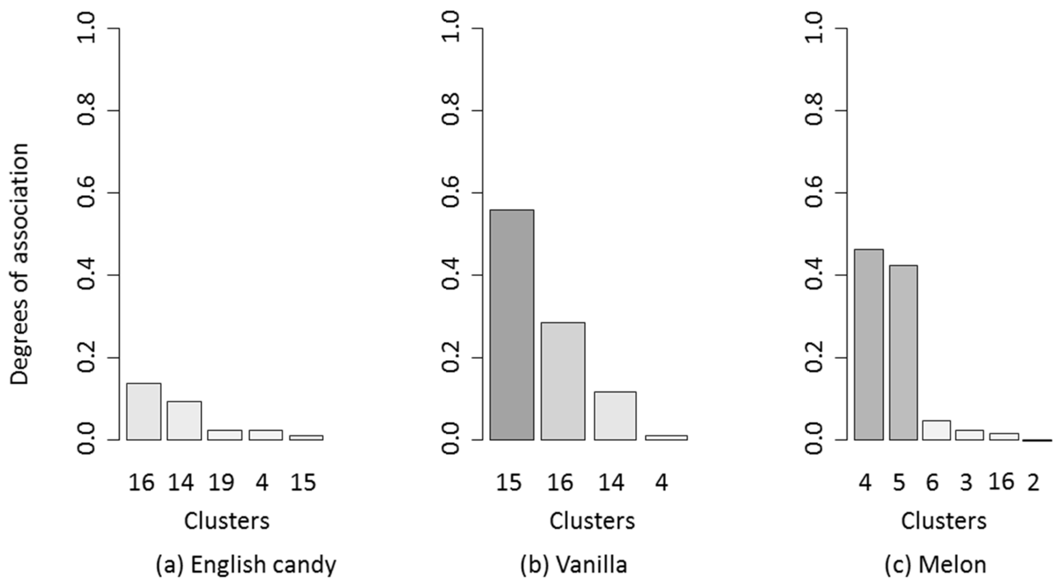

Fig. 9: Values of $D_{i,k}$ for (a) ‘English candy’, (b) ‘Vanilla’, and (c) ‘Melon’. The darker the rectangle, the higher the degree of association. Number in abscissa corresponds to the clusters’ number.

The analysis of each cluster of the partition on the basis of the cohesion and isolation criteria, together with the analysis of the degree of association of each term to the clusters made it possible to highlight different stability patterns among the various clusters of odor's terms.

On the one hand, clusters 1, 2, 3, 7, 8, and 11 represented highly stable clusters, with regard to both cohesion and isolation criteria. Moreover, the distributions of degrees of association, $D_{i,k}$, of each term i which belonged to those clusters were clearly uni-modal. Thus, we can conclude that these clusters were composed of terms that were highly associated with the terms belonging to the cluster and well separated from the terms belonging to the other ones.

On the other hand, some clusters were less stable, because of either a poor cohesion or a poor isolation value. The inspection of the degrees of association values of their terms made it possible to identify specific situations. For instance, terms from clusters 12 and 13 showed strong affinities. Merging both clusters would result in an improvement in their isolation values, however at the expense in the cohesion measure.

Besides, the stopping criterion used in the recursive algorithm was based on the homogeneity of the clusters which were formed. Some clusters were identified at the very first steps (e.g. cluster 1 was formed after two splitting steps) and were homogeneous enough not to be split further. On the contrary, the small-sized cluster 16 or the singletons 20 and 21, only reached adequate homogeneity in the very last steps. We may conclude that the later the cluster is formed in the partitioning procedure, the more likely this cluster is unstable.

5. Discussion

The first aim of this study was to introduce a versatile algorithm for determining a partition from an additive tree. This algorithm operates in a recursive way so that, for the tree/subtree under consideration at a given step, the edge corresponding to the highest *LengthRatio* value is cut and two new additive trees are then computed on the basis of the terms on either side of the cut. Overall, there were differences between the clusters formed by our algorithm (listed in Fig. 6) and the branching patterns of the original additive tree (Fig. 5). In other words, if the tree had been split only on the basis of its original branching pattern, the resulting clusters would have been different. We can notice that such a difference may be due to the size of the dataset considered in this study. Indeed, the large number of terms can result in a potentially unstable ranking of the set of the *LengthRatio* of each edge of the additive tree. In particular, at the first iteration, some terms may seem closed in opposition to the rest of the terms because of the number of terms. As we split the tree, the number of objects in each rebuilt tree is reduced, the *LengthRatio* computation became more precise and the ranking of actual *LengthRatio* values more trustworthy. By updating the tree's structure for each subset, and in turn updating the distribution of quality measures, we were able to select more relevant edges to split at each step.

In addition, the algorithm includes a stopping criterion to determine whether a subset of objects needs to be split further or whether it can be considered as a single cluster. A threshold value of 80% was set herein but it is worth noting that this value can be adapted depending on the dataset and the context. Other stopping rules could also be considered, including the number or the size of the clusters. For example, in Guénoche et al. (2012), the number of clusters was defined *a priori*. Such a criterion could be integrated in a straightforward way in the proposed algorithm.

A second objective was to assess the stability of the partition. Overall, the three measures proposed here—cohesion, isolation, and degree of association—generated complementary information that contributed to an improved assessment of the quality of the partition. The analysis of cohesion and isolation enabled us to assess the stability of the clusters: cohesion was a good indicator of a strong pattern within a cluster, and isolation yielded insight into other objects that might be related to a cluster. By analyzing the values for degree of association of the objects, we were able to explain how the stability of several clusters was affected by the objects they contained. It is worth noting that, unlike the Rand index, the use of both cohesion and isolation measures provided a more comprehensive understanding of the stability of the clusters. Finally, the inspection of the values of degree of association gave additional information on the objects which were responsible for poor stability.

This method generated a partition of odor terms that was largely similar to other categorizations proposed in the literature. Here, we refer more particularly to the “Wine Aroma Wheel” (Noble et al., 1984, 1987), which is currently considered as the standard in the wine sector. Some clusters highlighted herein mirror, to a large extent, those reported in the literature. For example, cluster 1, composed of the floral terms ‘Acacia’, ‘Orange blossom’, ‘Jasmine’, ‘Lilac’ and ‘Rose’, is consistent with a floral odor category usually identified in the literature (Noble et al., 1987). Cluster 2 of the partition obtained was composed of terms related to dry fruit and nutty odors that are separated in the “Wine Aroma Wheel” but merged together in other reports (Caillé et al., 2017; Coulon-Leroy et al., 2017; Esti et al., 2010). Clusters related to citrus fruits, exotic fruits and red fruits (clusters 3, 4, and 6, respectively), lactic, roasted, chemical and spice terms (clusters 7, 8, 10, and 14), and aromatic plants (cluster 19) were all in line with previous work in the literature. Most of these clusters were highly stable according to the measures of cohesion and isolation.

It is interesting to inspect the 'Melon' term more precisely. In the additive tree built from the initial distance matrix, ‘Melon’ was placed in a branch together with terms pertaining to cluster 5 (green, southeast position in Fig. 5). However, when the tree was rebuilt with the recursive algorithm, ‘Melon’ was lumped with cluster 4 (orange, east position in Fig. 5). From a wine perspective, ‘Melon’ is usually associated with the odors listed in cluster 4 (‘Pineapple’ or ‘Banana’) rather than with the odors belonging to cluster 5 (such as ‘Apple’ or ‘Pear’), as shown in the “Wine Aroma Wheel” of Noble et al. (1984). This result demonstrates the relevance of a strategy based on the recursive partitioning.

However, some of the clusters exhibited herein have not been reported in the literature. For example, cluster 5 (hereafter called ‘other fruits’) was found to be unstable and was particularly associated with a poor cohesion measure. In the literature, the terms belonging to this cluster are often grouped together, as they are all ‘fruity’. However, several publications combine these terms in different ways. The wheel of Noble et al. (1984) distinguishes ‘Tree fruit’ from ‘Berry’, as the work of Coulon-Leroy et al. (2017) does. In Esti et al. (2010) and in “The Master Sommelier Wine Aroma Wheel©” (Aromaster, 2010), the 'Pear' and 'Apple' terms are separated from the fruit terms and form their own category called ‘Pomaceous’ or ‘Pome fruits’. All in all, there is no consensus on the categorization of these ‘other fruits’, which may explain the poor cohesion measure of this cluster.

Another explanation for the poor stability of a cluster may be a high degree of variability among the subjects. Although we did not directly evaluate the degree of agreement within the panel, the bootstrapping strategy and the derived measures of stability provide a good perspective on this. As demonstrated by the stability values, all subjects were relatively consistent in their clustering of the following aromatic aspects: citrus, floral, tropical fruit, dried fruit, white fruit, red and black fruit, vegetal, empyreumatic, spicy, mineral, chemical, lactic, woody and earthy. This observation is consistent with Noble et al.’s aroma wheel (1984). It is also consistent with cross-cultural studies showing that the semantic categorization of odors is largely similar among subjects and is mainly based on odorant sources (Chrea et al., 2004, 2005).

Overall, the associated stability values were representative of the consistency of the clusters found in the literature, i.e., the clusters with poor stability values were also those that differed from one study to another. For example, cluster 16, containing ‘English candy’ and ‘Honey’ presented some discrepancies with the literature, especially for the former term. In fact, ‘English candy’ is a technical term used to describe the amylic odor of a wine (Lawrence et al., 2013) and typically associated with ‘Banana’.

As our panel was composed of both consumers and professionals, we assume that some of the subjects did not know the meaning of the term and therefore misclassified it, resulting in the poor stability of the cluster. To test this hypothesis, the procedure presented in the paper was performed on four homogeneous expertise segments (see Koenig et al., 2020 for more details on the segments). Overall, results were very similar between the different segments of expertise and it was possible to find clusters of odor terms expressing each of the following aromatic aspects: citrus, floral, tropical fruit, dry fruit, white fruit, red and black fruit, vegetal, empyreumatic, spicy, mineral, chemical, lactic, woody and earthy. However, some discrepancies appeared for some terms, as 'Banana'. Indeed, the less experts' segment assigned the term 'Banana' to the cluster of tropical fruits, while the most experts' segment associated it with the term 'English candy', forming a cluster characterized by amyl notes. 'English candy' clearly belongs to dedicated terminology related to wine tasting that is not or little used in everyday language. As a consequence, subjects with a low level of expertise are less familiar with this term and therefore have greater difficulty associating it with other terms.

From a practical point of view, the method presented in this paper provides guidelines for the determination of clusters from a wine-odor lexicon. Sensory lexicons are usually structured into groups of terms that are hierarchically embedded. Moreover, most of them are arranged as wheels derived from free sorting tasks (Gawel et al., 2000; Hayakawa et al., 2010; Koch et al., 2012; Spencer et al., 2016). Sorting procedures have gained ground among researchers in sensory analysis because products can be characterized without the need for subjects to undergo training (Faye et al., 2004; Santosa et al., 2010; Withers et al., 2014). This paper presents an innovative statistical strategy for highlighting a structure within a sensory lexicon, based on data from a sorting procedure. As argued by Noble et al. (1984) and Lawless & Civille (2013), the development of a lexicon serves the purposes of unifying the language used among stakeholders and of facilitating communication among winemakers, marketing personnel, wine researchers, and consumers. The choice of the subjects used to construct the structuration of sensory lexicon is also an important issue. In our case study, the inspection on the expertise led us to consider the whole panel of subjects as only minor differences occurred between different segments of expertise (Koenig et al., 2020). Moreover, these ones were mostly due to very specific terms such as 'English candy'. In light of these minor differences, we decided to deal with the whole dataset consisting in all the subjects. However, the choice of the subjects used is very important for the implementation of the method.

In addition, with the proposition of a structured lexicon, we also may suggest the use of intermediate terms to label the clusters, as in the aroma wheel proposed by Noble (1984). For some clusters, the title is obvious and may be a term of the cluster: for example, 'Citrus' for cluster 3 or 'Tropical fruit' for cluster 4. However, it may not be so obvious for others. In order to define clusters labels, the verbalization task performed at the end of the hierarchical sorting task (data non-shown), followed by a frequency analysis of the elicited words may be used. Verbalization generated at this step can be used to describe each cluster, based on an analysis of the frequency of the terms involved in each cluster.

Finally, according to the categorization theory presented by Rosch (1973), each category of objects has an internal structure based on the typicality of objects in the category. The internal structure of categories has, to the best of our knowledge, never been considered in the construction of a wine-odor lexicon. Our study confirmed that terms related to wine-odors cannot always be neatly assigned to a single cluster, and a term may present similar association values for different clusters. However, our method highlighted stable categories which can be considered more trustworthy as well as clusters that may be grouped to improve their isolation. The relationship between degree of association and typicality, as described in Rosch's theory, is not straightforward. Typicality is related rather to the distribution of the association degree of one term than a single value. Nevertheless, values of the degree of association can provide insight into the internal structure of a cluster and should be taken into consideration in efforts to consolidate a wine lexicon and ultimately to improve sensory tools such as aroma wheels.

6. Conclusion

Our aim was to propose a method of categorization that was adapted to both the semantic nature of the data set and the large number of terms under study. A partitioning procedure was developed to determine a set of clusters from an additive tree. Here, this algorithm was applied on a free sorting dataset corresponding to wine-odor terms. The algorithm relies on an additive tree representation of the data which appears to be more adapted when a large number of objects is considered. The approach proposed herein provide a partition of the objects from the additive tree representation. To better characterize the quality of the partition obtained, values of cohesion, isolation, and degree of association were computed using a bootstrapping strategy.

The partition highlighted by our strategy, and the measures of stability associated with it, were largely consistent with the literature, in particular compared to the aroma wheel of Noble et al. (1984). The stability measures applied here enabled a more precise assessment of the stability of the clusters and provided new perspectives for the creation of lexicons.

Finally, the entire strategy presented in this paper—categorization and assessment of stability—appears to be versatile and can be easily extended to any distance matrix. This offers various perspectives of its use for a much wider scope of applications than wine-odor lexicons.

Appendix

The equations related to cohesion and isolation measures and their connection with the Rand index are detailed below. Let us denote n_C the size of the cluster C, N the total number of objects, Co_C the cohesion value of the cluster C (Eq.(1)) and Is_C the isolation value of the cluster C (Eq.(2))

In order to estimate the cohesion, $Co_{\mathcal{P}}$ (resp. the isolation, $Is_{\mathcal{P}}$) of the global partition $\mathcal{P}$, a weighted sum of clusters' cohesion (resp. isolation) indices for each cluster is defined as in Eq. (4) (resp. Eq. (5)):

$$Co_{\mathcal{P}} = \frac{1}{T_1}\sum_{C\in\mathcal{P}}\frac{n_C(n_C-1)}{2}Co_C \quad \text{with } T_1 = \sum_{C\in\mathcal{P}}\frac{n_C(n_C-1)}{2} \tag{4}$$

$$Is_{\mathcal{P}} = \frac{1}{T_2}\sum_{C\in\mathcal{P}}\frac{n_C(N-n_C)}{2}Is_C \quad \text{with } T_2 = \sum_{C\in\mathcal{P}}\frac{n_C(N-n_C)}{2} \tag{5}$$

Herein, $T_1$is for the total number of pairs of objects being in the same cluster of $\mathcal{P}$, and T_2, the total number of pairs of objects not in the same cluster. We then have $T_1 + T_2 = N\ (N-1)/2$.

Let us remark, that both Co_C and Is_C are estimated as the average over all the B bootstrapped trials. If we define $Co_{C,\mathcal{P}_b}$ as the cohesion of the cluster C of $\mathcal{P}$ regarding the b^{th} "bootstrapped" partition $\mathcal{P}_b$ with $(b = 1, \dots, B)$, and, in the same vein, $Is_{C,\mathcal{P}_b}$ as the isolation of the cluster C regarding $\mathcal{P}_b$, it turns out that:

$$Co_C = \frac{1}{B}\sum_b Co_{C,\mathcal{P}_b} \quad \text{with } Co_{C,\mathcal{P}_b} = \frac{n_{11,b}^C}{n_C(n_C-1)/2}$$

$$Is_C = \frac{1}{B}\sum_b Is_{C,\mathcal{P}_b} \quad \text{with } Is_{C,\mathcal{P}_b} = \frac{n_{00,b}^C}{n_C(N-n_C)}$$

According to the definition of $n_{11,b}^C$ and $n_{00,b}^C$ given in section 3.3.1.

It follows that the cohesion and the isolation measures of $\mathcal{P}$ can be expressed as:

$$Co_{\mathcal{P}} = \frac{1}{B}\sum_b Co_{\mathcal{P},\mathcal{P}_b} \quad \text{with } Co_{\mathcal{P},\mathcal{P}_b} = \frac{\sum_C n_{11,b}^C}{T_1}$$

$$Is_{\mathcal{P}} = \frac{1}{B}\sum_b Is_{\mathcal{P},\mathcal{P}_b} \quad \text{with } Is_{\mathcal{P},\mathcal{P}_b} = \frac{\sum_C n_{00,b}^C}{2\,T_2}$$

Let us now consider the Rand index between the reference partition $\mathcal{P}$ and the b^{th} "bootstrapped" partition $\mathcal{P}_b$. By definition:

$$Rand_{\mathcal{P},\mathcal{P}_b} = \frac{N_{11,b}+N_{00,b}}{N\ (N-1)/2}$$

where $N_{11,b}$ is the number of pairs of objects put together in a same cluster in $\mathcal{P}$ and being also grouped in $\mathcal{P}_b$. But we have: $N_{11,b} = \sum_{C\in\mathcal{P}} n_{11,b}^C = T_1\ Co_{\mathcal{P},\mathcal{P}_b}$

In addition, $N_{00,b}$ is the number of pairs of objects separated in different clusters in $\mathcal{P}$ and also separated $\mathcal{P}_b$. But, we have: $N_{00,b} = \frac{1}{2}\sum_{C\in\mathcal{P}} n_{00,b}^C = \ T_2\ Is_{\mathcal{P},\mathcal{P}_b}$

So, the Rand index between partition $\mathcal{P}$ and the b^{th} partition $\mathcal{P}_b$ can be expressed as a weighted average of the cohesion and isolation of $\mathcal{P}$, confronted with $\mathcal{P}_b$, as specified in El Moubarki (2009). More precisely:

$$Rand_{\mathcal{P},\mathcal{P}_b} = \frac{T_1\ Co_{\mathcal{P},\mathcal{P}_b} + T_2\ Is_{\mathcal{P},\mathcal{P}_b}}{N(N-1)/2} = \frac{\sum_{C\in\mathcal{P}} n_C(n_C-1)Co_{\mathcal{P},\mathcal{P}_b} + \sum_{C\in\mathcal{P}} n_C(N-n_C)Is_{\mathcal{P},\mathcal{P}_b}}{N\ (N-1)} \tag{6}$$

This makes it possible to deduce that the average of the Rand index between the reference and "boostrapped" partitions $\mathcal{P}_b$ after B resampling trials can be expressed as a function the cohesion and isolation indices for the partition $\mathcal{P}$ as defined in Eq. (4) and (5).

Acknowledgments

This work was carried out with the financial support of the regional program "Food for Tomorrow / Cap Aliment, Research, Education, and Innovation in Pays de la Loire”, which is supported by the French Region Pays de la Loire and the European Regional Development Fund (FEDER), as well as the financial support of the Ecole Supérieure d’Agricultures in Angers. The study was carried out using the facilities of the Senso’Veg platform. We would like to thank Flore Lourtioux and Margaux Lecourtois, who participated in the recruitment of subjects and the acquisition of the data for the free-sorting task.

References

Abdi, H. (1990). Additive-tree representations. In *Trees and Hierarchical Structures* (pp. 43–59). Springer.

Abdi, H., Barthélémy, J.-P., & Luong, X. (1984). Tree Representations of Associative Structures in Semantic and Episodic Memory Research. In E. Degreef & J. Van Buggenhaut (Eds.), *Advances in Psychology* (Vol. 20, pp. 3–31). North-Holland. https://doi.org/10.1016/S0166-4115(08)62078-6

Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*, *18*(4), 627–640. https://doi.org/10.1016/j.foodqual.2006.09.003

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207–216. https://doi.org/10.1145/170035.170072

Aromaster. (2010). *Wine Aroma Wheel by Aromaster*. Retrieved 20 Novembre 2019. https://aromaster.com/fr/produit/roue-des-aromes-du-vin/

Bel Mufti, G., Bertrand, P., & El Moubarki, L. (2012). *Decomposition of the Rand index in order to assess both the stability and the number of clusters of a partition*. *hal-00707357*. Retrieved 26 February 2020. https://hal.archives-ouvertes.fr/hal-00707357

Beller, M. (1990). Tree Versus Geometric Representation of Tests and Items. *Applied Psychological Measurement*, *14*(1), 13–28. https://doi.org/10.1177/014662169001400102

Berland, A., Gaillard, P., Guidetti, M., & Barone, P. (2015). Perception of Everyday Sounds: A Developmental Study of a Free Sorting Task. *PLOS ONE*, *10*(2), e0115557. https://doi.org/10.1371/journal.pone.0115557

Bertrand, P., & Bel Mufti, G. (2006). Loevinger's measures of rule quality for assessing cluster stability. *Computational Statistics and Data Analysis*, *50*(4), 1220–1236.

Blancher, G., Chollet, S., Kesteloot, R., Hoang, D. N., Cuvelier, G., & Sieffermann, J.-M. (2007). French and Vietnamese: How do they describe texture characteristics of the same food? A case study with jellies. *Food Quality and Preference*, *18*(3), 560–575. https://doi.org/10.1016/j.foodqual.2006.07.006

Blancher, G., Clavier, B., Egoroff, C., Duineveld, K., & Parcon, J. (2012). A method to investigate the stability of a sorting map. *Food Quality and Preference*, *23*(1), 36–43. https://doi.org/10.1016/j.foodqual.2011.06.010

Buneman, P. (1971). Filiation of manuscripts. *Mathematics in the Archaeological and Historical Sciences*, 387–395.

Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality and Preference*, *28*(1), 106–115. https://doi.org/10.1016/j.foodqual.2012.09.005

Cadoret, M., Lê, S., & Pagès, J. (2009). A Factorial Approach for Sorting Task data (FAST). *Food Quality and Preference*, *20*(6), 410–417. https://doi.org/10.1016/j.foodqual.2009.02.010

Cadoret, M., Lê, S., & Pagès, J. (2011). Statistical analysis of hierarchical sorting data. *Journal of Sensory Studies*, *26*(2), 96–105. https://doi.org/10.1111/j.1745-459X.2010.00326.x

Caillé, S., Salmon, J.-M., Bouvier, N., Roland, A., & Samson, A. (2017). Modification of the olfactory sensory characteristics of Chardonnay wine through the increase in sotolon concentration. *Food Quality and Preference*, *56*(Part A), 225–230. https://doi.org/10.1016/j.foodqual.2016.08.001

Cariou, V., & Qannari, E. M. (2018). Statistical treatment of free sorting data by means of correspondence and cluster analyses. *Food Quality and Preference*, *68*, 1–11. https://doi.org/10.1016/j.foodqual.2018.01.011

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, *35*(3), 283–319. https://doi.org/10.1007/BF02310791

Carroll, J. D., & Pruzansky, S. (1975). Fitting of hierarchical tree structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares. US-Japan Seminar on Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques, San Diego, CA.

Chollet, S., Lelièvre, M., Abdi, H., & Valentin, D. (2011). Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality and Preference*, *22*(6), 507–520. https://doi.org/10.1016/j.foodqual.2011.02.004

Chrea, C., Valentin, D., Sulmont-Rossé, C., Ly Mai, H., Hoang Nguyen, D., & Abdi, H. (2004). Culture and odor categorization: agreement between cultures depends upon the odors. *Food Quality and Preference*, *15*(7), 669–679. https://doi.org/10.1016/j.foodqual.2003.10.005

Chrea, C., Valentin, D., Sulmont-Rossé, C., Nguyen, D., & Abdi, H. (2005). Semantic, typicality and odor representation: a cross-cultural study. *Chemical Senses*, *30*, 37–49. https://doi.org/10.1093/chemse/bjh255

Coulon-Leroy, C., Symoneaux, R., Lawrence, G., Mehinagic, E., & Maitre, I. (2017). Mixed Profiling: A new tool of sensory analysis in a professional context. Application to wines. *Food Quality and Preference*, *57*, 8–16. https://doi.org/10.1016/j.foodqual.2016.11.005

Courcoux, P., Faye, & Qannari, E. M. (2014). Determination of the consensus partition and cluster analysis of subjects in a free sorting task experiment. *Food Quality and Preference*, *32*, 107–112. https://doi.org/10.1016/j.foodqual.2013.05.004

Courcoux, P., Qannari, E., Taylor, Y., Buck, D., & Greenhoff, K. (2012). Taxonomic free sorting. *Food Quality and Preference*, *23*(1), 30–35. https://doi.org/10.1016/j.foodqual.2011.04.001

De Soete, G. (1983). A least squares algorithm for fitting additive trees to proximity data. Psychometrika, 48(4), 621–626. https://doi.org/10.1007/BF02293884

De Soete, G., & Carroll, J. D. (1996). Tree and other network models for representing proximity data. *Clustering and Classification*, 157–197.

Dehlholm, C. (2015). Free multiple sorting as a sensory profiling technique. In *Rapid Sensory Profiling Techniques* (pp. 187–196). Woodhead Publishing. https://doi.org/10.1533/9781782422587.2.187

Delarue, J., Lawlor, B., & Rogeaux, M. (2015). *Rapid Sensory profiling Techniques and Related Methods*. Woodhead Publishing.

Diday, E., & Simon, J. C. (1976). Clustering Analysis. In K. S. Fu (Ed.), *Digital Pattern Recognition* (pp. 47–94). Springer. https://doi.org/10.1007/978-3-642-96303-2_3

Dubois, D. (2000). Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, *1*(1), 35–68.

El Moubarki, L. (2009). *Décomposition et évaluation des mesures de stabilité d'un partitionnement* [Thesis, Paris 9]. http://www.theses.fr/2009PA090035

Esti, M., Airola, R. L. G., Moneta, E., Paperaio, M., & Sinesio, F. (2010). Qualitative data analysis for an exploratory sensory study of grechetto wine. *Analytica Chimica Acta*, *660*(1), 63–67. https://doi.org/10.1016/j.aca.2009.10.014

Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality and Preference*, *15*(7), 781–791. https://doi.org/10.1016/j.foodqual.2004.04.009

Gambette, P., Gala, N., & Nasr, A. (2012). Longueur de branches et arbres de mots. *Corpus*, *11*, Article 11. http://journals.openedition.org/corpus/2245

Gawel, R., Oberholster, A., & Francis, I. L. (2000). A 'Mouth-feel Wheel': terminology for communicating the mouth-feel characteristics of red wine. *Australian Journal of Grape and Wine Research*, *6*(3), 203–207.

Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology*, *61*(1), 54–63.

Guénoche, A. (2011). Consensus of partitions: a constructive approach. *Advances in Data Analysis and Classification*, *5*(3), 215–229.

Guénoche, A. (2012). Sur le consensus en catégorisation libre. *Mathématiques et Sciences Humaines. Mathematics and Social Sciences*, *197*, 65–82.

Guénoche, A., & Garreta, H. (2002). Representation and evaluation of partitions. In *Classification, Clustering, and Data Analysis* (pp. 131–138). Springer.

Hayakawa, F., Kazami, Y., Wakayama, H., Oboshi, R., Tanaka, H., Maeda, G., Hoshino, C., Iwawaki, H., & Miyabayashi, T. (2010). Sensory Lexicon of Brewed Coffee for Japanese Consumers, Untrained Coffee Professionals and Trained Coffee Tasters. *Journal of Sensory Studies*, *25*(6), 917–939. https://doi.org/10.1111/j.1745-459X.2010.00313.x

Honoré-Chedozeau, C., Lelièvre-Desmas, M., Ballester, J., Chollet, S., & Valentin, D. (2017). Knowledge representation among assessors through free hierarchical sorting and a semi-directed interview: Exploring Beaujolais wines. *Food Quality and Preference*, *57*(Supplement C), 17–31. https://doi.org/10.1016/j.foodqual.2016.11.008

King, M. C., Cliff, M. A., & Hall, J. W. (1998). Comparison of Projective Mapping and Sorting Data Collection and Multivariate Methodologies for Identification of Similarity-of-Use of Snack Bars1. *Journal of Sensory Studies*, *13*(3), 347–358. https://doi.org/10.1111/j.1745-459X.1998.tb00094.x

Koch, I. S., Muller, M., Joubert, E., van der Rijst, M., & Næs, T. (2012). Sensory characterization of rooibos tea and the development of a rooibos sensory wheel and lexicon. *Food Research International*, *46*(1), 217–228. https://doi.org/10.1016/j.foodres.2011.11.028

Koenig, L., Coulon-Leroy, C., Symoneaux, R., Cariou, V., & Vigneau, E. (2020). Influence of expertise on semantic categorization of wine odors. *Food Quality and Preference*, *83*. https://doi.org/10.1016/j.foodqual.2020.103923

Lahne, J. (2020). Sorting backbone analysis: A network-based method of extracting key actionable information from free-sorting task results. *Food Quality and Preference*, *82*, 103870. https://doi.org/10.1016/j.foodqual.2020.103870

Lawless, H. T., Sheng, N., & Knoops, S. S. C. P. (1995). Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference*, *6*(2), 91–98. https://doi.org/10.1016/0950-3293(95)98553-U

Lawless, L. J., & Civille, G. V. (2013). Developing lexicons: A review. *Journal of Sensory Studies*, *28*(4), 270–281.

Lawrence, G., Symoneaux, R., Maitre, I., Brossaud, F., Maestrojuan, M., & Mehinagic, E. (2013). Using the free comments method for sensory characterisation of Cabernet Franc wines: Comparison with classical profiling in a professional context. *Food Quality and Preference*, *30*(2), 145–155. https://doi.org/10.1016/j.foodqual.2013.04.005

Lenca, P., Meyer, P., Vaillant, B., & Lallich, S. (2008). On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operational Research*, *184*(2), 610–626. https://doi.org/10.1016/j.ejor.2006.10.059

Navarro, D. J., & Lee, M. D. (2002). Combining dimensions and features in similarity-based representations. *Proceedings of the 15th International Conference on Neural Information Processing Systems*, 67–74.

Noble, A. C., Arnold, R. A., Buechsenstein, J., Leach, E. J., Schmidt, J. O., & Stern, P. M. (1987). Modification of a Standardized System of Wine Aroma Terminology. *American Journal of Enology and Viticulture*, *38*(2), 143–146.

Noble, A. C., Arnold, R. A., Masuda, B. M., Pecore, S. D., Schmidt, J., & Stern, P. M. (1984). Progress towards a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, *35*(2), 107–109.

Orden, D., Fernández-Fernández, E., Rodríguez-Nogales, J. M., & Vila-Crespo, J. (2019). Testing SensoGraph, a geometric approach for fast sensory evaluation. *Food Quality and Preference*, *72*, 1–9. https://doi.org/10.1016/j.foodqual.2018.09.005

Orden, D., Fernández-Fernández, E., Tejedor-Romero, M., & Martínez-Moraian, A. (2021). Geometric and statistical techniques for projective mapping of chocolate chip cookies with a large number of consumers. *Food Quality and Preference*, *87*, 104068. https://doi.org/10.1016/j.foodqual.2020.104068

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England)*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika*, *47*(1), 3–24. https://doi.org/10.1007/BF02293848

Qannari, E. M., Courcoux, P., & Faye, P. (2014). Significance test of the adjusted Rand index. Application to the free sorting task. *Food Quality and Preference*, *32*, 93–97. https://doi.org/10.1016/j.foodqual.2013.05.005

Qannari, E. Mostafa., Cariou, V., Teillet, E., & Schlich, P. (2010). SORT-CC: A procedure for the statistical treatment of free sorting data. *Food Quality and Preference*, *21*(3), 302–308. https://doi.org/10.1016/j.foodqual.2009.01.003

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rossini, K., Verdun, S., Cariou, V., Qannari, E. M., & Fogliatto, F. S. (2012). PLS discriminant analysis applied to conventional sensory profiling data. *Food Quality and Preference*, *23*(1), 18–24. https://doi.org/10.1016/j.foodqual.2011.01.005

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Santosa, M., Abdi, H., & Guinard, J.-X. (2010). A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality and Preference*, *21*(7), 881–892. https://doi.org/10.1016/j.foodqual.2010.05.011

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, *42*(3), 319–345.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, *39*(4), 373–421. https://doi.org/10.1007/BF02291665

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, *86*(2), 87–123. https://doi.org/10.1037/0033-295X.86.2.87

Spencer, M., Sage, E., Velez, M., & Guinard, J.-X. (2016). Using Single Free Sorting and Multivariate Exploratory Methods to Design a New Coffee Taster's Flavor Wheel. *Journal of Food Science*, *81*(12), 2997–3005. https://doi.org/10.1111/1750-3841.13555

Steinberg, D. D. (1967). The Word Sort: An instrument for semantic analysis. *Psychonomic Science*, *8*(12), 541–542. https://doi.org/10.3758/BF03331741

Takane, Y. (1981). MDSORT: A special-purpose multidimensional scaling program for sorting data. *Behavior Research Methods & Instrumentation*, *13*(5), 698–698.

Takane, Y. (1982). IDSORT: An individual differences multidimensional scaling program for sorting data. *Behavior Research Methods & Instrumentation*, *14*(6), 546–546.

Van der Kloot, W. A., & Van Herk, H. (1991). Multidimensional Scaling of Sorting Data: A Comparison of Three Procedures. *Multivariate Behavioral Research*, *26*(4), 563–581. https://doi.org/10.1207/s15327906mbr2604_1

Vanpaemel, W., Verbeemen, T., Dry, M., Verguts, T., & Storms, G. (2010). Geometric and featural representations in semantic concepts. *Memory & Cognition*, *38*(7), 962–968. https://doi.org/10.3758/MC.38.7.962

Varela, P., & Ares, G. (2014). *Novel Techniques in Sensory Characterization and Consumer Profiling*. CRC Press.

Waterman, M. S., Smith, T. F., Singh, M., & Beyer, W. A. (1977). Additive evolutionary trees. *Journal of Theoretical Biology*, *64*(2), 199–213.

Withers, C., Methven, L., Qannari, E. M., Allen, V. J., Gosney, M. A., & MacFie, H. J. H. (2014). Taxonomic Free Sorting: A Successful Method with Older Consumers and a Novel Approach to Preference Mapping. *Journal of Sensory Studies*, *29*(3), 182–189. https://doi.org/10.1111/joss.12093