SensoGraph: Using proximity graphs for sensory analysis

David N. de Miguel*[1], David Orden†[2], Encarnación Fernández-Fernández‡[3], José M. Rodríguez-Nogales §[3], and Josefina Vila-Crespo ¶[4]

[1]Universidad de Alcalá, Spain.
[2]Departamento de Física y Matemáticas, Universidad de Alcalá, Spain.
[3]Departamento de Ingeniería Agrícola y Forestal, Universidad de Valladolid, Spain.
[4]Departamento de Anatomía Patológica, Medicina Preventiva y Salud Pública, Medicina Legal y Forense, Universidad de Valladolid, Spain.

Abstract

Sensory evaluation of foods is as important as chemical, physical or microbiological examinations, being specially relevant in food industries. Classical methods can be long and costly, making them less suitable for certain industries like the wine industry. Some alternatives have arisen recently, including Napping®, where the tasters represent the sensory distances between products by positioning them on a tablecloth; the more similar they perceive the products, the closer they position them on the tablecloth. This method uses multiple factor analysis (MFA) to process the data collected. The present paper introduces the software `SensoGraph`, which makes use of proximity graphs to analyze those data. The application is described and experimental results are presented in order to compare the performances of `SensoGraph` and Napping®, using eight wines from the Toro region and two groups of twelve tasters with different expertise.

1 Sensory analysis

The goal of sensory evaluation of foods is the study of the sensations they produce. When consuming a food, stimulus of several classes are perceived; visual (color, shape, brightness), tactile (at fingertips or mouth epithelium), odorous (at nose epithelium), gustatory (at taste buds), and even auditory (e.g., for crunchy food). The norm ISO 5492:2008 defines sensory analysis as the science related to the evaluation of organoleptic attributes of a product by the senses, and other ISO norms unify the tools and methods used for that evaluation. The sensory examination turns out to be as important as chemical, physical or microbiological examinations, being specially relevant in food industries.

The main tool in sensory analysis is a panel of tasters, either experts or consumers, who evaluate the products from an analytic and/or hedonic point of view. As any other instrument depends on its calibration, such a panel depends on human beings. Their perceptions are translated into quantifiable data, which is then treated by means of different methods.

The classical method is descriptive analysis, which aims to describe the sensory characteristics of a product and use them to quantify the sensory differences between products [11]. Different implementations of this method provide a quantitative description of the sensory attributes perceived by a group of expert tasters, chosen because of their sensory abilities, who are trained to describe and evaluate sensory differences among products. Such a training is a critical step in the creation of an expert panel of tasters, when tasters agree on the definitions of descriptors and the use of scales, in order to provide reliable and consistent results.

However, this training can be long and costly, making it less suitable for certain industries like the wine industry. There, sensory characterization is usually performed by the oenologist in charge of the winery, for whom it is difficult to enrol in a panel requiring a regular activity during a long time. Thus, in the last years several alternative methods have been proposed, aiming to provide a fast sensory positioning of a set of products, in order to avoid the most time-consuming steps in classical methods. A prominent one among these alternatives is Napping® [12]. In a single session, all the products are provided simultaneously to the tasters, who represent the sensory distances between products by positioning them on a tablecloth. Products which are perceived as similar should be positioned close to each other, while products perceived as different should be positioned far enough. Each

*Email: david.n.demiguel@gmail.com.
†Email: david.orden@uah.es. Research partially supported by MICINN Project MTM2011-22792, ESF EUROCORES programme EuroGIGA - ComPoSe IP04 - MICINN Project EUI-EURC-2011-4306 and Junta de Castilla y León Project VA172A12-2.
‡Email: effernan@iaf.uva.es.
§Email: rjosem@iaf.uva.es.
¶Email: jvila@pat.uva.es.

taster chooses the criteria and the relative importance given. These data are later processed using multiple factor analysis (MFA) [4], in order to take into account the criteria and relative importance of all the tasters. Despite having both advantages and disadvantages, Napping® has become a useful tool when some accuracy can be sacrificed for the sake of a faster study [13].

2 Proximity graphs

Given a set of points in the plane, a (geometric) *proximity graph* connects two of them according to a chosen proximity criterion. These graphs have been widely used in order to analyze the relative position of points, for instance looking for clusters or spanning structures. See [1, 7] and the references therein. Thus, it seems natural to use them in order to analyze the data collected by a tablecloth method for sensory analysis, like Napping®. Among the many different types of proximity graphs, we have chosen the following:

Nearest Neighbor Graph (NNG): Each point is joined to the closest among the remaining points [14].

k-Nearest Neighbor Graph (k-NNG): In this generalization of the NNG, each point is joined to the k closest among the remaining points.

Minimum Spanning Tree (MST): Among the trees passing through all the given points, the MST is the one which minimizes the sum of edge lenghts [10].

Relative Neighborhood Graph (RNG): This graph, introduced by Toussaint [15], joins two points P, Q if there is no point whose distances to both P and Q are smaller than the distance $d(P, Q)$.

k-Relative Neighborhood Graph (k-RNG): The generalization of the RNG which allows up to k points with distances to both P and Q smaller than $d(P, Q)$.

Gabriel Graph (GG): In this graph two points P, Q are joined if there is no other point inside the closed disk which has the segment $\overline{PQ}$ as diameter [5].

k-Gabriel Graph (k-GG): Generalization of the GG allowing up to k points to lie inside the closed disk.

Delaunay Triangulation (DT): Three points P, Q, R form a triangle precisely if their circumcircle does not contain any other point [3].

k-Delaunay Triangulation (k-DT): Generalization of the DT allowing up to k points to lie inside the closed disk.

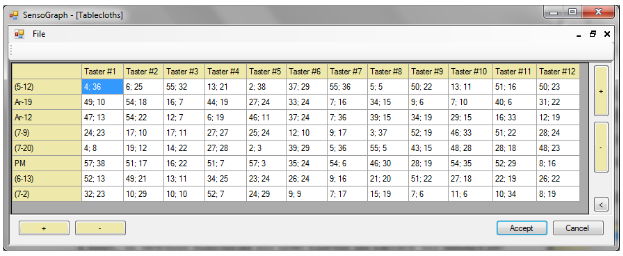
β-skeleton (β-SK): A family of proximity graphs, one for each value of $\beta \geq 0$, see [9] for more details. For $\beta = 1$ we get the GG. For $\beta = 2$ we get the RNG.

Unit Disk Graph (UDG): In this graph two points P, Q are joined if the distance $d(P, Q)$ between them is no greater than a fixed threshold [2].

3 The SensoGraph application

`SensoGraph` is an application, still under development, which aims to use proximity graphs for the analysis of data collected from tablecloth sensory methods like Napping®. The interface intends to be intuitive and easy to use, so that no special knowledge is needed.

Tasting data are stored in the form of an $m \times 2n$ matrix, in which there is a row per product and a pair of columns per taster, storing the two coordinates assigned to the corresponding product. In a first screen, this matrix can be manually created or inserted from a CSV file, according to the IETF RFC 4180 standard. After insertion, the matrix can be modified by adding or removing rows or columns, as well as by editing an individual entry. See Figure 1.



	Taster #1	Taster #2	Taster #3	Taster #4	Taster #5	Taster #6	Taster #7	Taster #8	Taster #9	Taster #10	Taster #11	Taster #12
(5-12)	4; 36	6; 25	55; 32	13; 21	2; 38	37; 29	55; 36	5; 5	50; 22	13; 11	51; 16	50; 23
Ar-19	49; 10	54; 18	16; 7	44; 19	27; 24	33; 24	7; 16	34; 15	9; 6	7; 10	40; 6	31; 22
Ar-12	47; 13	54; 22	12; 7	6; 19	46; 11	37; 24	7; 36	39; 15	34; 19	29; 15	16; 33	12; 19
(7-9)	24; 23	17; 10	17; 11	27; 27	25; 24	12; 10	9; 17	3; 37	52; 19	46; 33	51; 22	28; 24
(7-20)	4; 8	19; 12	14; 22	27; 28	2; 3	39; 29	5; 36	55; 5	43; 15	48; 28	28; 18	48; 23
PM	57; 38	51; 17	16; 22	51; 7	57; 3	35; 24	54; 6	46; 30	28; 19	54; 35	52; 29	8; 16
(6-13)	52; 13	49; 21	13; 11	34; 25	23; 24	26; 24	9; 16	21; 20	51; 22	27; 18	22; 19	26; 22
(7-2)	32; 23	10; 29	10; 10	52; 7	24; 29	9; 9	7; 17	15; 19	7; 6	11; 6	10; 34	8; 19

Figure 1: Matrix of tasting data.

After accepting the data matrix, a new screen is shown. There, the user can choose a type of proximity graph among the ones specified in Section 2. For those proximity graphs depending on a parameter, k-NNG, k-RNG, k-GG, k-DT, β-SK, and UDG, the user can change the value of the parameter. Furthermore, the application allows to intersect any of the graphs considered with the UDG, in case the user wants to avoid too long edges.

For the given choice of a type of proximity graph, the application generates the graph for each of the taster's tablecloths. The user can choose a taster and check its tablecloth and the resulting graph. When visualizing a tablecloth, it is also possible to change the type of proximity graph, in order to check the differences between them. See Figure 2.

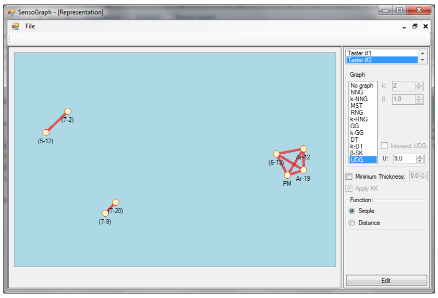


Figure 2: Tablecloth for taster number 2 with the UDG for radius 10.

Furthermore, chosen a type of proximity graph, the user can also merge all the tablecloths and their corresponding graphs into a single, global, picture. There, every edge is shown with a thickness, computed according to a *simple* function: The thickness corresponds to the number of tasters for which that edge appears in the proximity graph. This representation encodes the global opinion of the panel of tasters, so that those products perceived as similar are joined by thicker edges, while products considered different are joined by thinner edges (or even not joined at all).

In order to position the vertices of this global picture, SensoGraph considers the edges like springs which try to approach their endpoints, with a strength proportional to the edge thickness. Thus, vertices joined by thicker edges become closer than those joined by thinner edges. In this new picture, the products considered similar by the panel of tasters can be recognized not only by the thickness of the edge between them, but also by their mutual distance, see Figures 4, 6, and 8. Such a positioning is performed by a slight adaptation of the algorithm by Kamada and Kawai [8].

Since some types of proximity graphs insist in joining vertices which are far apart, SensoGraph offers a second way of computing the thickness of an edge. The *distance* function also looks only at the edges appearing in the proximity graph chosen but, in addition, it takes into account their length, decreasing the contribution of long edges to the total thickness. As mentioned above, the user can also choose to discard too long edges, by intersecting any of the proximity graphs with the UDG.

Furthermore, the user can also peel the graph in the global picture, by removing edges below any chosen thickness, in order to keep only the most relevant ones.

4 Experimental results

Eight wines from the Toro region, elaborated using different yeasts during the alcoholic fermentation, were considered. Two panels, of twelve non-trained tasters each, were selected. The *experts* panel was composed by people, mainly young, with some knowledge of the techniques for sensory analysis of wines. The *non-experts* panel was composed by plain consumers, with different ages and levels of knowledge. Each of the panels performed a session of Napping®, and the experts panel repeated for a second session, in order to check for improvements due to such a slight training.

The data from those three sessions was then processed both by multiple factor analysis (MFA) [6], as usual in Napping®, and by SensoGraph. Figures 3 to 8 show the results obtained, with SensoGraph using GG and the *simple* thickness function.

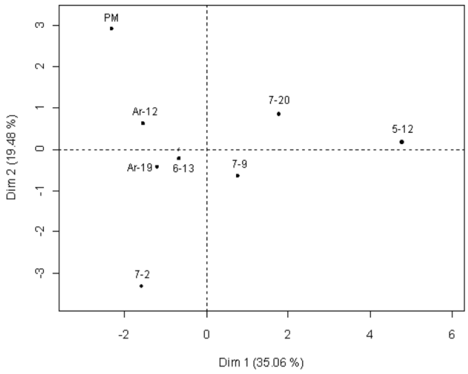


Figure 3: Experts panel, Napping®.

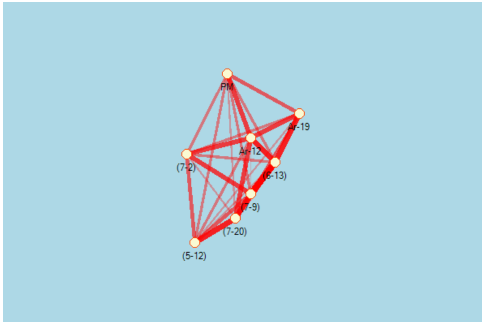


Figure 4: Experts panel, SensoGraph.

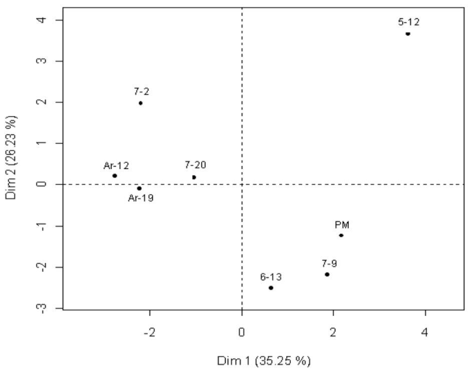


Figure 5: Repetition of experts panel, Napping®.

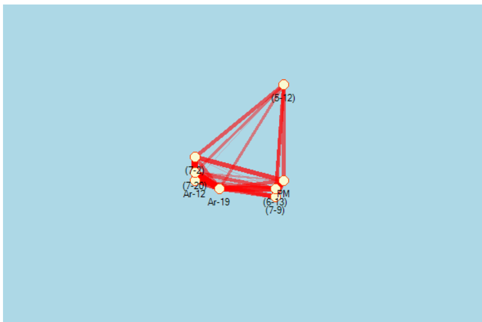


Figure 6: Repetition of experts panel, SensoGraph.

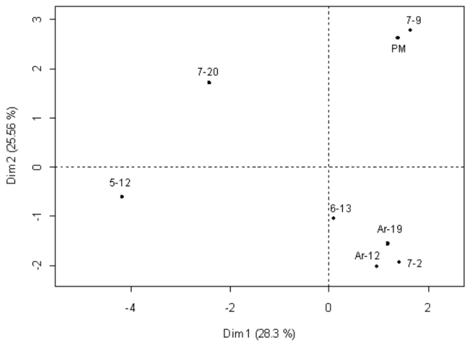


Figure 7: Non-experts panel, Napping®.

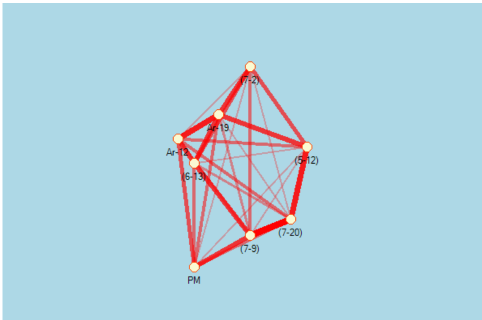


Figure 8: Non-experts panel, `SensoGraph`.

Among the three sessions performed, the most representative results of Napping® were those from the repetition of the experts panel. Figure 5 shows a 35.25% of the variation explained by the first dimension and a 26.23% of the remaining variation explained by the second dimension, for a total inertia of 61.48%. Being this the session for which Napping® is most reliable, it is the one chosen to compare with the results provided by `SensoGraph`.

For that session, using `SensoGraph` with the *simple* thickness function provides the same clusters as Napping® for all the graphs considered except for NNG, which has too few edges, and for k-DT, which has too many edges. Using the *distance* thickness function does not change the elements in the clusters, although the whole picture appears expanded and highlights only the strongest connections between different clusters. Furthermore, it improves the results for the extreme cases above, leading to the same clusters as Napping® for k-DT and palliating the differences for NNG.

A possible advantage of `SensoGraph` is giving other kind of information than Napping®. Apart from providing several types of proximity graphs and parameters to test with, `SensoGraph` shows how the different clusters are connected. For an example, Figures 5 and 6 lead to the same three clusters, but only the one from `SensoGraph` shows that they are actually quite connected, reflecting the fact that all the wines considered were actually quite similar [6].

5 Acknowledgements

The authors want to gratefully thank Ferran Hurtado for proposing them the use of proximity graphs in order to analyze data from tablecloth methods.

References

[1] J. Cardinal, S. Collette, and S. Langerman, Empty region graphs, *Computational Geometry: Theory and Applications*, **42** (2009), 183–195.

[2] B. N. Clark, C. J. Colbourn, and D. S. Johnson, Unit Disk Graphs, *Discrete Mathematics*, **86** (1990), 165–177.

[3] B. N. Delaunay, Sur la sphere vide, *Bulletin of the Academy of Sciences of the USSR*, **VII** (1934), 793–800.

[4] B. Escofier and J. Pagès, *Analyses factorielles simples et multiples*, Dunod, Paris, 1998.

[5] K. R. Gabriel and R. R. Sokal, A new statistical approach to geographic variation analysis, *Systematic Biology*, **18:3** (1969), 259–278.

[6] L. Gallego-Expósito, Nuevas técnicas de análisis sensorial de alimentos: Métodos espaciales, Degree thesis, 2011.

[7] J.W. Jaromczyk and G.T. Toussaint, Relative neighborhood graphs and their relatives, *Proceedings of the IEEE*, **80:9** (1992), 1502–1517.

[8] T. Kamada and S. Kawai, An algorithm for drawing general undirected graphs, *Information Processing Letters*, **31** (1989), 7–15.

[9] D. Kirkpatrick and J. Radke, *A framework for computational morphology*, in: Computational Geometry, Machine Intelligence and Pattern Recognition, 2, North-Holland, Amsterdam, 1985, 217—248.

[10] J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, *Problems of the American Mathematical Society*, **7** (1956), 48–50.

[11] H. T. Lawless and H. Heymann, *Principles of Sensory Evaluation*, Aspen Publishers, Gaithersburg, 1999.

[12] J. Pagès, Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley, *Food Quality and Preference*, **16** (2005), 642–649.

[13] L. Perrin , R. Symoneaux, I. Maître, C. Asselin, F. Jourjon , and J. Pagès, Comparison of three sensory methods for use with the Napping procedure: Case of ten wines from Loire valley, *Food Quality and Preference*, **19** (2008), 1–11.

[14] M. I. Shamos and D. Hoey, Closest-point problems, in *Proceedings of FOCS*, 1975, 151–162.

[15] G. T. Toussaint, The relative neighborhood graph of a finite planar set, *Pattern Recognition*, **12** (1980), 261–268.