Efficient Representation of Pairwise Sensory Information

John M. Ennis and Daniel M. Ennis

Background: Pairwise comparisons appear throughout sensory science – two products can be compared for statistical difference, equivalence, compatibility, and so on. When many products are compared, there can be hundreds of these comparisons. And if these comparisons occur on each of a large number of attributes, the total number of comparisons can be in the tens of thousands. An outstanding challenge is to represent this pairwise information as concisely as possible, and in this report we meet this challenge by sharing recent developments in the production of tables known as compact letter displays. These displays have been optimized using techniques from the mathematical field of graph theory – using these displays sensory scientists can represent pairwise information with maximal efficiency.

Scenario: You work for a manufacturer of whey protein based sports beverages and have just conducted a large-scale category appraisal of 22 products evaluated by 240 consumers in a nationwide study. The goal of this study is to fully understand the drivers of liking space for your product category. During the study, the consumers provided hedonic ratings of all 22 products, together with attribute intensity scores for the products on 18 sensory scales. In order to represent the statistically significant differences between the products on the various attributes, you decide to use letter displays. In these displays, product means are shown together with a set of letters – means with at least one letter in common are not significantly different from each other[1]. See Table 1.

Beverage	Sweetness
1	6.25 a
2	5.95 ab
3	5.92 abc
4	5.79 abc
5	5.75 abcd
6	5.71 bcde
7	5.68 bcde
8	5.65 bcdef
9	5.62 bcdef
10	5.59 bcdefg
11	5.54 bcdefgh
12	5.53 bcdefgh
13	5.52 bcdefghi
14	5.50 bcdefghi
15	5.42 bcdefghi
16	5.41 cdefghi
17	5.24 defghi
18	5.20 efghi
19	5.13 fghi
20	5.06 ghi
21	5.01 hi
22	5.00 i

Table 1. Means ranked from highest to lowest. Means with at least one letter in common are not significantly different at the 95% level.

Having created the full set of tables of product means for the 18 attributes in your study, you notice some inefficiencies in the letter assignments. For example, in Table 1, the letter 'b' placed on Beverage 3 carries redundant information – every product with a 'b' also has one of the letters 'a' or 'c.' Since Beverage 3 has also been assigned both 'a' and 'c,' these two letters by themselves inform you of all of the beverages that are not significantly different from Beverage 3 on sweetness. You begin to wonder how many more letters are not needed in your tables.

Compact Letter Displays: In the past few years, letter displays have been optimized with respect to several measures using tools from the mathematical field of graph theory[2,3,4]. Graph theory is the study of pairwise connections between objects, and in the case of putting letters on means we consider each mean to be an object in a graphical network. In this network, means that are not significantly different from each other are connected while means that are significantly different from each other are not. When means in a group all carry a letter in common, these means will all not be significantly different from each other. Thus, from a graphical standpoint, all of the means will be connected to each other. As explained in a previous technical report[5], a fully connected set of objects within a graphical network is called a 'clique.' For example, see Table 2 for a sample letter display and Figure 1 for the corresponding graph with cliques identified for each letter.

Product	Letter
1	a
2	ab
3	abc
4	bc
5	c

Table 2. A sample letter display for 5 products. For clarity, the means for the products are not listed.

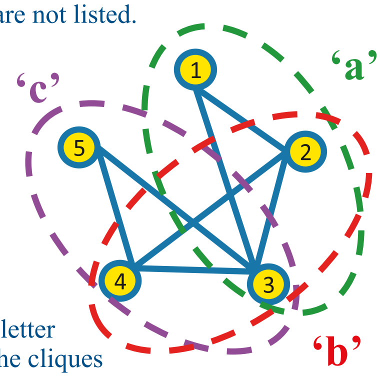

Figure 1. A graph corresponding to the letter display in Table 2. The cliques defined by the letters in Table 2 are labeled accordingly.

As in Table 1, there is inefficiency in Table 2 – the letter 'b' on Product 3 is not needed. Using graph theory we can understand this inefficiency in terms of the corresponding graphical network. Specifically, in order to capture all of the significant differences in a letter display, we need to find cliques that together cover all of the edges in the corresponding graph. Such collections of cliques are called 'clique coverings.'

Reducing Displays using Graph Theory: In the case of Figure 1, we see that the edge from Product 2 to Product 3 is covered by both of the cliques 'a' and 'b.' Similarly, the edge from Product 3 to Product 4 is covered by both of the cliques 'b' and 'c.' Thus we can remove Product 3 from Clique 'b' and still cover every edge of the graph. This removal gives the smaller clique covering shown in Figure 2 and the more concise letter display shown in Table 3.

Product	Letter
1	a
2	ab
3	ac
4	bc
5	c

Table 3. A more concise letter display representing the information in Table 2.

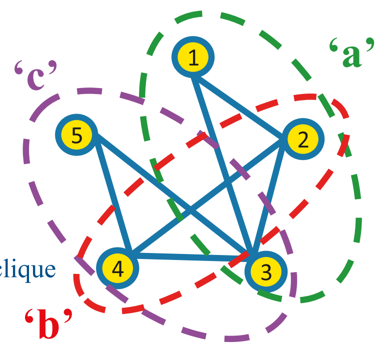

Figure 2. A smaller clique covering of the graph shown in Figure 1.

Beverage	Sweetness
1	6.25 a
2	5.95 ab
3	5.92 abc
4	5.79 abc
5	5.75 abcd
6	5.71 be
7	5.68 be
8	5.65 bf
9	5.62 bf
10	5.59 bfg
11	5.54 bfh
12	5.53 bfh
13	5.52 bi
14	5.50 bi
15	5.42 bi
16	5.41 cefi
17	5.24 defi
18	5.20 efi
19	5.13 fi
20	5.06 ghi
21	5.01 hi
22	5.00 i

Table 4. A compact letter display showing the statistically significant differences between the beverages on 'sweetness.'

This example shows we can obtain maximally concise letter displays by searching for smallest clique coverings of graphs. To find smallest clique coverings, there are two measures we can use. We can search for clique coverings that contain either: 1) as few cliques as possible or 2) as few assignments of objects to cliques as possible. These two approaches give letter displays that are respectively minimal in the number of: 1) unique letters throughout the display or 2) total number of letters used in the display. Displays that achieve either of these goals are called 'compact letter displays.' An algorithm to accomplish goal (1) was described by Gramm *et al*. (2007, 2009) while an algorithm to accomplish goal (2) is given in Ennis *et al*. (2012). In addition, Ennis *et al*. demonstrated through example that it is not always possible to achieve goals (1) and (2) simultaneously.

Return to Scenario: For the purposes of creating clearer tables for your report, you decide to pursue goal (2). Using the algorithm of Ennis *et al*., as implemented in the *IFPrograms*™ software, you reduce your letter displays for each attribute. Following this reduction, the statistical differences for the 22 products on all 18 attributes can be shown on only a few slides. For example, Table 4 shows the compact letter display for 'sweetness.' In Table 4, only 55 total letters are used instead of the 103 total letters in Table 1.

Conclusion: In this report we focused on the use of compact letter displays to concisely represent multiple statistical comparisons of products on attributes. When there are many products and many attributes, the use of such displays aids in the efficient representation of information. This efficiency brings the practical benefits of smaller tables and faster access to the information contained in the tables. And, in addition to this improvement in efficiency, graph theory also allows us to develop compact letter displays for means with different variances that cannot be formed from ranking and subsetting.

Recent advances in graph theory allow us to concisely represent pairwise information. Since pairwise information appears throughout sensory science – whether in the form of statistical differences, equivalency, compatibility, or otherwise – these advances help sensory researchers realize efficiency in displaying their data explorations.

References and Notes

1. Letter displays allow for many attributes to be displayed in the same table. Also, letter displays can accommodate unequal variance between product means on a single attribute. See Piepho[6,7] for more details.
2. Gramm, J., Guo, J., Huffner, F., and Niedermeier, R. (2009) Data reduction and exact algorithms for clique cover. *ACM Journal of Experimental Algorithmics*. **13**, 2.2:2-2.2:15.
3. Gramm, J., Guo, J., Huffner, F., Niedermeier, R., Piepho, H. P., and Schmid, R. (2007) Algorithms for compact letter displays: Comparison and evaluation. *Computational Statistics & Data Analysis*. **52**(2), 725-736.
4. Ennis, J. M., Ennis, D. M., and Fayle, C. M. (2012). Assignment-minimal clique coverings. *ACM Journal of Experimental Algorithmics*. **17**(1), 1.5:1-1.5:17.
5. Ennis, J. M., Fayle, C. M., and Ennis, D. M. (2011) From many to few: A graph theoretic screening tool for product developers. *IFPress*. **14**(2), 3-4.
6. Piepho, H. P. (2000). Multiple treatment comparisons in linear models when the standard error of a difference is not constant. *Biometrical Journal*. **42**(7), 823-835.
7. Piepho, H. P. (2004). An algorithm for a letter-based representation of all-pairwise comparisons. *Journal of Computational and Graphical Statistics*. **13**(2), 456-466.