



OPINION

Open Access

Network analysis and data mining in food science: the emergence of computational gastronomy

Sebastian E Ahnert

Abstract

The rapidly growing body of publicly available data on food chemistry and food usage can be analysed using data mining and network analysis methods. Here we discuss how these approaches can yield new insights both into the sensory perception of food and the anthropology of culinary practice. We also show that this development is part of a larger trend. Over the past two decades large-scale data analysis has revolutionized the biological sciences, which have experienced an explosion of experimental data as a result of the advent of high-throughput technology. Large datasets are also changing research methodologies in the social sciences due to the data generated by mobile communication technology and online social networks. Even the arts and humanities are seeing the establishment of 'digital humanities' research centres in order to cope with the increasing digitization of literary and historical sources. We argue that food science is likely to be one of the next beneficiaries of large-scale data analysis, perhaps resulting in fields such as 'computational gastronomy'.

Keywords: Networks, Data mining, Sensory science, Computational gastronomy, Flavour compounds

Large-scale data analysis

The past two decades have seen the advent of high-throughput technologies in biology, making it possible to sequence genomes cheaply and quickly, to measure gene expression for thousands of genes in parallel, and to test large numbers of potential regulatory interactions between genes in a single experiment. The large amounts of data created by these technologies have given rise to entire new research areas in biology, such as computational biology and systems biology. The latter, which attempts to understand biological processes at a 'systems' level, is particularly indicative of the potential advantage that large datasets and their analysis can offer to biology, and to other fields of research. This advantage is a 'birds-eye' perspective, which, with the right kind of analysis, can complement the more established research methods that take place 'on the ground' and investigate the system in much more detail. An example would be the analysis of high-throughput gene expression data of tumour tissues in order to highlight a set of potential candidate genes that may play a role

in causing a particular cancer. These candidates would then be investigated one by one, for instance by creating mutant organisms in which one of these genes is deactivated.

Similar large-scale data analysis methods have more recently arrived in the social sciences as a result of rapidly growing mobile communications networks and online social networking sites. Here too data analysis offers a birds-eye perspective of large social networks and the opportunity to study social dynamics and human mobility on an unprecedented scale. The most recent research areas to be transformed by information technology are the Arts and Humanities, which have witnessed the emergence of 'digital humanities'. As more and more literary and historical documents are digitized, it becomes possible to uncover fundamental relationships that underlie large corpora of literary texts, or long-term historical and political developments. A striking example is the discovery by Lieberman *et al.* [1] that the regularisation of verbs across 12 centuries of English is governed by a simple quantitative relationship between the frequency of verb usage and the speed at which it is regularised.

Correspondence: sea31@cam.ac.uk
Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK



© 2013 Ahnert; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

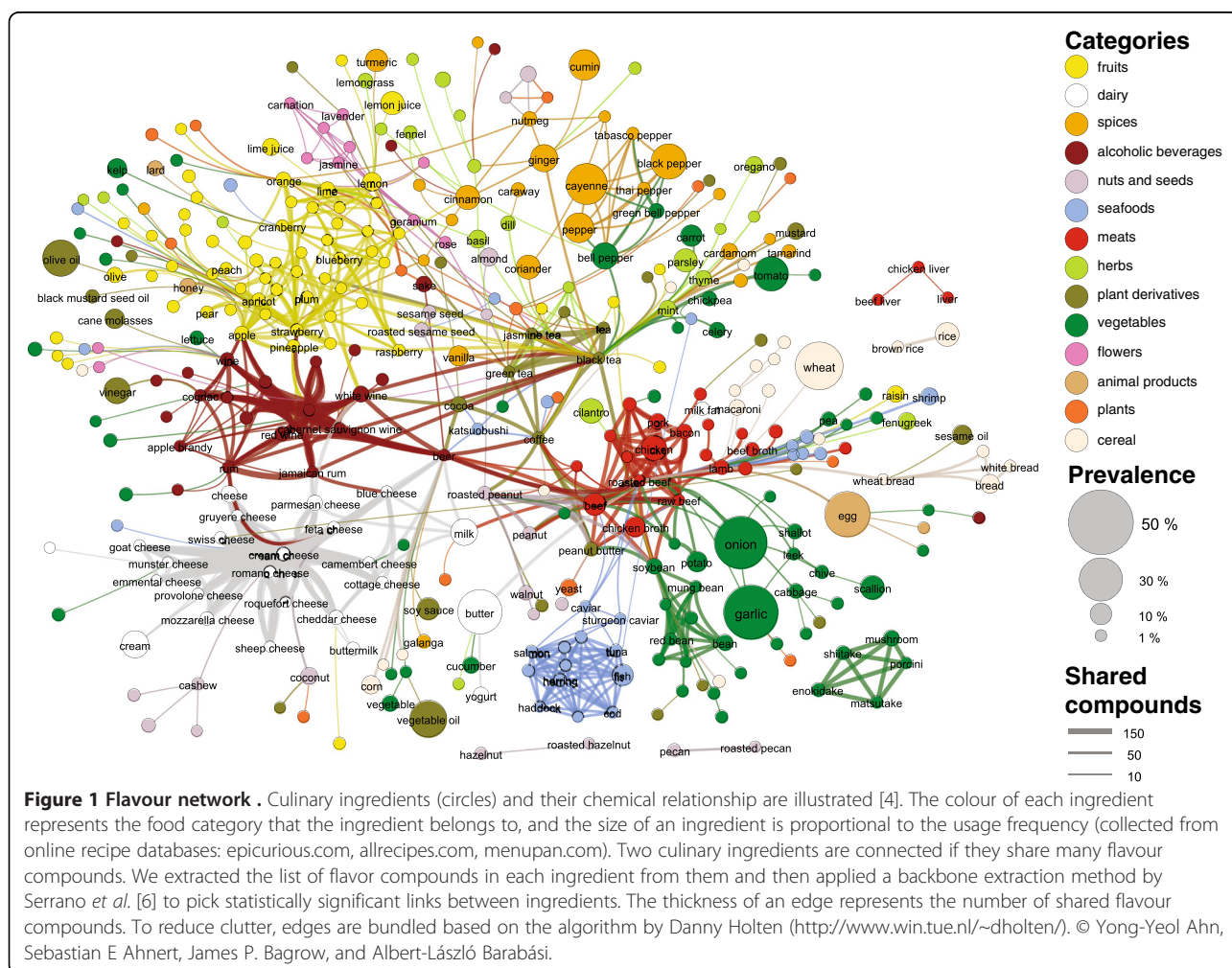
Network analysis of flavour compounds

The growing availability of network data in a wide variety of research disciplines has made complex network analysis a rapidly growing research area ever since two seminal publications in the late 1990s uncovered fundamental principles that underlie many real-world networks such as social networks, power grids, neural networks and genetic regulatory networks [2,3]. In recent work [4] we construct a bipartite network of chemical flavour compounds and food ingredients in which a link signifies the natural occurrence of a compound in an ingredient. These data were derived from *Fenaroli's Handbook of Flavour Ingredients* [5]. Using a one-mode projection the bipartite network is converted into a weighted network of ingredients only, in which the weight of a link between two ingredients is given by the number of flavour compounds they share. This weighted network shows a modular organization, with modules corresponding to food types such as fruits, vegetables and meats. While this might be expected, it is particularly interesting to see the location of these modules with respect to each other.

Meats for instance lie between fruits and vegetables, and closer to spices and herbs than seafood does. The backbone of this network, extracted using the method described in [6], is shown in Figure 1.

The chef Heston Blumenthal, together with flavour scientists, has suggested that two foods that share chemical flavour compounds are more likely to taste good in combination [7]. By comparing the network of ingredients to a body of 56,498 online recipes, downloaded from epicurious.com, allrecipes.com, and menupan.com, we were able to show that this hypothesis is confirmed in most Western cuisines, but not in Eastern ones. This result indicates that shared compounds may offer one of several possible mechanisms that can make two ingredients compatible.

Our network of ingredients and flavour compounds is just a first step towards a true network of shared flavour compound perception, which would have to include compound concentrations [8] and detection thresholds [9] in order to further investigate the shared compound hypothesis. Its most important purpose is to open up a



new way in which data analysis can aid sensory science and the study of culinary practice.

In a broader development the increasing availability of data on food usage, food chemistry and sensory biology is likely to result in the establishment of new research disciplines, such as 'computational gastronomy'.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

SEA is supported by the Royal Society, UK.

Received: 7 November 2012 Accepted: 8 November 2012

Published: 9 January 2013

References

1. Lieberman E, Michel JB, Jackson J, Tang T, Nowak MA: **Quantifying the evolutionary dynamics of language.** *Nature* 2007, **449**:713–716.
2. Watts D, Strogatz S: **Collective dynamics of "small-world" networks.** *Nature* 1998, **393**:440–442.
3. Barabási AL, Albert R: **Emergence of scaling in random networks.** *Science* 1999, **286**:509–512.
4. Ahn YY, Ahnert SE, Bagrow JP, Barabási AL: **Flavor network and the principles of food pairing.** *Sci Rep* 2011, **1**:196.
5. Burdock GA: *Fenaroli's Handbook of Flavor Ingredients.* 5th edition. Boca Raton: CRC Press; 2004.
6. Serrano MA, Boguñá M, Vespignani A: **Extracting the multiscale backbone of complex weighted networks.** *Proc Natl Acad Sci USA* 2009, **106**:6483–6488.
7. Blumenthal H: *The Big Fat Duck Cookbook.* London: Bloomsbury; 2008.
8. Nijssen LM, van Ingen-Visscher CA, Donders JJH: *VCF Volatile Compounds in Food: database. – Version 13.2 – Zeist.* The Netherlands: TNO Triskelion; 1963–2012.
9. van Gemert LJ: *Flavour threshold values in water and other media.* 2nd edition. Canton, GA: Leffingwell; 2011.

doi:10.1186/2044-7248-2-4

Cite this article as: Ahnert: Network analysis and data mining in food science: the emergence of computational gastronomy. *Flavour* 2013 **2**:4.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

