Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Forensic Science International xxx (2006) xxx–xxx
Forensic Science International
www.elsevier.com/locate/forsciint

Pattern detection in forensic case data using graph theory: Application to heroin cutting agents

Anne-Laure Terrettaz-Zufferey [a,1,*], Frédéric Ratle [b,2], Olivier Ribaux [a,1], Pierre Esseiva [a,1], Mikhail Kanevski [b,2]

[a] Institut de Police Scientifique et de Criminologie, Ecole des Sciences Criminelles, Université de Lausanne, Batochime, CH-1015, Switzerland
[b] Institut de Géomatique et d'Analyse du Risque, Faculté des Géosciences et de l'Environnement, Université de Lausanne, Amphipôle, CH-1015, Switzerland

Received 2 June 2006; accepted 14 June 2006

Abstract

Pattern recognition techniques can be very useful in forensic sciences to point out to relevant sets of events and potentially encourage an intelligence-led style of policing. In this study, these techniques have been applied to categorical data corresponding to cutting agents found in heroin seizures. An application of graph theoretic methods has been performed, in order to highlight the possible relationships between the location of seizures and co-occurrences of particular heroin cutting agents. An analysis of the co-occurrences to establish several main combinations has been done. Results illustrate the practical potential of mathematical models in forensic data analysis.

Keywords: Heroin seizures; Profiling; Adjacency matrix; Co-occurrences; Swiss cantons

1. Introduction

Pattern recognition techniques have been applied successfully in forensic sciences mainly to the problem of image processing [1,2]. Such methods are also used in the field of drug profiling, as described in ref. [3]. They provide information about links between seizures through the processing of the chemical profile of illicit drug seizures using statistical models of the data [4,5]. Other pattern recognition methods have been applied to illicit drug data using clustering and classification algorithms in refs. [6–8]. Alternative methods of clustering data with graphs is described in refs. [9,10]. An application of graph theory for modelling time intervals of real-case palaeontological data was made in ref. [12]. It is reasonable to suppose that such methods could be adapted and used in the field of drug profiling.

This study focuses on cutting agents in heroin seizures. They represent semi-quantitative data, collected in the same experimental analysis as the major constituents of heroin [4]. Their distribution in the four regions of investigation (canton of Vaud, Geneva, Neuchâtel and Tessin) is not homogeneous. Some products are present in all regions and other tend to be endemic. This fact encourages the study of the co-occurrences of cutting agents. This investigation aims to assess the performance of a combinatorial method to process cutting agents data. The application of mathematical models related to graph theory [11] can potentially help through visualisation the detection of regularities or patterns in the co-occurrences of cutting agents. On this basis, hypotheses can be developed about particular aspects of drug processing and distribution processes.

2. Methodology

2.1. The data

The studied area covers four Swiss cantons: Vaud, Geneva, Neuchâtel and Tessin (Fig. 1). In the framework of the profiling methodology [3], cutting agents are detected through gas chromatography analysis in the same way as major constituents. They are transformed into a binary format: the presence or absence of the agents in the samples are indicated, respectively, as 1s and 0s in a matrix. Eighteen cutting agents have been considered.

* Corresponding author. Tel.: +41 216924657.
E-mail address: anne-laure.terrettaz@unil.ch (A.-L. Terrettaz-Zufferey).
[1] Tel.: +41 21 692 56 57; fax: +41 21 692 46 05.
[2] Tel.: +41 21 692 35 31; fax: +41 21 692 35 35.

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2006.06.059

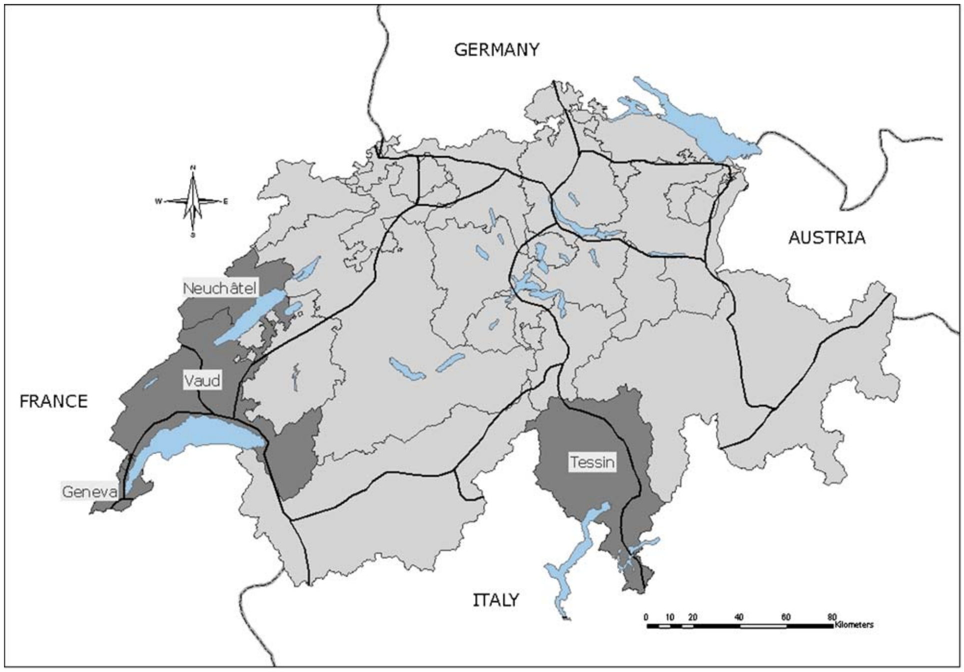
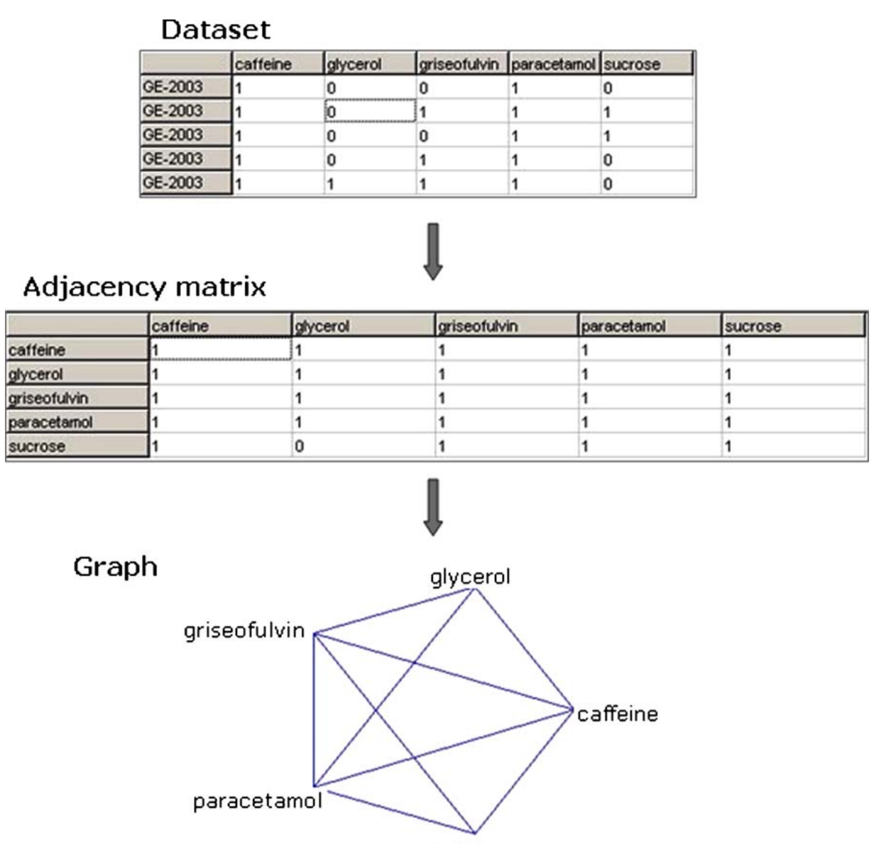


Fig. 1. In dark the regions of study.

2.2. *Basic definitions and notations*

A graph $G = (V, E)$ is specified by its vertex set, V, and edge set, E, that represents connections between each vertex. We usually define V as: $V = v_1, \ldots, v_n$. A matrix called the *adjacency matrix*, A_G, is associated to the graph, G, and its entries a_{ij} are given by

$$\alpha_{ij} = \begin{cases} 1, & \text{if } \{i, j\} \in E \\ 0, & \text{otherwise} \end{cases}$$



Dataset

	caffeine	glycerol	griseofulvin	paracetamol	sucrose
GE-2003	1	0	0	1	0
GE-2003	1	0	1	1	1
GE-2003	1	0	0	1	1
GE-2003	1	0	1	1	0
GE-2003	1	1	1	1	0

Adjacency matrix

	caffeine	glycerol	griseofulvin	paracetamol	sucrose
caffeine	1	1	1	1	1
glycerol	1	1	1	1	1
griseofulvin	1	1	1	1	1
paracetamol	1	1	1	1	1
sucrose	1	0	1	1	1

Fig. 2. Modelling process from the dataset to the graph. Only five cutting agents are presented for clarity reasons.

2.3. Method

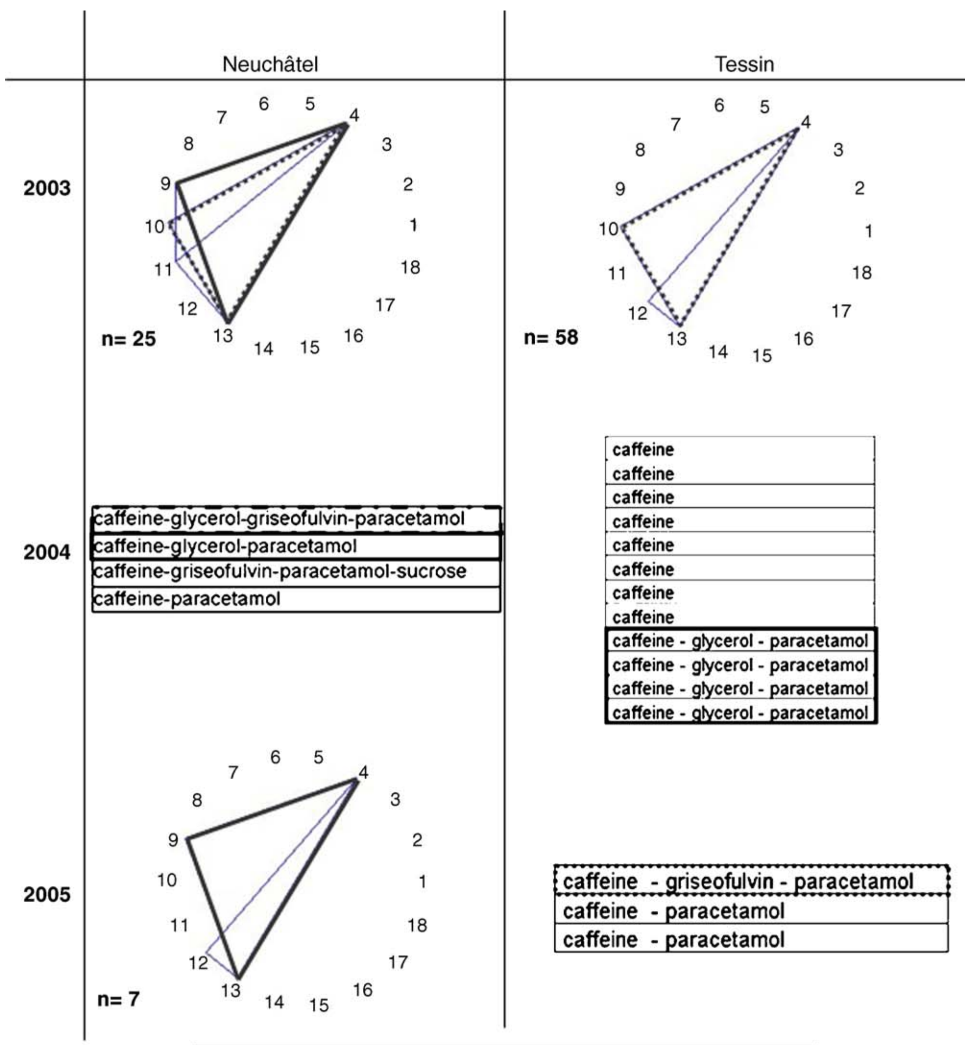
Fig. 2 illustrates, with a small subset of the data, the modelling process followed in this study:

(1) Each year of the binary data concerning 18 cutting agents extracted from the analysis seizures from one region is considered separately.
(2) $V = \{v_1, \ldots, v_{18}\}$ represents the list of the 18 detected cutting agents and corresponds to the vertex of the graph; $E = \{v_i, v_j\}$ represents the edges between the vertices, which appear if two cutting agents are found together at least one time during 1 year.
(3) Construction of the adjacency matrix A_G whose entries are $a_{ij} \in \{0, 1\}$, and of the corresponding graph.

Cutting agents are in the same alphabetical order in the dataset as on the graphs. No dependencies between products are represented with the graphs.

2.4. Tools

All experiments were done using UA-Graph [13] and Matlab. The graph theory implementation was taken from free toolboxes proposed on the Mathworks website (http://www.mathworks.com).



1	citric acid	7	fructose	13	paracetamol
2	fatty acid	8	glucose	14	phenobarbital
3	aspirin	9	glycerol	15	phosphate
4	caffeine	10	griseofulvin	16	piracetam
5	diazepam	11	lactose	17	procaine
6	diethylene	12	mannitol	18	sucrose

Fig. 3. Graphs of the co-occurrences for cantons of Neuchtel and Tessin [n is the number of samples in the dataset].

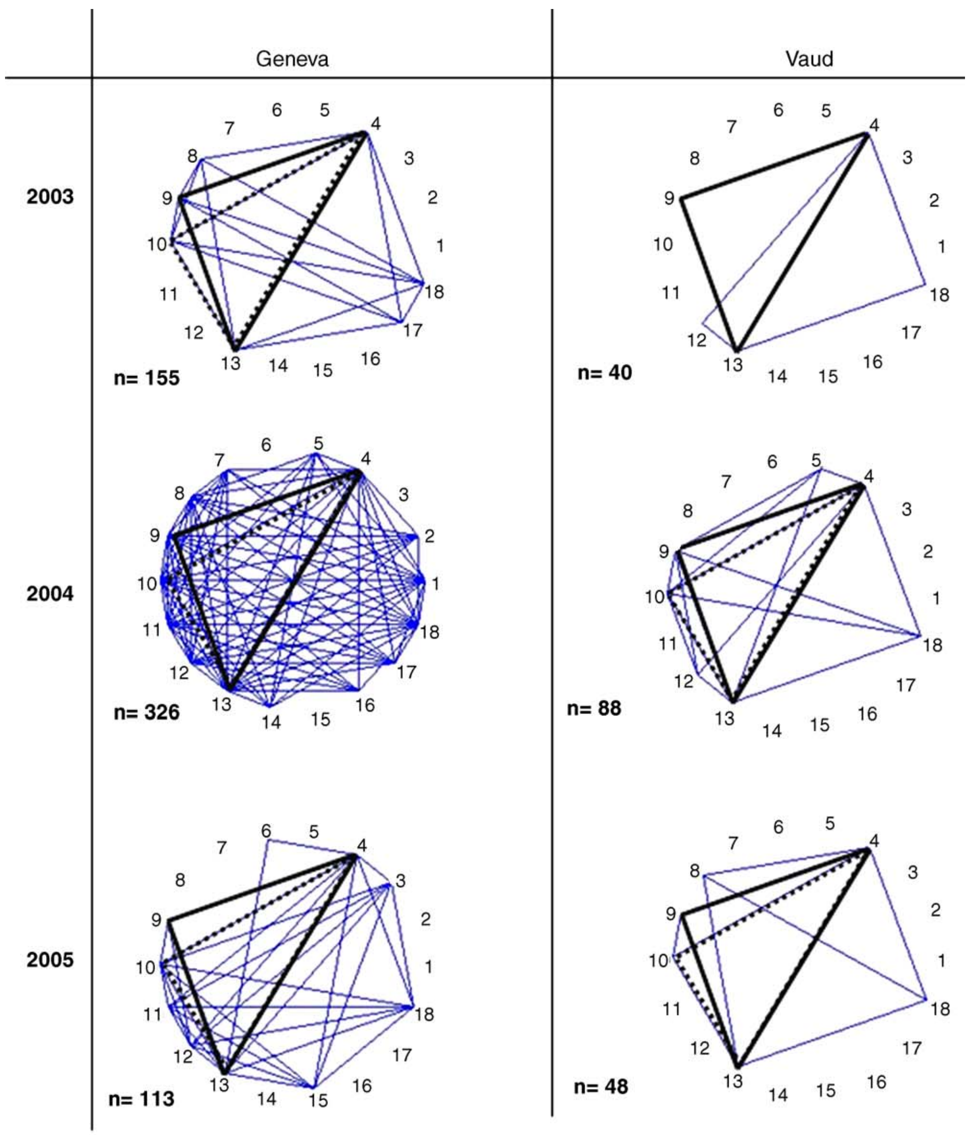
3. Results

The co-occurrence of cutting agents has been highlighted for the four studied cantons during 3 years. The adjacency matrix has been computed for years with enough relationships between products. The lack of co-occurences between cutting agents for cantons of Neuchâtel and Tessin does not allow the construction of an adjacency matrix, and thus of a graph. Figs. 3 and 4 show the structures of the graphs by years and cantons and the combination are mentionned for cases without graph representation.

4. Discussion

Graph structures range from very simple to highly complex. It can be observed that complexity does not depend solely on the quantity of seizures. Various interpretations are possible.

It is known that the data from Tessin (see Fig. 3) represents exclusively important seizures made at the border. They are articulated around the combination ''caffeine–paracetamol–griseofulvin'' (c–p–g). Data collected at Geneva and Vaud contains a broader variety of seizure's types (border, result of investigations and street seizures). They show more complex



1	citric acid	**7**	fructose	**13**	paracetamol
2	fatty acid	**8**	glucose	**14**	phenobarbital
3	aspirin	**9**	glycerol	**15**	phosphate
4	caffeine	**10**	griseofulvin	**16**	piracetam
5	diazepam	**11**	lactose	**17**	procaine
6	diethylene	**12**	mannitol	**18**	sucrose

Fig. 4. Graphs of the co-occurrences for cantons of Geneva and Vaud [n is the number of samples in the dataset].

structures (Fig. 4). In all the regions, mostly the same c–p–g mixture appears with other combinations of cutting agents. A plausible hypothesis is that the cutting process occurs also within the country, at a very local level, whereas a first cutting process of this nature is performed earlier in the distribution process. Consequently, those variables show a promising potential in an intelligence perspective for analysing the local distribution process of heroin.

The higher complexity of the graph structures representing data from the canton of Geneva can be explained by the strategic situation of the region, which entails an international airport, several customs and a highway bounding France and Switzerland. There is evidence that these factors influence the diversity of the illicit drug traffic that is reflected through the graphs.

The complexity of the graphs seems to evolve during the period under consideration, rather independently of the number of seizures. It could reveal a time-dependent change in the cutting process. This hypothesis is supported by the observed decrease in the global purity of the seized heroin in Switzerland (http://www.sgrm.ch/, last accessed 29/05/2006). A possible interpretation is that heroin was scarce during a particular period around 2004–2005. This could have led to a more intensive cutting, possibly earlier in the drug fabrication process, in order to fill market demand.

Differences can also be observed through graph comparisons. For instance, the presence of mannitol varies as a function of the year and the region under consideration. These types of variations can justify a spatio-temporal analysis of flows of specific cutting agents.

Graph visualisation has shown to be very promising for the detection of patterns leading to hypothesis that can be then tested. The intelligence potential of the approach is therefore validated. Perspectives for improving the method are numerous.

The number of co-occurrences of the cutting agents for the period under consideration is not represented. It could eventually be shown through the adaptation of the width of the edges. Moreover, co-occurrences of more than two products can be highlighted through edges colours. It can also be observed that some combinations of cutting agents are more frequent than others. This could lead to a reorganisation of the graphs in order to make them more readable, but they must stay comparable. Finally, the chosen period for each graph is somewhat arbitrary. Ideally, the influence of the length of the period would be worth investigating. The development of a dedicated software tool would improve the possibilities to explore and compare in a more comprehensive way those graphical structures on every aspect.

It is planned that the potential of this visualisation method will be assessed on other types of forensic case data.

5. Conclusion

Graph theory has been used to represent the co-occurrence and evolution of heroin cutting agents within specific regions. This visualisation method has helped to develop a great variety of hypotheses explaining the local cutting process.

From this preliminary study, ways for improvement have been suggested. Particularly, refining the visualisation process could be reached through the development of an exploratory software tool that would allow the detailed inspection of all aspects of the structures.

This promising method can be integrated into broader intelligence approaches and will be tested in order to detect patterns within other complex forensic datasets.

Acknowledgement

This work was supported by the Swiss National Science Foundation (grant no. 105211-107862).

References

[1] S.Z. Li, A.K. Jain (Eds.), Handbook of Face Recognition, Springer, 2005.
[2] J. Wayman, A. Jain, D. Maltoni, D. Maio (Eds.), Biometric System, Technology, Design and Performance Evaluation, Springer, 2005.
[3] O. Guéniat, P. Esseiva, Le Profilage de l'Héroïne et de la Cocaïne, Presses Polytechniques et Universitaires Romandes, Lausanne (2005).
[4] P. Esseiva, L. Dujourdy, F. Anglada, F. Taroni, P. Margot, A methodology for illicit drug intelligence perspective using large databases, Forensic Sci. Int. 132 (2003) 139–152.
[5] P. Esseiva, F. Anglada, L. Dujourdy, F. Taroni, P. Margot, E. Du Pasquier, M. Dawson, C. Roux, P. Doble, Chemical profiling and classification of illicit heroin by principal component analysis, calculation of inter sample correlation and artificial neural networks, Talanta 67 (2005) 360–367.
[6] F. Ratle, A.L. Terrettaz, M. Kanevski, P. Esseiva, O. Ribaux, Pattern analysis in illicit heroin seizures: a novel application of machine learning algorithms, in: European Symposium on Artificial Neural Networks, 2006.
[7] M.G. Madden, A.G. Ryder, Machine learning methods for quantitative analysis of Raman spectroscopy data, in: Proceedings of the International Society for Optical Engineering (SPIE 2002), vol. 4876, 2002, pp. 1130–1139.
[8] M.L. O'Connell, T. Howley, A.G. Ryder, M.G. Madden, Classification of a target analyte in solid mixtures using principal component analysis, support vector machines, and Raman spectroscopy, in: Proceedings of the International Society for Optical Engineering (SPIE 2005), vol. 5826, 2005, pp. 340–350.
[9] S. Yu, J. Shi, Multiclass spectral clustering, in: Proceedings of ICCV, 2003.
[10] M. Peter, M.J. Zaki, CLICK: Clustering Categorical Data using K-partite Maximal Cliques, Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180, http://www.citeseer.ist.psu.edu/678652.html, last access June 2006.
[11] D.B. West, Introduction to Graph Theory, 2nd ed., Prentice Hall, Upper Saddle River, 2001.
[12] J. Guex, E. Davaud, Unitary Associations Method: Use of Graph Theory and Computer Algorithm Comp. Geosc.,10/1, pp. 195–205.
[13] O. Hammer, J. Guex, J. Savary, UA-Graph, http://folk.uio.no/ohammer/uagraph, last access June 2006.