

Using formal concept analysis to detect and monitor organised crime

ANDREWS, Simon <<http://orcid.org/0000-0003-2094-7456>>, AKHGAR, Babak <<http://orcid.org/0000-0003-3684-6481>>, YATES, Simeon, STEDMON, Alex and HIRSCH, Laurence <<http://orcid.org/0000-0002-3589-9816>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/8608/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

ANDREWS, Simon, AKHGAR, Babak, YATES, Simeon, STEDMON, Alex and HIRSCH, Laurence (2013). Using formal concept analysis to detect and monitor organised crime. In: LARSEN, Henrik Legind, MARTIN-BAUTISTA, Maria J., VILA, María Amparo, ANDREASEN, Troels and CHRISTIANSEN, Henning, (eds.) Flexible Query Answering Systems. Lecture Notes in Computer Science (8132). Heidelberg, Springer, 124-133.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Using Formal Concept Analysis to Detect and Monitor Organised Crime

Simon Andrews, Babak Akhgar, Simeon Yates, Alex Stedmon, Laurence Hirsch

CENTRIC **

Sheffield Hallam University, Sheffield, UK

{s.andrews, b.akhgar, s.yates, a.stedmon, l.hirsch}@shu.ac.uk

Abstract. This paper describes some possible uses of Formal Concept Analysis in the detection and monitoring of Organised Crime. After describing FCA and its mathematical basis, the paper suggests, with some simple examples, ways in which FCA and some of its related disciplines can be applied to this problem domain. In particular, the paper proposes FCA-based approaches for finding multiple instances of an activity associated with Organised Crime, finding dependencies between Organised Crime attributes, and finding new indicators of Organised Crime from the analysis of existing data. The paper concludes by suggesting that these approaches will culminate in the creation and implementation of an Organised Crime ‘threat score card’, as part of an overall environmental scanning system that is being developed by the new European ePOOLICE project.

1 Introduction

Efficient and effective scanning of the environment for strategic early warning of Organised Crime (OC) is a significant challenge due to the large and increasing amount of potentially relevant information that is accessible [5, 18]. The types of question and analysis required are not always clear-cut or of a straightforward numerical/statistical nature, but rather necessitate a more conceptual or semantic approach. New developments in computational intelligence and analytics, have opened up new solutions for meeting this challenge. A theoretical development of particular interest for this purpose is Formal Concept Analysis (FCA), with its faculty for knowledge discovery and ability to intuitively visualise hidden meaning in data [3, 19]. This is particularly important in environmental scanning, where many of the signals are weak, with information that may be incomplete, imprecise or unclear.

The potential for FCA to reveal semantic information in large amounts of data is beginning to be realised by developments in efficient algorithms and their implementations [1, 13] and by the better appropriation of diverse data for FCA [2, 4].

** Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research

This paper describes some possible approaches for detecting and monitoring OC using these advances in FCA.

2 Formal Concept Analysis

Formal Concept Analysis (FCA) was introduced in the 1990s by Rudolf Wille and Bernhard Ganter [8], building on applied lattice and order theory developed by Birkhoff and others in the 1930s. It was initially developed as a subsection of Applied Mathematics based on the mathematisation of concepts and concepts hierarchy, where a concept is constituted by its *extension*, comprising of all objects which belong to the concept, and its *intension*, comprising of all attributes (properties, meanings) which apply to all objects of the extension. The set of objects and attributes, together with their relation to each other, form a *formal context*, which can be represented by a cross table.

| Airlines | Latin America | Europe | Canada | Asia Pacific | Middle east | Africa | Mexico | Caribbean | USA |
|-------------------|---------------|--------|--------|--------------|-------------|--------|--------|-----------|-----|
| Air Canada | × | × | × | × | × | | × | × | × |
| Air New Zealand | | × | × | × | | | | | × |
| Nippon Airways | | × | | × | | | | | × |
| Ansett Australia | | | | × | | | | | |
| Austrian Airlines | | × | × | × | × | × | | | × |

2.1 Formal Contexts

The cross-table above shows a formal context representing destinations for five airlines. The elements on the left side are formal objects; the elements at the top are formal attributes. If an object has a specific property (formal attribute), it is indicated by placing a cross in the corresponding cell of the table. An empty cell indicates that the corresponding object does not have the corresponding attribute. In the Airlines context above, Air Canada flies to Latin America (since the corresponding cell contains a cross) but does not fly to Africa (since the corresponding cell is empty).

In mathematical terms, a formal context is defined as a triple $\mathbb{K} := (G, M, I)$, with G being a set of objects, M a set of attributes and I a relation defined between G and M . The relation I is understood to be a subset of the cross product between the sets it relates, so $I \subseteq G \times M$. If an object g has an attribute m , then $g \in G$ relates to m by I , so we write $(g, m) \in I$, or gIm . For a subset of objects $A \subseteq G$, a derivation operator $'$ is defined to obtain the set of attributes, common to the objects in A , as follows:

$$A' = \{m \in M \mid \forall g \in A : gIm\}$$

Similarly, for a subset of attributes $B \subseteq M$, the derivation operator $'$ is defined to obtain the set of objects, common to the attributes in B , as follows:

$$B' = \{g \in G \mid \forall m \in B : gIm\}$$

2.2 Formal Concepts

Now, a pair (A, B) is a *Formal Concept* in a given formal context (G, M, I) only if $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. The set A is the extent of the concept and the set B is the intent of the concept. A formal concept is, therefore, a closed set of object/attribute relations, in that its extension contains all objects that have the attributes in its intension, and the intension contains all attributes shared by the objects in its extension. In the Airlines example, it can be seen from the cross-table that Air Canada and Austrian Airlines fly to both USA and Europe. However, this does not constitute a formal concept because both airlines also fly to Asia Pacific, Canada and the Middle East. Adding these destinations completes (closes) the formal concept:

$$(\{\text{Air Canada, Austrian Airlines}\}, \{\text{Europe, USA, Asia Pacific, Canada, Middle East}\}).$$

2.3 Galois Connections

Another central notion of FCA is a duality called a ‘Galois connection’, which is often observed between items that relate to each other in a given domain, such as objects and attributes. A Galois connection implies that “if one makes the sets of one type larger, they correspond to smaller sets of the other type, and vice versa” [15]. Using the formal concept above as an example, if Africa is added to the list of destinations, the set of airlines reduces to $\{\text{Austrian Airlines}\}$.

2.4 Concept Lattices

The Galois connections between the formal concepts of a formal context can be visualized in a *Concept Lattice* (Figure 1), which is an intuitive way of discovering hitherto undiscovered information in data and portraying the natural hierarchy of concepts that exist in a formal context.

A concept lattice consists of the set of concepts of a formal context and the subconcept-superconcept relation between the concepts. The nodes in Figure 1 represent formal concepts. It is conventional that formal objects are noted slightly below and formal attributes slightly above the nodes, which they label.

A concept lattice can provide valuable information when one knows how to read it. As an example, the node which is labeled with the formal attribute ‘Asia Pacific’ shall be referred to as *Concept A*. To retrieve the extension of Concept A (the objects which feature the attribute ‘Asia Pacific’), one begins at the node where the attribute is labeled and traces all paths which lead down from the node. Any objects one meets along the way are the objects which have that particular attribute. Looking at the lattice in Figure 1, if one takes the attribute

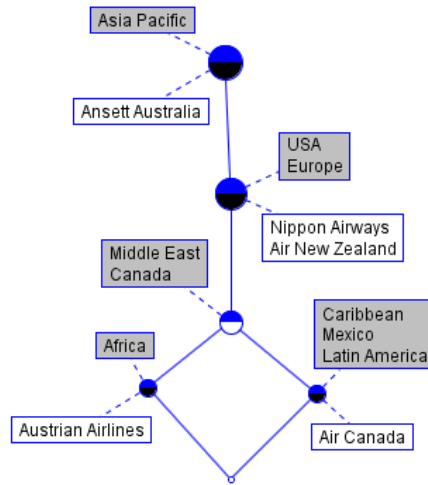


Fig. 1. A lattice corresponding to the Airlines context.

‘Asia Pacific’ and traces all paths which lead down from the node, one will collect all the objects. Thus Concept A can be interpreted as ‘All airlines fly to Asia Pacific’. Similarly, the node which is labeled with the formal object ‘Air New Zealand’ shall be referred to as *Concept B*. To retrieve the intension of Concept B (the attributes of ‘Air New Zealand’), one begins at the node where the object is labeled and traces all paths which lead up from the node. Any attributes one meets along the way, are the attributes of that particular object. Looking at the lattice once again, if one takes the object ‘Air New Zealand’ and traces all paths which lead up from the node, one will collect the attributes ‘USA’, ‘Europe’, and ‘Asia Pacific’. This can be interpreted as ‘The Air New Zealand airline flies to USA, Europe and Asia Pacific’. The concept that we formed previously by inspecting the cross-table, is the node in the center of the lattice; the one labeled with ‘Middle East’ and ‘Canada’. It becomes quite clear, for example, that although Air New Zealand and Nippon Airways also fly to Europe, USA and Asia Pacific, only Air Canada and Austrian Airlines fly to Canada and the Middle East as well.

Although the Airline context is a small example of FCA, visualising the formal context clearly shows that concept lattices provide richer information than from looking at the cross-table alone. This type of hierarchical intelligence that is gleaned from FCA is not so readily available from other forms of data analysis.

2.5 Representing Organised Crime with FCA

To represent Organised Crime (OC) with FCA it is necessary to consider what are suitable as the objects of study and what are attributes of those objects. For example, the objects could be instances of crime or types of crime and the attributes could be properties of these crimes. A formal context can be created from recorded instances of crime or from domain knowledge regarding the types of OC. Alternatively, for horizon scanning or situation assessment purposes, objects could be represented by activities or events that may be associated with OC. From appropriate data sources, formal contexts can be created using existing software tools and techniques [2, 4, 20]. Then, using the formalisms and tools available in FCA and its related disciplines, it will be possible to carry out analyses to detect and monitor OC:

- Finding multiple instances of an activity associated with OC based on Frequent Itemset Mining [10].
- Finding dependencies between OC attributes based on association rules [11, 12].
- Finding new OC indicators from existing data based on Machine Learning/Classification methods [6].
- Developing and using an OC 'threat score card', based on association rules (strength of association between an indicator and an OC).

The following sections illustrate these possibilities using simple examples.

3 Detecting OC Activities

Let us say that the purchasing of a certain type of fluorescent light tube is common in the cultivation of cannabis plants. An OC gang does not want to make its presence known by making large numbers of purchases from the same location/web site, so they make an effort to spread their purchases over several locations/sites. However, it may still be possible to detect this activity using Frequent Itemset Mining (FIMI) [9, 10]. This uses the notion of frequency of occurrence of a group of items (the so-called item-set). It is akin to FCA with objects being represented by instances. If we monitor the purchasing of tubes, FIMI can be used to automatically highlight possible clusters, thus alerting the possibility of OC. The itemset (attributes) need to be carefully considered and may be a combination of quantity of tubes purchased, location of purchase (stores or towns/areas for delivery from web sites) and time frames. Thus we may be automatically alerted of a number of purchases occurring in a particular time frame and in a particular geographical area.

Although the computation required to carry out the analysis is intensive, recent developments in high-performance concept mining tools [1, 13] mean that this type of monitoring could be carried out in real-time situation assessment. The outputs of the analysis would be suitable for visualising on a map (see figure 2) and end-users can be provided with the ability to alter parameters such as geographical area size and time frame, as well as being able to select different OC activities to analyse.

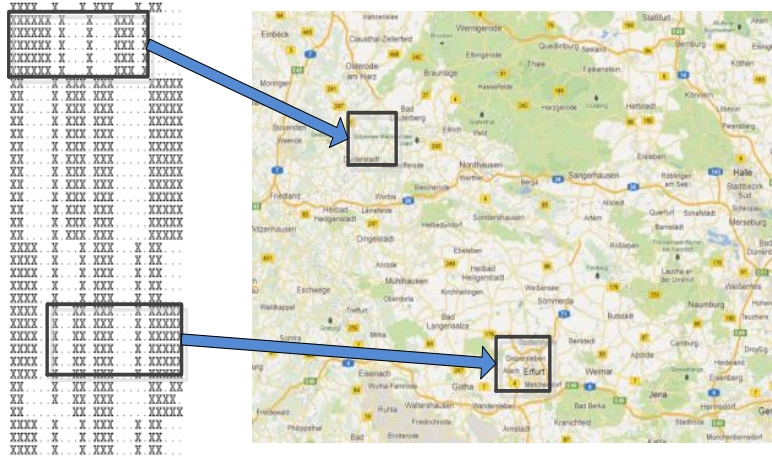


Fig. 2. Visualising OC Activities from Frequent Itemsets.

4 Finding OC Dependencies

By creating a formal context of Organised Crime using information such as that from the EU survey [16] it may be possible to reveal hidden dependencies between types of OC or between certain OC activities. Such dependencies are often called association rules and are inherent in FCA, being the ratio of the number of objects in concepts that have Galois connections. Using the simple airlines example above, one can say that if an airline flies to the USA (Air New Zealand, Nippon Airways, Austrian Airlines and Air Canada) then there is a 50% chance that it will also fly to Canada (Austrian Airlines and Air Canada). Similarly, if an airline flies to Africa then there is 100% chance it will also fly to the Middle East.

Now, if we take OC, we could investigate the association between drugs trafficking and the use of violence by OC gangs. Taking information from the same survey, FCA produces the lattice in figure 3. The numbers represent the number of OC gangs. Thus, perhaps surprisingly, in this sample of gangs at least, there is little difference in the use of violence by gangs who traffic drugs and those who do not. The profile of violence use is similar in both cases.

Using FCA tools such as ConExp [20] (which was used to produce the lattices in this paper), it is possible to investigate all associations between attributes in a formal context by calculating and listing the association rules. Typically, if we are carrying out an exploratory analysis of attributes, we will be interested in rules that show a strong association and that involve a (statistically) significant

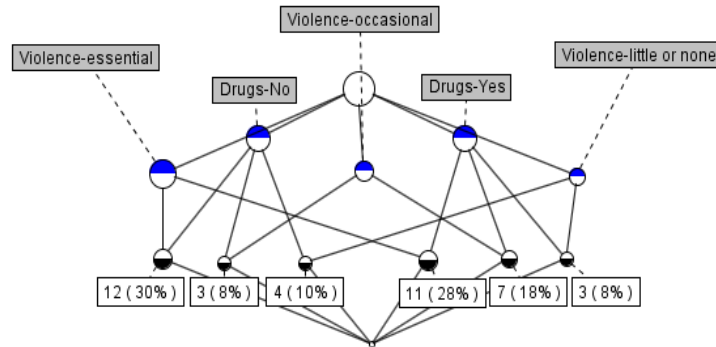


Fig. 3. A concept lattice showing the association between trafficking in drugs and the use of violence.

number of objects. The list below is a number of association rules generated by the information from the same OC gang (OCG) survey. The numbers in between angle brackets are the number of gangs involved and the percentages show the strength of the association. Thus, for example from rule 1, the use of violence is usually essential for gangs with multiple criminal activities; from rules 2 and 3, there appears to be a strong link between a low level of trans-border activity and lack of cooperation between OC gangs; from rules 6 and 10, local/regional political influence and extensive penetration into the legitimate economy usually imply that OCGs find the use of both violence and corruption essential.

1. <13>Activity-multiple = [85%] => <11>Violence-essential;
2. <12>OCG Cooperation-none = [83%] => <10>Trans-border Activity-1-2 counties;
3. <12>Trans-border Activity-1-2 counties = [83%] => <10>OCG Cooperation-none;
4. <10>Economy Penetration-none/limited = [90%] => <9>Political Influence-none;
5. <10>Political Influence-local/regional Economy Penetration-extensive = [90%] => <9>Violence-essential;
6. <11>Violence-essential Political Influence-local/regional = [82%] => <9>Economy Penetration-extensive;
7. <9>Structure-Rigid hierarchy Economy Penetration-extensive = [89%] => <8>Corruption-essential;
8. <9>Corruption-essential Political Influence-local/regional = [89%] => <8>Economy Penetration-extensive;
9. <9>Structure-Rigid hierarchy Corruption-essential = [89%] => <8>Economy Penetration-extensive;
10. <10>Political Influence-local/regional Economy Penetration-extensive = [80%] => <8>Corruption-essential;
11. <10>Violence-occasional = [80%] => <8>Activity-1 primary plus others;
12. <10>Activity-2-3 activities = [80%] => <8>Economy Penetration-extensive;
13. <10>Structure-Devolved hierarchy = [80%] => <8>Violence-essential;

5 Finding New Indicators for OC

The notions of dependency and association can be taken a step further by analysing the links between situations, events and activities and the occurrence or emergence of OC. There are many known indicators of OC [5, 17] but FCA may provide a means of discovering new, less obvious ones. The problem may be considered akin to a classification problem, either classification instances as OC or not OC, or by classifying instances as particular types of OC. Whilst there exist several well-known techniques of classification (such as those automated in the field of Machine Learning [6]), FCA has shown potential in this area [7] and, with the evolving of high-performance algorithms and software [1, 13] FCA may provide an approach that can be applied to large volumes of data in real time situation assessment. To illustrate the possibility an example is taken here using the well-known (in Machine Learning) data set of agaricus-lepiota mushroom, some of which are edible and some poisonous, with no obvious indicator for each. The data set contains a number a physical attributes of the mushrooms, such as stalk shape and cap colour and the issue is to find a reliable method of classifying the mushrooms as poisonous or edible. The concept lattices in figure 4 were produced from the data set and show some strong associations between various combinations of attributes and the classes edible and poisonous. The numbers are the number of mushrooms. The poisonous class also shows an interesting feature with zero mushrooms in the bottom concept - indicating that there appears to be two distinct and disjoint groups of poisonous mushroom, classified by two different sets of attributes. It is important to note that no single attribute (such as a foul odor) is a reliable indicator of a class. It is only in combination with other attributes (such as bulbous root and chocolate spore colour) that reliable sets of indicators are found.

With appropriate existing data (to be used as training data), a similar analysis should be possible for OC, to reveal possible sets of indicators for OC that can be used as part of a ‘horizon scanning’ system to detect or predict the emergence of OC.

6 Conclusion: Developing an OC ‘Threat Score Card’

Although the work presented here is mainly of a propositional nature, it shows potential for FCA to be applied in the domain of detecting and monitoring OC. The culmination of the FCA for OC may be in the creation and implementation of an OC ‘threat score card’. Using known and newly discovered indicators, association rules can be used to provide a weighting of the indicators. The resulting ‘score card’ can be implemented as part of an horizon scanning system for the detection of OC and OC types and situation assessment of the possible emergence of OC (OC types) if certain environmental conditions (indicators) pertain. Indeed, this is the proposed role of FCA in the new European ePOOLICE project [14] (grant agreement number: FP7-SEC-2012-312651), where it will play a part as one of several data analysis tools in a prototype pan-European OC monitoring and detection system.

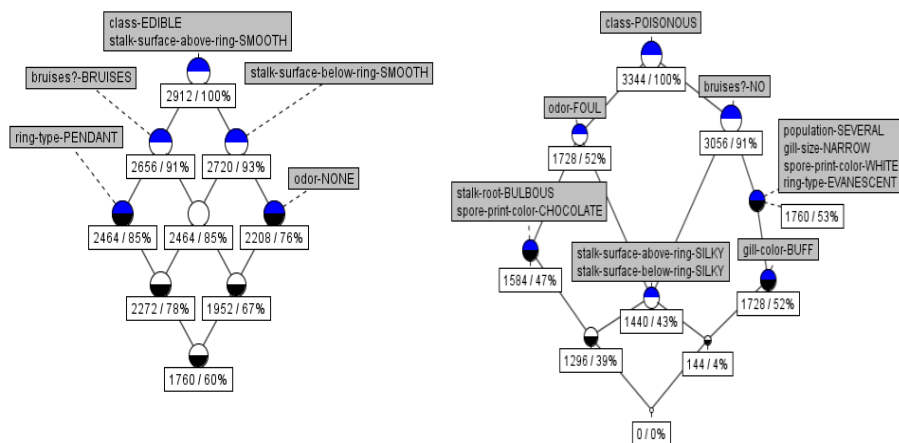


Fig. 4. Concept lattices showing indicators for edible and poisonous mushrooms

Acknowledgment The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n 312651.

References

1. S. Andrews. In-close2, a high performance formal concept miner. In S. Andrews, S. Polovina, R. Hill, and B. Akhgar, editors, *Conceptual Structures for Discovering Knowledge - Proceedings of the 19th International Conference on Conceptual Structures (ICCS)*, pages 50–62. Springer, 2011.
2. S. Andrews and C. Orphanides. Fcabedrock, a formal context creator. In M. Croitoru, S. Ferre, and D. Lukose, editors, *ICCS 2010*, volume 6208/2010 of *LNCS*. Springer, 2010.
3. S. Andrews and C. Orphanides. Knowledge discovery through creating formal contexts. pages 455–460. IEEE Computer Society, 2010.
4. P. Becker and J.H. Correia. *The ToscanaJ Suite for Implementing Conceptual Information Systems*, volume 3626 of *LNCS*, pages 324–348. Springer, 2005.
5. Europol. Eu organised crime threat assessment: Octa 2011. file no. 2530-274. Technical report, Europol, O2 Analysis & Knowledge, The Hague, 2011.
6. A. Frank and A. Asuncion. UCI machine learning repository: <http://archive.ics.uci.edu/ml>, 2010.
7. B. Ganter and S. O. Kuzntesov. Formalizing hypotheses with concepts. In *ICCS*, LNAI, pages 342–356. Springer-Verlag Berlin Heidelberg, 2000.
8. B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, 1998.
9. B. Goethals. Frequent itemset mining implementations repository. <http://fimi.ua.ac.be/>.
10. B. Goethals and M. Zaki. Advances in frequent itemset mining implementations: Report on fimi’03. *SIGKDD Explorations Newsletter*, 6(1):109–117, 2004.

11. S. Imberman and D. Domanski. Finding association rules from quantitative data using data booleanization, 1999.
12. S. O. Kuznetsov. Mathematical aspects of concept analysis. *Journal of Mathematical Science*, 18:1654–1698, 1996.
13. Jan Outrata and Vilem Vychodil. Fast algorithm for computing fixpoints of galois connections induced by object-attribute relational data. *Inf. Sci.*, 185(1):114–127, February 2012.
14. R. Pastor. epoolice: Early pusuit against organised crime using environmental scanning, the law and intelligence systems. <https://www.epoolice.eu/>, 2013.
15. U. Priss. Formal concept analysis in information science. *Annual Review of Information Science and Technology (ASIST)*, 40, 2008.
16. United Nations: Global programme against transnational organized crime. Results of a pilot survey of forty selected organized criminal groups in sixteen countries. Technical report, United Nations: Offcie on Drugs and Crime, 2002.
17. General Secretariat. Serious and organised crime threat assessment (socta) - methodology. Technical report, Council of the European Union, 2012.
18. CISC Strategic Criminal Analytical Services. Strategic early warning for criminal intelligence. Technical report, Criminal Intelligence Service Canada (CISC), Central Bureau, Ottawa, 2007.
19. P. Valtchev, R. Missaoui, and R. Godin. Formal concept analysis for knowledge discovery and data mining: The new challenges. In P. Eklund, editor, *Second International Conference on Formal Concept Analysis: Concept Lattices*, volume 2961 of *Lecture Notes in Computer Science*, pages 352–371. Springer Berlin Heidelberg, 2004.
20. S. A. Yevtushenko. System of data analysis ”concept explorer”. (in russian). In *Proceedings of the 7th national conference on Artificial Intelligence KII-2000*, pages 127–134, 2000.