A METHOD BASED ON TEMPORAL CONCEPT ANALYSIS FOR DETECTNG AND PROFILING HUMAN TRAFFICKING SUSPECTS

Jonas Poelmans¹, Paul Elzinga³, Stijn Viaene¹-², Guido Dedene¹-⁴

¹K.U.Leuven, Faculty of Business and Economics, Naamsestraat 69, 3000 Leuven, Belgium
²Vlerick Leuven Gent Management School, Vlamingenstraat 83, 3000 Leuven, Belgium
³Amsterdam-Amstelland Police, James Wattstraat 84, 1000 CG Amsterdam, The Netherlands
⁴Universiteit van Amsterdam Business School, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Paul.Elzinga@amsterdam.politie.nl

ABSTRACT

Human trafficking and forced prostitution are a serious problem for the Amsterdam-Amstelland police (the Netherlands). In this paper, we present a method based on Temporal Concept Analysis for detecting and profiling human trafficking suspects. Using traditional Formal Concept Analysis, we first build a lattice with early warning indicators to find persons who are potentially involved in human trafficking. These persons are then subjected to an in-depth investigation and we use Temporal Concept Analysis for constructing visual profiles of these suspects. Finally, the evolution of the social network surrounding these persons is visually analyzed using Temporal Concept Analysis.

1. Introduction

Human trafficking is defined as the recruitment, transportation, harboring and receipt of people for the purpose of slavery, forced labor and servitude [1]. Trafficking is considered as a serious problem by the police and government of the Netherlands [2]. Unfortunately, research has shown that only a small number of victims of human trafficking makes a statement to the police [3]. Recently, the Amsterdam-Amstelland police installed a human trafficking team to actively search for potential suspects.

Police databases contain a large number of reports describing observations made by officers during motor vehicle inspections, interventions, id controls, etc. These documents are often spread over different database systems, may have different identification keys and only limited information browsing functionality is provided. To analyze a potential suspect, officers currently have to search all these databases and manually inspect all the reports, because some of these observations might contain indications for human trafficking.

In this paper, we make use of the techniques known as Formal Concept Analysis (FCA) [4-6] and Temporal Concept Analysis (TCA) [7-12]. TCA was only recently introduced as an extension to FCA and can be used to analyze temporal data. First, we use FCA to identify a reliable set of early warning indicators and to extract potential suspects for in-depth analysis from the large amount of unstructured reports. Second, we showcase how TCA is used to automatically compose a comprehensible and visually appealing “profile” of these suspects that can be used by police officers to quickly decide whether or not this person should be monitored. Finally, we analyze the social network of the suspect and its evolution over time using TCA.

The remainder of this paper is composed as follows. In section 2, we give some background on human trafficking and the current way of working by the police. In section 3, we elaborate on the dataset used. In section 4, we discuss the human trafficking indicators. In section 5, the essentials of TCA theory are introduced. In section 6, the detection and profiling of human traffickers is explained. Finally, section 7 concludes the paper.

2. Background

2.1. Human trafficking

Human trafficking is the fastest growing criminal industry in the world, with the total annual revenue for trafficking in persons estimated to be between $5 billion and $9 billion [13]. The council of Europe states that “people trafficking has reached epidemic proportions over the past decade, with a global annual market of about $42.5 billion” [14].

Trafficking victims are typically recruited using coercion, deception, abuse of power or outright abduction. Threats, violence and economic leverage such as debt bondage often make a victim consent to exploitation. Exploitation includes forcing people and children into prostitution, forced labor or slavery.

Women are particularly at risk from sex trafficking. Criminals exploit lack of opportunities, promise good jobs or opportunities for study and then force the victims to become prostitutes.
Through agents and brokers who arrange the travel and job placements, women are escorted to their destinations and delivered to the employers. Upon reaching their destinations some women learn that they have been deceived about the nature of the work they will do; most have been lied to about the financial arrangements and conditions of their employment; and find themselves in coercive or abusive situations from which escape is both difficult and dangerous [15].

Due to the illegal nature of trafficking and differences in research methodology, the exact extent is unknown. Rough estimates suggest that 700,000 to 2 million women and girls are trafficked across international borders every year [16]. The majority of transnational victims are trafficked into commercial sexual exploitation. Since the fall of the Iron Curtain, the impoverished former Eastern bloc countries such as Albania, Moldova, Romania, Bulgaria, Russia, Belarus and Ukraine have been identified as major trafficking source countries for women and children. It is estimated that 2/3 of women trafficked for prostitution worldwide annually come from Eastern Europe, three quarters have never worked as prostitutes before [17].

For Amsterdam, official statistics say that the majority of victims have either the Dutch nationality, come from Eastern European countries or from West Africa [18].

2.2. Current situation
Victims of human trafficking rarely make an official statement to the police. The human trafficking team is installed to proactively search police databases for any signals of human trafficking. Unfortunately, this turns out to be a laborious task. The investigators have to manually read and analyze the police reports, one by one, because only an estimated 10% of the information containing human trafficking indications has been labeled as such by police officers.

As soon as the investigators find sufficient indications against a person, a document based on section 273f of the code of criminal law is composed for the person under scrutiny. Based on this report, a request is sent to the Public Prosecutor to start an in-depth investigation against the potential suspects. After permission is received from the Public Prosecutor, the use of special investigation techniques such as phone taps and observation team is allowed.

3. Dataset
Our dataset consists of 69788 general police reports from the year 2008. These general reports contain observations made by police officers during motor vehicle inspections, during a police patrol, when a known person was seen at a certain place, etc. An example of a report is displayed in Figure 4. We also have 1101 reports at our disposal, for which we know with certainty that they are related to human trafficking. These include arrestment reports, a statement made by a victim to the police, charges against a human trafficking suspect, etc. These reports date back to the years 2005-2008.

<table>
<thead>
<tr>
<th>Reporting date</th>
<th>Project code</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>23-03-2008</td>
<td>Prostitution related observation</td>
<td>Amsterdam Wallengebied yyyy</td>
</tr>
<tr>
<td>Suspect (male)</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>(18-45yr)</td>
<td>Amsterdam Wallengebied yyyy</td>
<td></td>
</tr>
<tr>
<td>Involved (female)</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>(18-45yr)</td>
<td>Amsterdam Wallengebied yyyy</td>
<td></td>
</tr>
</tbody>
</table>

4. Human trafficking indicators
The first step of the research consists of finding a set of indicators on the basis of which we can decide whether or not a person might be involved in human trafficking. We make a distinction between early and late indicators. Many of the strong human trafficking indicators available in some of the reports were only obtained after special investigation techniques such as phone taps or the use of observation teams. One of the main goals of the research consists of being able to automatically find suspects that are potentially related to human trafficking from the large set of observational reports. We thus aim at finding cheap, fast and reliable early warning indicators that give a good indication that a person might be involved in human trafficking. Table 1 shows some of the indicators we obtained after analyzing the data at hand. To detect these indicators in the unstructured text of police reports, we use a thesaurus. For each indicator, we have some terms in the thesaurus that indicate whether or not the indicator is present in the police report. To construct a thesaurus, find relevant terms and refine the set of indicators, we use the methodology presented in [21].

This method is an iterative knowledge discovery from unstructured text approach based on FCA and Emergent Self Organizing Map. During the successive iterations, the data is explored, interesting new concepts are distilled from the texts and conceptually analyzed in a lattice. The
focus lies on finding those concepts that help in identifying human trafficking suspects.

Table 1. Excerpt of the human trafficking indicators

<table>
<thead>
<tr>
<th>Static indicators</th>
<th>Indicators with time dimension</th>
<th>Indicators coming from social network</th>
</tr>
</thead>
<tbody>
<tr>
<td>- nationality</td>
<td>- red light district: Wallen</td>
<td>- seen with known suspect</td>
</tr>
<tr>
<td>- violence</td>
<td>- red light district: Ruysdaelkade</td>
<td></td>
</tr>
<tr>
<td>- minors involved</td>
<td>- regularly visiting suspicious club</td>
<td></td>
</tr>
<tr>
<td>- restriction of personal freedom</td>
<td>- regularly dropping of girls at club</td>
<td></td>
</tr>
<tr>
<td>- id-problems</td>
<td>- expensive car</td>
<td></td>
</tr>
<tr>
<td>- carrying large amount of money</td>
<td>- woman in car</td>
<td></td>
</tr>
<tr>
<td>- forcing person to work in bad condition</td>
<td>- vacation</td>
<td></td>
</tr>
<tr>
<td>- prostitute involved</td>
<td>- car trade</td>
<td></td>
</tr>
<tr>
<td>- dependency relationship</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- injury observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- woman not speaking</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We distinguish between three types of indicators. Static indicators typically contain socio-demographical information, such as the nationality of the person involved. We also have indicators with an inherent time dimension, such as for example when a person was signaled several times in a prostitution area and always declared to the police that he was on vacation. Finally, we also have indicators that come from analyzing the social network of the suspect, e.g. if he is seen together with a known criminal.

5. Temporal Concept Analysis

Temporal Concept Analysis (TCA) is a mathematical theory that was introduced in scientific literature about nine years ago [7-9]. TCA is based on Formal Concept Analysis (FCA) [4,5] and addresses the problem of conceptually representing time. TCA is particularly suited for the visual representation of discrete temporal phenomena. In the following sections, we first introduce the essentials of FCA theory. Then, we discuss the extension of FCA with a time dimension, i.e. TCA.

5.1. FCA essentials

FCA concept lattices are used to describe the conceptual structures inherent in data tables without loss of information by means of line diagrams yielding valuable visualizations of real data [20]. In a previous paper, we analysed the concept of domestic violence using FCA [21]. The main difference with domestic violence is that there is a time dimension involved in human trafficking. Suspects are often spotted several times by the police and it is important to incorporate this time dimension in the visualisation of the data. FCA can be used as an unsupervised clustering technique [19,22] and police reports containing terms from the same term clusters are grouped in concepts.

The starting point of the analysis is a database table consisting of rows $M$ (i.e. objects), columns $F$ (i.e. attributes) and crosses $T \subseteq M \times F$ (i.e. relationships between objects and attributes). The mathematical structure used to represent such a cross table is called a formal context $(T, M, F)$. An example of a cross table is displayed in Table 2. In this table, reports of police observations (i.e. the objects) are related (i.e. the crosses) to a number of terms (i.e. the attributes); here a report is related to a term if the report contains this term. The dataset in Table 2 is an excerpt of the one we used in our research. Given a formal context, FCA then derives all concepts from this context and orders them according to a subconcept-superconcept relation, which results in a line diagram (a.k.a. lattice).

The notion of concept is central to FCA. A concept is considered to be a unit of thought constituted of two parts: its extension and its intension, [4,23]. The extension consists of all objects belonging to the concept, while the intension comprises all attributes shared by those objects.

For a set of objects $O \subseteq M$, the common features, written $\sigma(O)$, can be identified via the following formula:

$$A = \sigma(O) = \{ f \in F \mid \forall o \in O : (o, f) \in T \}$$

By collecting all reports of this context that share these attributes, we get to a set $O \subseteq M$ consisting of reports. This set $O$ of objects is closely connected to set $A$ consisting of attributes.

$$O = \tau(A) = \{ i \in M \mid \forall f \in A : (i, f) \in T \}$$

Table 2. Example of a formal context

<table>
<thead>
<tr>
<th></th>
<th>Prostitution area</th>
<th>Expensive car</th>
<th>Bulgarian</th>
<th>ID problems</th>
<th>Large money amount</th>
<th>Women in car</th>
</tr>
</thead>
<tbody>
<tr>
<td>report 1</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>report 2</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>report 3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>report 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>report 5</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
That is, $O$ is the set of all objects sharing all attributes of $A$, and $A$ is the set of all attributes that are valid descriptions for all the objects contained in $O$. Each such pair $(O, A)$ is called a formal concept (or concept) of the given context. The set $A = \sigma(O)$ is called the intent, while $O = \tau(A)$ is called the extent of the concept $(O, A)$.

There is a natural hierarchical ordering relation between the concepts of a given context that is called the subconcept-superconcept relation:

$$(O, A) \subseteq (O', A') \iff (O \supseteq O' \land A' \subseteq A).$$

A concept $d = (O, A)$ is called a subconcept of a concept $e = (O, A)$ (or equivalently, $e$ is called a superconcept of a concept $d$) if the extent of $d$ is a subset of the extent of $e$ (or equivalently, if the intent of $d$ is a superset of the intent of $e$).

The set of all concepts of a formal context combined with the subconcept-superconcept relation defined for these concepts gives rise to the mathematical structure of a complete lattice, called the concept lattice of the context, which is made accessible to human reasoning by using the representation of a (labelled) line diagram. The circles or nodes in this line diagram represent the formal concepts. The shaded boxes (upward) linked to a node represent the attributes used to name the concept. The non-shaded boxes (downward) linked to a node represent the objects used to name the concept. The information contained in the formal context can be distilled from the line diagram in by applying the following reading rule: an object “$g$” is described by an attribute “$m$” if and only if there is an ascending path from the node named by “$g$” to the node named by “$m$”.

Retrieving the extension of a formal concept from a line diagram such as the one in Figure 2 implies collecting all objects on all paths leading down from the corresponding node. To retrieve the intent of a formal concept, one traces all paths leading up from the corresponding node in order to collect all attributes. The top and bottom concepts in the lattice are special: the top concept contains all objects on all paths leading down from the corresponding node. To retrieve the intent of a formal concept, one traces all paths leading up from the corresponding node in order to collect all attributes. The top and bottom concepts in the lattice are special: the top concept contains all objects on all paths leading down from the corresponding node, whereas the bottom concept contains all attributes in its intent. A concept is a subconcept of all concepts that can be reached by travelling upward. This concept will inherit all attributes associated with these superconcepts.

### 5.2. TCA essentials

The pivotal notion of TCA theory [10,11] is that of a conceptual time system [12]. An example of a data table of a conceptual time system is displayed in Table 3.

**Table 3.** Data table of a conceptual time system

<table>
<thead>
<tr>
<th>Time part</th>
<th>Event part</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time granule</td>
<td>Date</td>
</tr>
<tr>
<td>0</td>
<td>26/01/2008</td>
</tr>
<tr>
<td>1</td>
<td>21/02/2008</td>
</tr>
<tr>
<td>2</td>
<td>15/02/2008</td>
</tr>
<tr>
<td>3</td>
<td>13/03/2008</td>
</tr>
<tr>
<td>4</td>
<td>27/04/2008</td>
</tr>
<tr>
<td>5</td>
<td>01/05/2008</td>
</tr>
<tr>
<td>6</td>
<td>14/06/2008</td>
</tr>
<tr>
<td>7</td>
<td>18/07/2008</td>
</tr>
</tbody>
</table>

Table 3 contains the observations of one real person at several points of time. To make a single observation, police officers needed some time, varying from a few minutes to a few hours. We abstract from the duration of an observation and use the notion of a point of time, also called time granule. We thus start from a set of which the elements are time granules. In Table 3 for example, we have 8 time granules. For describing the observations, we use a single valued context with $G$ as its set of formal objects. This context consists of an event part and a time part. The indicators observed at each of these time granules are described in the event part of the data table. In contrast to [12], where a multi-valued context was used, we only need a single-valued context here. Formally, the conceptual time system we use can be described as follows.

Let $T := (G, M, I_T)$ and $C := (G, E, I_C)$ be two single valued contexts respectively on the same object set $G$. Then the pair $(T, C)$ is called a conceptual time system on the set $G$ of time granules. $T$ is called the time part and $C$ the event part or space part of $(T, C)$. The combination of $T$ and $C$ is denoted by $K_{TC} := TIC$. It is the context of the conceptual time system $(T, C)$. The object concepts of $K_{TC}$ are called situations, the object concepts of $C$ are called states and the object concepts of $T$ are called time states. The sets of situations, states and time states are called the situation space, the state space and the time state space of $(T, C)$ respectively. In the visualization of the data, we want to express the “natural temporal ordering” of the observations. In the TCA lattice, a time relation $R$ is introduced on the set $G$ of time granules of a conceptual time system. We speak of a conceptual time system with a time relation (CTST).

Let $(T, C)$ be a conceptual time system on $G$ and $R \subseteq G \times G$ . Then the triple $(T, C, R)$ is called a conceptual time system (on $G$) with a time relation. On the set $G :=\{0,1,2,3,4,5\}$ of time granules we introduce the relation $R :=\{(0,1), (1,2), (2,3), (3,4), (4,5)\}$ shortly described as $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$. We also need the notions of transitions and life tracks. The basic idea of transition is a “step from one point to another”. Each arrow in Figure 3 represents a “transition of the suspect” and corresponds to an element of $R$. The transitions in Figure 3 form a life track of the suspect.

### 6. Detecting and profiling human traffickers

The method we propose consists of 3 steps. First, we aim at extracting all persons that are potentially involved in human trafficking from the large set of observations. For this extraction process we make use of the visualization capabilities offered by FCA. Second, we construct a detailed profile of these persons using TCA. TCA allows
for comprehensible visualization of the indicators found and the evolution over time of these persons. Third, we use TCA to analyze the social network and its evolution over time around these persons.

6.1. Detecting suspects with FCA
For detecting potential suspects from the large amount of observations, we make use of an FCA lattice. The persons mentioned in the reports are the objects of the lattice. The indicators contained in the observational report of these persons are combined in one feature vector. For the static indicators and indicators coming from the social network analysis, we make use of a single-valued formal context. For the indicators with time dimension we make use of a many-valued formal context. This many-valued context is then scaled to a single valued context. This results in an FCA lattice as displayed in Figure 2.

From domain experts, we know that logistics is an important aspect of human trafficking and forced prostitution. Girls have to be transported from the place where they live (often under surveillance of their pimp) to the place where they work. Whereas it is difficult for the police to detect human trafficking and to infiltrate the underground world of forced prostitution, there are a lot of observations performed on the street amongst others during motor vehicle inspections that might give some clues on where to look. The lattice in Figure 2 for example contains some of these early warning indicators for forced prostitution and zooms in on some of the logistic aspects of human trafficking.

One target group we selected for in-depth analysis were the 11 persons who were seen in an expensive car with Bulgarian licence plate in a prostitution area. From domain experts, we know that Bulgarian human traffickers are very active in Amsterdam. We also selected the 10 persons who were involved with prostitutes and had problems with legitimating themselves or the women. We also retrieved the 4 Bulgarian persons who were seen with a large amount of money in the red light district. Finally, we investigated the 26 Bulgarian persons who were involved with prostitutes and committed violence. In the tool we developed, it is possible to click on the persons in which we are interested and automatically a TCA profile (see section 6.2) is constructed.

6.2. Profiling suspects with TCA
For composing a detailed profile of suspicious persons, we make use of TCA. The goal of these profiles is to collect all relevant information around a person and provide a conceptual overview of this person and his evolution over time. The TCA lattice serves as an intuitive knowledge browser making the interaction between the police officer and data more efficient. Based on this lattice, police officers can easily judge whether or not there is sufficient evidence available for the Public Prosecutor to start closely monitoring this person.

In Figure 3, a lattice profile of one of the potential suspects selected using the FCA lattice in Figure 2 is shown. The arrow in Figure 3 shows the life track connecting the successive observations of the person. One can see that the person with Bulgarian nationality was observed 8 times in Amsterdam by the police. During these 8 times, he declared 5 times that he was on vacation in Amsterdam. He was spotted 3 times in a prostitution area, 4 times in an expensive car and 3 times with women in his car. He once had problems with legitimating himself or the girls he was transporting and he once carried a large amount of money with him. These are all clear indications that this person is probably involved in human trafficking. Based on analysis of this lattice, it was decided to start an investigation into this person.

Figure 4 contains a TCA lattice profile of another human trafficking suspect discovered using the lattice in Figure 2. The lattice in Figure 4 shows that the reports related to this person contain some very strong indicators for forced prostitution. The Bulgarian person has been spotted several times during which he was involved with prostitutes and companies that have a dodgy reputation when it comes to forced prostitution. Moreover, there has been one incident in which the person restricted the personal freedom of a woman who was dependent of him. He also used violence against her. Again, these are clear indicators for human trafficking and an investigation was started against this person after analyzing this profile.

6.3. Network evolution analysis using TCA
For each suspect, there are multiple observations available. In each of these observations, one or more other persons may be mentioned. For each of these persons, we collected all data available and incorporated this into the lattice of the suspect. The mathematical foundation we used for this analysis is that of a “conceptual time system with actual objects and a time relation” (CTSOT) [12].

This time line helps us gain insight in the role of the person in the overall network, the evolution over time of his cooperation with other persons, the importance of his position, the role and evolution of the persons surrounding him etc. For example, consider the lattice in Figure 5. This is a simplified version of a lattice based on real life data. It contains the life tracks of 3 persons. Consider the life track of “Person 1”. At time granule 1, he was observed by the police in an expensive car and with two women at the back of his car. These women were probably prostitutes according to the officers. They seemed to be scared of the officers; they didn’t say much and didn’t carry their ID with them. Two months later, one of these women “Person 2” was observed in a prostitution area with a Bulgarian male “Person 3” and with bruises on her arm.
Based on these observations and the visualization of these relationships, we were able to gain insight in the structure of this network. “Person 1” was probably into the logistics part of the network. “Person 2” was probably a woman who was forced into prostitution and “Person 3” is probably her pimp. Based on these observations, police officers had enough indications to start an official investigation against “Person 1” and “Person 3”, whereas the lattice containing observations of “Person 1” alone did not suffice to start a case against “Person 1”.

7. Conclusions

In this paper, we have investigated the practical usefulness of Temporal Concept Analysis for analyzing police data containing an inherent time dimension. We first used a regular FCA lattice containing early warning indicators to extract potential human trafficking suspects from observational police reports. After finding potential suspects using an FCA lattice, TCA turned out to provide interesting visualization capabilities for composing a detailed profile of these persons and their social network.

For each point in time the person was spotted by the police, this TCA lattice contains the human trafficking indications that were observed. These observations are chronologically ordered in the lattice using the TCA notion of life track of a suspect. The visualization of the social network of suspects and its evolution over time using TCA also resulted in new insights during police investigations. In particular, this social network lattice provided insights in the different roles of the observed persons in the global network.

Future work includes the embedding into daily policing practice of the techniques presented in this paper.

8. Acknowledgements

The authors would like to thank the police of the region Amsterdam-Amstelland and in particular Deputy Chief Reinder Doeleman and Chief Hans Schönfeld for supporting this research. The authors are grateful to the Police Organisation Amsterdam-Amstelland, for providing us with the data. Jonas Poelmans is aspirant of the “Fonds voor Wetenschappelijk Onderzoek Vlaanderen” or “Research Foundation Flanders”.

Fig. 2. FCA lattice used for detecting potential human trafficking suspects
**Fig. 3.** TCA lattice used for profiling a human trafficking suspect

**Fig. 4.** TCA lattice profile of a human trafficking suspect
Fig. 5. TCA lattice used for gaining insight in social network of human trafficking suspects

9. References


