Organised crime and social media: detecting and corroborating weak signals of human trafficking online

ANDREWS, Simon <http://orcid.org/0000-0003-2094-7456>, BREWSTER, Benjamin <http://orcid.org/0000-0003-3536-5862> and DAY, Tony <http://orcid.org/0000-0002-3214-6667>

Organised Crime and Social Media; detecting and corroborating weak signals of Human Trafficking online

Simon Andrews[1,2], Ben Brewster[2], and Tony Day[2]

[1] Conceptual Structures Research Group
Department of Computing
[2] Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research
Communication and Computing Research Centre
Sheffield Hallam University, Sheffield, UK
{s.andrews, b.brewster, t.day}@shu,ac,uk

Abstract. This paper describes an approach for detecting the presence or emergence of Organised Crime (OC) signals on Social Media. It shows how words and phrases, used by members of the public in Social Media, can be treated as weak signals of OC, enabling information to be classified according to a taxonomy of OC. Formal Concept Analysis is used to group information sources, according to Crime and Location, thus providing a means of corroboration and creating OC Concepts that can be used to alert police analysts to the possible presence of OC. The analyst is able to 'drill down' into an OC Concept of interest, discovering additional information that may be pertinent to the crime. The paper describes the implementation of this approach into a fully-functional prototype software system, incorporating a Social Media Scanning System and a map-based user interface. The approach and system are illustrated using the Trafficking of Human Beings as an example. Real data is used to obtain results that show that weak signals of OC have been detected and corroborated, thus alerting to the possible presence of OC.

1 Introduction

The vociferous proliferation of the Internet, and more recently Social Media, into society and the everyday lives of its citizens has, over the last fifteen or so years, resulted in a sea-change in the behaviours and perceptions we have in relation to the information that is shared freely online [14]. Such behaviour has resulted in the creation of a vast repository of information that holds potential value for police investigations, and the emergence of the open-source researcher as a valuable skill-set within the analytical repertoire of the police and security agencies. Resources such as social media, RSS news feeds, interactive street-maps and online directory services all provide valuable stores of information that can be used to support existing investigative and analytical practices in response to serious and organised crime. This paper focuses on the identification, extraction and corroboration of data from social media using automated data acquisition, natural language processing and formal concept analysis (FCA), specifically in order to identify what we will refer to as 'weak signals' of human trafficking, and to transform these signals into corroborated alerts linked to the presence or emergence of human trafficking activity.

The concept of weak signals is abstracted from the Canadian Criminal Intelligence Service's (CISC) definitions of primary and secondary indicators [16], and the perception that in reality there is little tangible value to be extracted from isolated indicators as there is potential for them to be indicative of a variety of phenomena. However, when these indicators are grouped under certain conditions, such as proximity to a certain location and type of activity, they can begin to provide insights into to the presence or emergence of crime. It is with this definition, and the notion of 'weak signals' that we use as the basis of this paper and the approached presented within it.

Perhaps the greatest shift in the use of the internet over the last 10-15 years or so is the relative phenomenon which is the usage of Social Media among normal citizens. In the aftermath of the events which followed the killing of Mark Duggan in 2011, a HMIC commissioned review highlighted significant inefficiencies in the the way that authorities were equipped to deal with social media as an intelligence source [11]. Social media intelligence, or 'SOCMINT', provides opportunities for providing insights into events and groups, enhancing situational awareness, and enabling the identification of criminal intent [12].

In order to demonstrate the potentially utility of SOCMINT, in respect of identifying the presence and/or emergence of organised crime, the problem domain of Human Trafficking will be used as the exemplar throughout this paper. Human trafficking operates on a vast scale, impacting upon almost every country in the world as an origin, transit or destination location for the movement and exploitation of human beings. Trafficking is so defined by article 3 of the Palermo protocol as the 'recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation' [17].

Europol [5] has in the past acknowledged the growing criminal dependence on the internet and the increasingly trans-European perspective of serious and organised crime. These changes in the way that information is created and shared, combined with the diversity in the way that existing forms of criminality are being conducted provides the opportunity, and desire, for the development of new means to assist law enforcement in combating such crime. To provide one such approach to enable this, the tools described here facilitate the identification, extraction, processing, analysis and presentation of data from open sources, such as social media, that can provide insight into the emergence and presence of crime. While at one end of the scale international intelligence agencies such as the CIA's PRISM programme are facillitating the aquisitoin, fusion and analysis of vast amounts of data from disperate sources, the (known) resources and capability of law enforcement agencies (both locally, regionally and even internationally) are much more modest with the use of data from open sources and social media often a manual task, and the remit of just a few specialist analysts and officers within each force.

2 Taxonomy of Organised Crime

In beginning to model its constituent elements it is necessary to ascertain a thorough understanding of the actual problem domain - human trafficking, by drawing upon established definitions used to describe and diagnose the problem by the practitioner base. The UNODC [17] have defined, using the UN Palermo protocol as the basis, what they refer to as, the three constituent 'elements' of trafficking, these being the 'act', 'means' and 'purpose', see Figure 1. Firstly, the 'Act' refers to what is being done, this can include context such as whether and how the victim has been recruited, transported, transferred or harboured. The question of how this is being achieved is answered by the 'Means' which seeks to establish whether force is being used as the basis of manipulation, such as through kidnapping, abduction or the exploitation of vulnerabilities, or more subtle methods such as through fraud, imposing financial dependencies or coercion. The final element, the 'Purpose' establishes the reason why the act and means are taking place, or to put it simply - the form of exploitation behind the act and means, be it for forced labour, sexual exploitation and prostitution, organ harvesting or domestic servitude.

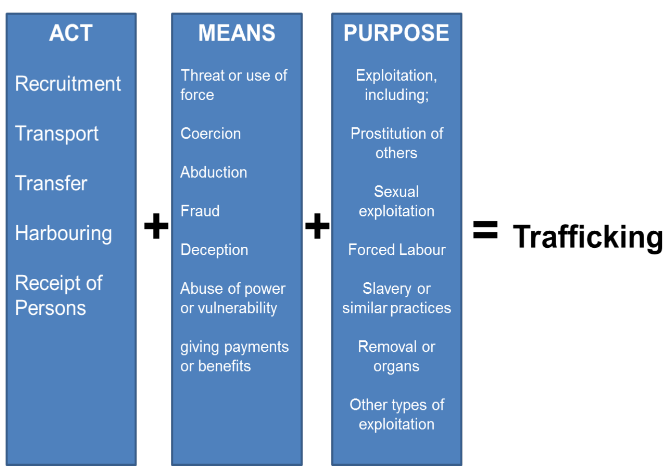

Fig. 1. The Elements of Human Trafficking

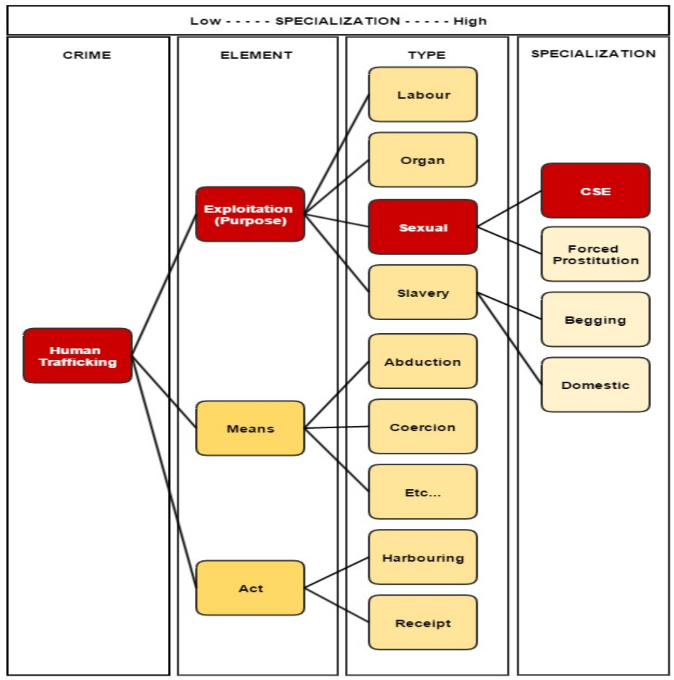

Fig. 2. Organised Crime Taxonomy (excerpt)

This definition and categorisation provides an ideal underpinning for the formation of a taxonomy of Human Trafficking that can be used to form the basis of an approach to automatically identify and extract valuable data from open sources (Figure 2). This taxonomy in actuality forms part of a larger model, consisting of elements of a broader range of organised crime threats, including the cultivation and distribution of illegal narcotics. The elements of the taxonomy are defined across for types, visualised here using vertical lanes. The first of these four lanes starting on the left is used as a high level categorisation used to separate between different crime types. The second lane deals with different elements within a specific type of crime - in this case one of the three component parts of trafficking, while subsequent lanes, type and element, are used to show further more specialised aspects of these elements.

Each of the nodes contained within the taxonomy represents a ruleset designed to determine content's, in this case a particular twitter posting, relevance to the subject matter - Human Trafficking. The level of specialisation of the rules themselves follows the structure of the taxonomy, moving from more generic words or phrases that may indicate Human Trafficking used at the higher levels of the taxonomy, where as as the more specialised end of the taxonomy, more nuanced rules that may allude to the presence of criminality are used.

3 Weak Signals of Organised Crime

To enable the development of a taxonomy to begin to model and structure the information deemed useful to extract we can refer to a wealth of literature from both academic and practitioner perspectives that provide insights into the factors that contribute to and indicate organised crime. These indicators vary from high level, secondary information such as Political, Economic, Socio-cultural, Technological, Legal and Environmental (PESTLE) factors, right down to operationally oriented information that helps us to identify potential victims of trafficking. Existing models to anticipate changes and developments in organised criminality accross geographic areas have focused on this kind of data alongside existing crime statistics [20].

In the past, and to some extent a problem that still exists, a lack of information and common understanding about what human trafficking is has hindered the impact and effectiveness of efforts to combat it [19]. Despite varied and wide-ranging counter-trafficking initiatives from NGO's, Law Enforcement and Governments, reliable information regarding the magnitude and nature of trafficking across regional and national borders is still hard to come by due to a number of issues around the sharing, fusion and understanding of data that is already being collected [9]. The purpose of the approach developed and described in this paper is not to provide a statistically accurate representation of the presence and emergence of trafficking but rather to increase the access and usability of data from previously untapped open-sources. In previous work, we have discussed indicators across a three-level model [2] moving from credible and accepted indicators of trafficking at level 1 of the model, through to the observations and content created online, including on social media, by citizens regarding these 'weak signals'.

In this paper we discuss the latter and more specifically the modelling and use of this information as 'weak signals' that allude to the presence and/or emergence of criminality in citizen generated content, whilst using the formal definitions and doctrine that exists to underpin the framework and organisation of the model itself. Perhaps the most comprehensive list of indicators comes from the UNODC [18] provide an extensive list of indicators categorised by different types of exploitation such as; domestic servitude, child, sexual exploitation, labour exploitation and begging/petty crime - the labels used in the taxonomy structure outlined in Fig. 2. Using sexual exploitation as an example indicators include things such as the appearance that persons are under the specific control of another, that the person(s) appear to own little clothing, appear to rely on their employer for basic amenities, transport and accommodation and more. Although in this form, these indicators are quite abstract and it can may be potentially difficult to see how they may manifest in real, open-source, data - it is possible to develop rules looking for keywords and phrases that can provide 'weak-signals' of their existence online.

In order to facilitate the identification and extraction of these weak signals in social media and other open-sources, what we will refer to as 'contextual extraction' methods [3] are used in order to identify, and subsequently extract, key entities and facts (i.e. previously unknown relationships between different entities) from the data. This approach to information extraction using natural-language processing builds upon the existing principles of template ased information extraction [15] , also sometimes referred to as 'Atomic Fact Extracton' [3] These 'facts' enable the extraction of entities within a specific context, i.e. locations in relation to an arrest or type of exploitation on a per sentence basis.

While in isolation, the extraction of these entities on their own does not necessarily provide much actionable information, it is possible using rules that attempt to infer relations between them to begin to make some assumptions about the data and its content. For example a single tweet may contain multiple locations and other entities, but without some means to establish a relationship between the two there is no way to automatically infer, with any confidence at least, that they are linked. Fortunately, through the use of contextual extraction, and the aforementioned 'facts' we can infer these relationships in a number of ways, using prepositions, parts-of-speech tagging and Boolean operators that specify distance between words and other parameters. As the examples in discussed in this paper refer to data from Twitter only, these relationships are done on a 'per sentance' basis. The following example shows how this works in practice: From these rules we can begin to make some assumptions about

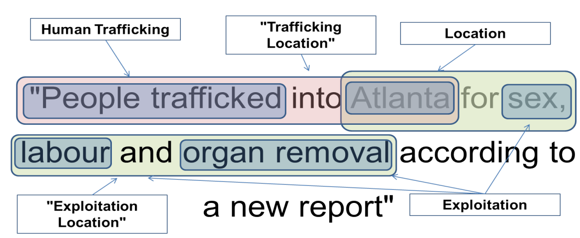

Fig. 3. Entity and Fact Extraction in Social Media

the entities being extracted. For example, it is now possible to ascertain with a degree of confidence that specific locations are in reference to a specific event. At this point, it is important to acknowledge the challenges posed by the use of SMS-language (textese) as communication via services such as twitter do not necessarily adhere to strict grammer or syntax conventions. Although a number of novel approaches to handle this type of language are in development (see, for example [13]) due to the use of examples that use accepted, formal teminology, we do not address this issue here.

4 Categorisation and Filtering

In addition to extracting facts and entities from the input data, similarly techniques are also utilised as the basis of a rule based approach to classify content against a number of pre-defined categories. Membership to one or more of these categories is then used as the basis for content filtering. If the content analysed by the crawler does not meet the criteria of at least one of the rules within the parameters defined in the content categorisation process, it is then disregarded. The categorisation model is defined using a similar approach to the entity and fact extraction model, utilising a number of 'hand-crafted' rules organised in a hierarchical structure, with the only key difference being that rather than being designed to identify and extract specific pieces of data and/or information they aim to discern the relevance of the content against the defined topics using the same taxonomy structure defined in Fig 2. The rules themselves use a range of techniques, again focusing on the identification of keywords and phrases.A number of examples of the phrases and keywords used as part of the categorisation taxonomy are shown in table 1.

Weak Signal	Keywords and Phrases
Physical Injury	subjected to violence timid forced to have sex women beaten
Physical Appearance	provocative dress live with a group of women unhappy
Unable to leave place of residence	afraid to leave under control of others financially dependant
Irregular movement of individuals	men come and go at all hours women do not appear to leave lots of activity at night

Table 1. Weak Signals - Keywords and Phrases

5 A Social Media Scanning System

To implement the content extraction and categorisation models, an integrated pipeline that facilitates the crawling of social media is put in place. This process manages and enables the seamless collection, restructuring, processing, filtering and output of the data in preparation for further analysis. The stages of the data preparation and processing pipeline is shown in Fig. 4.

Utilising the 'Search API' offered by Twitter [8], queries can be made against the service's index of recent and/or popular posts from the previous seven days, with only the most relevant tweets returned from during the time period. At the time of writing, the amount of data returned is limited by the API's rate limit, current set at 180 queries per 15 minutes. This amount is subject to change. In

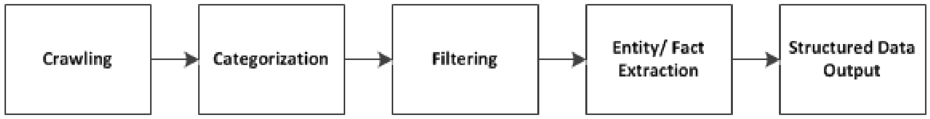

Fig. 4. Social Media Scanning Process

terms of the queries themselves, a number of pre-defined operators exist that allow for the matching of keywords, exact phrases and other operations.

6 Corroborating Information using Formal Concept Analysis

A single Tweet containing a weak signal of OC is not a sensible basis for Law Enforcement Agencies (LEAs) to take action. However, if a number of sources contain weak signals of the same element of OC from the same location, then this may form a credible basis to warrant further investigation. Such corroboration can be automated by the application of Formal Concept Analysis (FCA) [6] to the structured data extracted from the information sources. FCA can be used to cluster the sources into so-called *OC Threat Concepts* (or just *OC Concepts*) where one shared attribute is a location and another shared attribute is an element of OC. When mining the data for formal concepts, if a minimum support is set, say 10 information sources, only OC Threat Concepts with a least 10 sources will be obtained.

To carry out FCA, the structured data extracted from the information sources must be scaled into a formal context. For example, each location in the data becomes a formal attribute in the context. The weak signals of elements of OC are scaled using the Taxonomy. So, for example, a weak signal may indicate Human Trafficking, and thus the source in which the signal was contained will be labeled with the formal attribute *Crime-HumanTrafficking*. Several different weak signals may all point to Human Trafficking, and thus sources containing any of them would all be labeled with the attribute *Crime-HumanTrafficking*. Other weak signals may point to more specific elements of Human Trafficking, such as Exploitation which is a component of Human Trafficking (from the taxonomy). A source with such a weak signal will be labeled with both *Crime-HumanTrafficking* and *Element-Exploitation*. Thus the general 'is a part of' rule in a taxonomy becomes naturally scaled in FCA (see Figure 5).

Using a data set created from 29096 Tweets as information sources, obtained by scanning for Tweets containing weak signals of OC, a formal context was created by scaling the extracted structured data as above. Using a minimum support of 80, the context was mined for OC Threat Concepts using a modification of the In-Close concept miner [1]. The result is visualised as a formal concept tree in Figure 6.

OC Taxonomy	Human Trafficking	Exploitation	Sexual	CSE
weak signal of THB	×			
weak signal of Exploitation	×	×		
weak signal of Sexual Exploitation	×	×	×	
weak signal of CSE	×	×	×	×

Fig. 5. A Formal Context scaling part of the OC Taxonomy (CSE is Child Sexual Exploitation)

In the tree, the head node is the concept containing all the Tweets from concepts that satisfy the minimum support (5512 Tweets) and each of the branches is to an OC Concept - a concept where one attribute is a location and another is an OC. In this example, every OC is Human Trafficking as this was the type of OC being searched for by the Scanning System. The number inside each node is simply a concept ID number assigned by the concept miner. The number outside the node, below the list of attributes, is the object count (the number of Tweets contained in the concept) and in each case this is above the minimum support threshold of 80. Thus concept 53, for example, has the attributes *authorlocation-Atlanta* and *Crime-HumanTrafficking*, and has 185 objects. In other words, within the data set there are 185 Tweets that have the author location Atlanta and contain a weak signal of Human Trafficking. With this high level of corroboration, a police analyst will be alerted to investigate this further, and a possible next step in the investigation is automated by FCA in the form of a ‘drill-down’ to the OC Concept’s sub-concepts.

6.1 OC Concept Drill-Down

The OC Concepts in Figure 6 contain limited information - they only have a location and the OC Human Trafficking. However, individual Tweets in the OC Concept may contain further information pertinent to the OC. But examining 185 Tweets, for example, although far less work than examining 29096 Tweets, is nonetheless quite time consuming. However, several Tweets in the OC Concept may all share the same additional information and this can be divulged by examining the sub-concepts of an OC Concept. Each of the sub-concepts will have the same location and crime as the original OC Concept but with one or more additional attributes from the structured data extracted from the Tweets. Such a result can easily be obtained by mining the data for concepts that contain the attributes of the OC Concept and at the same time reducing the minimum support required.

Figure 7 shows a concept tree with the ‘Atlanta’ OC Concept from Figure 6 and its sub-concepts produced when the minimum support is set to 5.

In the tree, concept 3 shows that 10 of the 185 ‘Atlanta’ Tweets also contain a reference to the drug amphetamine. They may not all contain the actual word *amphetamine*, but they will all contain a word or phrase that is commonly used

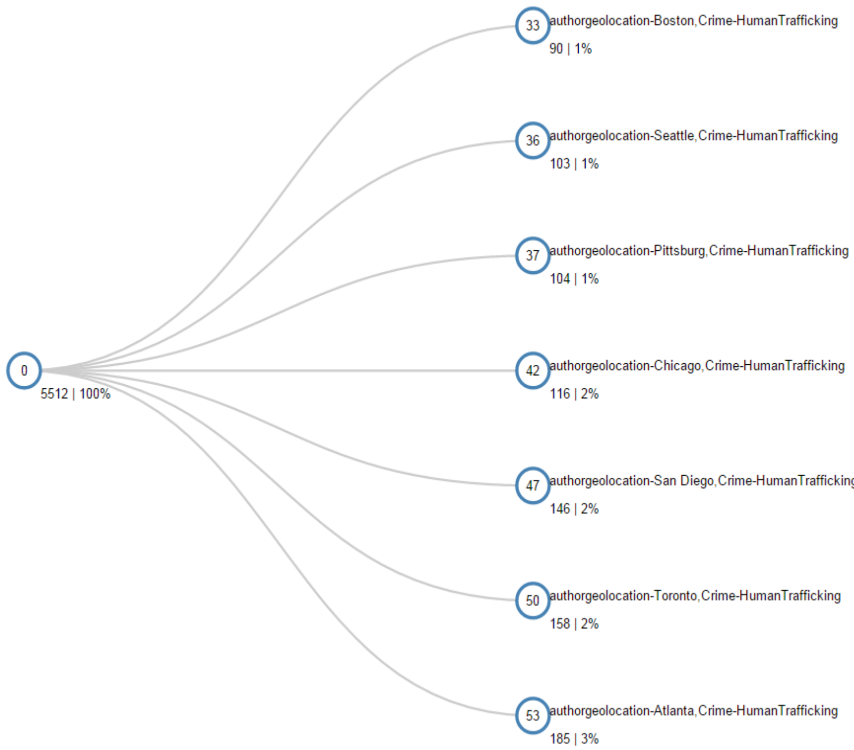

Fig. 6. Tree of OC Concepts

to mean or refer to amphetamine. But using lists of such words and phrases, the entity extraction process carried to produce the structured data will thus label each of these Tweets with the attribute *drug-amphetamine*, which in turn enables the FCA to group them together.

Concept 2 shows that 6 of the 185 Atlanta Tweets also contain the location Central America and the county Mexico. Furthermore, a semantic rule in the entity extraction process has determined that Mexico is being referred to in the Tweet as a trafficking location and thus these Tweets are labeled with the attribute *traffickinglocation-Mexico*.

Concept 1 shows that 10 of the 185 Atlanta Tweets contain weak signals of the OC Human Trafficking *element* Exploitation and the *exploitation type* Sexual. Furthermore, in 8 of those 10 Tweets there are weak signals of CSE, further specialising the OC.

Thus, through this simple automated process, the police analyst has potentially more information that may be pertinent to an OC and more specific information regarding the nature of the OC. Because the original OC Concept involved corroboration by a large number of sources, the analyst can gain some

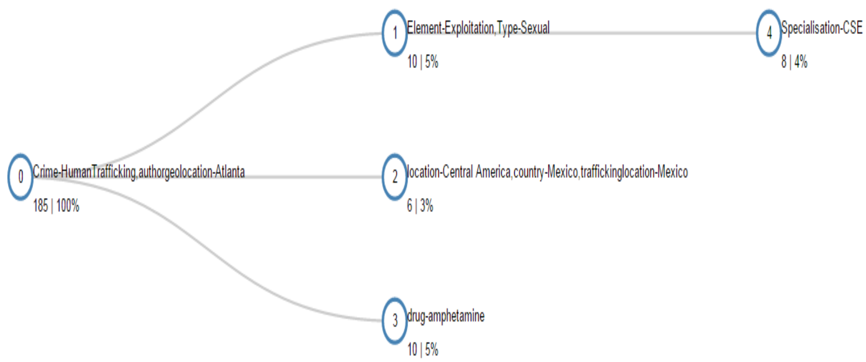

Fig. 7. Drill-Down for the 'Atlanta' OC Concept

confidence that further information contained in sub-sets of the Tweets has credibility. Indeed, the analyst may now want to trace back to the original Tweets (or to the text of these Tweets) and, because they have been grouped together by FCA, it is simple task to provide this facility.

7 Implementation

The processes and components described above were implemented as a part of the European ePOOLICE Project [4]. The OC Taxonomy and entity extraction components developed by the authors (with assistance from people acknowledged below) were implemented in the system to provide data to be consumed by various analytic components, one of which was the FCA 'OC Threat Corroboration' component described above. The user interface to the system was developed by other colleagues at Sheffield Hallam University (who are also acknowledged below). The system allows a police analyst to select a region and type of OC to scan for and then acquire sources on the Internet (such as Tweets) that match those search criteria. Structured data is extracted automatically from the sources, as described above, allowing the user to carry out a variety of analytic tasks and display the results in an appropriate visualisation. For many of the analytics, including the FCA Threat Corroboration a map-based visualisation is used. Figure 8 shows the user interface with the FCA 'Corroborated Threats' option selected. There are a number of other options listed and these components were developed by other members of the ePOOLICE consortium. It is out of scope for this paper to describe them here, but for more information please visit the project website [4].

The map of the USA in Figure 8 is displaying the San Diego and Atlanta OC Concepts from Figure 6. Various icons are used by the system to indicate types of OC and the one here is for Human Trafficking. The analyst is able to select an OC Concept to display its information and to drill down to its sub-concepts. The Atlanta OC Concept has been selected and is thus displaying its associated information, including its attributes (*crime: humantrafficking* and *location: atlanta*) and its objects, listed as URLs. The 'drill-down' sub-concepts (from Figure 7) are being displayed as icons below the main concept and Figure 9 shows the additional information shown when one of these sub-concepts is selected - in this case the attribute *drug: amphetamine*. The 10 sources that contain a reference to amphetamine are listed and at this point the analyst may wish to look at some of these. Clicking on a URL will take the user to the original source.

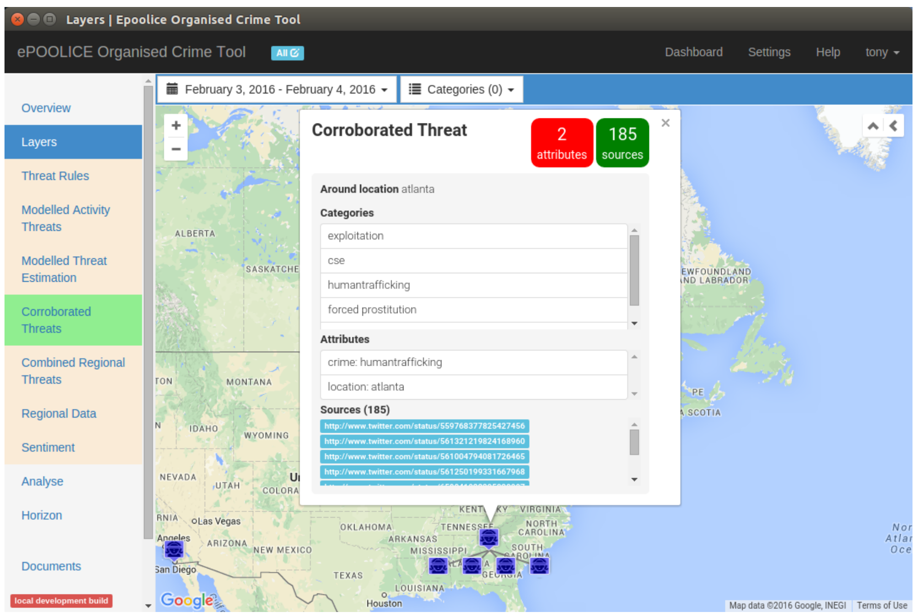

Fig. 8. ePOOLICE system showing San Diego and Atlanta OC Concepts and details of the Atlanta OC Concept

7.1 Evaluation and Concluding Remarks

Although difficult to evaluation in an operational sense (we cannot, for example, act as the police in investigating organised crime) it is possible to say something about the quality of the results in terms of the accuracy of the weak signals. A sample of 20 inferred OC categories were inspected against the original text sources, with 16 out of 20 correctly identified from weak signals as being crime

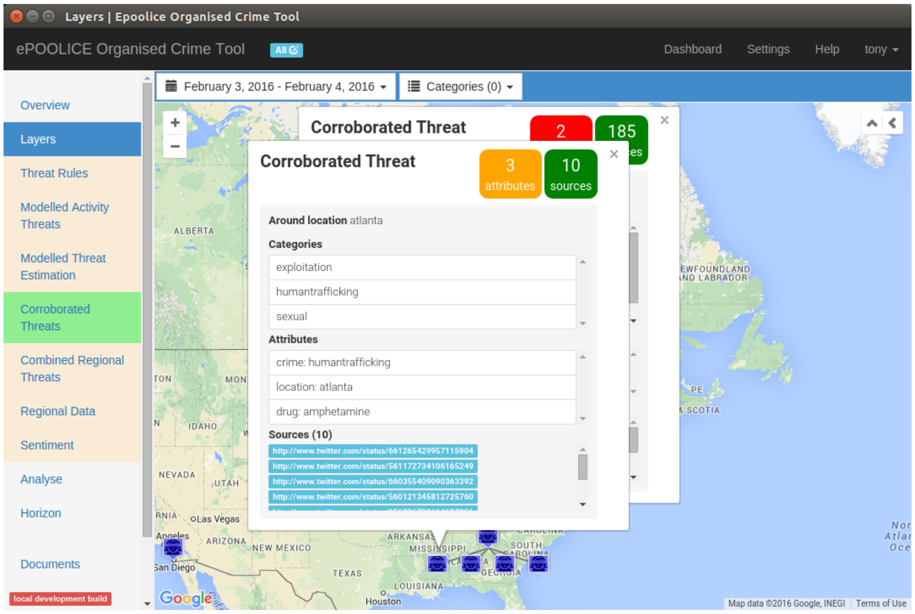

Fig. 9. ePOOLICE system showing Drill-Down information

related. In the other four cases, the context within which the identified words or phrases were used clearly indicated that the source was not referring to OC. Although only a small sample, this was an encouraging level of false positives. Further evaluation is required, however, on larger samples to produce a statistically significant result.

A qualitative evaluation was provided by 24 end-users during a hands-on feedback session held in December 2015. Various law enforcement agencies from within the EU were represented, including those from a regional, national and international organisations. The challenges for the use of the system operationally centred on factors such as the need to refer to personal and sensitive information about the origins of the information extracted and corroborated using the system, and the complexity of the way in which the systems outputs were presented via the map-based interface. However, the utility of the system in providing a means to show trends in current, and alluding to emerging, forms of criminality in different geographic areas was overwhelmingly positive, especially through the utilisation of previously untapped data sources such as social media.

Legal and Ethical Disclaimer No data that can or may be considered sensitive or personal has been handled as a result of the research undertaken. The authors do however acknowledge, despite being outside of the scope of the research present, that in practice the operational utility of such a system would be dependant on the use of data thay may be considered personal and/or sensitive.

References

1. Simon 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. Ben Brewster, Timothy Ingle, and Glynn Rankin. Crawling open-source data for indicators of human trafficking. In *Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing* [2], pages 714–719.
3. Goutam Chakraborty, Murali Pagolu, and Satish Garla. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*. In [3], 2014.
4. ePOOLICE. The epoolice project. Online, 2015.
5. Europol. Eu organised crime threat assessment: Octa 2011. file no. 2530-274. Technical report, Europol, O2 Analysis & Knowledge, The Hague, 2011.
6. Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, 1998.
7. Dan Gilgoff and Jane J Lee. Social media shapes boston bombings response. In *National Geographic* [7].
8. Twitter Inc. Twitter search api. Online, 2015.
9. Frank Laczko and Marco A Gramegna. Developing better indicators of human trafficking. In *Brown J. World Aff.* [9], page 179.
10. Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Where is this tweet from? inferring home locations of twitter users. In *ICWSM* [10], pages 511–514.
11. Her Majesty's Inspectorate of Constabulary. The rules of engagement: A review of the august 2011 disorders. Online, 2011.
12. David Omand, Jamie Bartlett, and Carl Miller. Introducing social media intelligence (socmint). In *Intelligence and National Security* [12], pages 801–823.
13. Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A Smith. Improved part-of-speech tagging for online conversational text with word clusters. Association for Computational Linguistics, 2013.
14. Andrew Perrin. Social media usage: 2005-2015. Technical report, Pew Research Center, 2015.
15. Alan Ritter, Oren Etzioni, Sam Clark, et al. Open domain event extraction from twitter. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1104–1112. ACM, 2012.
16. CISC Strategic Criminal Analytical Services. Strategic early warning for criminal intelligence. Technical report, Criminal Intelligence Service Canada (CISC), 2007.
17. United Nations. United nations convention against transnational organized crime and the protocols thereto. [17].
18. UNODC. Anti-human trafficking manual for criminal justice practitioners, 2009.
19. UNODC. Global report on trafficking in persons, 2009.
20. Phil Williams and Roy Godson. Anticipating organized and transnational crime. *Crime, Law and Social Change*, 37(4):311–355, 2002.