Security Informatics
a SpringerOpen Journal

RESEARCH Open Access

CrimeFighter Investigator: Integrating synthesis and sense-making for criminal network investigation

Rasmus Rosenqvist Petersen* and Uffe Kock Wiil

Abstract

Criminal network investigation involves a number of complex tasks and problems. Overall tasks include collection, processing, and analysis of information, in which analysis is the key to successful use of information since it transforms raw data into intelligence. Analysts have to deal with problems such as information volume and complexity which are typically resolved with more resources. This approach together with sequential thinking introduces compartmentalization, inhibits information sharing, and ultimately results in intelligence failure. We view analysis as an iterative and incremental process of creative synthesis and logic-based sense-making where all stakeholders participate and contribute. This paper presents a novel tool that supports a human-centered, target-centric model for criminal network investigation. The developed tool provides more comprehensive support for analysis tasks than existing tools and measures of performance indicate that the integration of synthesis and sense-making is feasible.

Introduction

Target-centric criminal network investigation involves a number of complex knowledge management tasks such as collection, processing, and analysis of information. The motivation for such work is well described in the training manual of the intelligence analysts of the London Metropolitan Police [1]: *Analysis is the key to successful use of information; it transforms raw data into intelligence. Without the ability to perform effective and useful analysis, the intelligence process is reduced to a simple storage and retrieval system for effectively unrelated data*. Synthesis and sense-making are core analysis tasks; analysts move pieces of information around, stop to look for patterns that can help them relate the information pieces, add new pieces of information and iteration after iteration an information structure appears. Synthesizing emerging and evolving information structures is a creative and cognitive process best performed by humans. Making sense of synthesized information structures (i.e., searching for patterns) is a logic-based process where computers outperform humans as information volume and complexity increase. Developing useful tool support for target-centric criminal network investigation requires integration of synthesis and sense-making.

We present a novel tool for target-centric criminal network investigation called CrimeFighter Investigator with focus on core investigative processes and tasks. CrimeFighter Investigator is part of the CrimeFighter toolbox for counterterrorism [2]. Besides CrimeFighter Investigator, the toolbox consists of the Explorer tool targeted at open source collection and processing and the Assistant tool targeted at advanced structural analysis and visualization.

The remainder of this paper is organized as follows: In the Section 'Criminal network investigation', we discuss this area of security informatics research and focus on four recurring problems. A generic process model for human-centered, target-centric criminal network investigation is proposed to embrace these problems. A list of investigation tasks is presented to guide the development of tool support. The Section 'CrimeFighter Investigator' describes how a selection of these tasks is supported by CrimeFighter Investigator. The Section 'Evaluation' evaluates CrimeFighter Investigator, and the Section 'Conclusions' summarizes our methods, contributions, and outlines future work.

*Correspondence: rasmusrosenqvistpetersen@gmail.com
The Maersk Mc-Kinney Moeller Institute, University of Southern Denmark, Odense, Denmark

Criminal network investigation

Criminal network investigation involves the collection, processing, and analysis of information related to specified targets. We use three investigation cases to identify criminal network investigation tasks and the challenges associated with these tasks. We propose a generic process model for human-centered, target-centric investigation to embrace the identified challenges. Finally, a list of specific investigative tasks are outlined to guide the development of useful software tool support.

A review and step-by-step reconstruction of three criminal network investigation cases[a] emphasized the following challenges: **Resources** (e.g., [3-5]) are inherently a challenge for criminal network investigations, for example not enough man power to follow up on leads. Contextual pressures such as time constraints, dynamism, and changing goals are typically resolved with more resources. Existing evidence suggests that decision-making and information processing abilities are often not optimal because the informational complexity of the world overwhelms human cognitive abilities and creates bias. **Information volume** challenges (e.g., [3,6-8]) includes information abundance and information scarcity. If information is abundant and the resources required to process the information are limited, potential suspects might not be discovered. On the other hand, if information is scarce, decisions might be based on uncorroborated intelligence later proven to be false. **Information complexity** (e.g., [8-10]) can be introduced by both emerging and evolving information or static collections of information where much of the information is unreliable. Information abundance or scarcity on its own does not necessarily make the network information more complex. The use of aliases, social complexity (e.g., culture and language) and the mix of different information types (e.g., audio, images, signals, video) are all factors that increase the complexity of information. **Information sharing** challenges (e.g., [7,11,12]) are often the consequence of a compartmentalized intelligence process, the culture within intelligence agencies and the trade craft of secret intelligence itself. Several reports have concluded that insufficient information sharing between intelligence agencies is often a root cause of intelligence failure. Examples include the 9/11 commission report and the UK report on whether or not the 7/7 bombings in London could have been prevented.

The main goal of our research is to understand criminal network investigation challenges, processes, and tasks and develop tools to assist the people working with these processes and tasks every day, to help minimize the impact of the challenges.

Investigation model

Based on a specific target-centric model for intelligence analysis [12], we propose a generic process model for human-centered, target-centric criminal network investigation [13,14] in Figure 1.

The customer requests information about a specific target. The investigators search through existing information and request information from the collectors (that may also be investigators). Information related to the target is acquired in disparate pieces over time. The investigators use the acquired information to build a model of the target (synthesis) and extract useful information from the model (sense-making). The extracted information added to new information coming in, results in changes to the model (synthesis). The sense-making - synthesis cycle is continued throughout the investigation as new information is acquired from investigation or extracted from the model. The investigators both work individually and cooperatively as a team. The investigation results are disseminated to the customer at the end of the investigation or at certain intervals (or as requested).

Criminal network investigation is a human-centered process. Investigators (and collectors) rely heavily on their past experience (tacit knowledge) when conducting investigations. Hence, these processes cannot be fully automated. The philosophy of the CrimeFighter Toolbox is that the humans (in this case the investigators) are in charge of the investigative tasks and the software tools are there to support them [2]. The tools should be controlled by the investigators and should support the complex intellectual work (e.g., synthesis and sense-making) to allow the investigators to reach better results faster. CrimeFighter Investigator focuses on providing human-centered, target-centric support for criminal network investigations (acquisition, synthesis, sense-making,

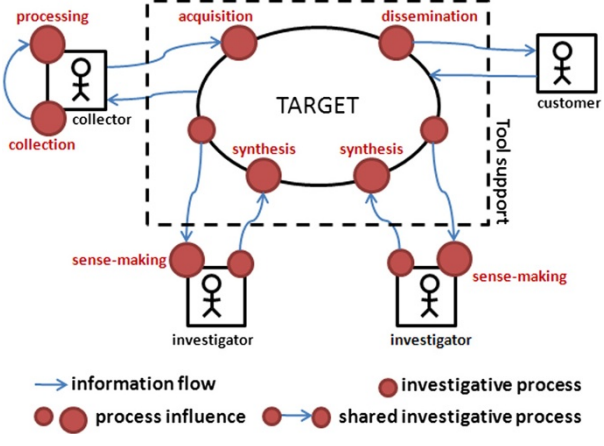

Figure 1 Human-centered, target-centric criminal network investigation model.

and dissemination). Tool support for collection and processing is beyond the scope of this paper: the CrimeFighter Explorer tool focuses on that. Tool support for advanced structural analysis and visualization of the target model generated is also beyond the scope of this paper: the CrimeFighter Assistant tool focuses on that.

Criminal network investigation tasks

Based on cases and observations of criminal network investigation, contact with experienced end-users from various communities, examination of existing process models and existing tools for criminal network investigation (e.g., [1,10,12,15-24]), and our own ideas for tool support, we maintain a list of investigation tasks for each of the processes: acquisition, synthesis, sense-making, and dissemination. The task lists can be seen as wish lists of requirements for what a tool for criminal network investigation should support; the lists serve as the basis for our tool development efforts. The lists are not exhaustive; we expect to uncover additional requirements for all four processes over time.

Acquisition

Some information may be available at the beginning of an investigation, but new information tends to dribble in over time in disparate pieces of varying size and complexity.

- **Acquisition methods.** Electronic information arrives from various sources and should be easy to insert into the investigation tool using methods such as import, drag-and-drop, and cut-and-paste.
- **Dynamic attributes** are required to support acquisition of various data sets formatted using for example graph markup language (GraphML) or comma separated values (CSV), which are likely to have attributes different from those already in an investigation.
- **Attribute mapping.** To support dynamic attributes it is necessary to map attributes in the acquired information to the investigation data model. For example, mapping attributes to i an information element's visual labels.

Synthesis tasks assist investigators in enhancing their particular criminal network model:

- Creating, editing, and deleting **entities**. Investigators think in terms of people, places, objects, their relationships and groups.
- Creating, editing, and deleting **relationships**. Descriptive associations between entities help discover similarities and ultimately solve criminal network investigations. The impact of link (relation) analysis on the creation of the target model is crucial.
- **Re-structuring.** Information structures typically evolve and new structures emerge during the investigation, through continuous re-structuring of entities and their associations.
- **Grouping.** Investigators often group entities using symbols like color and co-location (weak association), or they use more encapsulating symbols like labeled boxes (strong association).
- **Collapsing and expanding** information is essential since the space available for manipulating information is limited physically, perceptually, and cognitively.
- **Brainstorming** is often used during the early phases of an investigation to get an initial overview of the target and the investigation at hand. Brainstorming is an example of a task that involves both synthesis and sense-making activities. Brainstorming is often supported by tools that allow information elements to be organized in a hierarchical manner.
- **Information types.** Multimedia support is helpful when investigators want to add known locations of individuals to a map or link persons to different segments within an audio file.
- **Emerging attributes.** New attributes are added to entities during synthesis and when importing network information into ongoing investigations.

Sense-making tasks assist investigators in extracting useful information from the synthesized target model:

- **Retracing the steps.** Investigators often retrace the steps of their investigation to see what might have been missed and where to direct resources in the ongoing investigation.
- **Creating hypotheses.** Generating sets of alternative hypotheses is a core task of any investigation that involves making claims and finding supporting and opposing evidence.
- **Adaptive modeling.** Representing the expected structure of networks for pattern and missing link detection is a proactive sense-making task. Adaptive modeling embeds the tacit knowledge of investigators in network models for prediction and analysis.
- **Prediction.** The ability to determine the presence or absence of relationships between and groupings of people, places, and other entity types is invaluable when investigating a case. Prediction at different information levels, i.e., attribute-, entity-, and group-level is often required.
- **Alias detection.** Network structures may contain duplicate or nearly duplicate entities. Alias detection can be used to identify multiple overlapping representations of the same real world object.
- **Exploring perspectives.** To reduce the cognitive biases associated with a particular mind set, exploring

different perspectives (views) of the information is a key criminal network investigation task.

- **Decision-making.** During an investigation, decisions have to be made such as selecting among competing hypotheses, or where to allocate resources on the ground.
- **Social network analysis.** Network centrality measures such as degree, betweenness, closeness, and eigenvector can provide important investigation insights.

Dissemination tasks help investigators to formulate their accumulated knowledge for the customer:

- **Storytelling.** Investigators ultimately need to 'tell stories'. Organizing evidence by events and source documents are important tasks, since the alleged story behind the evidence can then be presented.
- **Report generation** involves graphics, complete reports, parts of investigations, etc. Being able to produce reports fast is important in relation to time-critical environments and frequent briefing summaries, which otherwise take up too much valuable investigation time.

CrimeFighter Investigator

CrimeFighter Investigator is based on a number of overall concepts (see Figure 2). At the center is a shared information space. Spatial hypertext research has inspired the features of the shared information space including the support of investigation history [13]. The view concept provides investigators with different perspectives on the information in the space and provides alternative interaction options with information (hierarchical view to the left; spatial view at the center; algorithm output view to the right). Finally, a structural parser assists the investigators by relating otherwise unrelated information in different ways, either based on the entities themselves or by applying algorithms to analyze them. In the following, core CrimeFighter Investigator features are presented, but unfortunately space limitations mean that details are limited and not all supported features can be presented (see [13,14,25-27] for more details).

Acquisition support

CrimeFighter Investigator supports the import of network information formatted as comma separated values. Relations are imported as either an adjancy matrix or a list of information element pairs for large criminal networks. When importing criminal network information into investigations, it is necessary to map all network-dependent variables of the existing data model to attributes of the imported entities. Figure 3 shows the entity attributes for a data set containing *Person* information elements. *Person* entities have a label that displays the value of a particular attribute below the graphical abstraction of the entity (in this case, a stick man figure). When importing data, the user is requested to select the attribute to link to the display label by selecting and dragging the attribute from a list of all the entity's attributes to a label reference area. In Figure 3, the attribute *Short Name* is being dragged to the label reference area for person entities.

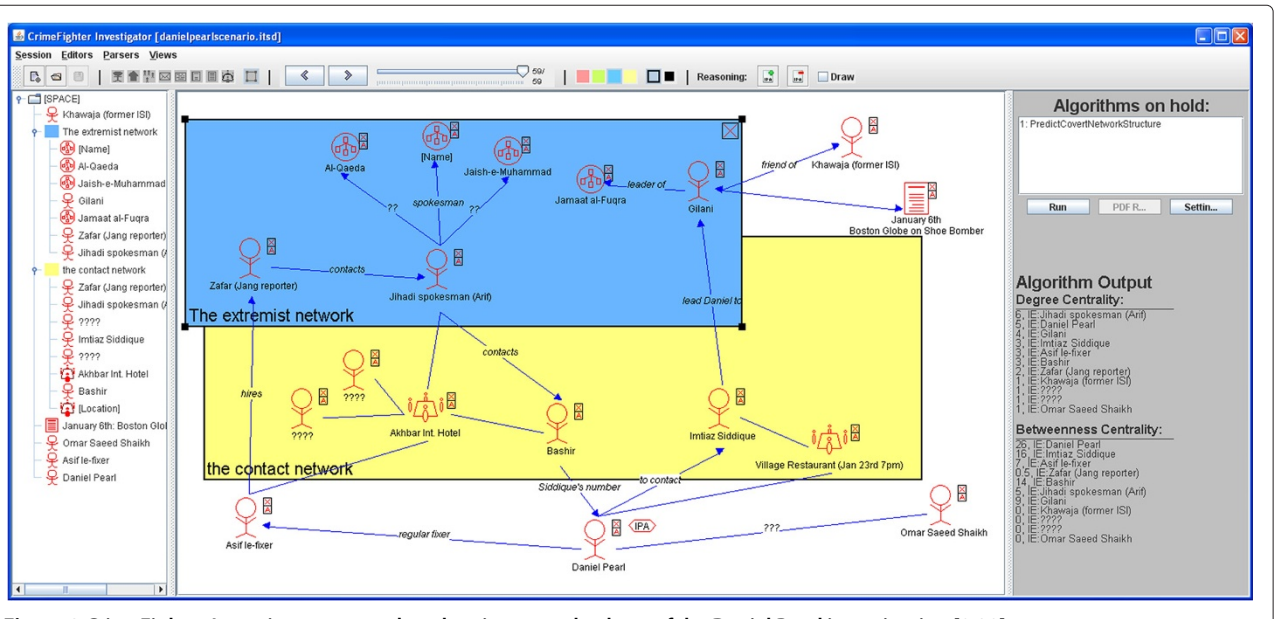

Figure 2 CrimeFighter Investigator screenshot showing an early phase of the Daniel Pearl investigation [9,28].

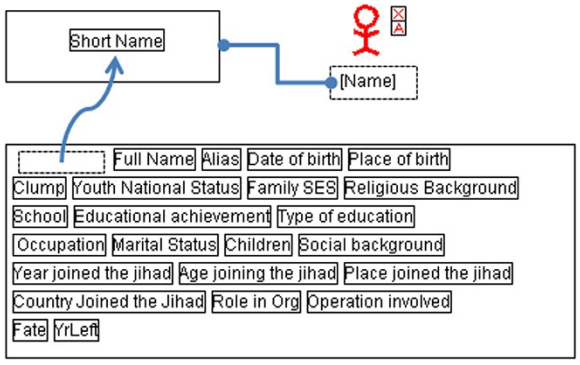

Figure 3 (semi mock-up) Mapping entity attribute to *person* entity's visual label.

Synthesis support

Synthesis of entities and their associations is done using well-known interaction metaphors. CrimeFighter Investigator supports three first class entities: **Information elements**, **relations** (between information elements) and **composites** (for grouping information elements and/or relations). Entities are created using a simple mouse drag gesture within the investigation space. Once created, delete and edit functionalities are available from a button menu attached to the entities as shown in Figure 2. Interactive labels provide another way to edit entity attributes linked to entity display labels. The direction and the label of relations are both edited by clicking the relation label. All entities are deleted using a menu button (☒) positioned relative to the entity. The color of a composite can be set before and after its creation (Figure 2, top).

Our criminal network investigation entities are first class and therefore support continuous **restructuring** of network information. When an information element with multiple relations is deleted, the relation endpoints are considered empty and can be moved freely in the common information space. The investigator can delete the relation endpoints or reconnect them to other entities if desired using a drag and drop gesture. The hierarchical view (Figure 2, left) is used for classification by moving information elements in the hierarchically displayed structure. Different types of composites can be used to group information. The relation composite allows investigators to **group** multiple relations between two entities (such as multiple emails or phone calls between two persons) into a single visible entity (composite). Relation composites group relations by inclusion. Another type of composite supports **collapsing and expanding**. This type of composite groups all information elements by inclusion. Relations that are internal to the composite (have both endpoints inside) are also included, while external relations (at least one endpoint outside) are referenced. This type of composite supports the concept of a subspace that allows the investigators to work in detail with a portion of the complete network.

Sense-making support

Retracing the steps of criminal network investigations is facilitated by a history feature. Recording investigation history allows the investigative team to review the path or progress of their investigation or to retrieve information that previously had been deemed irrelevant or deleted, but then found to have greater significance due to new incoming information. The user interface of the investigation history feature is embedded in the tool bar (see Figure 2). Buttons allow navigation of recorded events, and current events are visualized using a slider as well as a label showing both the current event and the total number of events (e.g., 59/59). The history feature records all the interactions that investigators have with entities as events, e.g., "*create* information element", "*resize* composite", "*move* information element", and so on. Each event is given a time stamp and added to the sequential history.

Prediction support includes covert network structure and missing links prediction algorithms [5,29]. Examples of **social network analysis** support are centrality measures such as degree, closeness, and betweenness [30].

An example of two centrality measures running simultaneously can be seen in Figure 2 (right). A structural parser is used to select the algorithms that the investigator wants to run, if they should run sequentially or simultaneously, the order in which they should run, and how to output the results of algorithms. Customized combinations of different algorithms can be created for frequent use. For example, we developed a custom node removal algorithm which can be used by criminal network investigators to ask 'what-if' questions about the secondary effects of removing a key individual from a criminal network [25].

To evaluate the prediction algorithms, we developed three measures of performance and used two different data sets (see 'Evaluation'). One of these data sets is a network of al-Qaeda individuals who were an active part of the organization up to the beginning of 2003. A missing links prediction on a sampled version of 20 individuals from the al-Qaeda network is shown in Figure 4. The investigator can decide whether to append the predictions to the network or simply discard them.

Creating hypotheses is facilitated by the issue-based argumentation view shown in Figure 5. Criminal network investigators use factual evidence or inferential judgments to reason about the issues they come across in their work. Inferential judgments typically require detailed reasoning involving several positions and even more 'pro' and 'con' arguments, while fact-based reasoning typically is done

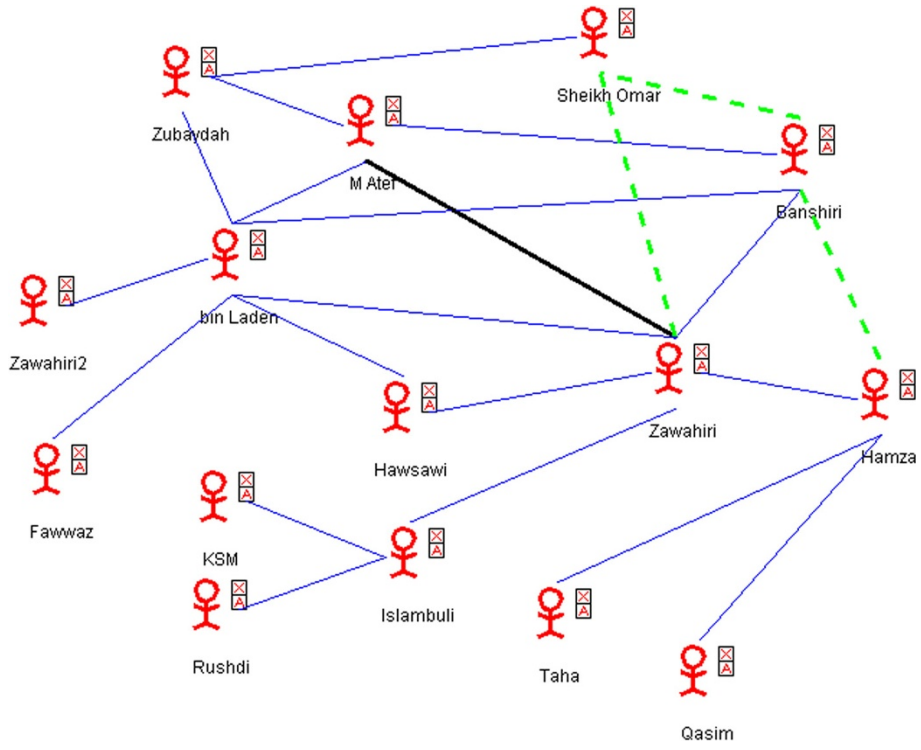

Figure 4 The result of a missing links prediction on a sampled version of 20 al-Qaeda individuals. Blue lines are previously known links. Green dashed lines are predicted links that were ultimately shown to be false. The thick black line is a predicted link that was true.

by creating relations between pieces of evidence (see the person information element 'Daniel Pearl' in Figure 2).

Hypothesis reasoning can be attached to any entity. A small hexagon icon with the text 'IPA' is used to show that reasoning is attached, and clicking the icon opens the issue-based argumentation shown in Figure 5. Reasoning can be used for several purposes: (1) to capture and visualize disagreement in an analysis situation, ensuring that all positions and arguments are heard; (2) to reason convincingly during storytelling (e.g., a senior police officer is creating a briefing based on a recently concluded investigation);

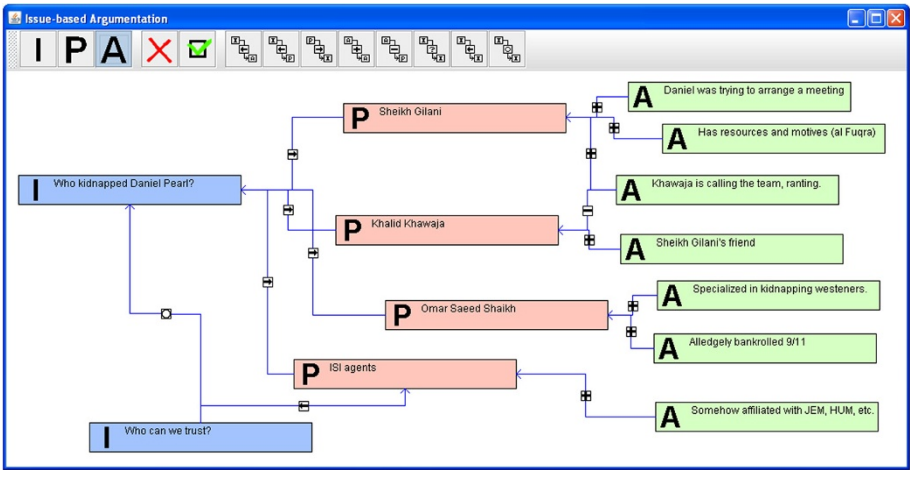

Figure 5 Issue-based argumentation view from the investigation of a kidnapping in Pakistan. The tool bar has buttons for creating issues (I), positions (P), arguments (A), and relations between them.

and (3) to create and explore (competing) hypotheses. According to the IBIS[b] model [31], we have adopted the following predefined relations: is-suggested-by (←), responds-to (→), supports (+), objects-to (−), questions (?), and generalizes or specializes (◯). The relation direction can be both ways in all cases. These predefined relations aid the criminal network investigators in controlling the mapping of their dialogue about issues, positions, and arguments.

Decision-making is currently only supported in the issue-based argumentation view. A "decision" includes a position, the issue it responds to, and associated arguments.

Dissemination support

A history editor is implemented in CrimeFighter Investigator to support **Storytelling**. The granularity of system level history events is often too fine for telling a story. The history editor allows the investigators to group history events that are relevant for the story individually, but when grouped together they explain one important step of the investigation. The investigators can delete events (e.g., if an entity was created by mistake and then deleted), annotate events or groups of events if they feel that the system-generated description is not sufficient, and move events up and down in order to match a time line of real events (e.g., people join a group at a different time from the investigation finding their group membership).

Report generation is available at all stages of a criminal network investigation. All CrimeFighter Investigator features implement a report interface that facilitates the addition or removal of individual report elements. The order in which elements are added to the report is also dynamic. This makes it easier to create reports targeting specific usages. For example, after a prediction is made, reports with or without detailed calculations can be retrieved using the algorithm view (Figure 2, right).

Evaluation

Current criminal network investigation tools aiming to integrate synthesis and sense-making utilize a variety of technologies. Therefore various approaches to evaluation are required. We apply three different evaluation methods. The results of two of these methods are shown in Table 1 and Table 2.

We compared the capabilities of CrimeFighter Investigator with other prominent commercial and research tools for criminal network investigation. In the future, we plan to involve more researchers and end-users in such comparisons. We are currently designing structured usability experiments following [32,33] for evaluation of specific CrimeFighter Investigator features. For now, we present usability feedback gathered from semi-structured interviews with a number of end-users from various criminal network investigation domains. Finally, a sense-making task has been evaluated by developing and testing measures of performance relevant for criminal network investigators.

Capability comparisons

In this section, various tools that support criminal network investigation are compared with CrimeFighter Investigator. The chosen tools include both prominent commercial tools and research prototypes, so that both current and imminent future state-of-the-art is considered. The following three commercial tools have been selected:

- **i2 Analyst's Notebook 8.5** supports a large set of analysis and visualization capabilities support analysts in quickly turning large sets of disparate information into high-quality and actionable intelligence to prevent crime and terrorism [19].
- **Palantir Government 3.0** is a platform for information analysis designed for environments where the fragments of data which an analyst combines to tell the larger story, are spread across a vast set of starting material. Palantir is currently used in various domains such as intelligence, defense, and cyber security [20].
- **Xanalys Link Explorer 6.0** (previously Watson [1]) allows investigators to apply powerful query and analysis techniques to their data, presenting the answers in a range of visualizations such as link charts, time lines, maps, and reports [24].

Also, three research prototype tools have been selected for the comparison:

- **Aruvi** is a prototype implementation of an information visualization framework that supports the analytical reasoning process [22].
- **Sandbox** is a flexible and expressive thinking environment that supports both ad-hoc and formal analytical tasks [23].
- **POLESTAR** (POLicy Explanation using STories and ARguments) is an integrated suite of knowledge management and collaboration tools for intelligence analysts [21].

The evaluation and comparison of the selected tools was made separately for each of the tasks listed in the section 'Criminal network investi- gation'. A thorough examination of each tool has been made by the authors based on the available research literature, books, manuals, and other publicly available information. The results can be seen in Table 1.

Table 1 Overview of tool capability comparison based on support of criminal network investigation processes and tasks

	AN 8.5*	PG 3.0*	XLE 6.0*	Aruvi	Sandbox	POLESTAR	CFI*
Aquisition	**3**	**4**	**4**	**2**	**2**	**3**	**2**
Acquisition methods	■	■	■	■	■	■	■
Dynamic attributes	■	■	■	■	■	■	■
Attribute mapping	■	■	■	□	□	■	■
Synthesis	**2**	**3**	**2**	**1**	**3**	**2**	**4**
Entities	■	■	■	■	■	■	■
Associations	■	■	■	□	■	■	□
Re-structuring	■	■	■	□	■	■	■
Grouping	■	■	□	■	■	■	■
Collapsing and expanding	□	■	□	■	■	□	■
Brainstorming	■	■	■	□	■	■	■
Information types	■	□	■	□	□	□	□
Emerging attributes	■	■	■	□	□	□	□
Sense-making	**1**	**3**	**1**	**2**	**2**	**2**	**4**
Retracing the steps	□	■	□	■	□	□	■
Creating hypotheses	■	■	□	■	■	■	■
Adaptive modeling	□	□	□	□	□	□	□
Prediction	□	□	□	□	□	□	■
Alias detection	□	□	□	□	□	□	□
Exploring perspectives	■	■	■	■	■	■	■
Decision-making	■	■	■	■	■	■	■
Social network analysis	■	■	■	□	□	□	■
Dissemination	**3**	**4**	**2**	**3**	**2**	**1**	**4**
Storytelling	□	■	□	■	□	■	■
Report generation	■	■	■	■	■	□	■

*Tool abbreviations (AN: NB! Analyst's Notebook, PG: Palantir Government, XLE: Xanalys Link Explorer, CFI: CrimeFighter Investigator). Investigative processes (0: no support, 1: fragmentary support, 5: full support). Investigative tasks (■: supported, ■: partially supported, □: not supported).

Each tool is rated against each task in the list. A judgment has been made whether the tool provides full support, partial support, or no support for the task. This is indicated by different icons in the table. Based on the support for individual tasks, each tool has been given a score for each process based on a judgment of the number of tasks they support. This score is either 0 (no support), 1 (fragmentary support), 2-4 (partial support), or 5 (full support). Fragmentary support means that the core task is in theory supported by the tool through the combination of various features, but it is found to be too time-consuming to be really useful.

CrimeFighter Investigator scores well on three of the four evaluated processes synthesis, sense-making, and dissemination, but scores lower on aquisition. In fact, CrimeFighter Investigator is found to provide more comprehensive support of synthesis and sense-making than any of the other tools, and the same high level of support on dissemination.

Usability feedback

We have received usability feedback from a number of people experienced in investigating criminal networks from various fields such as investigative journalism, counterterrorism, and policing (see Table 3). For the usability feedback interviews (individuals or groups) we followed three steps: First, we gave a general introduction to and demonstration of CrimeFighter Investigator. Second, the criminal network investigators were asked to describe their background and characteristics of their ongoing network investigations. Third, we discussed which CrimeFighter Investigator features would be useful for the criminal network investigators in their work.

Table 2 Overview of end user feedback on the usability and relevance of supporting criminal network investigation processes and tasks within each end user's investigation domain

	Investigative journalism	Counterterrorism	Policing	Research and industry
Aquisition				
Acquisition methods	+	+	+	+
Dynamic attributes	+	+	+	+
Attribute mapping	+	+	+	+
Synthesis				
Entities	+	+	+	+
Associations	+	+	+	+
Re-structuring	+	+	−	+
Grouping	−	−	−	+
Collapsing and expanding	−	+	−	+
Brainstorming	+	−	−	+
Information types	−	+	+	+
Emerging attributes	−	+	+	+
Sense-making				
Retracing the steps	−	+	+	−
Creating hypotheses	+	+	+	−
Adaptive modeling	−	+	−	−
Prediction	+	+	−	+
Alias detection	−	+	−	−
Exploring perspectives	−	+	+	−
Decision-making	−	−	+	−
Social network analysis	+	+	+	−
Dissemination				
Storytelling	+	−	+	−
Report generation	−	−	+	−

To indicate the relevance of supporting the investigative task for end users from a particular investigation domain we use a plus sign (+). A minus sign (−) is used to indicate when support is not a priority for the end users from that domain.

To exemplify our interview approach we provide extracts of an interview held with historian and investigative journalist Alex Strick van Linschoten. The example demonstrates the value of CrimeFighter Investigator usability feedback for both development of future features and evaluation of existing features. At the time of the interview, Alex is investigating the alleged links between al-Qaeda and the Afghan Taliban, and he has observed several interesting network characteristics (marked using a bold font in the paragraph below).

Table 3 Investigators interviewed from secret and public criminal network investigation organizations

	Secret	Public	Both
Investigative journalists	0	1	1
Intelligence analysts	1	5	6
Police officers	3	0	3
Research community	0	7	7
TOTAL	**4**	**14**	**24**

Alex's data set on the Afghan Taliban spans the **time-period** 1970–2011. As of 2011 the data set had **500–600 individuals**, a network he claims to have more or less memorized. The data set is based on interviews with Taliban members who were asked who they fought with in the '80s, their andiwaal groups (friend groups normally formed by Afghans in their teenage years) and other **relations**. **Reports** on Afghanistan by the International Security Assistance Force ISAF also contribute to the data set. 70 percent of the **relations** in the network are based on rumors, which is indicated using relation **weights**. When Alex interviews

Taliban members he notes down **attributes** such as 'name', 'date of birth', 'place of birth', 'tribe', 'ethnicity' and 'andiwaal group'. Alex uses Tinderbox [34], a spatial hypertext **tool**, to **record** and **structure** the network information he **collects**.

Alex processes the network information in a number of different ways and has in general many ideas for how a tool such as CrimeFighter Investigator could be applied to organize the information he retrieves. Alex studies the evolution of the network through time (a historical evolution **perspective**). He believes that knowledge about an individual's andiwaal **group** could be used to **predict** who that person might be fighting with in future operations. Alex is **searching** for different **patterns** in the data set like for example changes in age or gender.

Alex has encountered a number of problems for which specialized **tool support** would be an advantage, for instance, a social network analysis tool for support of an actual time line (Tinderbox only supports snapshots of the network). At the time of interview he was **analyzing** the network data to see if there were any important observations that he might have missed. Alex mentions that different **layout** functionality would be useful for this, e.g., laying out nodes according to **betweenness centrality**. Finally, if Alex **exports** information from Tinderbox [34] to **import** it into Analyst's Notebook [19] to create a special visualization, it is not possible to get that visualization back into Tinderbox. The **interchange** of information is not facilitated both ways.

The feedback from potential end users of CrimeFighter Investigator on the usability of the features presented to them, has helped us prioritize those features to focus on the most important ones. In general, the end users found the features supported by CrimeFighter Investigator to be applicable to their work, and in some cases our integration of features was found to be more useful than existing tools (such as i2 Analyst's Notebook, see summary of evaluated tools earlier in this section).

Measures of performance

Our measures of performance (MoPs) focus on the internal structure, characteristics, and nature of criminal network sense-making. We have developed three measures to help us evaluate how CrimeFighter Investigator sense-making performs in terms of 'information volume', 'attribute completeness', and 'attribute accuracy'. In the longer term, these MoPs will help us build a process that criminal network investigators can have confidence in, when going before a decision maker in their organization [35]. We need to make sure that our algorithm-supported sense-making tasks can perform on the criminal networks that investigators are dealing with on a daily basis. More specifically, we want to evaluate if the integration of synthesis and sense-making tasks is feasible.

In this section, we test our support for one sense-making task, prediction, by evaluating the predict missing links algorithm. We use two criminal networks for our evaluation: November 17 and al-Qaeda. The data set of the (believed defunct) Greek terrorist group November 17 (N17) was derived from open source reporting [36]. The N17 group was a small close knit organization of 22 individuals with 63 links out of a potential 231 links. The links of the dataset were obtained from open sources and report some connection between two individuals at some point in the past, but no specific weightings of the links are given [29] (i.e., the link weights are all the same).

The second dataset is the al-Qaeda network at the beginning of 2003. All the network information was gathered from public domain sources: 'documents and transcripts of legal proceedings involving global Salafi mujahedin and their organizations, government documents, press and scholarly articles, and Internet articles' [37]. We have included *acquaintance*, *friend*, and *post joining jihad* relations, but the algorithm does not differentiate between them. *Nuclear family*, *relatives*, *religious leader*, and *ties not in sample* links are excluded from our version of the data set. The topology of all networks are presented in Table 4.

We use sampled versions of the full networks for our evaluations, created by removing either 50 or 25 percent of the links in the network and then seeing what is left. The number of nodes and links alone directly affect algorithm performance in terms of speed. The number of attributes that each node has does not impact the performance of the predict missing links, since tests are run with four attributes every time. We define the complexity of node attributes as the average of valid enumerated values per attribute. Link density is the ratio between the number of links and the number of potential links and indicates for example the connectivity and covertness of the given network.

We logged three variables for each test. Time is the seconds it takes to predict missing links. True positives are predicted links that exist in the non-sampled version of the data set. False positives are predicted links that do not exist in the non-sampled version of the data set. The predict missing links algorithm was customized in the same way for each sampled data set before each test as described in Table 5. The al-Qaeda attributes are selected to match the number of enumerated values for each November 17 attribute.

We evaluate the predict missing links algorithm against all the data sets using the three measures of performance. The results are listed in Table 6.

Table 4 The November 17 and al-Qaeda datasets

	al-Qaeda			November 17	
version →	*full*	*full*	*id 1–20*	*full*	*full*
sampling →	*100%*	*25%*	*50%*	*100%*	*50%*
Nodes	366	256	15	22	17
Attributes	17	17	17	11	11
Complexity*	9.53	9.53	9.53	2.09	2.09
Links	999	249	18	63	32
Link density	0.015	0.008	0.17	0.27	0.24

*Complexity indicates the average number of enumerated values (text strings) for each entity attribute.

Information volume

This measure of performance is based on the change in processing time and true and false positive ratios when the number of nodes and links increases across the three sampled data sets.

We observe that the sampled al-Qaeda data set increases the time required to process the prediction significantly (as expected). However, in the worst case the logged time is only 63 seconds and this value does not raise any operational concerns for most criminal network investigations. We realize that the network can be much larger, and expect the required time to increase also for the tested data set if attributes with more enumerated values were selected. But it is our experience that for very large networks, criminal network investigators will request predictions within subgroups mostly and not the whole network. And if the network gets very large, then the investigators will be prepared to wait longer for assistance from the algorithm.

Attribute accuracy

The predict missing links algorithm assumes that attribute content (text string) is machine-recognizable, i.e., the content should be one of a list of predefined text strings (e.g., role [LEADERSHIP, OPERATIONAL] or degree centrality [HIGH, MIDDLE, LOW]). We have decreased the attribute accuracy of the sampled data set by scrambling a percentage of the attribute values.

Table 5 Algorithm setup for the November 17 and al-Qaeda data sets

Data set →	November 17	al-Qaeda
L Cutoff	2.5	2.5
Attribute 1	*Role*	*Children*
Attribute 2	*Faction*	*Clump*
Attribute 3	*Resources*	*Fate*
Attribute 4	*Degree centrality*	*Degree centrality*

Decreasing the accuracy of attribute content to simulate the data having reduced reliability clearly impacts on the number of predicted links, but the ratio between true and false positives does not change, indicating some robustness of the predict missing links algorithm. The time actually decreases together with the decreasing accuracy of attributes; a decrease in predicted links (due to less attribute content matching up) can more easily be processed by the algorithm. One interesting observation here is that the ratio of true positives dropped significantly for the N17 data set at 70% acurracy to 1 (from 5 at 90% and 9 at 100%). We expect this is caused by N17 having fewer attributes than the al-Qaeda data set, making it more vulnerable to random scrambling of attribute values.

Attribute completeness

End user requirements and usability feedback indicated a need to support dynamic and emerging entity attributes, since limited information is typically available about the individuals in criminal networks. To simulate this, we delete attribute values from the data sets and replace them with empty values.

Like attribute accuracy the total number of predicted links decreases as the number of non-empty attribute values increases but the ratios stay more or less the same. We anticipated this similarity between the accuracy and completeness MoPs as CrimeFighter Investigator does currently not include technology that could improve the attribute accuracy by correcting for example typographical spelling errors.

We chose the predict missing links algorithm to evaluate our developed measures of performance, because we found it promising for investigation of criminal networks of sizes equal to or smaller than the al-Qaeda data set. It is our experience, that criminal network investigators do not synthesize and apply sense-making algorithms to networks significantly bigger than the al-Qaeda data set [37].

Conclusions

The CrimeFighter Investigator approach to target-centric criminal network investigation has been developed based on three types of analysis work:

- **Exploring methods.** We have explored analytical practices, processes, and techniques related to investigative journalism, counterterrorism and policing.
- **Studying related work.** We have found inspiration from existing tools supporting criminal network investigation and from other relevant investigations.
- **Evaluation.** We have compared the capabilities of six existing tools with the CrimeFighter Investigator. We have collected feedback from various criminal network investigation communities regarding which CrimeFighter Investigator features are useful and usable. Finally, we have developed three measures of performance for prediction algorithms.

Together, this analysis work resulted in a list of tasks that guided our development. Many of the tasks envisioned are already now supported. The main contributions to useful integration of criminal network synthesis and sense-making are:

Process model. We have developed a target-centric process model for criminal network investigation, splitting the responsibilities between investigators and tool, empowering humans to make more informed decisions.

Task list. We have outlined and evolved criminal network investigation tasks spanning acquisition, synthesis, sense-making, and dissemination processes, helping us understand how to integrate these tasks.

Tool. We have developed a tool that assists criminal network investigators in target-model synthesis and sense-making, to produce useful intelligence products for their customers.

Evaluation. Our evaluation results from capability comparisons, usability feedback, and measures of performance, indicate that we are on the right path to integrate a broad range of criminal network synthesis and sense-making tasks in one tool. We have observed that existing tools typically focus on either synthesis or sense-making tasks.

Together, this has resulted in a novel tool that combines knowledge from various research domains (including hypertext, knowledge management, software engineering, and criminal network analysis) to address criminal network investigation challenges, processes, and tasks.

As part of our near term future work, we will provide support for the remaining tasks related to synthesis, sense-making, and dissemination. We then plan to thoroughly test the tool in cooperation with experienced investigators. In the longer term, we plan to include

Table 6 Measures of performance for the predict missing links algorithm

	Data set	Version	Sampling	Time (s)	TP*#	TP%	FP*#	FP%
Full data set								
	November 17	(full)	(50%)	0.219	9	42.9	12	57.1
100%	al-Qaeda	(id 1–20)	(50%)	0.078	7	35.0	13	65.0
	al-Qaeda	(full)	(25%)	63.093	288	4.9	5547	95.1
Attribute accuracy								
	November 17	(full)	(50%)	0.235	5	35.7	9	64.3
90%	al-Qaeda	(id 1–20)	(50%)	0.79	6	46.2	7	53.8
	al-Qaeda	(full)	(25%)	37.562	165	5.1	3052	94.9
	November 17	(full)	(50%)	0.124	1	16.7	5	83.3
70%	al-Qaeda	(id 1–20)	(50%)	0.62	5	45.5	6	54.5
	al-Qaeda	(full)	(25%)	24.656	167	5.0	3171	95.0
Attribute completeness								
	November 17	(full)	(50%)	0.282	5	45.5	6	54.5
90%	al-Qaeda	(id 1–20)	(50%)	0.094	7	41.2	10	58.8
	al-Qaeda	(full)	(25%)	41.344	197	4.8	3939	95.2
	November 17	(full)	(50%)	0.531	5	45.5	6	54.5
70%	al-Qaeda	(id 1–20)	(50%)	0.079	5	41.7	7	58.3
	al-Qaeda	(full)	(25%)	24.328	146	4.4	3167	95.6

*TP = true positives, FP = false positives.

support for cooperation beyond the shared information space and provide integration with the CrimeFighter Explorer tool for better acquisition support.

Endnotes

[a] The kidnapping of Daniel Pearl [9,28], the intelligence used for the United States case against Iraq concerning their (alleged) wmd programme [6,11], and the links between Operation Crevice and the 7/7 bombings in United Kingdom [3,4].

[b] Issue Based Information Systems.

Competing interests

Both authors declare that they have no competing interests.

Authors' contributions

RRP carried out the analysis, design, and implementation of support for criminal network investigation processes and tasks in CrimeFighter Investigator. RRP also carried out the analysis that formed the basis of evaluations, designed the experiments for each of the three evaluation methods, and subsequently carried out these experiments. UKW supervised the whole process. In collaboration, UKW and RRP synthesed the capability comparison table and developed the target-centric model for criminal network investigation. Both authors read and approved the final manuscript.

Acknowledgements

This work is based on previous publications by the authors in hypertext and security informatics conference proceedings [13,14,25].

Received: 15 November 2011 Accepted: 22 March 2013
Published: 1 May 2013

References

1. MK Sparrow, The application of network analysis to criminal intelligence: An assessment of the prospects. Soc. Netw. **13**, 251–274 (1991)
2. UK Wiil, N Memon, J Gniadek, CrimeFighter: A toolbox for counterterrorism. Lect. Notes Commun. Comput Inf. Sci. (Knowl. Discov., Knowl. Eng, Knowl. Manage). **128**, 337–350 (2011)
3. SecurityCommittee Intelligence and, *Could 7/7 have been Prevented? Review of the Intelligence on the London Terrorist Attacks on 7 July 2005*, (UK, 2009)
4. G Woo, in *Mathematical Methods in Counterterrorism*, ed. by N Memon, T Farley, JD Hicks, and DL Rosenorn. Intelligence constraints on terrorist network plots (Springer Wien, 2009), pp. 205–214
5. CJ Rhodes, P Jones, Inferring missing links in partially observed social networks. J. Oper. Res. Soc. **60**(10), 1373–1383 (2009)
6. B Drogin, *Curveball*. (Ebury Press, 2008)
7. National commission on terrorist attacks upon the United States, *The 9/11, Commission Report (Executive Summary)* (USA, 2004). http://www.9-11commission.gov/report/911Report_Exec.pdf
8. MR Kebbell, DA Muller, K Martin, Understanding and managing bias Dealing Uncertainties Policing Serious Crime. (Australian National University, Canberra, 2010), pp. 87–97
9. BH Levy, *Who Killed Daniel Pearl?*. (Melville House Publishing, Brooklyn, 2003)
10. C Atzenbeck, DL Hicks, N Memon, Supporting reasoning and communication for intelligence officers. Int. J. Netw. **8**(1/2), 15–36 (2011). Virtual Organisations
11. T Weiner, *Legacy of Ashes: The History of the CIA*. (Anchor Books, New York, 2008)
12. R Clark, *Intelligence Analysis: A Target-Centric Approach*. (CQ Press, California, 2007)
13. RP Petersen, UK Wiil, in *Proceedings of the 22nd ACM Conference on Hypertext*. Hypertext structures for investigative teams (ACM Press New York, 2011), pp. 123–132
14. RR Petersen, UK Wiil, in *Proceedings of European Intelligence and Security Informatics Conference*. CrimeFighter Investigator: a novel tool for criminal network investigation (IEEE, 2011), pp. 360–365
15. R Adderly, P Musgrove, Police crime recording and investigation systems - A user's view. Int. J. Police Strateg. Manage. **24**, 100–114 (2001)
16. RV Badalamente, FL Greitzer, in *Proceedings of International Conference on Intelligence Analysis*. Top ten needs for intelligence analysis tool development, (2005)
17. EA Bier, SK Card, JW Bodnar, Principles and tools for collaborative entity-based intelligence analysis. IEEE Trans. Vis. Comput. Graph. **16**(2), 178–191 (2010)
18. G Dean, P Gottschalk, *Knowledge Management in Policing and Law Enforcement*. (Oxford University Press, 2007)
19. i2, Analyst's Notebook (2011). http://www.i2group.com/
20. Palantir Government (2011). http://www.palantirtech.com/government
21. NJ Pioch, JO Everett, in *Proceedings of the International Conference on Information and Knowledge Management*. POLESTAR: collaborative knowledge management and sensemaking tools for intelligence analysts (ACM Press New York, 2006), pp. 513–521
22. YB Shrinivasan, JJ Wijk, in *Proceedings of the 26th Conference on Human Factors in Computing Systems*. Supporting the analytical reasoning process in information visualization (ACM Press New York, 2008)
23. W Wright, D Schroh, P Proulx, A Skaburskis, B Cort, in *Proceedings of the Conference on Human Factors in Computing Systems*. The Sandbox for analysis: concepts and methods (ACM Press, 2006), pp. 801–810
24. Xanalys (2011). http://www.xanalys.com/
25. RR Petersen, CJ Rhodes, UK Wiil, in *Proceedings of European Intelligence and Security Informatics Conference*. Node removal in criminal networks (IEEE Computer Society Washington, 2011), pp. 360–365
26. RR Petersen, UK Wiil, in *Handbook of Computational Approaches to Counterterrorism*, ed. by Subrahmanian. V S. CrimeFighter Investigator: criminal network sense-making (Springer New York, 2013), pp. 323–359
27. RR Petersen, in *Criminal Network Investigation: Processes, Tools, and Techniques* (University of Southern Denmark, 2012). Ph.D. dissertation
28. M Pearl, *A Mighty Heart*. (Virago Press, 2004)
29. CJ Rhodes, CMJ Keefe, Social network topology: a Bayesian approach. J. Oper. Res. Soc. **58**(12), 1605–1611 (2007)
30. S Wasserman, K Faust, in *Social Network Analysis: Methods and Applications* (Cambridge University Press Cambridge, 1994)
31. J Conklin, ML Begeman, gIBIS: a hypertext tool for exploratory policy discussion. ACM Trans Inf Syst. **6**(4), 303–331 (1988)
32. A Field, G Hole. How to Design and Report Experiments (Sage Publications Ltd London, 2003)
33. C Atzenbeck. WildDocs - Investigating Construction of Metaphors in Office Work (Aalborg University, 2006). PhD thesis
34. M Bernstein, *The Tinderbox Way*. (Eastgate Systems, Watertown, 2006)
35. JP Stenbit, IL Wells, DS Alberts, *NATO code of best practice for C2 assessment, [Chapter 5: Measures of Merit]*. (CCRP, Washington, 2002)
36. C Irwin, C Roberts, N Mee, in *Defence Science and Technology Laboratory*. Counter Terrorism Overseas, (2002). Dstl/CD053271/1.1 UK
37. M Sageman. Understanding Terrorist Networks (University of Pennsylvania Press (PENN) Philadelphia, Pensylvania, 2004)

doi:10.1186/2190-8532-2-10
Cite this article as: Petersen and Wiil: **CrimeFighter Investigator: Integrating synthesis and sense-making for criminal network investigation.** *Security Informatics* 2013 **2**:10.