Behavioural & Tempo-Spatial Knowledge Graph for Crime matching through Associative Questioning and Graph theory

Nadeem Qazi
School of Science & Technology
Department of Computer Science
Middlesex University
Email: n.qazi@mdx.ac.uk

B.L. William Wong
School of Science & Technology
Department of Computer Science
Middlesex University
Email: w.wong@mdx.ac.uk

Neesha Kodagoda
School of Science & Technology
Department of Computer Science
Middlesex University
Email: N.Kodagoda@mdx.ac.uk

***Abstract*—Crime matching process usually faces the time tedious and information intensive task of eliciting plausible linkages in the criminal data. It involves the extraction of possible associations among actors of crimes to identify potential suspects. Exploratory data analysis in the criminal investigation, though increasing, yet research towards crime matching is still in its infancy stage. Aiming towards the assistance of this procedure, we in this paper have exhibited the utilization of associative search; a relatively new search mining instrument to evoke conceivable associations from the information. We have demonstrated the use of three-dimensional, i.e. spatial, and temporal and modus operandi based similarity matching of crime pattern to establish hierarchical associations among the crime entities. Later we used these associations to extract plausible suspect list for an unsolved crime and the co-offender network to facilitate the crime matching process. The consideration of suspects in this list is built on the associative questioning relating the similarity of crime pattern of the unsolved crime to that of the suspect. Additionally, a similarity score is calculated to rank the suspect in the plausible list.knowledge graph consisting of tree structure coupled with iconic graphic are used to visualize the plausible list and co-offender network.We have conducted several prototype evaluation sessions with the groups of end-users (police intelligence analysts) and found very positive feedback. We also have estimated the performance of the associative search for generating the knowledge graph.The proposed visualization aims to assist in hypothesis formulation reducing computational influence in the decision making of criminal matching process.**

***keywords*—** Data mining,exploratory data analysis,associative questioning,data visualization,knowledge graph,graph/network,Linked Analysis.

1. Introduction

Crime matching is the process of assigning criminals to the previously solved or unsolved crimes [1].Reference [2] however, describes it as the ability to link or connect crimes in ways that enables the identification of potential suspect. Inspired by these two thoughts we describe the crime matching as a process that either involves of assigning the unsolved crime to a known offender or alternatively it investigates to find what possible unsolved crimes have been committed by some prolific offenders on the basic of similarity matching of his crime pattern with that of unsolved crime.Crime matching in either case is the information intensive and time tedious process. It requires enough relevance based information to connect the series of clues gathered from the crime scene for sense making to facilitate hypothesis formulation.

In our related last studies [3] we had located that police analysts follow "associative questioning" to learn more about the diverse nature of the occurred crime. They elicit associations among criminal entities through variety of contextual questions such as who else has used the same modus operandi in the similar crime, who are the other offenders who commit crimes like this,what are the similarities between the past and current crimes etc.This practice indeed requires them to painstakingly perform extensive and complex database searches, reading crime reports looking for clues for criminal associations among criminal entities such as criminals, vehicles, weapons, bank accounts, and organizations.This tedious time consuming process of associations extraction,coupled with lack of proper visualization suffers the efficiency of the crime matching process. Therefore its valuable important and challenging to use relevance for collecting similar cases , finding associations between them and present these computationally unconnected but operationally plausible associated data together in the same visual field to facilitate hypothesis formulation.

The purpose of this paper is to describe our adopted approach in answering these challenges to support the analytical reasoning processes engaged during the crime matching process. our focus is on how the concept of hierarchical associations can be implemented to support the analytical reasoning process of associative memory activation [4].We have employed 5WH query model to reveal the associations among crime entities, thus making it feasible for a machine to suggest the possibility of linkages between data for sense making. Our proposed data exploratory based technique introduce the concept of hierarchical associations elicited through associative questioning and unified spatial-temporal and behavioral crime pattern matching from the linked crime data. Later we utilize these associations to identify the solved crimes from unsolved crimes, associated offenders of the solved crimes, the co-occurrence network of offenders and most importantly to extract a list of plausible suspects for a given unsolved crime.This extracted suspect list rank each of its members on the basis of the frequency and similarity score of their committed solved crime with that of unsolved crime. In addition to this, we also have attempted to present interactive visualizations of this suspect lists and their network in parent-child relationship connecting each spatial, temporal and behavioral component of the given crime pattern to that of the offenders.The extraction of suspect list is a stepping step to start a crime matching process and require tedious time-consuming effort, the proposed visualization aims to expedite this process.

This research is the part of the VALCRI project - Visual Analytic for Sense-making in Criminal Intelligence Analysis.It is a 17 organization research consortium with the goal to research, design and develop a next-generation technology for information exploitation by police intelligence analysts. VALCRI aims to facilitate human reasoning and analytic discourse for intelligence analysis through a semi-automated human-mediated semantic knowledge extraction capability. In VALCRI we take the view that humans are still superior to machines especially in complex and dynamic environments where there is no or limited, missing, incomplete, uncertain or ambiguous data.

The rest of this paper is structured as follows: In Section 2 we present a selection of state of art literature on association analysis in criminology, followed by brief discussion of the used dataset in Section 3. We describe our proposed methodology including associative query manager and the association extraction module in Section 4, followed by the visualization in Section 5. A preliminary evaluation of the visualization is described in the sixth section and conclusion is made in the last section of this paper.

2. Literature Review on association analysis in Criminology

We have performed a detailed literature review towards the search mechanism and visualization that has been adopted by the researchers in the criminal analysis. The research in this field is still in its infancy stage, researchers have demonstrated the separate use of temporal, spatial and behavioral features of the committed crimes in identifying criminal hot spots, extracting associations and criminal networks etc with no or very minimal visualizations. They have employed various techniques such as concurrent analysis, shortest path algorithms, clustering, artificial neural network, self-organizing map.

We here reference few projects such as COPLINK, Prep-Search, (RECAP) and CrimeLink Explorer etc. COPLINK[5], [6], [7] one of the earlier projects started in 1997 is a framework for text mining, classification, and clustering of crime data aiming to accomplish relatively complex crime analysis visualized through a hyperbolic tree view. Reference [7] over the same framework, has employed co-occurrence analysis to identify, weighted criminal associations among five types of entities i.e. person, vehicle, organization, location, and incident/crime type. Another criminal search mechanism Prep-Search [8] answers only WHO and WHERE questions identifying suspects and their locations. The methodology is based on the concepts of geographic profiling with social network analysis, crime patterns, and fuzzy matching.

The Regional Crime Analysis Program (RECAP) has employed data fusion and spatial data mining techniques for crime analysis[9]. It allows the user to search with respect to location, time and geography. Reference [10] however has employed modus operandi component of the crime to establish associations among the crime cases and chronic criminals on burglary and robbery dataset. The employed association rule is based on similarity of the modus operandi used in the crime to that of the criminal profile of the criminal. Reference [11] have proposed a technique that uses shortest-path algorithms, priority-firstsearch (PFS) and two-tree PFS, to identify the strongest association paths between entities in a criminal network.

The CrimeLink Explorer by [12] identifies associations between people only, without revealing any associations information of other entity types such as location or vehicle. They employed shortest path algorithm, co-occurrence analysis, and a heuristic approach over the structured crime incident data extracted from the Tucson Police Department(TPD) Records Management System. The associations among these crime entities are visualized through lines with a thickness of the line proportional to the weight of the associations. They have also compared visualization as offered by some of the commercial software packages for link analysis such as Analysts Notebook, Netmap, Crime Link, Orion, and Visual ink. They concluded that all these software lack knowledge discovery ability for association identification and require data for visualization.

Recently Criminal network analysis has received great attention from researchers. The pioneer work in criminal network analysis is from [13], who have extracted criminal relations from a large volume of police departments incident summaries through data mining techniques. They determined the weight of relationships between pairs of criminals using co-occurrence frequency.Reference [14] demonstrated a method to extract criminal networks from web sites that provide blogging services by using a topic-specific exploration mechanism. Introducing notion of prominent criminal communities, [15] have proposed a social network mining method to extract social groups from unstructured textual data achieved from a suspects hard drive. Their proposed method can discover prominent communities of indefinite size identifying both direct and indirect relationships. Reference [16] have proposed an outliers-detection-based methodology over categorical variables to establish associations between crime and the criminal implemented over robbery data. The purposed association is based on the assumption that crime generated from the distinct template are more likely to be committed by the same criminal and is evaluated through measuring the distinctiveness, the higher the distinctiveness the more the association between crime and criminals.

Its clear from the above literature review that so far there is no any single envelop available in criminal analysis literature with adequate visualization that utilizes spatial-temporal and modus operandi similarities of the committed crime to extract multi-level association among crime entities. Additionally the existing visual criminal analysis tools lack knowledge discovery ability for association identification and require associations information to be fed in by the user for its visualization.

We in this work have utilized a combined feature vector consisting of multi-valued temporal, spatial and behavioral information of the committed crime to extract plausible suspects for the given unsolved crime based on associative questioning.Additionally using same spatial temporal and modus operandi similarity we elicit co-offender network. This combine approach of spatial temporal and modus operandi based similarity matching has never been demonstrated in criminal investigation literature for extracting criminal associations.

3. Data-Set

Anonymized Burglary Dataset was collected from UK Law Enforcement Agency for the experiment to evaluate the performance of the proposed associative search mechanism. The police utilize several kinds of updated information including a reported crime, Modus operandi description, Stop & Search, Forensic and intelligence data gathered from different sources.

A crime report document is generally defined to be a logical unit of textual data consisting of the crime reference, offender information, modus operandi description, offense category, location, time and date of the crime occurred along with other related informations. Modus operandi is the method of the operation adopted to commit a crime, it can be preparation actions, crime methods, weapon used, position of entry etc. Stop & search information contains data of a person or vehicle whenever they are stopped for a specific reason, this information holds the name, address of the person along with car number plate, location and reason of the stopped and search and details of the duty officer.

For this work, we have used crime reports consisting of 164800 record cases of five different types of burglary along with other crime entities such as nominals, modus operandi description and stopped and search from the provided dataset to generate the linked data. Twelve modus operandi variables each having a set of pre-defined values were used for each of the crime cases.The information about the role (e.g., the victim, suspect, the defendant , and witness) of each person associated with the crime case along with their names, location, date of birth etc. was also used for making linked dataset.

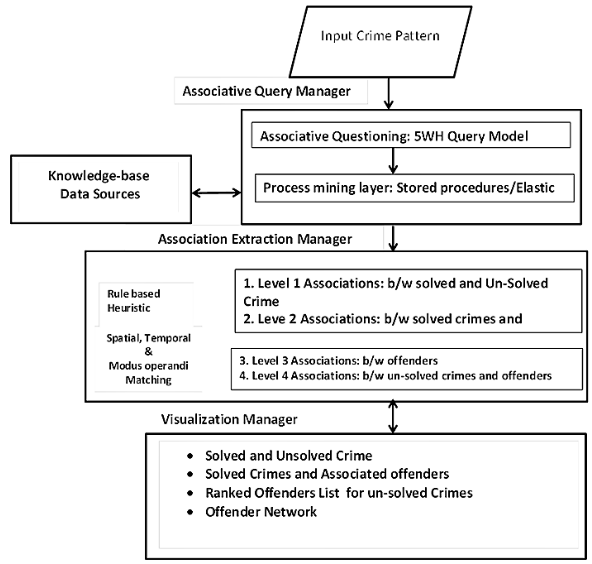

Figure 1. Visualization Scheme for Association Extraction.

4. Methodology

We now describe our adopted methodology for eliciting and visualization of the hierarchical associations. We have followed a layered architecture shown in Figure 1 consisting of Associative Query manager that performs process mining based on associative questioning derived from 5WH model, the association extractor that elicits the associations from data generated from the associative questioning layer and a visualization manager that visualizes the hierarchical associations uses iconic graphic and chain structure.

4.1. Associative Search Query Manager

The associative search query manager utilizes a 5WH search model which is described in our last research [17]. The Associative search 5WH model Figure 2 is based on associative concepts linked with each other through a set of properties or attributes. We employed these concepts in associative questioning to reveal the relationship, association or relevancy among these crime entities.

Associative search, unlike keyword and semantic based search, searches along the networks of associations between objects such as people, places, other organizations, products, events, services, and so forth. Some researchers [18] and [19] have related it with the human memory activation process.Reference [20] has implemented the concepts of association over mobile call data. He has visualized this linked based search through a semantic network of mobile call data including GPS information, SMS, picture viewer, MP3, charging and photo tagging. References [21], [22] and [23] however have used statistical and ontology measures to determine the associations from textual data.

Our proposed associative search is based on the cognitive thinking process consisting of associative questioning based on Crime Triangle and the Routine Activity Theory [24]. Working towards group detection, we have elicited temporal-spatial and behavioral (modus-operandi) associations for identifying the co-offender network and plausible suspect lists. The plausible suspect list groups suspects for a given unsolved crime, whereas the co-offender network groups the offenders/suspects/perpetrators for a given offenders linking temporal-spatial and modus operandi associations. We now present the associative questioning for extraction of these two groups.

4.1.1. Associative queries for extraction co-offender network:.

- What are those crimes in which two or more offenders have committed crime together i.e. their names have appeared together in a crime report.
- Where two or more offenders have committed crime/s together or separately?(spatial similarity).
- What are the temporal information of the crimes committed by two or more offenders?(temporal similarity)
- How two or more offenders have committed offenses in same area together or separately? (modus operandi similarity)
- Who are those offenders who have committed similar crimes using exactly the same MO?
- Who are the offenders who have committed similar offense using combination of the MO?

4.1.2. Associative queries for establishing associations between un-solved crimes and offenders:.

- What are other solved crimes,that are similar to crime pattern of the given unsolved crime?
- What are spatial,temporal and modus operandi similarity of those solved crimes and given unsolved crime?
- Who are the offenders of those solved crimes?
- How many similar solved crimes did an offender has committed in the past.
- What is the level of similarity between solved crimes and unsolved crimes?

These associative questions are then converted in to query language in process mining layer to retrieve the relevant data and then forwarded to the extraction manager which is discussed in the next section.

4.2. Association Extraction Manager

We establish the associations among crime entities through two basic principles, one is based on the domain expert knowledge which is encapsulated as a heuristic rule and another one is based on the associations criteria given by Scottish philosopher David Hume. He has given three principles of associations which are resemblance contiguity

Figure 2. Associative Search 5WH Model.

in time and place and causality. References [10] also has emphasized the use of modus operandi to determine the association with a chronic offender. It has also been shown that after a criminal gets used to a certain method of operation, he/she will use the same modus operandi again in committing other cases. The modus operandi thus can be used a signature of the perpetrators.

Therefore in addition to the spatial-temporal information, we also have included modus operandi component of the crime pattern as similarity matching criteria for establishing associations for crime matching process. We extract multi-level associations based on the heuristic rules and spatial-temporal and modus operandi similarity of the crime as shown in the Figure 1.The first two level of associations is used for identification tasks,whereas The next two levels of associations are employed to identify groups who may be thought to connected based on the similarity matching.

4.2.1. Object Identification Tasks. It involves to identify crime objects including solved crime,unsolved crime and associated offenders through establishing associations between solved and unsolved crimes and that of between solved crimes and offender. The association of solved and unsolved crime employ heuristic rule that any crime for which the offenders is not known is a unsolved crime otherwise a solved crime.The association between solved and offender is based on the heuristic that a offenders who has committed a crime is the associated offender of that crime,we also defined the role of associated offender such as defendant, victim suspect to it.

4.2.2. Group Identification Tasks.

The last two levels includes associations between offenders and un-solved crimes and that of between offenders.These associations are extracted through employing heuristic rule and crime pattern matching. Both of these types elicit the group of perpetrators;the former generate a plausible suspect list for an unsolved crime, whereas the later extract a co-offender network which are described in detail in the next section.

4.2.3. Plausible Suspect list through Associations between unsolved crime and offender.

We extract the plausible suspect list for a given unsolved crime employing similarity matching of crime pattern. We resolved the crime pattern of un-solved crime in to its modus operandi, temporal and spatial component and compared each of this component to that of the suspect's crime pattern components. This similarity matching process through associative questioning (described earlier) has generated group of plausible suspects for each component of the crime pattern.Each member of this list are those offenders who have committed at least one crime having similarity in one or all the components of given unsolved crime.In addition to this,each member is also ranked based on the similarity score of their committed crime with that of given unsolved crime. The similarity score was measured through cosine similarity for each of the crime component.

Mathematically Let C_1, C_2, C_3, C_n be the solved crimes committed by suspects that have shown the similarity to the each of the component of the given unsolved crime. Lets also assumed that the S_1, S_2, S_3, S_nbe the corresponding similarity scores of the solved crimes to these unsolved crimes. Lets $N_1, N_2, N_3 N_n$be the full names of the suspects who have committed any or all of the solved crimes. if suppose N_1 has committed the crimes $C_1$and C_2, N_2 has committed the crime $C_2, C_3 and N_3$ has committed the crimes $C1, C2, C3$,the rank of each of these suspects in their respective group ,then can be calculated using following equation:

Rank of N_1= $S_1 * C_1 + S_2 * C_2$
Rank of N_2= $S_1 * C_2 + S_3 * C_3$
Rank of N_3= $S_1 * C_1 + S_2 * C_2 + S_3 * C_3$.

Now suppose if $S_1 < S_2 < S_3$ which means that given unsolved crime is more similar to the C_3 and least similar to the C_1 , then based on the similarity score though N_1 has appeared in two crimes so is the N_2 but since C_3 is more similar to the given input crime due to high value of the S_3 hence it will rank high on plausible suspects list.

A suspect in plausible list may have similarity in one or all of the components of the given crime pattern. In other words a crime having same similarity score may have different combination of similarity in various component of the crime pattern. In order to depict this situation we also calculated an adjacency matrix of order nXm with n numbers of suspects and m numbers of the crime attributes of the crime pattern. This adjacency matrix records the similarity of various attributes of the crime pattern of the member suspect to that of unsolved crime pattern. The top row of this matrix contain all the attribute of a crime pattern i.e. spatial component including district and street, temporal component i.e. time of event and lastly all the elements of modus operandi. The first column of this matrix contains name of all the suspects in the list. We tag each cell of the matrix corresponding to each suspect and the crime pattern component either 0 or 1 to record the presence or absence of the similarity. Mathematically if we represent each cell of this matrix as S_{ij} then S_{ij} =0 when for absence of any similarity and its equal to 1 for presence of the similarity.

4.2.4. Associations between offenders and offender.

The similarity matching of the crime pattern among the offenders generate interesting associations between offenders ,we have named it as co-offender network and have used heuristic rule to identify it. When two or more offenders exhibit similarity in their crime pattern in any of spatial ,temporal and modus operandi component of the crime pattern, we detect these associations between them and group all these offenders tagging it as spatial, temporal and modus operandi (STM) network.This network is constructed on the basis of following associative queries:

- Who are the offenders who have committed similar or different crimes in the similar area together or separately?
- Who are the offenders who have committed similar or different crimes within same time however at other location?
- Who are the offenders who employ similar modus operandi or a combination of MO to commit a crime?
- What is the level of similarity (i.e. similarity score) of those solved crimes to the unsolved crimes?

Graph theory defines network as a set of ordered pairs of two or more entities called as nodes, linked through some connections known as edges of the network. We employed graph theory to structure the STM network as nodes of offenders connected through nodes of similarities containing crimes locations ,modus operandi elements and time of event. the edges between these nodes thus represent the type of spatial-temporal and modus operandi similarities. Two or more offenders in the STM network are thought to be connected with each other if they share any or all of three similarities.Two offenders however, may also connect indirectly if they exhibit any of these similarities with a common offender.Therefore like plausible suspect list,we also calculated an adjacency matrix of order nXn to depict the connection of an offender with all of the other offenders in the network. This co-offender network thus may be useful to analyst in hypothesis formulation in identifying offenders who may be working in groups.

5. Visualization of plausible suspect list and co-offender network

According to visualization literature nodes and links in a tree can signify relations among objects without the constraint of mapping variables onto multi-dimensional axes.

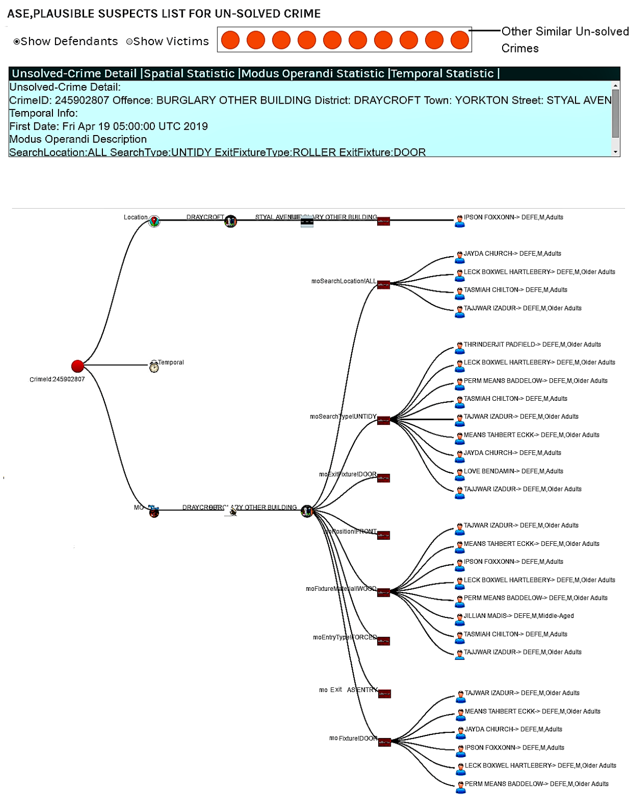

Figure 3. Knowledge Graph:Plausible Suspects List for Un-solved Crime

Thus, a tree has the potential of using space more effectively than a map. Consequently we present the associations relationship between the crime entities through a tree structure displaying parent-child relationship. The tree weaves a hierarchical multi-level knowledge graph showing linkages between a chosen crime entity such as unsolved crime or an offender and related other offenders/suspects. The related information of each of the suspect such as age group, ethnicity and gender is also presented. The root and child nodes of the knowledge graph are represented through iconic graphics while thick lines are used to show the relationship.Each of the node of the graph is collapsible and expandable which means a user can click a node of interest to view its underlying children while closing any other node so that only relevant/desired information is placed on the screen.In addition to this all the associated unsolved crime to the given unsolved crimes are presented in orange circles so that user can also view suspect list of other associated unsolved crime.

In the Figure 3 the root node for plausible suspect list represents the unsolved crime represented through red circle.The root node(selected unsolved crime) is branched into three child nodes each for the spatial temporal and modus operandi component. The spatial component is further resolved up to three levels narrowing down the similarity of the spatial information over district, street and offense. All suspects who has committed the crime in similar district and street are presented using iconic graphic at fourth level of the graph.Likewise the temporal node of the knowledge graph has its underlying three levels consisting of the date district and offense type details of the unsolved crime.The fourth level as before groups the entire suspect who has committed the similar crime in the same district with one week before and after the give date.Lastly the Modus Operandi node of the knowledge graph compares the behavioral component of the unsolved crime with that of suspect pattern and in its underlying three levels it narrows down the crime pattern for each element of the modus operandi with district and type of the offense.The fourth level as before shows suspects along with their gender ethnicity and age group.

We now present an estimation of number of queries needed to weave knowledge graph of the plausible offender list.The total number of queries needed would be equal to $Q_1 + Q_2 + Q_3 + Q_4$; Where Q_1 is query required to extract the spatial-temporal and modus operandi detail of the unsolved crime for similarity matching,while Q_2, Q_3, Q_4 are the queries need to extract the suspects for the spatial,temporal and modus operandi component of the crime pattern. However each of the spatial ,temporal and modus operandi queries in turn also need further sub queries corresponding to the their child nodes.for example district node of spatial node may have in turn contain several numbers of branches corresponding to child streets and grandchildren offenses within the vicinity of a street where the unsolved was occurred.The number of the queries required to complete spatial node can be thus be multiple of "n" number of the distinct streets in the district. This will extract the suspects for "n" unique combination of district street and offense. Likewise the temporal node may have "m" distinct district and will execute "m" number of queries for temporal component.The Modus operandi node execute queries corresponding to each of its element used in the unsolved crime pattern. For example if unsolved crime has modus operandi information for all of its 12 elements as defined in dataset section than 12 queries would be needed to weaved this part of knowledge graph. In addition to this each of this modus operandi component may contain data for several districts "D" so the total number of queries for the whole knowledge graph is given by Total Queries=$Q_1 + n * Q_2 + m * Q_3 + \sum_{mo=1}^{12} mo_i * D_i$.so if we say put n=4 ,m=4 and D=3 then total number of the queries would be 45.Thus in total to visualize the plausible suspects list with their role and other related information such as age group gender ethnicity an analyst with a moderate normalize database design has to query 45 times for each associated suspects.

In our preliminary evaluation of this visualization widget , our end user has mentioned that It usually takes about a day or two by the analyst to collect this data without any proper visualization.however our Association extraction does this job within one or two minutes with proper visualization.If analyst finds an interesting suspect in the list he may wish to know more about him particularly if he is working with any

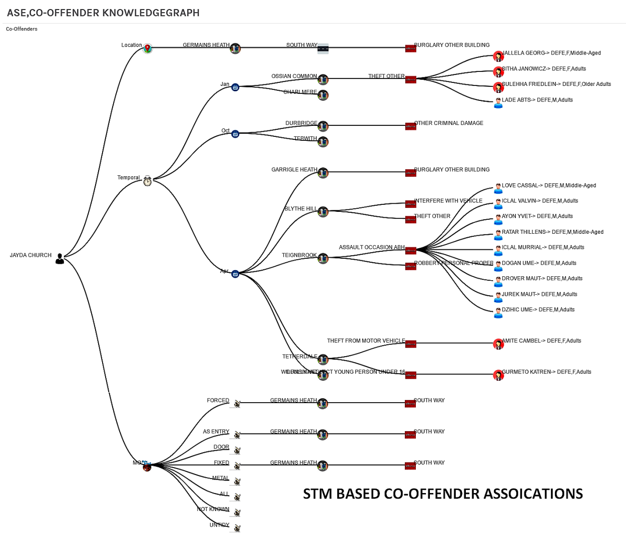

Figure 4. Knowledge Graph:Co-offender Network of a offender/Suspect

network or who could be his possible co-offender/s, this is where our proposed co-offender widgets comes in. this widget as described earlier groups the spatial,temporal and modus operandi similarities of the offenders. On clicking on any of suspect its co-offender network visualized following the similar chain structured is displayed as shown in Figure 4

6. Preliminary Evaluation

We conducted a preliminary evaluation with our police analyst end-users to elicit subjective feedback on our proposed associative search mechanism for the suspect list and co-offender network visualizations. The evaluation involved five qualitative focus groups of analysts from a different police organization. The prototype was demonstrated, illustrating the visualization for different association tasks to each of the focus group. Each group had 45 minutes for the demonstration and feedback. A separate group of observers recorder notes, ideas, and feedback from the end-users. We now report on the feedback as recorded by the observers, based on five questions:

- **Question 1: What is the use and value of knowledge graph (suspect list and co-offender network) to the user?**
  The end-users found that, the presented tools enables them to quickly discover the possible suspects for given un-solved crime.It has high potential of saving their time for collecting the relevant suspects list.
- **Question 2: Does proposed visualization help them to find the associations.?**
  The end users graded this as a perfect tool for their work.they liked the idea of splitting the similarity of crime pattern in three components and its tree based visualization.it gives quick retrieval of a large spectrum of possible suspects and then co-offender network gives them opportunity to generate and test new hypothesis.
- **Question 3: What features or functionalities would End Users like added, changed or removed?**
  The analysts wish that there should be an option to combine the modus operandi similarity with location and temporal and there should be a group in visualization tree that should display all the suspect who are visible in all of the three components.They also wish that those suspect who are in jailed should be highlighted in the suspect list. It would also be useful if dotted lines are used to show the associations rather than solid lines.
- **Question 4: Have they experience such a tool previously?**
  The end user think that they have not used such visualization tool for extracting suspect list and co-offender network in their existing criminal analysis practice.Some however think it has some resemblance to IBM i2.
- **Question 5: Overall, is the End User groups assessment positive, negative or neutral?** All the analysts found that the different aspect of the visualized associations could add value to what they are currently doing to make more effective decision making for crime matching process.

7. Conclusion

In this work we have presented two widgets (plausible offender list for unsolved crime and co-offender network) to assist crime and expedite crime matching process.These two widgets are developed utilizing multi-level association based on associative questioning and 5WH questions to expedite a crime matching process.The key to this research is the belief that there exists possible relationships within the various dataset used by the a police analyst, and simple visualization of these associations can be helpful for analytical reasoning during a crime matching process. The important characteristic of the presented visualization prototype is that it enables analyst to make assessment rather than recommendation.Our proposed approach of using temporal, spatial and behavioral pattern of a crime scene to find the association among the crime entities has never been demonstrated before in the crime investigation literature. The police analyst has given a positive feedback indicating that this prototype have potential to improve the efficiency of the crime investigation process. As a part of future work,we have planned to measure the accuracy and speed of the proposed visualization through a detailed evaluation.

Acknowledgment

The research leading to the results reported here has received funding from the European Union Seventh Framework Programme through Project VALCRI, European Commission Grant Agreement N FP7-IP- 608142, awarded to Middlesex University and partners.

References

[1] M. R. Keyvanpour, M. Javideh, and M. R. Ebrahimi, “Detecting and investigating crime by means of data mining: a general crime matching framework,” *Procedia Computer Science*, vol. 3, pp. 872 – 880, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050910005181

[2] D. G. C. Oatley, P. J. Zeleznikow, and D. B. W. Ewart, “Matching and predicting crimes,” in *In Proceedings of the Twenty-fourth SGAI International Conference on Knowledge Based Systems and Applications of Artificial Intelligence (AI2004*, 2004.

[3] B. W. Wong and N. Kodagoda, “How analysts think,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, pp. 178–182, 2016. [Online]. Available: http://dx.doi.org/10.1177/1541931213601040

[4] D. Kahneman. Penguin Books, London, England, 2012, p. 512.

[5] R. V. Hauck, H. Atabakhsh, P. Ongvasith, H. Gupta, and H. Chen, “Using coplink to analyze criminal-justice data,” *Computer*, vol. 35, no. 3, pp. 30–37, Mar. 2002. [Online]. Available: http://dx.doi.org/10.1109/2.989927

[6] Y. Xiang, M. Chau, H. Atabakhsh, and H. Chen, “Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list,” *Decision Support Systems*, vol. 41, no. 1, pp. 69 – 83, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923604001125

[7] H. Chen, D. Zeng, H. Atabakhsh, W. Wyzga, and J. Schroeder, “Coplink: Managing law enforcement data and knowledge,” *Commun. ACM*, vol. 46, no. 1, pp. 28–34, Jan. 2003. [Online]. Available: http://doi.acm.org/10.1145/602421.602441

[8] L. Ding, D. Steil, M. Hudnall, B. Dixon, R. Smith, D. Brown, and A. Parrish, “Perpsearch: An integrated crime detection system,” in *Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics*, ser. ISI’09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 161–163. [Online]. Available: http://dl.acm.org/citation.cfm?id=1706428.1706456

[9] D. E. Brown, “The regional crime analysis program (recap): a framework for mining data to catch criminals,” in *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, vol. 3. IEEE, 1998, pp. 2848–2853.

[10] J.-H. Wang and C.-L. Lin, “An association model based on modus operandi mining for implicit crime link construction,” in *Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining*, ser. ASONAM ’11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 548–550. [Online]. Available: http://dx.doi.org/10.1109/ASONAM.2011.34

[11] J. J. Xu and H. Chen, “Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks,” *Decision Support Systems*, vol. 38, no. 3, pp. 473–487, 2004.

[12] J. Schroeder, J. Xu, H. Chen, and M. Chau, “Automated criminal link analysis based on domain knowledge: Research articles,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 6, pp. 842–855, Apr. 2007. [Online]. Available: http://dx.doi.org/10.1002/asi.v58:6

[13] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, “Crime data mining: A general framework and some examples,” *Computer*, vol. 37, no. 4, pp. 50–56, Apr. 2004. [Online]. Available: http://dx.doi.org/10.1109/MC.2004.1297301

[14] C. C. Yang and T. D. Ng, “Terrorism and crime related weblog social network: Link, content analysis and information visualization.”

[15] R. Al-Zaidy, B. C. Fung, A. M. Youssef, and F. Fortin, “Mining criminal networks from unstructured text documents,” *Digital Investigation*, vol. 8, no. 34, pp. 147 – 160, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287612000023

[16] S. Lin and D. E. Brown, “An outlier-based data association method for linking criminal incidents,” *Decision Support Systems*, vol. 41, no. 3, pp. 604 – 615, 2006, intelligence and security informatics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923604001344

[17] N. Qazi, B. W. Wong, N. Kodagoda, and R. Adderley, “Associative search through formal concept analysis in criminal intelligence analysis,” in *2016 IEEE International Conference on Systems, Man, and Cybernetics SMC 2016 — October 9-12, 2016 Budapest, Hungary*, 2016, pp. 1917–1922.

[18] M. LLC. Associative search and the semantic web: The next step beyond natural language search.

[19] J. G. Raaijmakers and R. M. Shiffrin, “Search of associative memory.” *Psychological review*, vol. 88, no. 2, p. 93, 1981.

[20] K. Oh and S.-B. Cho, “Semantic networks of mobile lifelog for associative search based on activity theory,” in *Proceedings of the 11th Pacific Rim International Conference on Trends in Artificial Intelligence*, ser. PRICAI’10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 643–648. [Online]. Available: http://dl.acm.org/citation.cfm?id=1884293.1884360

[21] K. Qian, S. Hirokawa, K. Ejima, and X. Du, “A fast associative mining system based on search engine and concept graph for large-scale financial report texts,” in *Information and Financial Engineering (ICIFE), 2010 2nd IEEE International Conference on*. IEEE, 2010, pp. 675–679.

[22] K. Viriyayudhakorn and M. Ogawa, “Associative search on shogi game records,” *Information and Media Technologies*, vol. 6, no. 3, pp. 833–845, 2011.

[23] I. F. Moawad, M. Abdeen, and M. M. Aref, “Ontology-based architecture for an arabic semantic search engine,” in *The Tenth Conference. On Language Engineering Organized by Egyptian Society of Language Engineering (ESOLEC’2010)*, 2010, pp. 15–16.

[24] L. E. Cohen and M. Felson, “Social change and crime rate trends: a routine activity approach,” *American Sociological Review*, pp. 588–608, 1979.

[25] H. Sarvari, E. Abozinadah, A. Mbaziira, and D. Mccoy, “Constructing and analyzing criminal networks,” in *Proceedings of the 2014 IEEE Security and Privacy Workshops*, ser. SPW ’14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 84–91. [Online]. Available: http://dx.doi.org/10.1109/SPW.2014.22