Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: https://www.tandfonline.com/loi/tbit20

Sense-making strategies in explorative intelligence analysis of network evolutions

Johanna Doppler Haider, Bettina Gastecker, Margit Pohl, Patrick Seidler, Neesha Kodagoda & B. L. William Wong

To cite this article: Johanna Doppler Haider, Bettina Gastecker, Margit Pohl, Patrick Seidler, Neesha Kodagoda & B. L. William Wong (2019) Sense-making strategies in explorative intelligence analysis of network evolutions, Behaviour & Information Technology, 38:2, 198-215, DOI: 10.1080/0144929X.2018.1519036

To link to this article: https://doi.org/10.1080/0144929X.2018.1519036

Published online: 11 Sep 2018.

BEHAVIOUR & INFORMATION TECHNOLOGY
2019, VOL. 38, NO. 2, 198–215
https://doi.org/10.1080/0144929X.2018.1519036

Sense-making strategies in explorative intelligence analysis of network evolutions

Johanna Doppler Haider[a], Bettina Gastecker[a], Margit Pohl[a], Patrick Seidler[b], Neesha Kodagoda[b] and B. L. William Wong[b]

[a]Institute of Visual Computing and Human-Centered Technology, Vienna University of Technology, Vienna, Austria; [b]Computer Science, Middlesex University London, London, UK

ABSTRACT
Visualising how social networks evolve is important in intelligence analysis in order to detect and monitor issues, such as emerging crime patterns or rapidly growing groups of offenders. It remains an open research question how this type of information should be presented for visual exploration. To get a sense of how users work with different types of visualisations, we evaluate a matrix and a node-link diagram in a controlled thinking aloud study. We describe the sense-making strategies that users adopted during explorative and realistic tasks. Thereby, we focus on the user behaviour in switching between the two visualisations and propose a set of nine strategies. Based on a qualitative and quantitative content analysis we show which visualisation supports which strategy better. We find that the two visualisations clearly support intelligence tasks and that for some tasks the combined use is more advantageous than the use of an individual visualisation.

ARTICLE HISTORY
Received 30 April 2018
Accepted 28 August 2018

KEYWORDS
Sense-making; temporal network evolution; criminal intelligence; evaluation

1. Introduction

Crime groups operate in criminal ecosystems (Felson 2006) that form complex networks of inter-related and inter-dependent crime activities. They represent organic structures that can transcend local, regional, national and policing boundaries, and evolve over time and space (Archambault et al. 2014; Bach et al. 2015; Wasserman and Faust 1994) and often propagate their activities across different networks. Petty theft, e.g. pick pocketing of credit cards, and selling drugs, may be used to create opportunities for bigger, more profitable crimes such as credit card fraud or identity theft. These crimes in turn provide criminals with funding to become more adept at planning, executing and concealing further crimes. The intelligence analyst depends on understanding the internal dynamics of those networks (Robins and Kashima 2008). Therefore, it is essential to piece together a more complete picture of how the networks operate, how they are controlled, who controls them, and how information is communicated to plan, coordinate, execute and conceal their criminal operations.

Visualisation supports exploratory analysis and can additionally aid in communicating findings. However, a lack of appropriate visualisations suitable to transfer the given information correctly while providing the scalability that is needed for large dynamic networks can be identified (Bach et al. 2015). Current automated network analysis frameworks (Xu and Chen 2005) rely on Social Network Analysis (SNA) measures, which do not support the actual analytical tasks on understanding criminal behaviour efficiently, specifically in presenting the temporal dimension. Therefore, those systems do not cover current intelligence requirements, such as tracking crime type behaviour, number of crimes or re-occurring events in time. As a consequence, an operational gap between criminal network analysis and police operations remains (Johnson and Reitzel 2011), which possibly can be bridged by visualisations that assist in identifying relevant aspects of network developments without overwhelming the analyst.

Sense-making and insight provides a comprehension of a situation by the unconscious synthesis of prior knowledge and experience with newly collected data to create an unexpected, dramatic realisation. We think of insights as 'sudden unexpected thoughts that solve problems' (Hogarth 2001, 251), or an unexpected shift in the way we understand things (Klein et al. 2007) and sense-making as the process that leads to insights. Research on sense-making can inform the design of visualisations as, for example, an overview can be given by incorporating time steps in an integrated view instead of animating the data or splitting it up in multiple views (Khurana et al. 2011; Zhu, Watts, and Chen 2010), thus, supporting the construction of a mental model by reducing the cognitive load.

We conducted an empirical investigation to analyse the utility of two visualisations and, in addition, the sense-

CONTACT Johanna Doppler Haider johanna.haider@igw.tuwien.ac.at Institute of Visual Computing and Human-Centered Technology, Vienna University of Technology, Argentinierstraße 8, 2nd floor, 1040 Vienna, Austria

making processes the participants of the study adopted to solve realistic tasks. Both visualisations, a node-link (NL) diagram and a matrix (M is used in tables and figures), represent the same data. We developed these visualisation techniques for weighted networks based on the needs of intelligence analysts. An important aspect of our techniques is the use of weights for the links between co-offenders (two criminals who committed a crime together), which represent the seriousness of crimes in a specific time frame. Another important aspect is the representation of indirect (2nd degree) relationships of two offenders mediated by a third person, i.e. possible acquaintances, who can be involved in further criminal activities.

We were especially interested in how useful it is to use both visualisations in combination. We could show that matrices are superior for the representation of temporal developments and NL diagrams for structural tasks. In this way we create novel *starting points for the analysis* in form of interesting networks.

We also identified and analysed the sense-making strategies the participants of our study engaged in. To achieve this, we applied content analysis on top of thinking aloud protocols that were recorded from the participants during our study. The tasks the participants of the study had to solve were realistic and exploratory. We created the tasks in cooperation with experts during an extensive requirements elicitation phase of the goals of the intelligence analysts. The main contribution of the paper is this extensive analysis of users' sense-making processes while using information visualisation systems for intelligence analysis.

Hence, the contributions of this paper are twofold:

- A qualitative and quantitative analysis (N = 31) of two complementary visualisations, an NL diagram and a matrix, which outlines advantages and disadvantages, user preferences, and visualisation capabilities of each visualisation evaluated through different tasks. We specifically evaluate the representation of temporal developments in NL diagrams and matrices.
- A description of nine sense-making strategies of how thirteen participants gained insights including the frequency with which these strategies occur in realistic tasks.

2. Related work

Crime analysis is typically done with volume crimes, i.e. crimes that occur in quantity (e.g. burglaries or petty theft), represented by areas, days of the week or time of the day. Crime and crime pattern analysis includes hot-spot analysis, statistical process control charting, crime profiling and network analysis of offenders (Heuer and Pherson 2014; UK: National Policing Improvement Agency on behalf of the Association of Chief Police Officers 2008). Hot-spot analysis maps crimes to geographical areas, whereas statistical process control charting compares crimes across periods and calculates statistically significant differences.

Still, detection of crime patterns on top of large data sets remains a difficult task and makes volume crime typically difficult to solve. Here, the support of IT systems is especially appropriate. The investigation reported in this paper deals with one area of criminal intelligence analysis, that is the analysis of co-offender networks. Apart from the seriousness of crimes (weights of connections), analysts are interested in the temporal development of such networks. As yet, there is little research addressing this specific application area.

2.1. Visualisation of temporal evolution

Various methods exist for visualising temporal evolution. Timelines can be juxtaposed (small multiples approach), superimposed (stacked approach) and integrated, i.e. combined time steps are inseparable without changing the layout (Beck et al. 2014). Horizontally stacked timelines (Burch and Munz 2015), for example, can be used to colour-code time varying weighted digraphs to provide an overview of dynamic graphs. MultiPiles (Bach et al. 2015) is a hybrid between small multiples and a timeline with aggregated snapshots which are piled up to represent time intervals with little changes. NetEvViz (Khurana et al. 2011) uses a time slider to select two time points in a network and shows the differences in an NL visualisation. Visualisations including a matrix representation are the MatrixExplorer (Henry and Fekete 2006) and Nodetrix (Henry, Fekete, and Mcguffin 2007). Both juxtapose them with NL diagrams. NL diagrams with multivariate edges (Ko et al. 2014) were presented as multiple threads (parallel coloured lines). However, juxtaposition uses a lot of screen space, and superposition increases the cognitive load on users. One of the few empirical studies addressing the cognitive processes involved to understand change over time showed that linked, juxtaposed views were more effective than each view on its own (Sedig, Rowhani, and Liang 2005).

To show temporal evolution of networks, most commonly a series of diagrams gets animated or is shown next to each other as small multiples (Archambault et al. 2014; Beck et al. 2014). Rufiange and Melançon's (2014) taxonomy of dynamic networks includes glyphs that show small summaries of the evolution of edges. They propose extended glyphs for the visualisation of multivariate edges (e.g. by using animation or stacked bar charts) and introduce AniMatrix, an animated matrix-based software evolution representation. To understand changes over time, mental representations need to be preserved, especially in animations, as we depend on memorising previous information (Archambault et al. 2014). This is problematic due to a high cognitive demand on the working memory and perceptual effects like change blindness (Nowell, Hetzler, and Tanasse 2001). Animation in general causes problems with short-term memory, therefore it should be used with care (Kriglstein, Pohl, and Smuc 2014). Analysts who investigate co-offender networks usually only compare a limited number of points in time. Therefore, we decided to use an integrated approach showing all the time-steps in one visualisation.

2.2. NL diagrams and matrices

NL diagrams and matrices have been compared several times to identify their advantages and disadvantages. Ghoniem, Fekete, and Castagliola (2004) used simple, generic tasks and found out that matrices are especially useful for larger, denser networks. Graphs are suited for path related tasks. Keller, Eckert, and Clarkson (2006) supported these results. In the domain of brain connectivity and comparison tasks Alper et al. (2013) could show that matrices are more effective for encoding edge weights than NL diagrams. Henry and Fekete (2007) developed MatLink, a hybrid tool consisting of NL diagram and matrix, combined the matrix with links overlaid on its border. In an evaluation they showed that MatLink is superior to both. Alper et al. (2013) compared weighted NL diagrams and matrices and also found that matrix representations in general are more efficient than NL diagrams.

Those studies used generic and fairly simple tasks (e.g. find a node or identify a path between two nodes). For our study, we developed fairly complex and explorative tasks with intelligence analysts (Doppler Haider et al. 2017). One of the tasks, for example, requires the participants to identify groups of co-offenders whose criminal activity increases over time. Participants have to scan the visualisations repeatedly to find such groups and decide whether these groups are good examples even if there is some contradictory evidence (e.g. criminal activity decreases only in one of the time periods). One of our goals was to analyse whether such tasks also yield similar results as reported above.

2.3. Sense-making

Working with information visualisations has often been described as a sense-making process. One of the most well-known theories in this context has been developed by Pirolli and Card (2005). Nevertheless, this approach has scarcely been used for empirical research because of its fairly restrictive character. Klein, Moon, and Hoffman (2006a, 2006b) developed an alternative approach – the Data/Frame model. Klein's goal was to describe sense-making processes under naturalistic settings. The model assumes that people develop schematic representations called frames. Frames can be supported by the data, they can be questioned or rejected. Klein (2013) extended this model to explain how insights can be achieved. He developed the Triple Path Model that contains three possible ways to gain insights – *Connection, Contradiction* and *Creative desperation*.

Pirolli and Card's model (2005) and Klein's models (2013; Klein, Moon, and Hoffman 2006a, 2006b) are general approaches and have influenced the research in this area. Nevertheless, little research exists on analysing sense-making processes in detail. Sedig and Parsons (2013) provide a comprehensive framework for the analysis of interaction processes with visualisations. Their framework is based on a literature review. Pohl, Wallner, and Kriglstein (2016) and Reda et al. (2014) investigate interaction processes with information visualisations. The latter point out that the emphasis so far has been on the analysis of outcomes rather than on the sense-making process itself. Sedig et al. (2016) studied complementary interactions – that is, interactions that occur in conjunction with each other to improve the users' performance. Doppler Haider et al. (2017) used the Data/Frame model to analyse users' interaction log data. The goal of studying interaction processes is to clarify how users make sense of the information provided by visualisations. Sense-making can be analysed by studying either interactions or thinking aloud protocols. The studies cited above have primarily used interactions as source of analysis. The research described in this paper uses thinking aloud protocols. Another strand of sense-making research addresses storytelling with visualisations. Segel and Heer (2010) classified approaches and strategies used by news media to tell stories visually. The effectiveness of storytelling and its implications for designing visualisations remains an open research question.

In general, the sense-making processes of users of information visualisations are still not understood, but the above-mentioned studies provide first results in that area.

3. Requirements in intelligence analysis

This research is conducted as part of the project VALCRI that has the aim to create a Visual Analytics-based sense-making capability for criminal intelligence analysis by integrating a number of technologies into a coherent working environment. One of the aims is to investigate how design supports insight generation. A challenge in intelligence analysis lies in identifying and investigating networks aligned to police priorities to be able to focus on specific crime types. Analysts can start investigating a known offender and expand the network along the leads. In a strategic setting, in contrast, an analyst needs to extract and rank all networks of offenders with relevant activities as the network properties are directly related to the starting point they are created from; a strategy not feasible without specialised computational support. Current systems do not support this kind of network analysis. To identify and extract a priority-ranked list of networks we defined a prioritisation strategy and evaluated a prototype with domain experts (Doppler Haider et al. 2017; Seidler et al. 2016). Force priorities, for example, can imply that certain crime types get an increased weight compared to other crime types (e.g. drug related crimes).

An extensive requirements analysis was conducted with end-users from police forces in Belgium and the UK. One requirement is the visualisation of pairs (co-offender) or small groups committing specific crimes. Criminals who only work on their own are not represented in the network. The main goal is to see how highly relevant criminal activities, for example, violent crime, develop over time. Crimes that are seen as especially harmful have a higher weight than others. Hence, the visualisation should emphasise weighted crime types in a social network.

A specific characteristic of co-offender networks is the sparsity of the data (Adderley, Badii, and Wu 2008). It is obvious that offenders tend to keep their relationships secret, therefore the networks are typically not dense. In addition, intelligence analysts are primarily interested in the relationships of the individuals and the identification of small groups of offenders, not in larger groups. Consequently, the visualisations presented here are different to other network visualisations studied in information visualisation, e.g. social graphs (Wakita, Takami, and Hosobe 2015). A requirement from the end-users was to show the development of these networks over time. Again, in contrast to other network visualisations, our end-users are not interested in many points in time but only in the most recent developments (three to five points in time, e.g. years). In the co-offender networks described here development over time is only related to the relationships between the co-offenders (that is, the links between the nodes). This is because once two offenders have jointly committed a crime they will always have a relationship which analysts have to consider. Hence, the number of nodes is constant over time and only the number of edges increases. Another requirement was to present 2nd degree relationships, that is, relationships between two offenders which are not direct but mediated by a third person. Such relationships are highly relevant in criminal intelligence analysis because they can indicate possible connections between criminal gangs. In general, the goal of the visualisations is to provide an overview of criminal activities in a certain period of time to intelligence analysts.

4. User study

Our main research question is how to visually support analysts in their sense-making of network developments in criminal intelligence. This work is based on a pilot study with experts that showed acceptance and usefulness of a weighted network calculation of co-offenders, which is visualised using two techniques (Seidler et al. 2016). In a first step, we analysed the results of the user study in a qualitative approach using the emergent themes analysis (Doppler Haider et al. 2017). In this work we extend the analysis by a systematic quantitative and qualitative content analysis of the thinking aloud protocols. In the following we introduce the features from the visualisations to show how they can be used to solve seven realistic tasks.

4.1. Description of the visualisations

Both visualisations show the crimes committed jointly by pairs of co-offenders over the time span of three years, attempting to support the temporal analysis of large, weighted networks. A key difference to previous approaches is that they enable both, an overview of all and a detailed view on individual relationships over three aggregated time steps in an integrated view (compare Figure 1). Although NL diagrams and matrices have a long tradition, neither an NL diagram incorporating a temporal summary view within multiple parallel lines nor an adjacency matrix with mediated 2nd degree relations have been evaluated systematically yet.

The two visualisations address three requirements of criminal investigation: temporal development, weighted relationships and scalability.

4.1.1. Temporal evolution

The criminal activity over three years is comprised of the sum of crimes multiplied by the weight of the offence type, which is set by law enforcement agencies, depending on the current strategic priority. Time is represented in the NL diagram as three parallel lines and in the matrix as three bars in each cell. The NL diagram uses

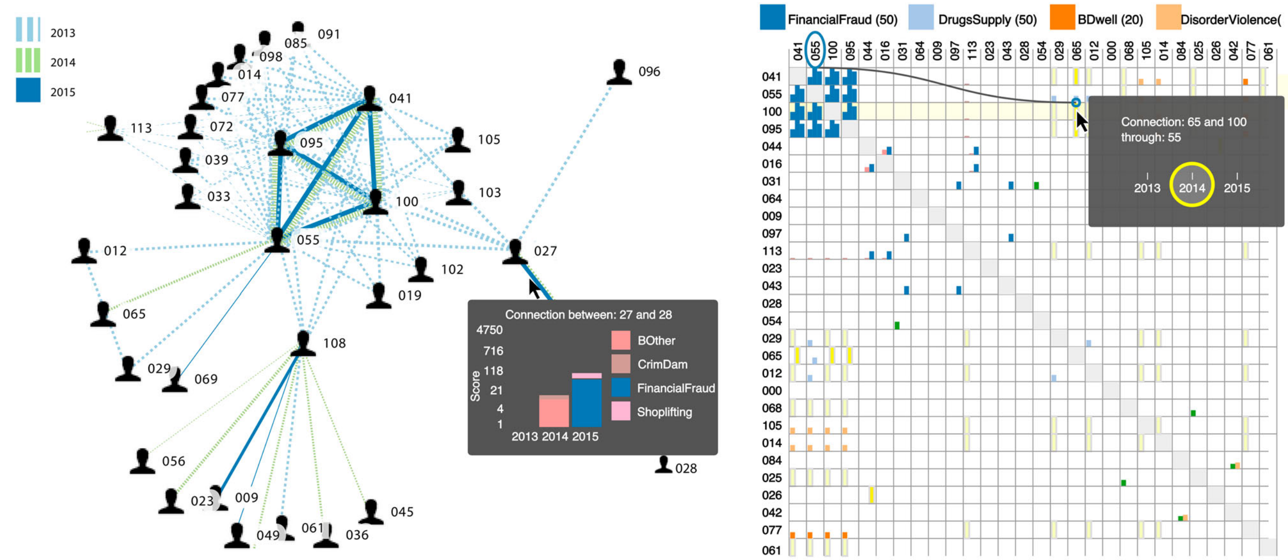


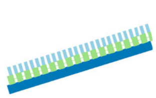
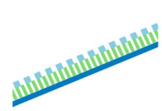
Figure 1. An evolution of 3 years shown in an NL diagram via multiple colour-coded lines and bar charts in the co-offender matrix (Doppler Haider et al. 2017) with details on demand (on hover). Yellow boxes indicate 2nd degree relations in the matrix.

different colours as well as different line styles for the multiple links. Because older relations are less important we use the metaphor of ageing, with older links being encoded in lighter colours. The multiple parallel lines represent the years. They are additionally encoded by colour because the layout algorithm does not ensure that all lines are unambiguously interpretable through a reading direction, e.g. left to right, because lines can be horizontally aligned as well as vertically. To support the ageing metaphor, the line style changes from continuous to dashed lines for older relations, see Table 1.

Hence, we encode the temporal information in the NL diagram as stacked coloured lines, like multiple threads (Ko et al. 2014), with an additional differentiation in the line style. Only the latest time interval is a continuous line, whereas the earlier years are dotted lines with decreasing dash lengths representing recency. We chose a colour-blind safe, 3-class-paired colour scheme recommended by colorBrewer (Harrower and Brewer 2003) that can be distinguished easily. A continuous colour scheme proved to be ineffective in practice because participants could not distinguish the different colour hues, especially, when the width of the lines differed considerably. In addition, it is difficult to unambiguously identify the hue in single lines, which occurs when the criminal relation is only in one year. Another reason for the double encoding of time in the NL diagram is that we are dealing with different link widths due to the weighting of crime types, see Section 4.1.2. Hence, some perceptual problems can arise. For example, a thin line can be hard to spot next to a stronger line. With the differentiation in the line style we emphasise the difference in time. We also think this has the same effect as to increase the transparency with link age. We acknowledge that the colour encoding can be a problem in combination with the crime types but we did not observe any influences on the tasks.

To show the temporal development of joint criminal activities in the matrix we use bar charts. The x-axis in the bar chart indicates the time (1st, 2nd or 3rd year respectively). To show the 2nd degree relationship yellow boxes are used. Time here is double-coded too. The boxes occupy the respective position on the x-axis and the yellow hue fades to a lighter hue for earlier years

Table 1. Encoding of time and weights in the NL diagram: Encoding of time by hue and line style per year from dashed (older) to continuous (most recent); Number of crimes by width.

Time	Number of crimes	Weighted crimes over time
Dashes + Hue	Width	Dashes + Hue + Width
		

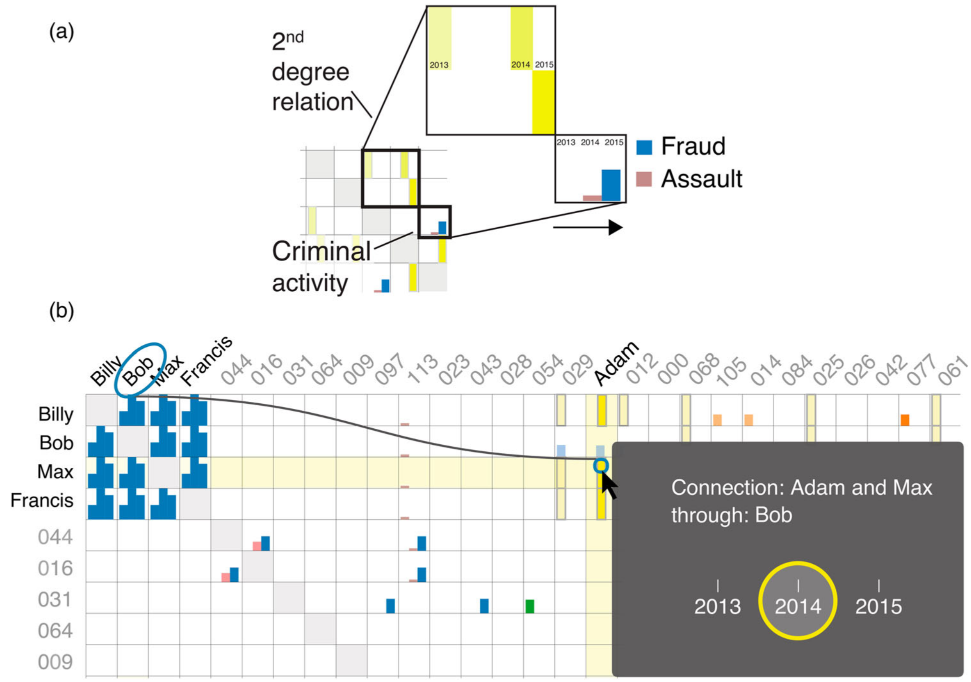


Figure 2. Encoding in matrix: Bar charts show three consecutive periods of convicted crimes. (a) The colours in each bar represent the type of crime; Yellow boxes indicate a 2nd degree relation. The time of occurrence is encoded by position and fading yellow hues. (b) Hovering on yellow boxes shows a line to the mediator and a pop-up. Adam is linked to Max through Bob, hence, likely connected to the fully connected main group (Billy, Bob, Max and Francis) from 2014 on.

(Figure 2(a)). On mouse-over a line between the mediator and the respective co-offenders and additional information related to the crimes is shown, see Figure 2 (b). This approach supports the perception of 2nd degree neighbours in matrix-based network representations.

4.1.2. Weighted relationships

The visualisations are supposed to provide an overview of the development of criminal activities. To show criminal activity the number of weighted crimes is encoded as the width of lines in the NL diagram and the height of the bar charts in the matrix. The type of crime is encoded by colour as this is a common practice in intelligence analysis. In our data set the emphasis lies on 'Financial Fraud' and 'Drug Supply', having assigned the highest weights.

The offender's criminal activity equals the sum of all co-offences multiplied with the codified offence type score. Prioritised crime types, therefore, are emphasised with thicker line widths in the NL diagram. The sum of scores for each offender over all time periods determines the position of the actors. This moves the focus of the investigation to the weighted relations showing up further in the top left corner of the matrix and in the centre of the graph the more prioritised they are (which is financial fraud in this case).

The matrix visualisation is inspired by the TimeMatrix (Yi, Elmqvist, and Lee 2010) visualisation, using intra-cell bar charts to visualise time, see Figure 2. The colours in our stacked bar charts indicate different crime types, but our data set includes only a couple of co-offenders committing crimes of different type in one period, e.g. 'Shoplifting' and 'Financial Fraud' in 2015, see Figure 1. We extended this concept by encoding new 2nd degree neighbours in the matrix with yellow boxes when a relationship begins. Hence, if two offenders have not committed a crime together but with a person who has relations to other criminals, a 2nd degree relation is represented by a yellow box in the cell at the intersection of the respective row and column.

4.1.3. Scalability

The choices made for this design do not fit to every scenario, because they are based on the requirements analysis in the law enforcement agencies where scalability is only an issue with regard to the network size. More than three time-steps are not relevant for law enforcement agencies. Especially because of this small amount of three time-steps we chose an integrated view to support analysts to gain a quick overview.

We limit our system design to basic principles with few interaction possibilities in this first stage of the evaluation to focus on sense-making and insights generated by the representations. Supported interactions are panning (click & drag), zooming, hovering and dragging nodes.

Both representations provide a detailed view of the temporal development in a pop-up layer when hovering over the co-offences, including actor information and the crime types.

4.2. Research questions

We formulated a comprehensive research question RQ1 based on the Triple Path Model of Insight (Klein 2013) and previous work on sense-making in intelligence analysis (Haider et al. 2015), and research questions RQ2-RQ5 on the basis of the requirements in intelligence analysis.

The main goal of the investigation is to clarify how intelligence analysts can be supported by a system presenting co-offender networks. This has two main aspects: the usage of sense-making strategies and the utility of the system. Research question RQ1 covers the sense-making aspect, research questions RQ2-RQ5 the aspect of utility. The utility aspect is about the specific characteristics of the visual analytics system that are supposed to help analysts to gain insights. We also assume that the analysis of sense-making strategies yields results that can help to design better systems. We also analyse how sense-making strategies are related to the utility aspects. The two visualisation techniques especially support the detection of relationships and temporal trends, therefore, we investigate whether the participants of the study engage in appropriate sense-making strategies (e.g. *Connections* and *Trend assessment*).

4.2.1. RQ1 – sense-making and insight generation

How does the visualisation design support insight generation? Which strategies do analysts use to generate insights?

To describe how a representation enables insights we want to observe participants in the tool use and compare applied strategies and performance in explorative tasks. The analysis of sense-making strategies can also provide information about how the visualisation system should be designed to support the analyst. We use a qualitative content analysis approach (Schreier 2012) for this analysis.

4.2.2. RQ2 – relationships

How does the visualisation design help understand co-offender relationships?

We want to investigate structural aspects of the two visualisations, e.g. links between nodes or the detection of groups.

4.2.3. RQ3 – weighted attributes

How does the visualisation design help understand co-offender activity?

To assess criminal activity, we focus on the number of accumulated crimes and the crime category, which are both expected to shift over time (e.g. criminals getting involved with other crime types). We look at the combination of multiple edges and edge weights in an NL diagram and stacked bar charts in a matrix and evaluate the two visualisation techniques with complex, explorative tasks.

4.2.4. RQ4 – temporal development

How does the visualisation design help understand co-offender evolution?

We think that for a small number of timesteps summary views that integrate or superimpose the temporal information, e.g. by encoding recency by transparency (older fading out) or layer ordering (newer on top of older ones) is appropriate and want to test this assumption.

4.2.5. RQ5 – visualisation capabilities

What are the capabilities of different visualisations?

Previous research (Ghoniem, Fekete, and Castagliola 2004; Henry and Fekete 2007) indicates that matrices are superior to graphs for most generic tasks. We want to determine whether this also holds for domain related, explorative tasks. We use the analysis of swaps between the two different visualisations to analyse which visualisations are predominantly used for which tasks. Each visualisation runs full screen; therefore, participants have to switch from one visualisation to the other explicitly. This makes it easier to analyse swaps and provides us with a more objective measure of preference for the visualisation type compared to post-test questionnaires. In addition, we analyse the results of the tasks to check whether participants arrived at correct or plausible answers and which visualisations they used. We also want to analyse whether the combination of the two different visualisations has specific advantages for the complex, explorative tasks we use.

An overview of the research questions is given in Table 2.

Table 2. Research questions of the user study.

Research Question
RQ1 Sense-making: How does the visualisation design support insight generation? Which strategies do analysts use to generate insights?
RQ2 Relations: How does the visualisation design help understand co-offender relationships?
RQ3 Weights: How does the visualisation design help understand co-offender activity?
RQ4 Time: How does the visualisation design help understand co-offender evolution?
RQ5 Visualisation: What are the capabilities of different visualisations? Can weaknesses be overcome by new visualisations?

4.3. Tasks

To simulate a realistic setting with non-experts we developed appropriate tasks based on expert requirements and checked them with the end-user partners/experts who assured us that they reflect typical activities of criminal intelligence analysts. The tasks were chosen to be realistic and are, therefore, rather complex and time consuming. By using realistic tasks, we wanted to elicit typical sense-making strategies from the participants of the study. We recruited non-experts as it was impossible to recruit experts who have background knowledge, experience and training in large quantities for this kind of rigorous testing.

We asked participants to start with looking for increased criminal activities on a pairwise level *over time* (task T1), which concerns the forth research question, and continue on a cluster level for growth of harm for the community via *weighted crimes* (task T2), which concerns the third research question. With task T3 and task T4 we address the second research question and the offender *relationships* and similarities. We ask the participants, for example, to identify well-connected actors in the network (task T4). Characteristics of individual actors can show up relationships too. In the context of crime analysis possible substitutes are of interest. If a person is geographically co-located and has a similar profile as a person targeted by law enforcement, then this person could simply replace the target once the target is arrested (task T3). Task T5 and task T6 address the research questions RQ3 and RQ4 by looking at changing *weights over time*. This can, for example, occur as a result of mitigation strategies of police forces or of changing priorities. In both cases this should be displayed through indicators going down. Finally, we are interested in the visual characteristics of the representations (RQ5) and ask for patterns and interesting observations in task T7. An overview of the tasks and addressed research questions is given in Table 3.

Table 3. Tasks of the user study.

Task	Description	Research question
T1	Identify co-offenders whose criminal activity increases over the given time period.	RQ4
T2	Identify a group (three or more offenders) whose criminal activity increases or crime type worsens over the shown time period.	RQ3
T3	Identify connectors, i.e. actors that connect two clusters, between groups and their possible successors, if the connector is removed from the network.	RQ2
T4	Identify a remarkably well-connected offender who has many 2nd degree connections in the last year.	RQ2
T5	Identify two problematic crime types that cause most problems in the network.	RQ3
T6	What seems to be the overall tendency of the network in terms of crime activity and harm to community?	RQ4
T7	Are there any patterns in the network or did you make any interesting observations?	RQ5

5. Evaluation

The aim of this evaluation on the one hand is to assess the suitability of the representations in relation to the tasks and, on the other hand, to identify the applied sense-making strategies while using the visualisations. We conducted a controlled experiment to investigate suitability and sense-making on a general level, based on a pilot study with domain experts (Seidler et al. 2016).

We used a within-subject experimental thinking aloud design with seven tasks and two techniques (NL diagram and M). Participants could decide on the technique they wanted to use, which provides insights into user expectations and preferences for each of the visualisations. There was no time limit. On completion of the tasks, we captured demographic information, such as, sex, age, perceptual issues (visual impairment), and familiarity to visualisations in general. Finally, we interviewed the participants about preference and utility of the visualisations in a semi-structured way.

5.1. Procedure and setting

We collected thinking aloud protocols with audio and video capture of the screens. The experimenter further took field notes during an observation. Participants were informed about and asked to sign a confidentiality agreement. Since all were new to the domain and visualisations they could gain experience in a training task before they solved seven explorative tasks (Table 3). For the training we asked participants to look up a specific offender and his/her direct and indirect neighbours to get familiar with the concepts and tool use. This training period lasted approximately 10 min.

The tasks were printed on a sheet and handed to the participants. The participants could work in their own pace and refer to the task description at any time. We asked them to start with task T1, read one description at a time and follow the tasks in the given order. All participants got the tasks in the same order. For the first task we randomly opened one visualisation, reminding the participants that they can swap visualisation as they wished at any time. On average, the sessions lasted 42 min and the follow-up interviews 5 min. Including the introduction and training, the experiment never lasted longer than one hour.

Participants used a 24-inch monitor with a resolution of 1920 × 1200 pixels and an aspect ratio of 16:10 (model

Dell U2415). The visualisations were shown in full screen to present a maximum of the data set but panning and scrolling was necessary in zoomed-in perspectives. A standard keyboard and scroll wheel mouse were provided to operate the visualisations. Quick swapping was enabled via keyboard shortcuts. Instructions were attached to the bottom of the screen.

5.2. Data analysis and coding

We conducted a qualitative content analysis (Schreier 2012) of the participants' verbal reports during seven explorative tasks. We used the codes from the previous analysis (Doppler Haider et al. 2017) as a starting point. The codes are partly based on Klein's Triple Path Model – a top-down approach (Klein 2013) and partly on codes resulting from previous research in a bottom-up approach (Doppler Haider et al. 2017). Based on that three researchers adapted the codes to the material at hand. The analysis included the creation of keywords for coding, counting the appearances of codes per visualisation and task and finally, comparing the numbers with the correctness of given answers.

While one researcher coded the statements from the thinking aloud protocols and counted them per visualisation, a second researcher checked them independently and coded one task per participant at random as practiced by Kodagoda et al. (2017). Table 5 gives an overview of the number of statements falling within each category. In Section 6 we discuss each category as a sense-making strategy in more detail. Exemplary quotes were translated from German. Due to technical reasons this analysis was done with thirteen protocols (N = 13). The quantitative analysis of the user interaction, i.e. completion times and swapping behaviour, as well as the user preference from the interview is reported from all participants (N = 31).

5.3. Data set

We used a social network with 121 nodes and 996 edges with a 3% edge density that was sanitised from original police data, containing data on victims and offenders, crimes and intelligence of three years (Seidler et al. 2016). It comprises 20 high-level crime types aggregated from the full set of 839 offence types, which are prioritised by a weighting score for more or less harmful crimes. Figure 1 shows how both visualisations represent approximately one-third of the given network.

5.4. Participants

We recruited 31 students (18 male, 13 female) from two Universities in Vienna and London, aged between 24 and 34 years (mean age 26.51) to participate in the experiment. Of the 31, 24 had basic to moderate knowledge in visualisations, and the remaining 7 participants were highly familiar with visualisations. All participants reported normal colour vision.

6. Results

We first report our analysis of participants' sense-making strategies. We provide exemplary quotes from the thinking aloud protocols as well as preferences from the interviews and describe their relation to the employed cognitive strategies. Table 4 shows how we address the research questions with the data from the experiment and the follow-up interview. Secondly, we describe usage per visualisation and task-specific results which address the research questions RQ2 Relations, RQ3 Weights and RQ4 Time. Finally, we report the completion times of the two visualisations, i.e. the length of time they viewed each and the frequency of switching between the two, and how the usage is related to the task performance. The sense-making strategies and the

Table 4. We quantitatively analyse the use of the visualisations and the quality of the results per task, to show which visualisation caused more plausible answers.

Analysis	Data	RQ1: Sense-making	RQ2: Relations	RQ3: Weights	RQ4: Time	RQ5: Visualisation
Quant.	T1				*	
	T2			M		
	T3		NL			
	T4		*			
	T5			NL		
	T6				M	
	T7					*
	Completion times					*
Qual.	Swapping					X
	Preference from interview					X
	Content analysis	X				

Note: The overview of the addressed research questions shows that sometimes both techniques yield plausible results. The qualitative analysis includes swapping behaviour between the two visualisations and preferred visualisations collected from interviews of 31 participants and the content analysis of thinking aloud protocols of 13 participants.
*Representing equally plausible results for matrix (M) and NL diagram

Table 5. The sense-making categories are sorted according to descending rate of coding frequency from selected transcripts (N= 13).

Category	Category description	Coding frequency	Agreement rate
1. Connections[a]	Relations in the network	156	79%
2. Trend assessment	Temporal evolution of a person	98	75%
3. Storytelling	Explain behaviour of a person, give meaning to it	48	100%
4. Comparison	Comparison of persons/attributes	47	50%
5. Elimination	Elimination of elements from search space	35	77%
6. Verification	Verify with the other view	29	75%
7. Coincidental Aha's[a]	New, sudden ideas	28	100%
8. Creative desperation[a]	Impasse situation	25	n.a.
9. Contradiction[a]	Something's odd	9	n.a.

[a]Represents strategies that are described in Klein's Triple Path Model of Insight (2013).

usage behaviour build the basis to discuss RQ1 Sense-making and RQ5 Visualisation types.

6.1. RQ1 – sense-making strategies

In the qualitative data analysis, we identified nine sense-making strategies that our participants used during the tasks. The coding was based on the Triple Path Model developed by Klein (2013). We quickly noticed that this model was not exhaustive and that we needed additional categories. These were developed in a bottom-up manner. We used screen recordings and thinking aloud protocol to analyse comments and observations (Saraiya, North, and Duca 2005). We conducted a full content analysis (Schreier 2012) of thirteen screen captures to see how the visualisations were interacted with and to enable meaningful coding. The resulting codes show a high inter-coder agreement (Cohen's Kappa κ = .82, compare Table 5).

In the following we describe the sense-making strategies in more detail and report their occurrences per visualisation. Overall and relative frequencies per visualisation of the applied strategies are shown in Figure 3. Figure 4 shows how strategy usage relates to our research questions per visualisation.

6.1.1. Connections – relationships

This is the main insight pathway of Klein's Triple Path Model of Insight (2013). We use this model because it fits to our realistic setting and tasks. Due to the nature of our tasks (looking at the relations of a co-offender network) participants often reported on looking at relations; here, crimes that two offenders committed together or are connected to indirectly (2nd degree). Example: *'This is the man in the middle connecting these networks'*.

Usage per visualisation:

Participants mainly looked for *Connections* in the NL diagram. Overall this was the most frequent strategy, used in every task but dominantly in the relational tasks (task T3 and task T4).

6.1.2. Trend assessment – looking for trends in the data over the three years

In the NL diagram, participants looked for multiple lines between co-offenders, which represented the years 2013 to 2015. Lines are double coded by line style (dashed to continuous) and colour (light blue, green and dark blue). Example: '*Here I look at the lines to see whether they were active all the time*'.

In the matrix, intra-cell stacked bar charts show the crime development over time, e.g. an ascending bar chart shows that the criminal activity is on the rise. Example: '*I can see in 2013 they have not committed a crime, but it is increasing, if you look at 2014 and 2015*'.

Usage per visualisation:

Trend assessment is a strategy that was used in both visualisations at the same rate. Asking for temporal developments (task T1 & task T6) the strategy was

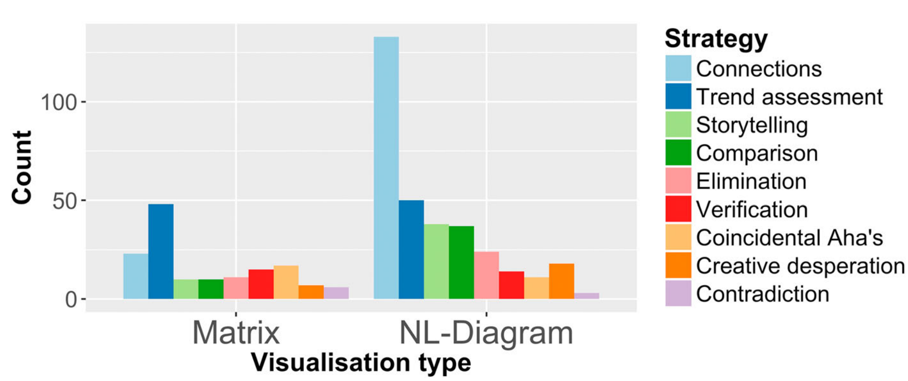


Figure 3. Strategies used per visualisation in descending order of total frequency.

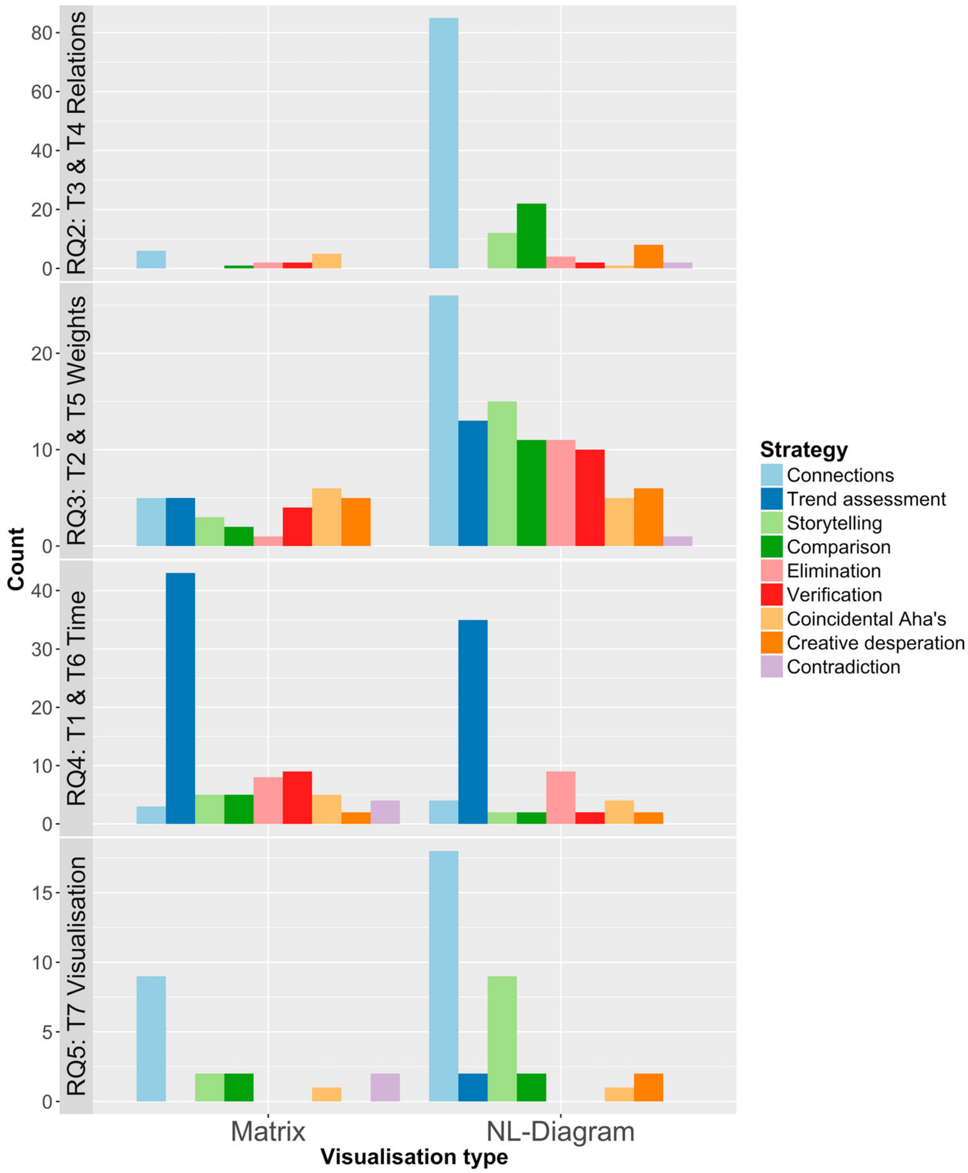


Figure 4. The frequency of the nine strategies used in the experimental setup per research question RQ2-RQ5 and visualisation technique matrix and NL diagram. We used two tasks for each research question. RQ2 is addressed by task T3 & task T4, RQ3 by task T2 & task T5, RQ4 by task T1 & task T6 and RQ5 by task T7.

used more often in the matrix than in the NL diagram. However, in tasks regarding weighted crime activities it was employed the other way around, i.e. the NL diagram was used more often to assess trends regarding crime types (compare Figure 4 RQ3).

6.1.3. Storytelling – explaining a situation with hypothesising

Participants construct a story by explaining the behaviour of criminals, crimes, time intervals and relationships they observed within the data and use their experience or imagination to add information that make sense to them. They elaborate explanations to create bridges to a new understanding. Participants inspect an actor of interest, for example, by looking at the number of 1st and 2nd degree connections (to get an idea whether the offender works with few or more criminals), crime types (what type of crimes does the actor commit), crime scores (is the score low or high), time intervals (are the crimes recent or old), and other associations to their neighbour's criminal activity. They enrich what they see, likely taking their prior experiences, knowledge and creativity into account. Some participants were using this strategy extensively with great imagination, whereas others did not use it at all and stuck to the presented facts. Example: '*They are committing crimes over all these years, so I guess they must be very good friends or successful because they didn't get caught*'.

Usage per visualisation:

Storytelling was the third most frequent strategy employed overall and much more often used in the NL diagram. There it notably occurred in all tasks, but least in those regarding time, whereas in the matrix *Storytelling* was used most frequently in the temporal tasks (compare Figure 4 RQ4).

6.1.4. Comparison – comparing indicators of crimes or criminals with initial set

This strategy was used to identify similarities or differences between actors of the network (offenders) depending on the task. One participant who previously identified a set of actors was then making a comparison with another set to assess their activity. Example: '*Here we have financial fraud … here they did something else …* '.

Usage per visualisation:

Participants used *Comparison* more often in the NL diagram than in the matrix, especially in the tasks regarding relationships and weights. In the matrix it was used more often for time related tasks.

6.1.5. Elimination – filtering the search space by eliminating data considered not relevant

This strategy generates a new understanding of a situation as part of the solution of a task. First the participant identifies possible answers and then eliminates those that do not fit the requirements. In that way mismatches get eliminated from the search space quickly and are not considered any more. Example: *'Here I have to search for blue lines, they were active in one year only'*.

Usage per visualisation:

Participants used *Elimination* to a similar extent in the matrix and in the NL diagram. On a task basis it was used as often in both views for time related tasks (task T1 & task T6); in all other tasks it was dominantly used in the NL diagram.

6.1.6. Verification – coming to a conclusion

Participants consulted both representations for verification when they came to the end of a task. To infer how long an offender was active, for example, participants looked up multiple lines in the NL diagram (line style and colour) and verified their findings by switching to the matrix and its bar charts. Example: '*I am going to the matrix as it can show the crime times in the visualisation. The matrix confirms my discovery*'. Within one visualisation participants also re-validated their findings by looking once more. Examples: *'So let me check again', 'Did I forget something?'*.

Usage per visualisation:

Participants *verified* results only one time more in the matrix than in the NL diagram. In the NL diagram it was mostly used for weights (RQ3) and in the matrix for time (compare Figure 4 RQ4).

6.1.7. Creative desperation – experiencing an impasse

This category is also part of the Triple Path Model developed by Klein (2013). It describes the situation of not knowing what to do next and the feeling of being stuck in an impasse. One can only escape by changing the direction and following a new idea. Examples: '*I have no clue*', *'nothing sticks out here', 'too little information', 'too complex', 'I don't find anybody'*.

Usage per visualisation:

Creative desperation happened in all tasks in the NL diagram, where it was overall observed more often, but primarily for relations and weights. In the matrix it did not occur for relational tasks at all (likely because it was barely used in those tasks) but happened more often with regard to weights than for time.

6.1.8. Coincidental Aha's – new, sudden ideas

Aha-moments indicate spontaneous comprehension of a problem and the emergence of a solution or an insight. Coincidences are coded as such if it is not clear to the observer why an action was taken, or where a new idea came from. When, for example, a new insight occurs where the origin is untraceable, the statement is coded as coincidental. Examples: '*Aha …* ', *'I think this could be … ', 'Now here I found that … '*.

Usage per visualisation:

Coincidental Aha-Moments occurred more often in the matrix in every task. In the NL diagram it occurred mostly with regard to weights and time, but in the matrix it occurred in relational tasks as well.

6.1.9. Contradiction – 'Something's odd'

This category is also part of the Triple Path Model developed by Klein (Klein 2013). When there are obvious mismatches in what is hypothesised or thought of being true, this contradiction can be the trigger of a new insight. Examples: '*Mhhh, this is different here … (wondering)*', *'No, this can't be right', 'There aren't as many here anymore'*.

Usage per visualisation:

Contradictions occurred rarely, but twice as much in the matrix than in the NL diagram (compare Figure 3). It happened most often in time related tasks (RQ4) in the matrix and relational tasks (RQ2) in the NL diagram (compare Figure 4).

6.2. RQ2 – relationships (Task T3 & task T4: connectors and popular offenders)

Task T3 required a detailed analysis of the offence types two persons are convicted for. Participants could easily choose connectors via the NL diagram but they sometimes had difficulties to find a successor. All participants predominantly used the NL diagram to find connectors and possible successors (least swaps and matrix use). 90% used the NL diagram and 10% used both views. Most frequently an offender with a high in-betweenness factor was named (37 times); a plausible connector of two big subgroups. Participants found two reasonable substitutes due to shared criminal activity. In contrast to that, only 19% of the participants identified the node with the highest in-betweenness centrality and had difficulties to find a possible successor.

There was one prominent new connector (task T4) in the latest year but many participants discarded the temporal requirement of this task and simply considered all years. 43% found the latest most popular connector in the NL diagram, which only 31% found in the matrix, hence, the NL diagram was slightly superior. Slightly more participants (45%) used only the NL diagram, than only the matrix (42%); 13% used both visualisations. The participants used the matrix slightly longer in this task (compare task T4 in Figure 5(a)). In both visualisations we observed that participants struggled to choose among plausible offenders. Participants wanted to get an exact rank and count, which was difficult in dense areas of the NL diagram as well as in the large matrix.

6.3. RQ3 – weights (Task T2 & task T5: crime type development and dominant crime types)

The top scored crime types in decreasing order are 'Financial Fraud', 'Disorder Violence', 'Drugs Supply', and 'Assault Minor'. To assess the importance of occurring crime types their weights need to be taken into consideration. Several cases in the network are fitting answers for task T2 and 90% of the participants found plausible groups whose criminal activity increased or crime type worsened. Participants explored shifting crime types in the NL diagram much longer (Figure 5 (a)) and swapped visualisations most because they looked for groups of more than two and matched pairs found in the NL diagram in the matrix and vice versa. They leveraged both views, e.g. one participant started in the matrix but was unsure if the identified co-offenders build a group and, therefore, switched to the NL diagram to cross-check and decide. However, participants had difficulties memorising the identifiers and often jumped back to look them up in the other visualisation again. This resulted in short visits (1–5 sec) per visualisation and a high swapping number (~0.5 per minute; AVG = 0.35), compare Figure 5(b).

In task T5 participants identified two crime types that cause most problems in the network. All participants identified 'Financial Fraud' (100%) as a major problem, which is a correct assessment because it has highest priority and occurrences. 19% correctly identified 'Disorder violence'. 58% perceived 'Assault Minor' as more dominant than 'Drugs Supply', which is a wrong assessment because although 'Drug Supply' occurs 5x less often it has a higher weight and subsequently a higher score.

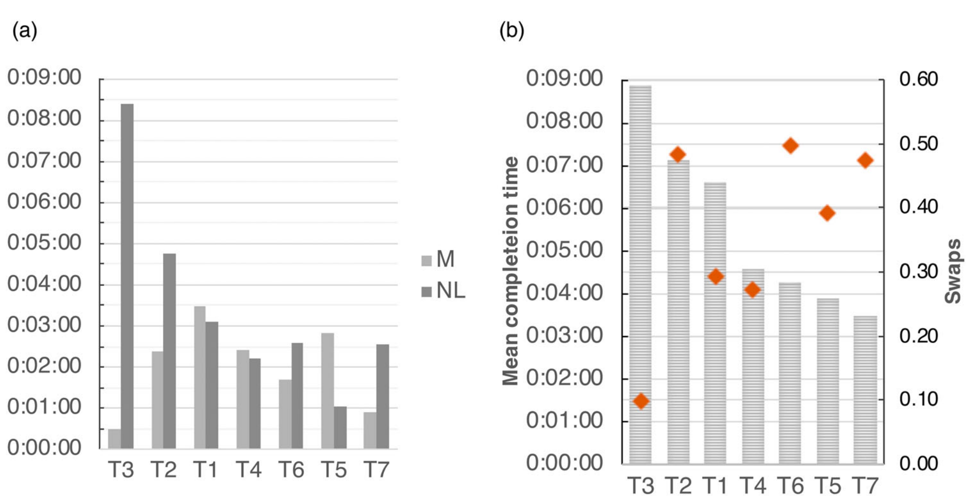


Figure 5. Mean completion time per task in descending order of the total completion time; Overall, the NL diagram was used 60% of the time and the matrix 40%. Figure (a) shows the distribution per visualisation technique, and (b) the mean swaps per minute. Participants swapped least in the task T3 with the longest completion time.

Here, the NL diagram conveyed the weights better. 67% of the participants identified a higher weighted crime in the NL diagram than they did using the matrix. However, only 10% of the participants used the graph alone, 77% used the matrix alone and 13% used both visualisations. Participants used the matrix much longer than the NL diagram (compare task T5 in Figure 5(a)). Despite the fact that the matrix was consulted longer and preferred for this task, participants using the NL diagram applied more diverse sense-making strategies. A participant using the NL diagram, for example, chose a crime type with a low weight as s/he figured there must be many crimes in numbers as a result. '*The crime type "Minor Assault" is more frequently represented but has a lower weight assigned than both "Disorder Violence" and "Burglary Other"*'.

6.4. RQ4 – time (Task T1 & task T6: trend assessment and overall tendency of development)

Seven of the 121 co-offender pairs increased their activity over the years. Overall it took participants more time to complete task T1 in the matrix, but performance was better using the matrix (66% correct answers), compared to using the NL diagram alone (49%) and consulting both (60%). 35% of the participants used only the matrix, 29% only the NL diagram and 35% used both.

The overall criminal activity decreases in the used network (task T6). More than half of the participants (61%) correctly reported a decrease. Consulting both views led to the best performance (71%). Using the matrix led to 67% accurate results and the NL diagram showed poorest performance with 45% accuracy. 45% of the participants used both visualisations to complete task T6, 35% used only the NL diagram and 19% only the matrix. Those who saw an increase in criminal behaviour interpreted the thickness of lines in the NL diagram as more important. These participants reported shifts to more harmful crime types based on individual cases, e.g. from a non-violent to a violent crime, and argued that the network's development worsens to a more violent situation.

6.5. RQ5 – visualisation

We addressed this research question on the one hand by using the results from task T7 and the interview, and on the other by analysing completion times per task and visualisation technique. From the follow-up interviews we collected preferences with regard to the visualisation. In a semi-structured way, we asked which visualisation they preferred, which visualisation was better suited for which task and what their advantages and disadvantages were.

6.5.1. Task T7: interesting observations

To find interesting patterns the participants used the NL diagram much longer than the matrix (compare Figure 5 (a)). Almost all participants switched to the other visualisation at least once. In the NL diagram participants first noticed the dominant group in the centre. They also observed that single offenders often connect bigger subgroups. In the matrix participants observed that the same crime types could build a group and more prominent groups lie close to the diagonal (an effect of the symmetric matrix).

6.5.2. Completion times and swaps

Participants spent more time on some tasks (e.g. task T3). Sometimes they spent more time with the matrix (e.g. in task T5), and sometimes with the NL diagram (e.g. in task T3). This is most likely due to the nature of the tasks. Task T3 asks for relationships between co-offenders, therefore, the NL diagram is more appropriate. In contrast to that, the structured overview of the matrix provides better information for temporal trends.

We were interested in the number of swaps between visualisations to identify which tasks are best solved with which visualisation. Furthermore, this usage can be compared to the preferences collected in the follow-up interview to get an impression of the participant's general preference for any of the given visualisations.

Participants swapped the visualisations least in the most time-consuming task T3 and the most in the second most time-consuming task T2 (see Figure 5). The visualisation usage time T does not correlate with the number of swaps S ($p(T,S) = .006$). An overview of completion times per task and swaps per minute is given in Figure 5(b).

6.5.3. Preferences

52% liked the NL diagram more, 19% the matrix, and as many stated that it depended on the task. 10% were undecided or liked both equally. Asking for task-dependencies, 52% see the NL diagram better suited for structural tasks and 68% see the matrix better suited for tasks concerning the criminal activity. This preference resembles the predominant usage of the NL diagram in task T3 and the matrix in task T5 (see Figure 5(a)).

7. Discussion

There are several studies addressing the differences between NL diagrams and matrices. The study at hand is different due to more complex (explorative) and realistic tasks and, as a consequence, our results differ from that of previous studies. Previous studies indicate that matrices in general are superior to NL diagrams (see the discussion in Section 2 Related work). The results of our study are more differentiated. We found that the matrix was more appropriate for the identification of temporal developments and the NL diagram for structural tasks. When analysing weights, the best results came from participants who looked at both visualisations. One of the reasons for the fact that the matrix is not clearly superior to the NL diagram in our investigation might be that the network we investigated is quite sparse. Therefore, the specific strength of matrices for dense networks is not as relevant. Another reason might be that we used real-world, explorative tasks, not the short generic tasks used in previous investigations, e.g. Ghoniem, Fekete, and Castagliola 2004; Henry and Fekete 2007. We also found that participants tended to prefer the NL diagrams, but this also depended on the task at hand. Many participants mentioned, for example, that the NL diagram was more appropriate for structural tasks and the matrix for crime types.

To the best of our knowledge, no comparison studies of the representation of temporal developments in NL diagrams vs. matrices have yet been conducted. Our results indicate that matrices might be better to represent temporal developments in graphs. We think that this investigation adds to the knowledge about the specific advantages and disadvantages of matrices vs. NL diagrams because it indicates that there are contexts where NL diagrams are superior to matrices and that there are situations where a combination of both visualisations is advantageous. Especially when *assessing trends*, participants chose to use both visualisations and gained suitable results with this strategy.

To investigate visualisation capabilities we let participants choose their preferred visualisation and switch freely to the other to support sense-making. Most frequent reasons for swapping were looking for groups in the NL diagram and looking for crime types in the matrix. Participants often switched from one visualisation to the other for *verification* purposes. Swapping behaviour throughout all tasks showed that participants understood the capabilities of the views and quickly learned their advantages. Immediate swaps can be explained with the thinking aloud protocols, e.g. '*this I can explore very well in the graph*', or '*this task I will better answer with the matrix*'.

The indication of 2nd degree relations was understood and participants made sense of them in task T4, e.g. one participant recalled '*the position of the yellow bars again represents the year*' and formulated a pattern they could look for. Another participant concluded '*this means I have to look for dark yellow bars in the rightmost part of the cells*'.

To make sense of the data, the participants used a group of strategies depending on the visualisations and tasks so as to understand the temporal evolution in the network – *Connections* (showing relationships is intuitively done using connecting lines) and *Trend assessment* (a bar chart can easily show a timeline on the x-axis). Hence, *Connections* were mostly searched via the NL diagram and the matrix was used more often for *Trend assessments*. More interestingly, however, is the third most common strategy, *Storytelling*. Here, the participant elaborated her understanding of the situation through own imagination in order to tell a coherent story. Therefore, participants created hypothesises about possible explanations or filled gaps with information that is not in the data. In our study, participants could freely choose which visualisation to use and reported reasons for switching the visualisation.

They gained new knowledge using strategies such as *Comparison, Creative desperation, Contradictions* and *Coincidental Aha's*. Most *Coincidental Aha's* were gathered in the matrix. Interestingly, participants indeed use 'Aha' as an expression for new insights.

To come to a conclusion, they applied the strategies *Elimination* and *Verification* and moved from uncertainty to certainty by reassuring themselves of the correctness of their answers to a task.

To support *Elimination* a visual analytics system needs to enable focusing on specific details. The NL diagram enables to focus on single connections as participants can investigate single lines easily in a sparsely connected network.

Providing several visualisations supports *Verification*. Participants used one visualisation to check the results gained from the other. Since the NL diagram was used 10% more often than the matrix, it is especially interesting that the matrix shows a higher frequency of strategy usage for *Contradictions*, *Coincidental Aha's*, and *Verification*. It might be interesting to follow up if the matrix supports these strategies better, although the difference in our study is not noticeable.

In general, we could observe that the participants of our study used the strategies *Connections* and *Trend assessment* most often. This is not surprising because the goal of visualisations in general is to help users to observe patterns in the data. Apparently, the visualisations we tested fulfil this affordance. Sometimes, it might be appropriate to support additional strategies, e.g. *Verification*. Combining two visualisations for the same data set (as in our study) does support this strategy, but additional incentives might be helpful to motivate users to engage in this strategy more often. It would be interesting to test the addition of linking and brushing capabilities and to study whether such capabilities increase the *Verification* strategy. Another strategy that is used fairly often is *Storytelling*. Some users adopt it frequently, some not at all. It is an open question whether this strategy is helpful and should be supported specifically or not, e.g. by visualisations presenting storylines. Results concerning design recommendations are still tentative, but we think we can provide some ideas concerning future research. One important aspect would be to analyse which sense-making strategy is successful in which context and how these strategies can best be supported.

In a study based on the research presented in this paper, we could show that the sense-making strategies described in this paper also hold for other visualisations (Doppler Haider et al. 2019). We also found that detecting *Connections* and *Trends* is the predominant activity when making sense of visualisations.

8. Conclusion

In this work, we have investigated how visualisations are able to support insight generation in the analysis of criminal networks through the sense-making strategies that analysts adopt. The specific strengths of our approach include the use of weighted networks (with the weights based on law enforcement's strategic priorities), the integration of the temporal evolutionary component into our study visualisations, and the possibility to specifically analyse 2nd degree relations. We were able to show that the combination of these features inside a single visualisation clearly supports intelligence tasks as plausible answers could be found in realistic tasks. Based on a qualitative analysis, we propose a set of nine strategies that are applied by visual analysts in order to arrive at relevant insights.

This research also addresses the respective capabilities of the two visualisations: an NL diagram and a matrix. We assumed that the NL diagram would be the preferred visualisation, specifically for relationships and structural issues. However, this proved not to be a clear-cut case. In the present study, the NL diagram proved to be superior for problematic relations, while the matrix provided a more plausible overview specifically for temporal evolution. However, for relationship details, participants had to study each and every cell in the matrix carefully. This can be seen as a strong argument for giving the analyst both visualisations as a general design recommendation.

Providing both visualisations strongly supported one specific strategy, namely the *Verification* strategy, as participants used both perspectives on the data to verify their conclusions. To support the *Connections* and the *Comparison* strategies better, we found that it is necessary to provide specific search options (e.g. highlighting or filtering). Furthermore, in situations where participants used the *Elimination* strategy, a clear outcome was that participants want to reduce laborious work via automatic filtering mechanisms.

In general, it can be said that participants predominantly look for patterns in the data (*Connections, Trends*). This is no surprise as the main purpose of visualisations is to support exactly these strategies. Other strategies therefore are used less often. Nevertheless, adopting strategies like *Verification* also seems to be beneficial and could be supported with appropriate interaction possibilities. The adoption of strategies also depends on the kind of task and on the visualisation used. There is no single optimal sense-making strategy, but study participants adapted the strategies to the task at hand and to the tool they used. Some of these choices are obvious, for example, that participants used the *Connections* strategy to identify relationships between co-offenders. Some of these choices are less obvious, for example, that *Storytelling* is more often used with the NL diagram than with the matrix.

9. Limitations and future work

The results indicate one major limitation of our two techniques. For testing purposes, we restricted the functionalities of the visualisations to be able to test the visualisations as such. These interaction possibilities are important for the practical work of analysts. Another limitation is the fact that we did not present both visualisations as multiple views. This restricted the *verification* process of the participants because they could not look at both visualisations at the same time. This was a decision based on the one hand, on the available monitor sizes in law enforcement agencies, and on the other hand, on the research question and methodology. We used two views in isolation to be able to collect insights per visualisation. An alternative method might be developed to track participants' insights by, e.g. using eye-tracking. However, the interpretation of eye-tracking data for sense-making is challenging. Biofeedback measures and questionnaires might help to counteract this problem.

There are still many open questions regarding the analysis of sense-making strategies. Individual differences and the generalisability to other domains are examples for such open questions. Providing additional interaction techniques such as filtering, highlighting, etc., possibly influences these sense-making strategies and it is worth to investigate this in a detailed study. In addition, it is still an open question which sense-making strategies lead to correct or plausible results and should be supported more efficiently. Furthermore, providing additional information such as background information on single offenders can support the creation of extended mental models. Finally, sense-making strategies in other domains and with other types of visualisations can be investigated. Future work along these lines can provide more precise design recommendations.

Acknowledgments

The research reported in this paper has received funding from the European Union 7th Framework Programme FP7/2007-2013, through the VALCRI project under grant agreement no. FP7-IP-608142, awarded to B.L. William Wong, Middlesex University London, and Partners. The authors acknowledge the TU Wien Library for financial support through its Open Access Funding Program.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Seventh Framework Programme: [Grant Number FP7-IP- 608142].

References

Adderley, Richard, Atta Badii, and Chaoxin Wu. 2008. “The Automatic Identification and Prioritisation of Criminal Networks from Police Crime Data.” In *Intelligence and Security Informatics*, 5–14. Berlin, Heidelberg: Springer.

Alper, Basak, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. 2013. “Weighted Graph Comparison Techniques for Brain Connectivity Analysis.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 483–492. New York: ACM.

Archambault, Daniel, James Abello, Jessie Kennedy, Stephen Kobourov, Kwan-Liu Ma, Silvia Miksch, Chris Muelder, and Alexandru C. Telea. 2014. “Temporal Multivariate Networks.” *Multivariate Network Visualization: Dagstuhl Seminar* #13201, Dagstuhl Castle, Germany, May 12–17, 2013, Revised Discussions, edited by Andreas Kerren, Helen C. Purchase, and Matthew O. Ward, 151–174. Cham: Springer International Publishing.

Bach, Benjamin, Nathalie Henry-Riche, Tim Dwyer, Tara Madhyastha, Jean-Daniel Fekete, and Thomas Grabowski. 2015. “Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks.” *Computer Graphics Forum* 34: 31–40.

Beck, Fabian, Michael Burch, Stephan Diehl, and Daniel Weiskopf. 2014. “The State of the Art in Visualizing Dynamic Graphs.” In *EuroVis – STARs*, 83–103. Swansea: EuroVis. Eurographics Association.

Burch, Michael, and Tanja Munz. 2015. “Edge-Stacked Timelines for Visualizing Dynamic Weighted Digraphs.” *Proceedings of the 6th International Conference on Information Visualization Theory and Applications*, VISIGRAPP, 93-100.

Doppler Haider, Johanna, Isabel Dallinger, Margit Pohl, Patrick Seidler, Neesha Kodagoda, and B. L. William Wong. 2019. “Sensemaking Strategies vs. Quality of Insights: Investigating Analysis Processes of Multiple Tiles”. Manuscript submitted for publication.

Doppler Haider, Johanna, Patrick Seidler, Margit Pohl, Neesha Kodagoda, Rick Adderley, and B. L. William Wong. 2017. “How Analysts Think: Sense-Making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61: 193–197.

Felson, Marcus. 2006. *The Ecosystem for Organized Crime. HEUNI Paper No. 26*. Helsinki: European Institute for Crime Prevention and Control, affiliated with the United Nations.

Ghoniem, Mohammad, Jean-Daniel Fekete, and Philippe Castagliola. 2004. “A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations.” *10th IEEE Symposium on Information Visualization*, (InfoVis).

Haider, Johanna, Margit Pohl, Eva-Catherine Hillemann, Alexander Nussbaumer, Simon Attfield, Peter Passmore, and B. L. William Wong. 2015. “Exploring the Challenges of Implementing Guidelines for the Design of Visual Analytics Systems.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59: 259–263.

Harrower, Mark, and Cynthia A. Brewer. 2003. “ColorBrewer.Org: An Online Tool for Selecting Colour Schemes for Maps.” *The Cartographic Journal* 40 (1): 27–37.

Henry, Nathalie, and Jean-Daniel Fekete. 2006. “MatrixExplorer: A Dual-Representation System to Explore Social Networks.” *IEEE Transactions on Visualization and Computer Graphics* 12 (5): 677–684.

Henry, Nathalie, and Jean-Daniel Fekete. 2007. “Matlink: Enhanced Matrix Visualization for Analyzing Social Networks.” In *Human-Computer Interaction (INTERACT)*, 288–302. Berlin, Heidelberg: Springer.

Henry, Nathalie, Jean-Daniel Fekete, and Michael Mcguffin. 2007. “NodeTrix: A Hybrid Visualization of Social Networks.” *IEEE Transactions on Visualization and Computer Graphics* 13 (6): 1302–1309.

Heuer Jr, Richards J., and Randolph H. Pherson. 2014. *Structured Analytic Techniques for Intelligence Analysis*. 2nd ed. Los Angeles, CA: SAGE CQ Press.

Hogarth, Robin M. 2001. *Educating Intuition*. Chicago, IL: University of Chicago Press.

Johnson, Jennifer A., and John David Reitzel. 2011. “Social Network Analysis in an Operational Environment.” In *International Police Executive Symposium*. Vol. 39. Virginia: Virginia Commonwealth University.

Keller, René, Claudia M. Eckert, and P. John Clarkson. 2006. “Matrices or Node-Link Diagrams: Which Visual Representation Is Better for Visualising Connectivity Models?” *Information Visualization* 5 (1): 62–76.

Khurana, Udayan, Viet-An Nguyen, Hsueh-Chien Cheng, Jae-wook Ahn, Xi Chen, and Ben Shneiderman. 2011. “Visual Analysis of Temporal Trends in Social Networks Using Edge Color Coding and Metric Timelines.”

Third International Conference on Privacy, Security, Risk and Tust (PASSAT)/ Social Computing (SocialCom).

Klein, Gary. 2013. *Seeing What Others Don't – The Remarkable Ways We Gain Insights*. New York: PublicAffairs, a Member of the Perseus Book Group.

Klein, Gary, Brian Moon, and Robert R. Hoffman. 2006a. "Making Sense of Sensemaking 1: Alternative Perspectives." *IEEE Intelligent Systems* 21 (4): 70–73.

Klein, Gary, Brian Moon, and Robert R. Hoffman. 2006b. "Making Sense of Sensemaking 2: A Macrocognitive Model." *IEEE Intelligent Systems* 21 (5): 88–92.

Klein, Gary, Jennifer K. Phillips, Erica L. Rall, and Deborah A. Peluso. 2007. "A Data Frame Theory of Sensemaking." In *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, 113–155. New York, NY, USA: Lawrence Erlbaum.

Ko, Sungahnn, Shehzad Afzal, Simon Walton, Yang Yang, Junghoon Chae, Abish Malik, Yun Jang, Min Chen, and David Ebert. 2014. "Analyzing High-Dimensional Multivariate Network Links with Integrated Anomaly Detection, Highlighting and Exploration." *IEEE Conference on Visual Analytics Science and Technology (VAST).*

Kodagoda, Neesha, Sheila Pontis, Donal Simmie, Simon Attfield, B. L. William Wong, Ann Blandford, and Chris Hankin. 2017. "Using Machine Learning to Infer Reasoning Provenance From User Interaction Log Data." *Journal of Cognitive Engineering and Decision Making* 11: 23–41.

Kriglstein, Simone, Margit Pohl, and Michael Smuc. 2014. "Pep up Your Time Machine: Recommendations for the Design of Information Visualizations of Time-Dependent Data." In *Handbook of Human Centric Visualization*, 203–225. New York, NY: Springer.

Nowell, Lucy, Elizabeth Hetzler, and Ted Tanasse. 2001. "Change Blindness in Information Visualization: A Case Study." In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'01)*, 15–22. San Diego: IEEE Computer Society.

Pirolli, Peter, and Stuart Card. 2005. "The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis." *Proceedings of International Conference on Intelligence Analysis.*

Pohl, Margit, Günter Wallner, and Simone Kriglstein. 2016. "Using lag-Sequential Analysis for Understanding Interaction Sequences in Visualizations." *International Journal of Human-Computer Studies* 96 (July): 54–66.

Reda, Khairi, Andrew E. Johnson, Jason Leigh, and Michael E. Papka. 2014. "Evaluating User Behavior and Strategy During Visual Exploration." In *Proceedings of the 5th Workshop BELIV*, 41–45. Paris: ACM.

Robins, Garry, and Yoshi Kashima. 2008. "Social Psychology and Social Networks: Individuals and Social Systems." *Asian Journal Of Social Psychology* 11 (1): 1–12.

Rufiange, Sébastien, and Guy Melançon. 2014. "AniMatrix: A Matrix-Based Visualization of Software Evolution." 2014 second IEEE working conference on software visualization (VISSOFT), 137–146.

Saraiya, Purvi, Chris North, and Karen Duca. 2005. "An Insight-Based Methodology for Evaluating Bioinformatics Visualizations." *IEEE Transactions on Visualization and Computer Graphics* 11 (4): 443–456.

Schreier, Margrit. 2012. *Qualitative Content Analysis in Practice*. London: SAGE Publications.

Sedig, Kamran, and Paul Parsons. 2013. "Interaction Design for Complex Cognitive Activities with Visual Representations: A Pattern-Based Approach." *AIS Transactions on Human-Computer Interaction* 5 (2): 84–133.

Sedig, Kamran, Paul Parsons, Hai-Ning Liang, and Jim Morey. 2016. "Supporting Sensemaking of Complex Objects with Visualizations: Visibility and Complementarity of Interactions." *Informatics* 3 (4): 20. Multidisciplinary Digital Publishing Institute.

Sedig, Kamran, Sonja Rowhani, and Hai-Ning Liang. 2005. "Designing Interfaces That Support Formation of Cognitive Maps of Transitional Processes: An Empirical Study." *Interacting with Computers* 17 (4): 419–452.

Segel, Edward, and Heer Jeffrey. 2010. "Narrative Visualization: Telling Stories with Data." *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1139–1148.

Seidler, Patrick, Johanna Haider, Neesha Kodagoda, B. L. William Wong, Margit Pohl, and Richard Adderley. 2016. "Design for Intelligence Analysis of Complex Systems: Evolution of Criminal Networks." *2016 European Intelligence and Security Informatics Conference (EISIC).*

UK: National Policing Improvement Agency on behalf of the Association of Chief Police Officers. 2008. *Practice Advice on Analysis.*

Wakita, Ken, Masanori Takami, and Hiroshi Hosobe. 2015. "Interactive High-Dimensional Visualization of Social Graphs." *IEEE Pacific Visualization Symposium*, (PacificVis).

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. 1st edition. Cambridge, NY: Cambridge University Press.

Xu, Jennifer, and Hsinchun Chen. 2005. "Criminal Network Analysis and Visualization." *In Communications of the ACM* 48 (June): 100–107.

Yi, Ji Soo, Niklas Elmqvist, and Seungyoon Lee. 2010. "TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations." *International Journal of Human-Computer Interaction* 26 (11–12): 1031–1051.

Zhu, Bin, Stephanie Watts, and Hsinchun Chen. 2010. "Visualizing Social Network Concepts." *Decision Support Systems* 49 (2): 151–161.