

How Analysts Think: Sense-making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities

Johanna Doppler Haider¹, Patrick Seidler², Margit Pohl¹, Neesha Kodagoda², Rick Adderley³,
B. L. William Wong²

¹Vienna University of Technology, Austria

²Middlesex University London, UK

³A E Solutions (BI) Ltd, Badsey, UK

Analysis of criminal activity based on offenders' social networks is an established procedure in intelligence analysis. The complexity of the data poses an obstacle for analysts to gauge network developments, e.g. detect emerging problems. Visualization is a powerful tool to achieve this, but it is essential to know how the analysts' sense-making strategies can be supported most efficiently. Based on a think aloud study we identified ten cognitive strategies on a general level to be useful for designers. We also provide some examples how these strategies can be supported through appropriate visualizations.

Keywords

User strategies, analytical reasoning, inference making

INTRODUCTION

Crime groups operate in criminal ecosystems that form complex networks of interrelated and interdependent crime activities (Felson, 2006). They represent organic structures that can transcend local, regional, national and policing boundaries, evolve over time and space (Archambault, 2014; Bach, 2015; Wasserman, 1994), and often propagate their activities across different networks.

For intelligence analysis, it is essential to piece together a concise picture of how these networks operate, how they are controlled, who controls them, and how information is communicated to plan, coordinate, execute and conceal their criminal operations. Police intelligence analysts collaborating with us explain that while network visualizations can be helpful, they are however of limited use. Three specific analyst challenges that need to be addressed include 1) Capturing the build-up of harmful developments like growing subgroups or groups becoming more violent, 2) Identifying and monitoring problematic offenders, and 3) Monitoring the effectiveness of current policing strategies. The purpose of our research is to investigate better forms of visualizations to improve analytical reasoning in the context of temporally evolving crime networks – both in terms of changes in their structures and relationships, and the nature of their harmful impacts over time. In this paper we focus on describing the cognitive strategies used by participants in the study for understanding, assessing, discovery and problem solving.

Current automated network analysis frameworks, e.g., Xu and Chen (2005) rely on Social Network Analysis measures, which do not or only have cumbersome methods for supporting the analysis of changes in criminal behavior over time. Current systems are inadequate in addressing intelligence requirements such as tracking evolving crime behavior, crime membership, or re-occurring events. This has led to operational gaps between criminal network analysis and police operations (Johnson & Reitzel, 2011). Research on sense-making can inform the design of such a system. Visualizations can provide overviews of network behavior by

integrating its evolution instead of animating the data or dividing it into multiple views (Khurana et al., 2011; Zhu, Watts & Chen, 2010), thus, reducing cognitive load.

We developed a combined node-link and matrix visualization system (see Figure 1, and later explanation of the design) for weighted networks with a number of key features. First, in our network visualization, links between co-offenders, i.e., criminals who commit crimes together, are weighted to represent the seriousness of the crime in a specific time frame. Seriousness can be calculated based on the concept of a harm index (Seidler & Adderley, 2013). Another feature 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. Finally, the visualization can show the temporal evolution of these networks over several points in time. It can also handle a large number of nodes and represent different crime types simultaneously. In this way we anticipate that this network visualization will create emergent cues that can suggest the build-up of harmful situations, cues for recognizing problematic offenders and emerging problematic offenders, and cues for observing changes in the harmfulness of various crime and criminal networks over time.

In this paper we report on the cognitive strategies employed by participants in the study as they think, infer and problem-solve with node-link and matrix visualizations to achieve the three analysts' challenges. We observed that these cognitive strategies can be employed repetitively and in any order, depending on the task at hand, the data available and what they aspire to achieve. We suggest that any system designed to support such analytic investigations should also possess the variety of features that can be fluidly interchanged while carrying out those tasks.

RELATED WORK

Network visualization research faces the challenge of scalability and interaction design as a major factor to provide a successful visual analysis tool (Pienta, Abello, Kahng & Chau, 2015). Most commonly a series of diagrams gets animated or are shown as small multiples (Archambault et al., 2014; Beck, Burch, Diehl & Weiskopf, 2014). Bach et al. (2015) state that

visualization of multi-node networks supports exploratory analysis of networks very well and can aid in communicating findings. However, there is a lack of appropriate network visualizations suitable to scale to several thousand nodes especially in dynamically changing networks.

Sense-making as the deliberate effort to understand events (Klein & Phillips, 2007) is mainly influenced by cognitive psychology. The well-known model of sense-making for intelligence analysis by Pirolli and Card (2005) lacks the tool interaction and knowledge of the user. Klein and Phillips (2007), on the other hand, state that we have to shift an anchor in our prior beliefs to gain new insights. Insight has been defined as, "sudden unexpected thoughts that solve problems" (Hogarth 2001, p.251), or "an unexpected shift in the way we understand things" (Klein, 2013). It provides a comprehension of a situation by the unconscious synthesis of prior knowledge and experience with newly collected data to create an unexpected, dramatic realization.

Visualizations including a matrix representation are the *MatrixExplorer* (Henry & Fekete, 2006) and *NodeTriX* (Henry, Fekete & McGuffin, 2007). Node-link diagrams with multivariate edges were presented as multiple threads, parallel colored lines (Ko et al., 2014). Node-link diagrams and matrices have been compared to identify their advantages and disadvantages. Using simple, generic tasks, matrices have been found to be especially useful for larger, denser networks, with graphs more suited for path-related tasks (Ghoniem, Fekete & Castagliola, 2004; Keller, Eckert & Clarkson, 2006). Henry and Fekete (2007) developed *MatLink*, a hybrid tool that combined the matrix with links overlaid on its border. *MatLink* was found to be superior to either node-link diagram or matrix. Matrix representations in general are more efficient than weighted node-link diagrams (Alper, Bach, Henry Riche,

Isenberg & Fekete, 2013). However, these studies used generic and fairly simple tasks. Rather, it is important to represent the real complexity of the challenges intelligence analysts face to better understand their requirements for visualizations.

SYSTEM: THE NODE-LINK/MATRIX REPRESENTATION OF CRIME AND CRIMINAL NETWORKS

We use a node-link (NL) diagram and a matrix representation that show crimes committed jointly by pairs of co-offenders (see Figure 1). The visualizations are supposed to provide an overview of the development of criminal activities. A key difference of this system to previous approaches is to support both overview and individual relationships of three aggregated time steps in an integrated view to support the temporal analysis of large, weighted networks. Indirect, i.e., 2nd degree, neighbors are easy to detect in NL diagrams. In the matrix, the relationship has to be explicitly encoded (e.g., by a yellow box in the cell at the intersection of the respective row and column. The number of crimes is encoded in 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 color as is common practice in intelligence analysis.

Time is represented in the NL diagram as three parallel lines, and double-coded; we use different colors as well as different line styles because it is difficult to distinguish three different lines in a large NL diagram and due to color vision deficiency we do not rely on color alone. On mouse-over, a pop-up field appears in the NL diagram showing a more detailed representation of the temporal development of criminal activities in a timeline and in case of a 2nd degree connection it is shown in which year the relation was established.

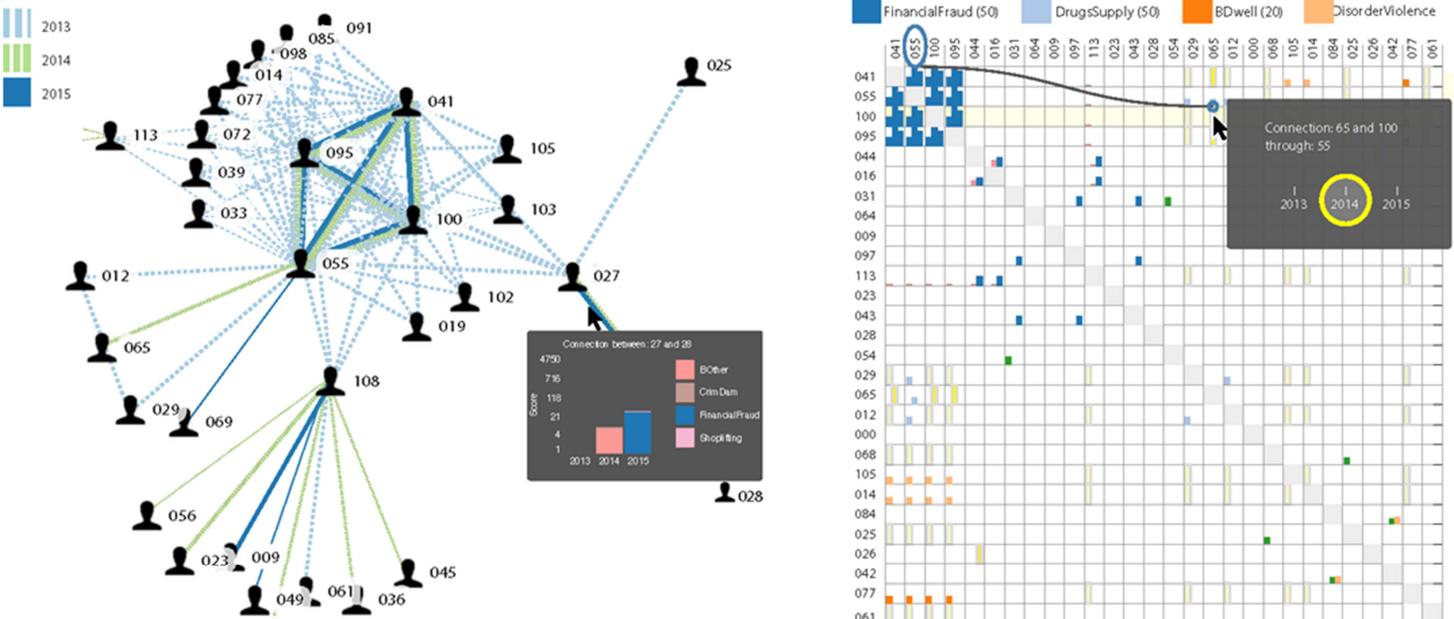


Figure 1: Network visualization via node-link diagram (left) and equivalent matrix representation with yellow 2nd degree boxes (right). The whole network comprises 121 nodes and 996 edges with a 3% edge density and is derived from a real database. Nodes represent offenders, while edges represent the undirected relationships between two offenders when their names appear in the same crime report. Co-offences are grouped over three time intervals representing the years in which the crime took place.

In the matrix, temporal development is shown as a bar chart in the cells. The x-axis of the bar chart indicates the time (1st, 2nd, 3rd year respectively), also in case of the yellow box indicating 2nd degree relationships. When hovering over these boxes a line between the middleman and the co-offenders shows the relation and additional information related to the crimes is shown. Zooming and panning enables detailed exploration of the data. Technical details of the system as well as the application scenario are described in a previous work (Seidler, Haider, Pohl, Kodagoda & Wong, 2016).

DESCRIPTION OF THE STUDY

We carried out a two-part study to investigate if and how the node-link and matrix network visualization supported or hindered analysts in detecting and understanding the temporal development of co-offender networks and network activities over time, and, more generally, which sense-making strategies analysts use to achieve this. The first part of the study (reported in Seidler et al., 2016) ascertained the efficacy of the designs with respect to the analysts' task goals. The second part of the study used a think aloud, concurrent protocol analysis method to identify participants' cognitive strategies when using the node-link and matrix network visualizations in order to understand how the proposed designs affected the way they think and reason. We recruited 31 undergraduate and postgraduate students (18 male, 13 female) from two European Universities aged between 24 and 34 years (mean age 26.5 years) to participate in the experiment. Of the 31, 24 had basic to moderate knowledge in information visualization, and the remaining 17 participants were very familiar with the subject. All participants reported normal color vision. All participants were new to the system and were introduced to the system design and trained on its interaction methods.

Method and Tasks

The participants were asked to carry out a number of tasks as quickly as possible while thinking aloud (Ericsson & Simon, 1993) and report what they were looking at, considering or pondering. They were randomly assigned a starting view and were told they could swap to any view at any time during the study. Each session lasted an average of 42 minutes with a follow-up interview. Including the introduction and training, the experiment lasted no longer than one hour. Interactions with the system were recorded via screen capture software, while an audio recorder recorded the verbal reports. The data analysis was carried out by two researchers and consequently checked for inconsistencies.

The tasks used were based on the three main analyst goals: 1) Capture the build-up of harmful developments like growing subgroups or groups becoming more violent, 2) Identify and monitor problematic offenders, 3) Monitor the effectiveness of current mitigation strategies. Seven types of realistic tasks were used, including *Identify increase of criminal activity*, *Identify groups with an increased criminal activity*, *Identify possible relationships between offenders*, and *Identify any overall tendencies or trends in criminal activities*. We use these complex tasks because sense-making strategies, which are necessary in realistic contexts, cannot be easily

observed with simple tasks. In contrast to that, to understand how intelligence analysts achieve their goals it is necessary to reproduce the real complexity of the challenges they face.

Equipment

We used a 24inch monitor with a resolution of 1920 x 1200 pixels (aspect ratio 16:10, model Dell U2415). The visualizations were designed to show all the data space within a single view, although panning and scrolling was necessary once a user zoomed in. Users operated with a standard keyboard and a scroll wheel mouse. Quick swapping was enabled via keyboard shortcuts. Instructions were attached to the bottom of the screen to remind the users.

RESULTS

We analyzed the think-aloud protocols from all participants as they engaged in their tasks. Using the Emergent Themes Analysis (Wong & Blandford, 2002) data analysis method, we identified ten sense-making strategies. Each cognitive strategy is described below and supported with exemplary quotes from the participants. These sense-making strategies can be grouped according to five purposes: (a) *to gain overview*, (b) *to gain new knowledge*, (c) *to elaborate*, (d) *to improve certainty*, and (e) *to create an overarching explanation* (see Figure 2).

A. To gain overview

Observations of changes in criminal activity. Looking for trends in the data. Once participants identified a set of criminals who have committed crimes over the years, participants focused on identifying increase in their criminal activity over time. NL users looked for cues such as multiple lines between co-offenders which represented years. The lines were double-coded by line style (dashed to continuous) and color (light blue, green and dark blue) representing 2013 to 2015. P5: So I am looking for people who have got more than one color line. So I am looking for the same color line blue, green, light blue, so it shows me the time period. So this is an increment.

Matrix users identified cues such as the pre-organized intra-cell stacked bar charts shown at the top left corner and of crimes the network is associated with and the severity of the crimes they committed over time.

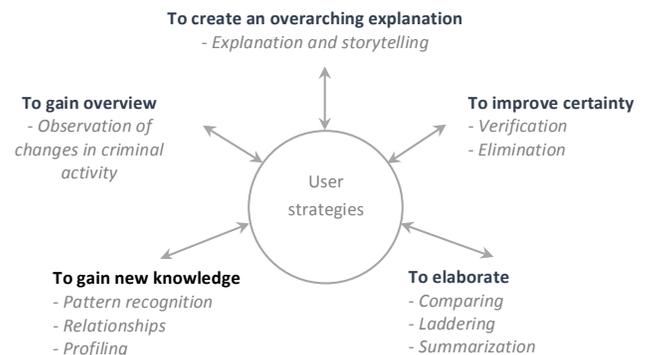


Figure 2: The interactions between the sense-making strategies are shown; they are used in a chaotic and cyclic manner.

Using the NL diagram, cues were the thickness of the line between actors provided a starting point, followed by hovering over the lines between co-offenders. P3: I can see that the numbers of crimes are increasing if the lines are thicker. 2013 light blue dash line thick, then green 2014 dash line thicker and the dark blue line thicker than the previous. In the matrix, cues such as the stacked bar chart color (crime type) and height (score) across one actor helped them determine the crime evolution and the severity. P2: It is very visible these small columns their rising or falling. P3: I already see these squares are in different color. So if I see on the left they have different color and the right different color (in a cell within the matrix) this means the crime is changed. So for example no 18 & 44, in 2014 they did "Burglary Other", and in 2015 jumped to Financial Fraud.

B. To gain new knowledge

Pattern recognition. Recognition of crimes, criminals time intervals, direct and indirect relationship patterns. Participant 7, for example, observed a pattern shift in crime type behavior. P7: Again 16 when committing crime with 113 in 2014 both of them committed Criminal Damage, but both of them moved to Financial Fraud and have a lot of activity. So you can see with whoever he is connected with maybe he influenced them, in a way to commit Financial Fraud.

Relationships. All participants looked at co-offender relations and the crimes they committed for direct (1st degree) or indirect neighbors (2nd degree). P1: The other interesting thing in this graph is that this guy 31 did some work robbery, and other crimes in 2013 (54, 67, 79, 53), then he switched to financial fraud with 8 in 2014 and then in 2015 started doing financial fraud with 97 & 43. I think the graph is good to see the pattern in changes over time.

Profiling. Characterize crimes or criminals based on features and relations they have observed over time. Here, participant 8 analyses that actors in the network are prone towards committing Financial Fraud. P8: I think, one crime which causes problems is Financial Fraud; P9: Financial Fraud seems to have considerably higher scores, in all three years across the network.

C. To elaborate and gain insight

Comparing. Comparing indicators of crimes or criminals with an initial set to identify if similarities or differences exist. Participant 9 previously identified a set of actors and consequently compared them with another set to assess their the trend of the bar chart, e.g., ascending over the time intervals showed if the criminal activity between a pair was on the rise. P9: I will look up the matrix because the top left corner has the highest score, for the crime types. P2: I can see in 2013 they have not committed crime, but it is increasing, if you look at 2014 and more.

Laddering. All participants developed an understanding of the situation based on initial cues they attended to. Explanations were elaborated to create bridges to new data to create a new understanding. Participant 3 first explored offender 16, and then identified 42 as a co-offender. As the pair's crimes over 2014 & 2015 shows an increase, it was used as an anchor to bridge to 42's criminal activities. P3: Even 16 (& 42) the same, in 2014 there was about 10 crimes, and then 2015 there was over 21, so it increased. You have Number 42 who has an increase (84 & 42). Then also 114 (& 42 looks at the row). Another is 18 (& 42). This means this actor seems to show an increase in criminal activity.

Summarization. Aggregation of crimes, offenders, and time intervals with direct or indirect connections. Participants were observed aggregating information, for example, offender 16 the number of crimes grouped by year. P3: 16 in

2014 there was about 10 crimes, and then 2015 there was over 21, so it increased.

D. To improve certainty

Elimination. Generating new understanding by eliminating data considered as not relevant. At first participant 4 identified four actors who are very active, which seemed interesting for the task at hand. However, then s/he identified that there is a mismatch in what s/he is looking for and the set got eliminated quickly. P4: These are the ones at the top left corner. We have 45, 55, 95 & 100. As you can see it is not constantly increasing, 2013 to 2014 there is a rise from 2014 to 2015 there is a drop. We are looking for something, which has an increase. This one does 84 & 42, robbery and disorder similar for 42 & 114, this one again 44 & 18 Criminal Damage they seem to be on the rise.

Verification. Participants consulted both representations for verification. When looking at multiple lines in the NL diagram to infer how long an offender was active, for example, the conclusions were verified by looking at the stacked bar chart's time interval bars in the matrix. P5: I am going to the matrix as it can show the crime times in the visualization. The matrix confirms my discovery. The bar graph is shown in different colors. I think that's all of them.

E. To create an overarching explanation

Explanation and storytelling. Constructing a story by explaining the behavior of crimes, criminal, time intervals and relationships they have observed within the data and also using their experience. After participants profiled offenders based on their observation, they elaborated further with explanations and storytelling. For an actor of interest some participants looked at cues, such as number of 1st and 2nd degree connections to see if the offender works with few or more criminals. Other cues were crimes 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 neighbor's criminal activity. Participant 9 enriched the given information with prior experiences, knowledge and creativity. P9: It kind of suggests that people's perception that they are able to get away easily with Financial Fraud then commit other physical crimes.

DISCUSSION

The work of intelligence analysts is very challenging and complex and cannot be reduced to simple lookup tasks. There is a lack of research to analyze these processes in more detail. In our research we tried to close this gap. We identified ten sense-making strategies, which are especially relevant for intelligence analysis. Nevertheless, we think they can be generalized to other domains as well. In general, the strategies we identified represent very complex behavior, which goes beyond simple identification of data points. Participants, on the one hand, adopted strategies **to gain overview** of the data, such as *Observations of changes in criminal activity*. They created **new knowledge** using strategies like *Pattern recognition*, *Relationships* and *Profiling*. As analysts interpret the data, they **elaborate** their understanding and create new insights by *Comparing indicator sets*, *Laddering* and *Summarization*. To **improve certainty**, participants used the strategies *Elimination* and *Verification* when they moved from

low uncertainty to high certainty. To **create overarching explanation**, participants adopted a strategy of *Explanation and storytelling*. These strategies took place in a chaotic and cyclic manner depending on the available information and the goals they needed to satisfy to gain cognitive traction.

Results from this study also indicate how sense-making strategies can be supported by identifying how changes in the data can be supported by a visualization. How to show such developments at a glance instead of forcing the analyst to jump from one visualization to another again and again. An important point was that participants used the combination of two visualizations for verification purposes, to make sure that the results they found in one visualization can also be observed in the other visualization. This indicates that the juxtaposition of two visualization of that type can be used to support verification strategies. These are good examples for the usefulness of analyzing sense-making strategies for design purposes. If designers know which sense-making strategies the potential users adopt they can adapt the system accordingly.

Nevertheless, further research is certainly necessary. We need to identify how frequent the involved strategies appear and whether they are applied in many different contexts or, in contrast to that, if they are only adopted for specific kinds of tasks. We also need to know whether these sense-making strategies are efficient or not. It is also necessary to check whether the sense-making strategies we found for intelligence analysis will also hold for other domains.

CONCLUSION

In this paper, we describe a system consisting of a node-link and matrix visualization for intelligence analysts who want to investigate social networks of co-offenders and their temporal evolution. We looked at how participants used the system during exploration and which strategies they developed to solve realistic tasks. These sense-making strategies can be used to inform the work of designers of such systems by supporting their structure, layout and interaction possibilities.

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.

REFERENCES

- Alper, B., Bach, B., Henry Riche, N., Isenberg, T., & Fekete, J.-D. (2013). Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 483–492). ACM.
- Archambault, D., Abello, J., Kennedy, J., Kobourov, S., Ma, K.-L., Miksch, S., Muelder, C. & Telea, A. C. (2014). *Temporal Multivariate Networks*. In A. Kerren, H. C. Purchase, & M. O. Ward (Eds.), *Multivariate Network Visualization* (pp. 151–174). Springer International Publishing.
- Bach, B., Henry-Riche, N., Dwyer, T., Madhyastha, T., Fekete, J. D., & Grabowski, T. (2015). Small MultiPiles: Piling time to explore temporal patterns in dynamic networks. In *Computer Graphics Forum* (Vol. 34, pp. 31–40).
- Beck, F., Burch, M., Diehl, S., & Weiskopf, D. (2014). The state of the art in visualizing dynamic graphs. In *EuroVis - STARs* (pp. 83–103). Eurographics Association.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. MIT press Cambridge, MA.
- Felson, M. (2006). *The Ecosystem for organized crime*. European Institute for Crime Prevention and Control, affiliated with the United Nations.
- Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2004). A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization. INFOVIS 2004* (pp. 17–24).
- Henry, N., Fekete, J.-D., & McGuffin, M. J. (2007). NodeTriX: a hybrid visualization of social networks. In *IEEE Transactions on Visualization and Computer Graphics*, 13(6), (pp. 1302–1309).
- Henry, N., & Fekete, J.-D. (2006). MatrixExplorer: A dual-representation system to explore social networks. In *IEEE Transactions on Visualization and Computer Graphics*, 12(5), (pp. 677–684).
- Henry, N., & Fekete, J.-D. (2007). MatLink: Enhanced matrix visualization for analyzing social networks. In C. Baranauskas, P. Palanque, J. Abascal, & S. D. J. Barbosa (Eds.), *Human-Computer Interaction – INTERACT 2007* (pp. 288–302). Springer Berlin Heidelberg.
- Hogarth, R. M. (2001). *Educating intuition*. Chicago: University of Chicago Press.
- Johnson, J. A., & Reitzel, J. D. (2011). Social network analysis in an operational environment: Defining the utility of a network approach for crime analysis using the Richmond City Police Department as a case study. In *International Police Executive Symposium*.
- Keller, R., Eckert, C. M., & Clarkson, P. J. (2006). Matrices or node-link diagrams: Which visual representation is better for visualising connectivity models? In *Information Visualization*, 5(1), (pp. 62–76).
- Khurana, U., Nguyen, V.-A., Cheng, H.-C., Ahn, J., Chen, X., & Shneiderman, B. (2011). Visual analysis of temporal trends in social networks using edge color coding and metric timelines. In *IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and IEEE Third International Conference on Social Computing (SocialCom)*, (pp. 549–554).
- Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (pp. 113-155). New York, NY, USA: Lawrence Erlbaum.
- Klein, G. (2013). *Seeing what others don't: The remarkable ways we gain insights*. New York, USA: PublicAffairs, a Member of the Perseus Book Group.
- Ko, S., Afzal, S., Walton, S., Yang, Y., Chae, J., Malik, A., Jang, Y., Chen, M., & Ebert, D. (2014). Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, (pp. 83–92).
- Pienta, R., Abello, J., Kahng, M., & Chau, D. H. (2015). Scalable graph exploration and visualization: Sensemaking challenges and opportunities. In *IEEE International Conference on Big Data and Smart Computing (BigComp)*, (pp. 271–278).
- Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis* (Vol. 5, pp. 2–4).
- Seidler, P., & Adderley, R. (2013). Criminal network analysis inside law enforcement agencies: A data mining system approach under the National Intelligence Model, *International Journal of Police Science and Management* (vol. 15, no. 4, pp. 323–337).
- Seidler, P., Haider, J., Kodagoda, N., Wong, B. L. W., Pohl, M., & Adderley, R. (2016). Design for intelligence analysis of complex systems: Evolution of criminal networks. In *2016 European Intelligence and Security Informatics Conference (EISIC)* (pp. 140–143).
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. (Vol. 8). Cambridge; New York: Cambridge University Press.
- Xu, J. & Chen, H. (2005). Criminal network analysis and visualization. *Communications of the ACM*, 48(6), (pp. 100–107).
- Zhu, B., Watts, S., & Chen, H. (2010). Visualizing social network concepts. *Decision Support Systems*, 49(2), (pp. 151–161).