Original Article

JDMS

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
2014, Vol. 11(2) 175–199

DOI: 10.1177/1548512912464532
dms.sagepub.com
SAGE

Visual Analytics for cyber security and intelligence

Valérie Lavigne[1] and Denis Gouin[1]

Abstract

In the context of modern defense and security operations, analysts are faced with a continuously growing set of information of different nature that causes significant information overload problems and prevents developing good situation awareness. Fortunately, Visual Analytics (VA) has emerged as an efficient way of handling and making sense of massive datasets by exploiting interactive visualization technologies and human cognitive abilities. Defence R&D Canada has conducted a review of the applicability of VA to support military and security operations. This paper is meant to provide someone new to this area with a quick overview of the current state of the art in VA. We introduce the important scientific visualization, interaction and reasoning concepts supporting VA, followed by VA advanced techniques. Then, we describe how VA can contribute to the cyber security and intelligence analysis application domains, along with promising research projects and commercial software. Finally, we discuss the future of VA research.

Keywords

Visual Analytics, cyber security, intelligence, counterterrorism, counter-insurgency, interactive information visualization

1. Introduction

In the era of the information age, decision makers and first responders in defense and security are faced with increasing amounts of dynamic information originating from a wide variety of sources and in a wide variety of formats, which they need to analyze in order to understand a situation and react promptly. This is the case, for example, in situation management following a natural or man-made disaster, a terrorist attack, a military conflict, a pandemic flu or a series of criminal activities. In developing situation awareness, analysts must understand how a situation has developed and how it may develop. They need to identify trends and patterns.

Although information fusion and rule-based systems have shown their value in helping make sense of information and providing situation awareness, during the last decade, a new science and technology called Visual Analytics (VA) has rapidly emerged in helping users understand a situation by cleverly representing the information and providing mechanisms to interact with it.

This paper is meant to provide someone new to this area with a quick overview of the current state of the art in VA. We begin with an introduction of the important scientific visualization, interaction and reasoning concepts supporting VA and we present some VA advanced techniques. Then, we describe how VA can contribute to the cyber security and intelligence analysis application domains, along with promising research projects and commercial software. Finally, we discuss the future of VA research.

2. Visual Analytics

''Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.'' This widespread definition comes from the US research agenda that launched the field: Illuminating the Path.[1] More recently, researchers from the European Union Coordination Action, in their book ''Vismaster: Mastering the Visualization

[1]Defence R&D Canada, Canada

Corresponding author:
Valérie Lavigne, Defence R&D Canada – Valcartier, 2459 Pie-XI Blvd. North, Quebec, Canada G3J 1X5.
Email: valerie.lavigne@drdc-rddc.gc.ca

Age'', stated that VA has become ''the medium of a semi-automated analytical process, where humans and machines cooperate using their respective, distinct capabilities for the most effective results''.[2]

VA is a multidisciplinary field that combines various related research areas where much valuable prior work has been done. Figure 1 shows a non-exhaustive list of scientific disciplines that are related to VA.[3] The challenge of making visual analysis effective calls for advancement in a variety of scientific fields. Visualization, interaction science and analytical reasoning are three key areas of VA research that bring highly important scientific concepts to VA and these are described in the following sections.

2.1. Visualization

Humans discovered a long time ago that they could enhance their cognitive abilities by using external representation aids.[4] The use of visualization to present information is not a new phenomenon. It has been used in maps, scientific drawings and data plots for over a thousand years. The intent of visualization is not merely to display information using pictures. The visual representation should be designed in a meaningful way in order to provide insight to the user. The optimal choice is highly dependent on the data involved and the task to be performed.

Information can be presented visually using points, lines, shapes, colors, intensity, textures, motion, etc. To select effective visual cues for data representation, we can use results obtained from the study of human perception. Pre-attentive features form a set of visual properties that are detected very rapidly and accurately by the low-level visual system. These properties were initially called pre-attentive, since their detection seems to precede focused attention. This process is effortless, meaning that it does not demand attentional resources from a human. In each case presented in Figure 2, a unique visual property in the target allows it to ''pop out'' of the display. Examples of tasks that can be performed using pre-attentive features are target detection, boundary detection, region tacking and counting and estimation.[5]

In Figure 3, the use of semantic depth of field makes some chess pieces more salient to guide the user's attention.[6]

Color in information presentation is mostly used to distinguish one element from another.[7] Contrasting colors are different and draw attention, while analogous colors are similar and group elements (see Figure 4). As Tufte[8] puts it: ''avoiding catastrophe becomes the first principle in bringing color to information: Above all, do no harm.'' Chosen poorly, colors can obscure the meaning of information.

The Gestalt laws of organization describe how people perceive visual components as organized patterns or wholes, instead of many different parts.[9,10] According to this theory, there are six main factors that determine how we group things according to visual perception: closure, similarity, proximity, symmetry, continuity and common fate.

Inattentional blindness, also known as perceptual blindness, refers to the inability to perceive features in a visual

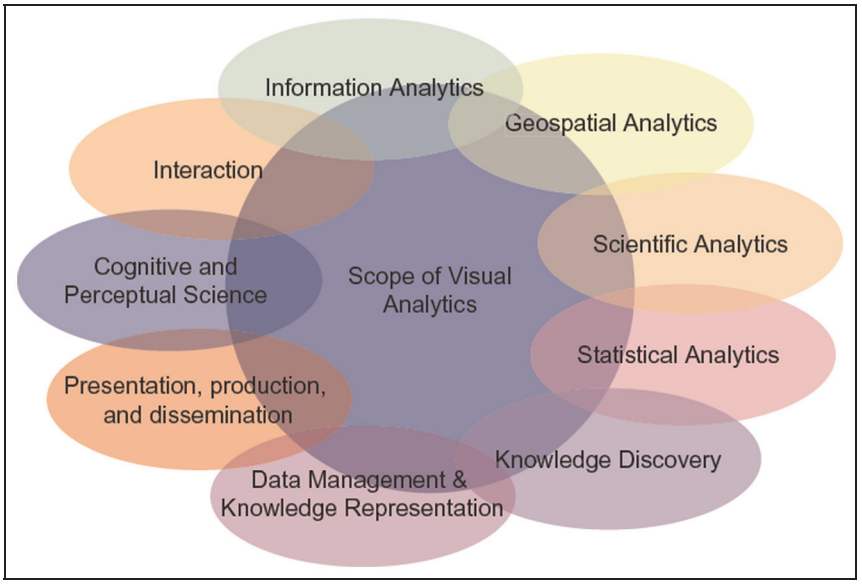

Figure 1. Visual analytics as a highly interdisciplinary field of research.[3]

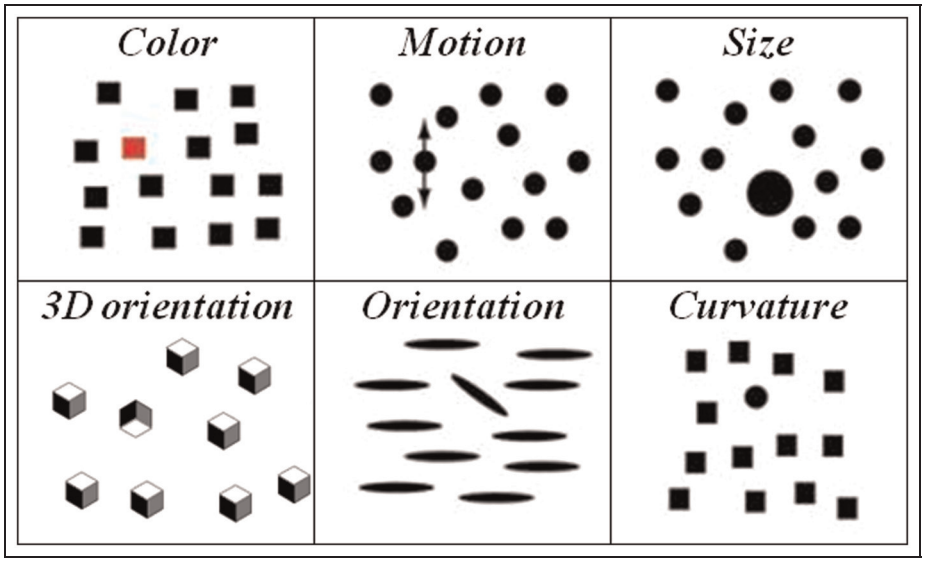

Figure 2. Examples of pre-attentive visual features (adapted from Healey[5]).

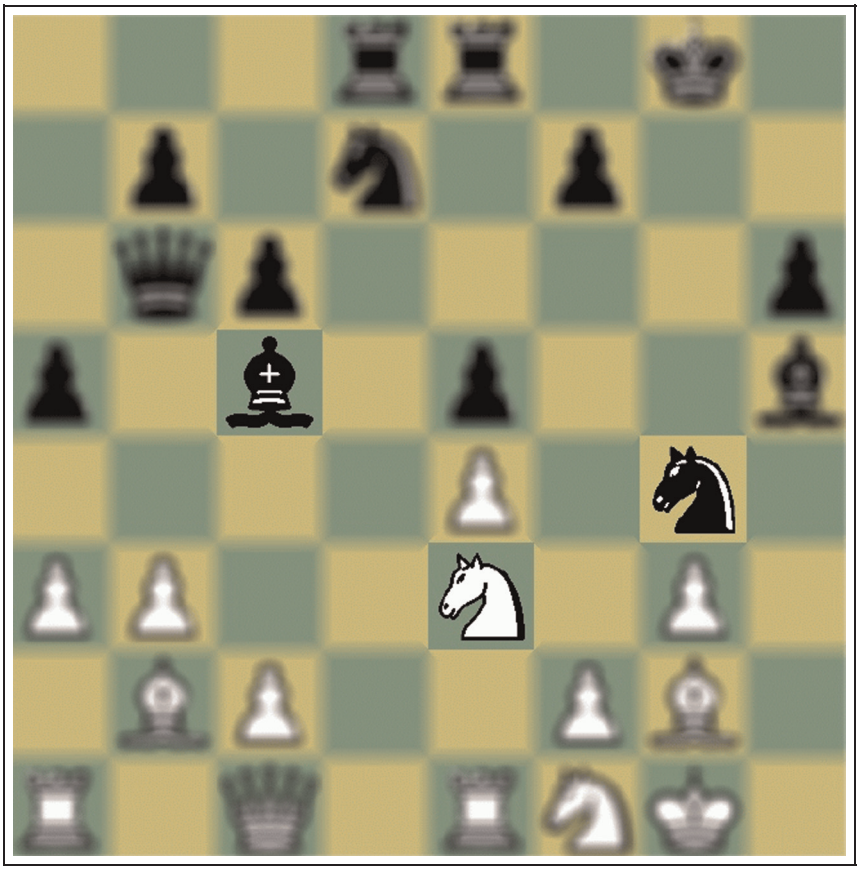

Figure 3. Semantic depth of field relies on pre-attentive features to offer a focus + context view. Focusing effects can highlight information.[6]

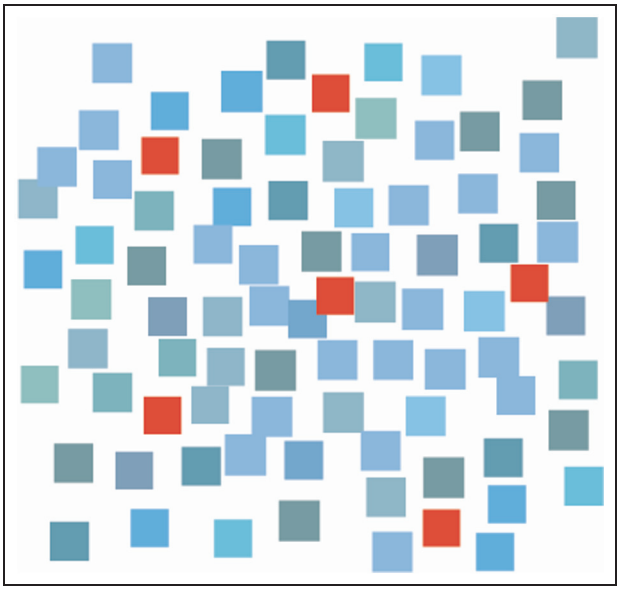

Figure 4. Contrast and analogy.[7]

scene when the observer is not expecting them. Salient features within the visual field will not be observed if not processed by attention because the amount of information processed at any particular time is limited. In an experiment, where subjects were asked to watch a short video and count the number of passes between two groups of basket ball players, 50% of subjects failed to notice that a person wearing a gorilla suit had walked through the scene.[11]

2.2. Interaction

Interaction enables the user to explore the data, try out hypotheses, drill into data, gain insight and collect knowledge. Human–computer interaction has been an active research field for many years and is the study of interaction between people and computers. Over the last years, the number of available devices for interacting with digital information has grown significantly, creating a variety of new challenges for interaction design. Visualization is considered interactive if control of some aspect of the information presented is available through a human input and the changes made by the user are incorporated in a timely manner.

Three categories of responsiveness (0.1, 1 and 10 s) have been suggested to give an order of magnitude of the required response time for interactivity.[12,13] 0.1 s is the upper limit for the system response to feel instantaneous. After more than 1 s, the user's flow of thought is interrupted and the user loses the feeling of operating directly on the data. For delays longer than 10 s, users will want to perform other tasks while waiting for the computer calculation to complete.

The famous visual information-seeking mantra for designing advanced graphical user interfaces ''overview first, zoom/filter, details on demand''[14] was modified to propose a VA mantra: ''analyse first - show the important - zoom, filter and analyse further - details on demand''.[15]

''To be most effective, visual analytics tools must support the fluent and flexible use of visualizations at rates resonant with the pace of human thought''.[16] For this purpose, a taxonomy of interactive dynamics that contribute to successful analytic dialogues, consisting of 12 interactions, organized into three categories, was developed and is presented in Table 1.[16]

2.3. Analytical reasoning

VA is intended to be an active, engaging exploratory process of discovery. This human–information discourse is between the analysts and their data. It supports three goals: assessment (understand current situation and explain past events); forecasting (estimate future capabilities, threats, vulnerabilities and opportunities); and planning (develop options, create possible scenarios, prepare reactions to

Table 1. Taxonomy of interactive dynamics for visual analysis.[16]

Data & view specification	**Visualize** data by choosing visual encodings.
	Filter out data to focus on relevant items.
	Sort items to expose patterns.
	Derive values or models from source data.
View manipulation	**Select** items to highlight, filter or manipulate them.
	Navigate to examine high-level patterns and low-level detail.
	Coordinate views for linked, multi-dimensional exploration.
	Organize multiple windows and workspaces.
Process & provenance	**Record** analysis histories for revisitation, review and sharing.
	Annotate patterns to document findings.
	Share views and annotations to enable collaboration.
	Guide users through analysis tasks or stories.

potential events). Analysts apply reasoning techniques in order to achieve these goals.

Data overload can be defined as ''a condition where a practitioner, supported by artifacts and other practitioners finds it extremely challenging to focus in on, assemble, and synthesize the significant subset of data for the problem context into a coherent situation assessment, where the subset is a small portion of a vast data field''.[17] Insight is ''thought to arise when a solver breaks free of unwarranted assumptions, or forms novel, task-related connections between existing concepts or skills''.[18] It is beyond the scope of this paper to cover the cognitive foundations of VA extensively, but a comprehensive review is provided by Greitzer et al.[19] VA is meant to facilitate high-quality analysis with limited user's time. Six basic ways were identified in how information visualization can expand human cognition:[4,20]

- increased resources: high-bandwidth hierarchical interaction, parallel conceptual processing, offload of work from cognitive to perceptual system, expanded working memory and expanded storage of information;
- reduced search: locality of processing, high data density and spatially indexed addressing;
- enhanced recognition of patterns: recognition instead of recall, abstraction and aggregation, visual schemata for organization and enhanced patterns and trends;
- perceptual inference: visual representations make some problems obvious, and complex specialized, graphical computations can be enabled;
- perceptual monitoring: visualizations can allow monitoring of a large number of potential events.
- manipulation medium: visualizations can allow exploration of a space of parameter values and amplify user operations.

The model of sense making for intelligence analysis,[21] presented in Figure 5, was derived from a study using cognitive task analysis and verbal protocols. According to this commonly used model, there are two main loops of activities that can occur iteratively:

- a foraging loop that involves processes aimed at seeking information, searching and filtering it, and reading and extracting information[22] possibly into some schema, and
- a sense-making loop[23] that involves iterative development of a mental model (a conceptualization) from the schema that best fits the evidence.

This model is not incompatible with the Data/Frame Theory, ''that posits a closed-loop transition sequence

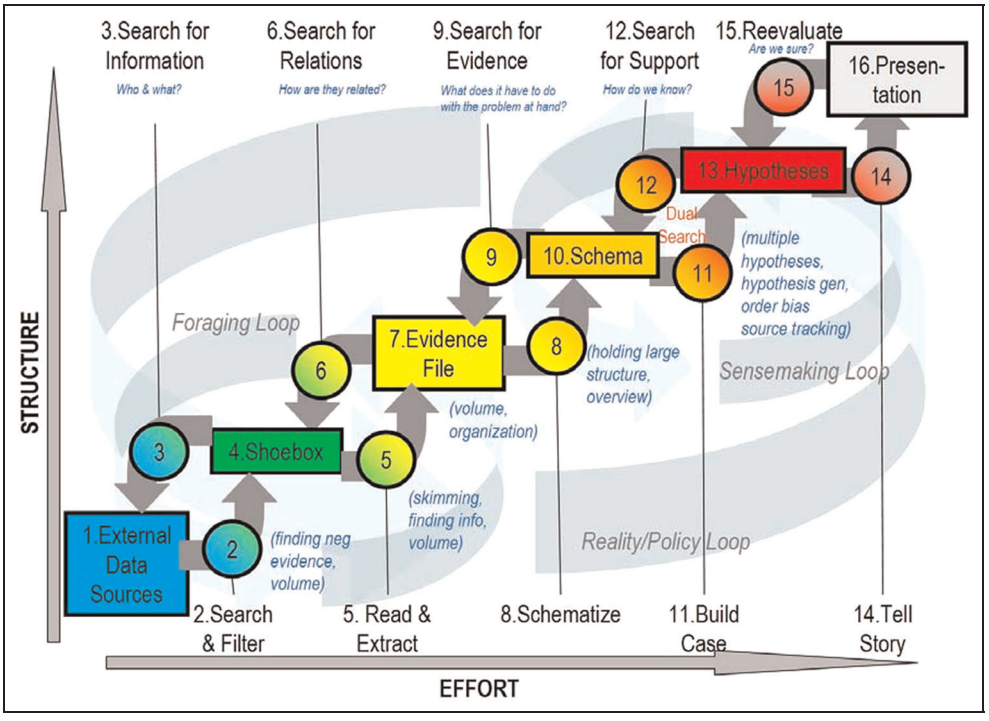

Figure 5. Notional model of sense-making loop for intelligence analysis.[21]

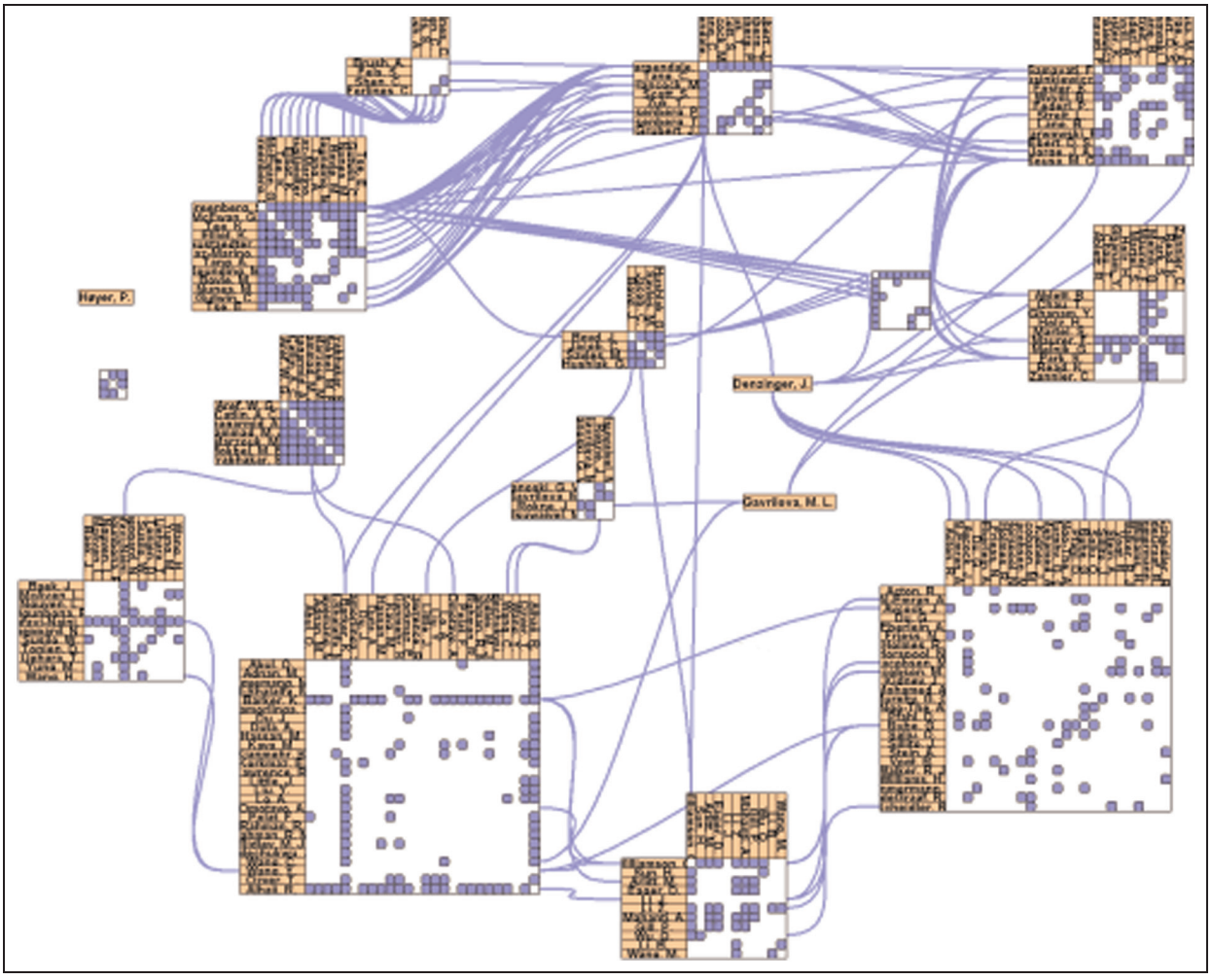

Figure 6. NodeTrix social network visualization.[25]

between a mental model formation (which is backward looking and explanatory), and mental simulation (which is forward looking and anticipatory)''.[24]

3. Advanced Visual Analytics concepts and techniques

This section presents a variety of advanced VA concepts applied to various information types to illustrate the breadth of VA techniques.

In network information, be it a graph or a hierarchical representation, the relevant aspect is mainly the links between the data elements. However, in many situations, the number of links can quickly grow and result in an undecipherable mesh of data relationships. Innovative visualizations are required to untangle these spider webs of links and make sense of the information presented.

NodeTrix is a hybrid approach to social network visualization.[25] Node-link diagrams are used to show the global structure of a network, while arbitrary portions of the network can be shown as adjacency matrices to better support the analysis of communities. It is especially useful in the case of globally sparse but locally dense social networks (Figure 6).

Temporal data analysis is useful to detect trends and recurring events over time. The study of event sequences also enables the identification of links between individual observations and possible causes for some events. The use of timelines provides an overview of what happened in a given time interval, while a time slider can be used to show unfolding events in ascending or descending time order. Time sliders are especially effective when one is faced with multiple dimensions that must be represented, as is the case when combining geospatial and temporal data, for example.

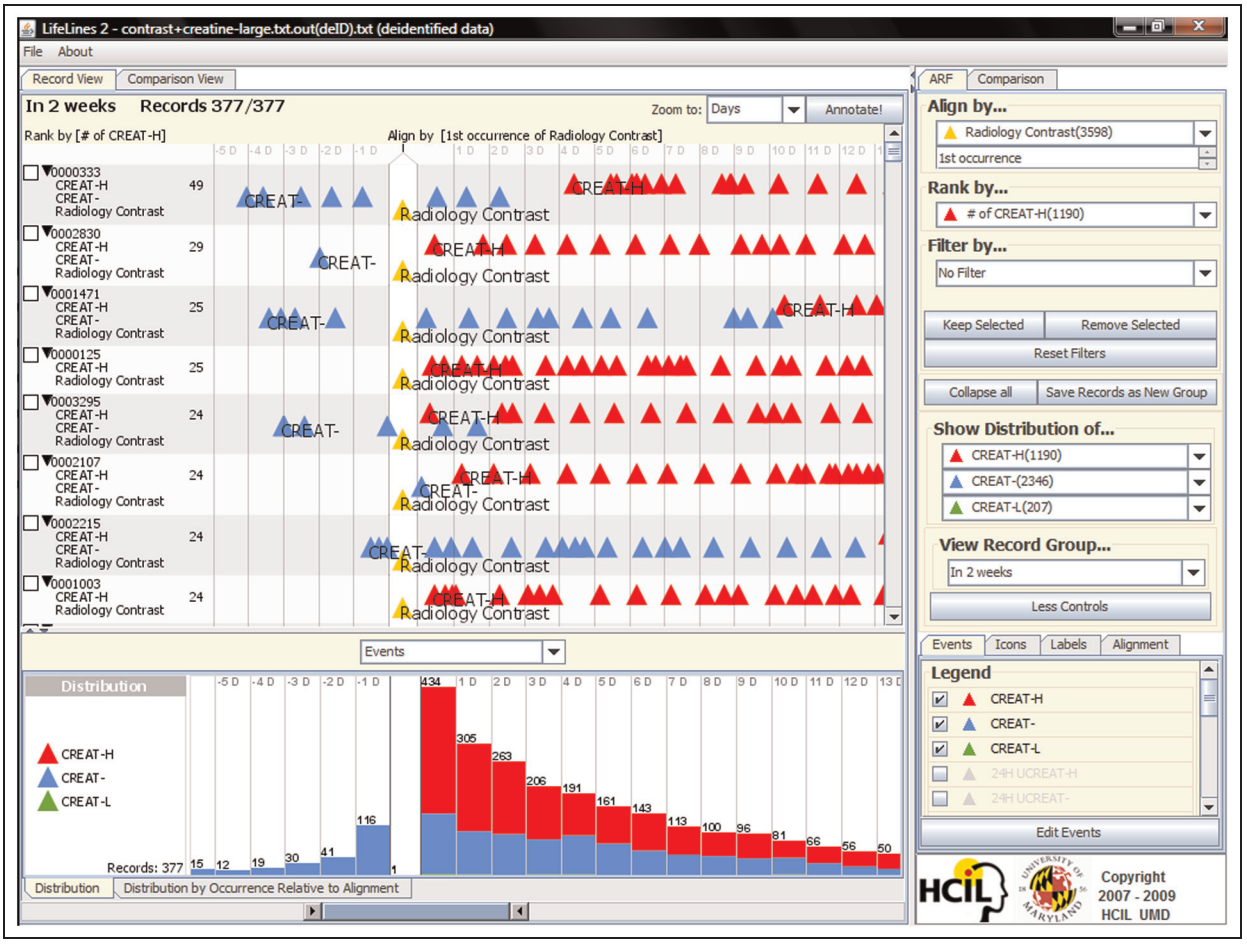

Figure 7. Lifelines2 visualization of multiple electronic health records.[26]

Lifelines 2[26] is an interactive visualization tool that organizes electronic record information in a temporal display to allow the discovery of patterns across multiple records, hypothesis generation and finding cause-and-effect relationships in a population (Figure 7).

Many hierarchical datasets can be displayed with the techniques used for graph and network data. However, some visualizations, such as treemaps, take advantage of the specific characteristics of hierarchical datasets.

Treemaps show data as a set of rectangles where the smaller rectangles contained inside a larger rectangle are sub-branches. The market tree map visualization in Figure 8 shows the changes in various financial areas for over 500 stocks. It is a tool that can be used to spot trends and investment opportunities. The rectangle size is proportional to the company's market cap, while the color shows price performance. The overview that this treemap visualization offers makes it obvious that on November 23, 2010, the economy was not going very well in the US. Each region can be selected to view more details about it. For example, the very bright rectangle in the bottom left part is the New York Times stock, which increased by 6.7%.While most treemaps are based on rectangular regions, other space division schemes were explored such as circular treemaps (inefficient use of space) and Voronoi treemaps.

Datasets with more than three dimensions can be very difficult to represent. A few techniques were developed in order to reduce the dimensionality of the data while being able to show the important characteristics of the datasets.

Parallel coordinates is a common way of visualizing high-dimensional data and dates back to 1885.[28] This view represents variables with parallel axes and data elements are plotted by crossing these lines at the height corresponding to the value of each attribute (Figure 9). Reordering the dimensions or filtering data through interactive queries can help pattern finding.[29] This concept has been extended to three dimensions using parallel planes (Figure 10).[30]

The Dust & Magnets metaphor represents individual cases as particles of iron dust, and dataset variables are

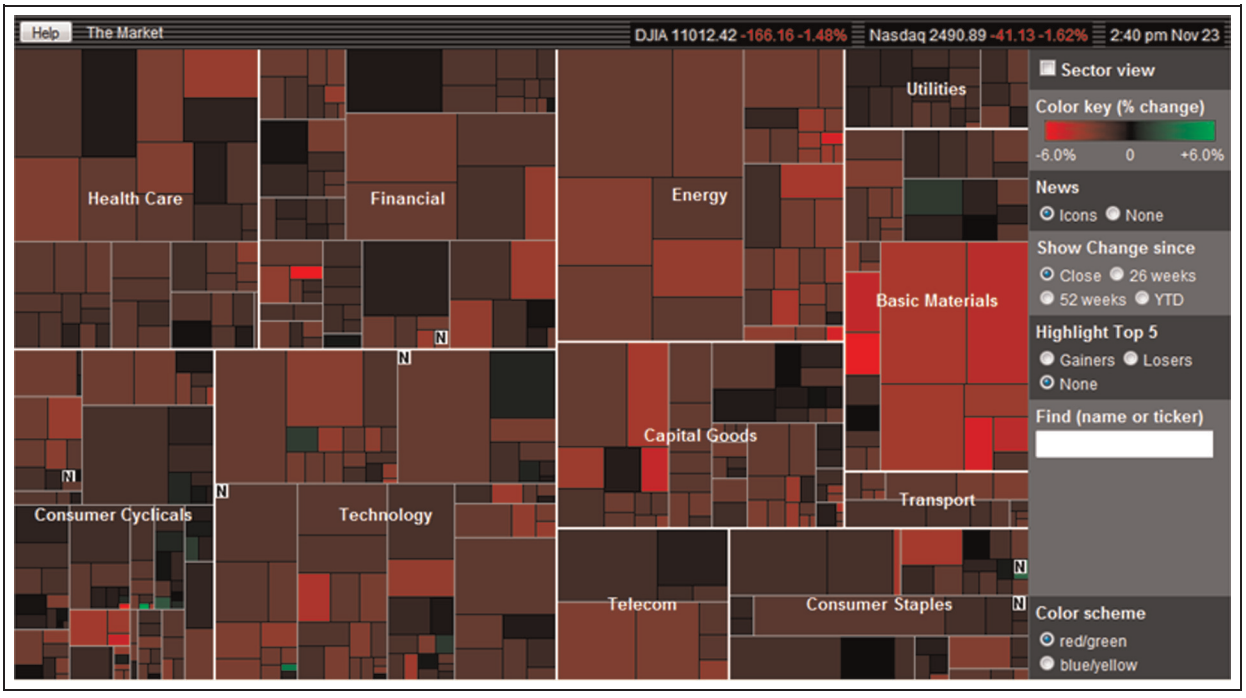

Figure 8. SmartMoney's treemap view of markets on November 23, 2010.[27]

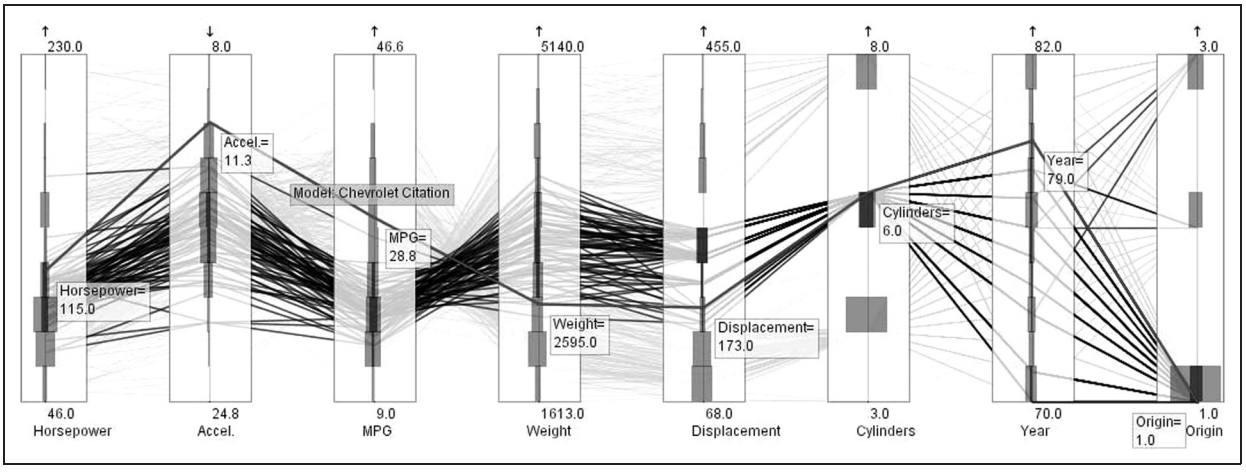

Figure 9. Extended parallel coordinates view representing car attributes.[29]

represented as magnets.[31] This enables the user to manipulate the magnets and see the dust particles move accordingly. When a magnet is dragged, individual dust particles are attracted to the magnet based on the value of the attribute corresponding to the magnet. The dataset characteristics are exposed through the interaction with the magnets, enabling the user to get a feel of the importance of each variable and of the relations between them (Figure 11).

When multiple organizational views of the same information are used, as the user interacts with information in any view, the relational changes are visualized in all views. Multiple views are a great tool to help human cognition.

> First, humans perceive information in a variety of ways including through the filter of their own assumptions, patterns are more likely to be discovered if represented in multiple

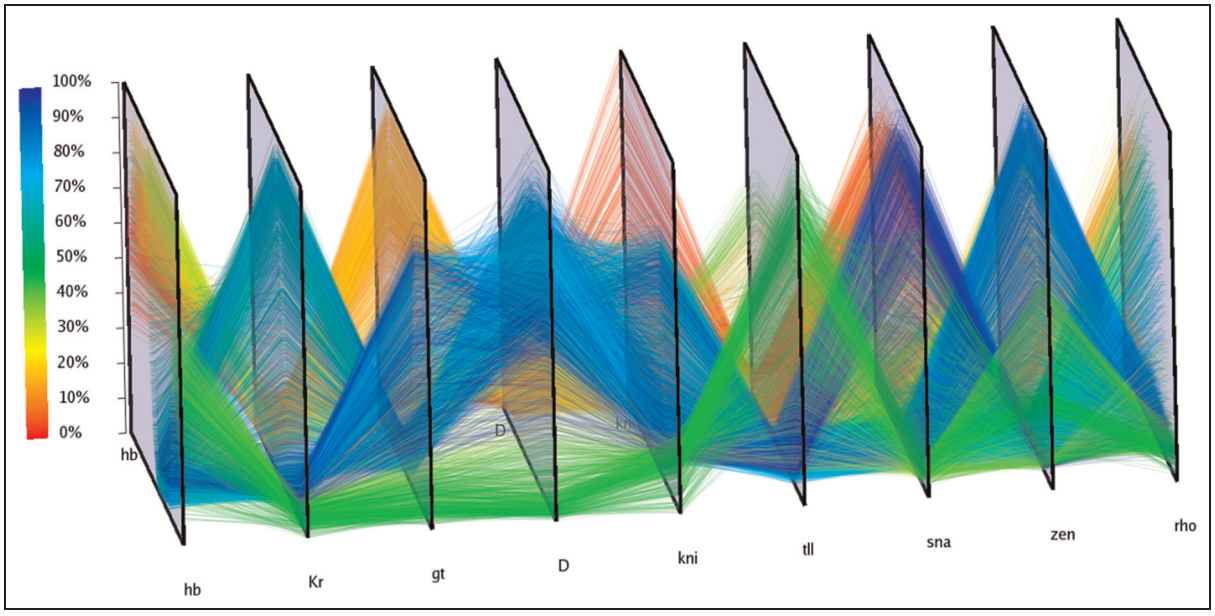

Figure 10. Three-dimensional parallel coordinates example with cells and genes data.[30]

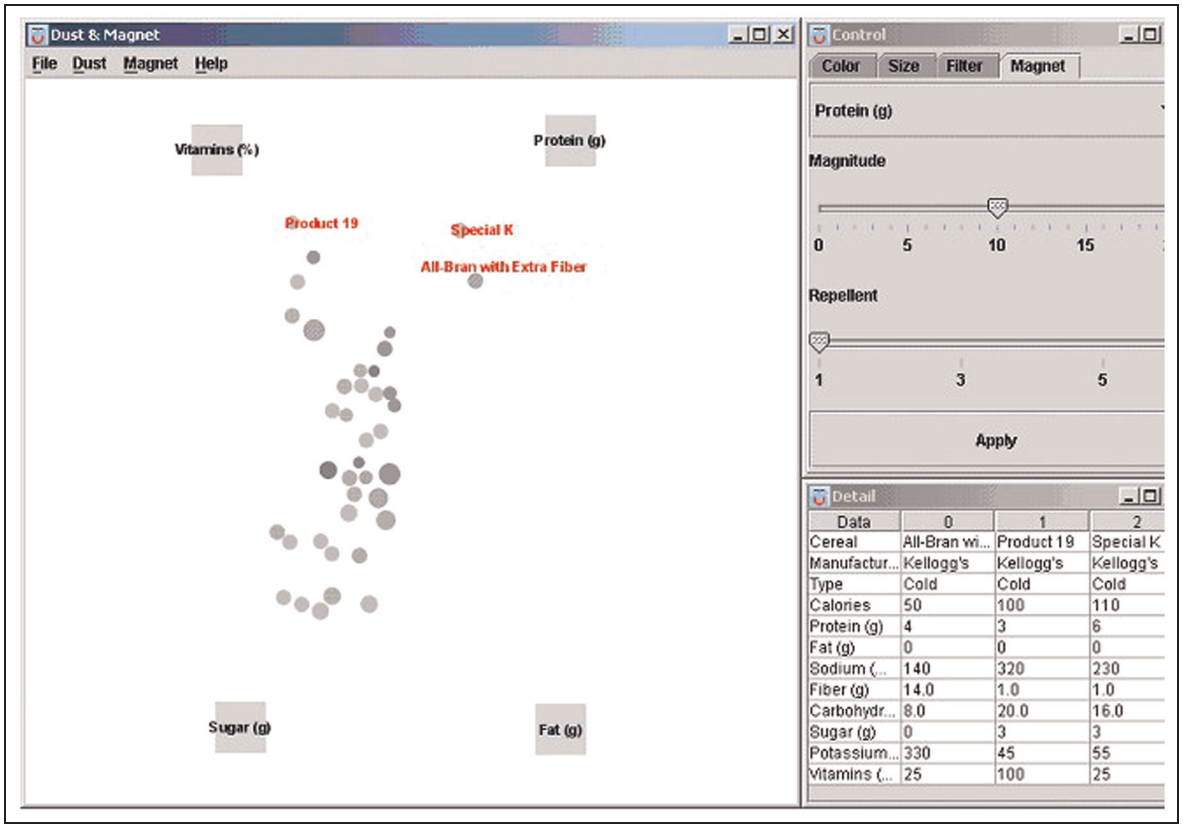

Figure 11. Dust & Magnet example using a cereal dataset.[31]

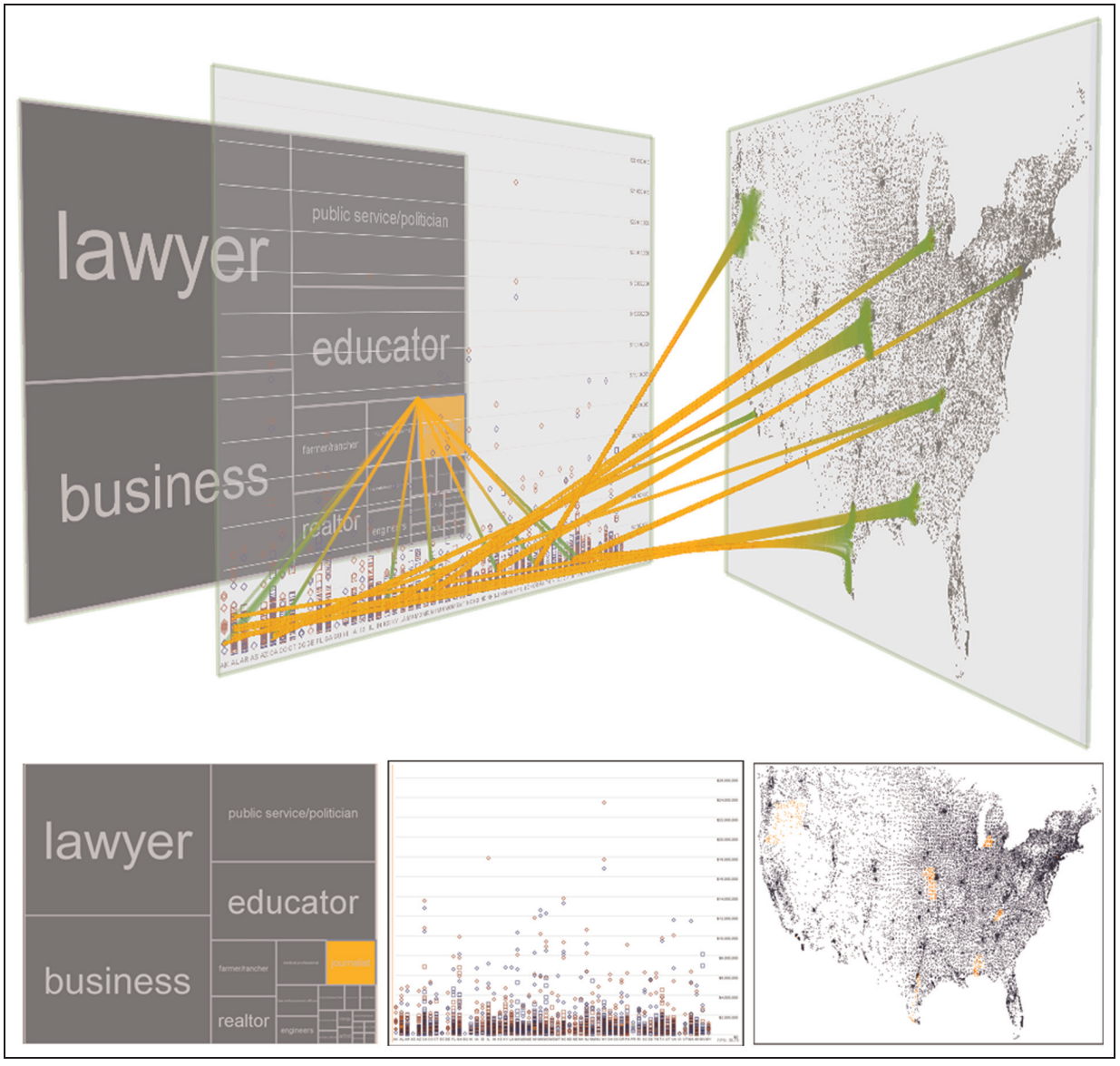

Figure 12. Application of VisLink to bridge existing visualizations. Views of the constituent visualizations, from the two-dimensional equivalency mode, are shown along the bottom.[33]

ways, each tuned to particular, important aspects of the data. Secondly, humans prefer to narrow down the field of choices by eliminating those that do not possess desired attributes. This is usually done before utilizing more complicated heuristics. Multiple views make the process easier; multiple layers of relational attributes are readily knowable without additional search. Thirdly, multiple views enable more intuitive manipulation. Humans themselves do not interact with information in one dimension; humans are capable of multi-layered processing: perceptual, emotional and higher-cognitive. Indeed, of all the guidelines we will discuss, use of multiple views is the one most likely to lead to spontaneous insight.[32]

VisLink is a method by which visualizations and the relationships between them can be interactively explored. It supports multiple visualizations, empowers inter-representational queries and enables the reuse of the spatial variables, thus supporting efficient information encoding and providing for powerful visualization bridging.[33] This approach uses multiple two-dimensional (2D) layouts, drawing each one in its own plane. These planes can then be placed and re-positioned in three-dimensional (3D) space: side by side, in parallel or in chosen placements (Figure 12). Relationships, connections and patterns between visualizations can be revealed and explored using

a variety of interaction techniques, including spreading activation and search filters.

4. Cyber security

One of the defense and security domains that VA can contribute to is cyber security. As the modern world is relying heavily on computers and networks to conduct day-to-day activities, these computers and networks have become increasingly a target of choice for countries conducting spying or disruptive operations, terrorist and criminal organizations or simply hackers. The impact of cyber attacks on a country, an organization or individuals can be severe and costly. Moreover, network attacks are increasingly sophisticated and unpredictable.

''The broad area of cyber security encompasses policy and configuration decisions, virus scanning, monitoring strategies, detection and reactions''.[34] Security tools are necessary to properly manage computer networks and prevent and detect intrusions. This includes tools to analyze service usage in a network, detect a distributed attack and investigate hosts in a network that communicate with suspect external internet protocols (IPs). One key requirement for these tools is the ability to process and filter massive amounts of information.

Through an extensive study of how computer defense network analyses perform their work,[34,35] six broad analysis roles are identified:

- triage analysis: weed out false positives, escalate suspicious activity for further analysis;
- escalation analysis: analyze data over longer time then triage, incorporate multiple data sources;
- correlation analysis: look for patterns and trends, assess similarity to related incidents;
- threat analysis: characterize attackers (identification, modus operandi, motivation, location);
- incident response: recommend, implement courses of action; support law enforcement investigations;
- forensic analysis: collect and preserve evidence, support law enforcement investigations.

Three important criteria for cyber defense tools are also identified: ''ease of input and output from the tool, support for report building, and management of evidence and analysis''.[34] Cyber analysts are better served by using a tool box of different visualizations.[36] Figure 13 shows five categories of uses of visualization and how they relate to the three stages of situational awareness.

Globally, VA can improve cyber security with capabilities to:[37]

- recognize risks and protect against cyber threats, allowing for more effective attack prevention and faster isolation and mitigation of attacks;
- enable key aspects of the digital forensic process, including data collection, discovery, investigation, examination, analysis and reporting capabilities for information discovery;
- allow information discovery, processing and visualization (tactics that apply across many applications for computer security and forensics).

> Incorporating visual analytics into an organization's best practices allows computer security professionals to quickly identify threats to their own organizations. By doing so earlier and more comprehensively than their competitors, this leads to significant competitive advantage in the face of increasing threats and daily attacks.[37]

VA techniques have been explored and put in service to counter cyber security. There are many VA applications related to network analysis.

> Visualization is often appropriate when human intelligence and domain knowledge must be combined with automated methods. This is certainly the situation with monitoring and exploring network traffic patterns. The sheer number of alerts and the sophistication of attacks require a symbiosis of Intrusion Detection Systems (IDS) algorithms and human analysis to fight new adversaries.[38]

The NFlowVis Network visualization tool provides a number of views used to perform large-scale network

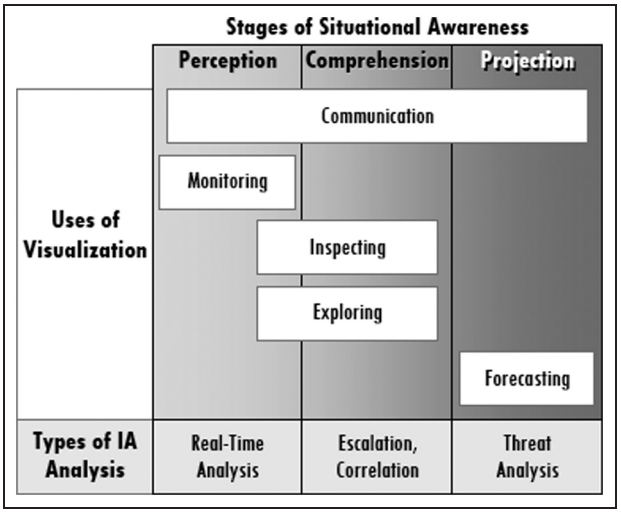

Figure 13. Relationship between the stages of situational awareness, the uses of visualization and the types of analysis performed.[36]
IA= Information Assurance.

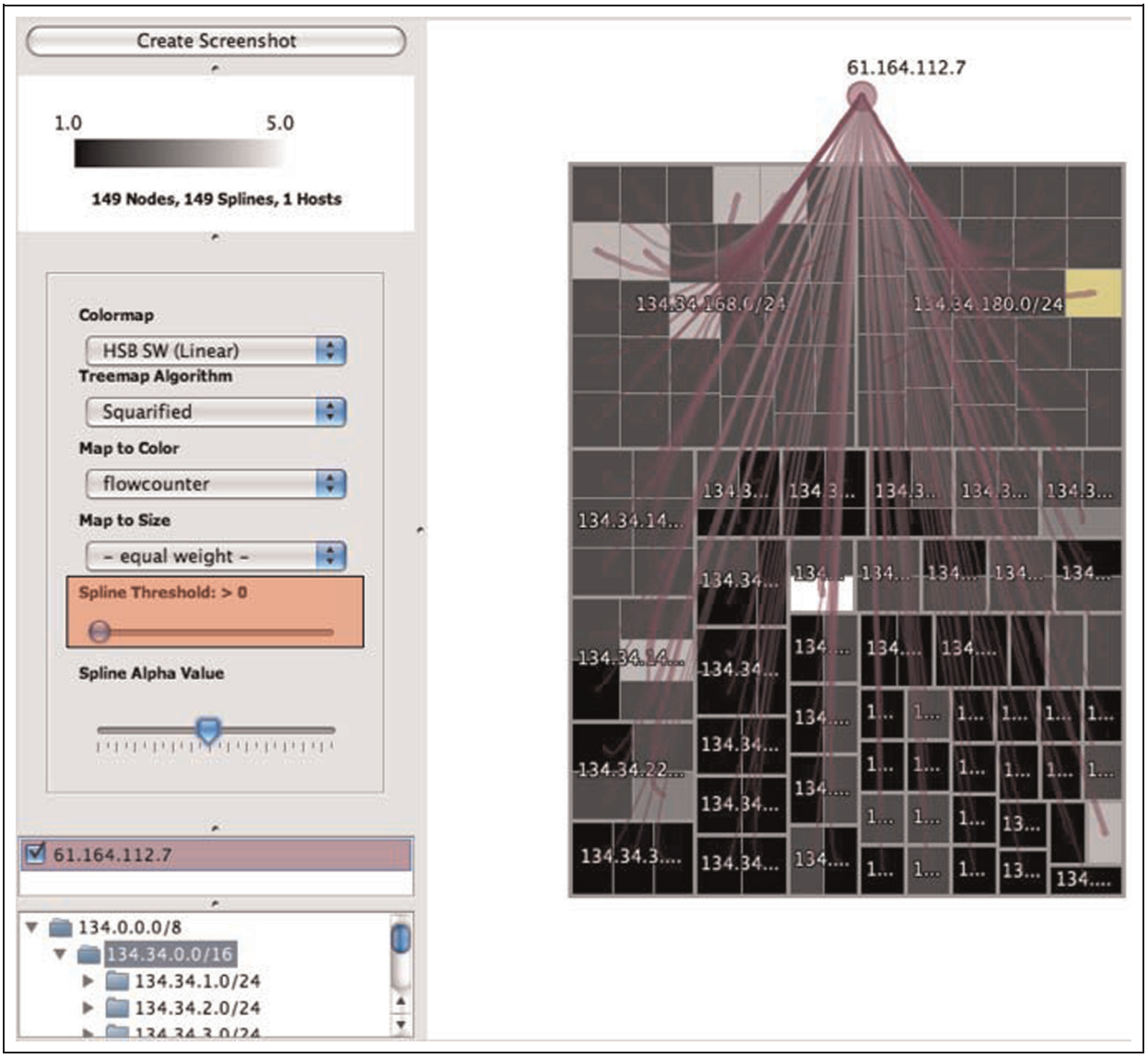

Figure 14. Example of NFlowVis showing the identification of compromised hosts using threshold adjustments.[38]

traffic monitoring, to detect distributed attacks and to analyze intrusion detection events. For example, as shown in Figure 14, a treemap is used to represent compromised hosts in the center of the display and selecting attacking hosts arranged at the borders of the display. Figure 15 is a graph visualization showing communication flows between source and destinations hosts.

The design of the Visual assistant for Information Asssurance analysis (VIAssist) was informed by cognitive tasks analysis activities.[35] This visual analysis platform (Figure 16) provides one view for in-depth event analysis and a dashboard view for global activity.[39] Different kinds of visualizations are provided to enable the analysis of events in network, temporal and geographic contexts. Multiple visualizations are linked together to facilitate exploration and discovery. It is worth noting that VIAssist was transitioned into operational use by the Department of Homeland Security (DHS).[40]

ManyNets is a network visualization tool with tabular interface designed to visualize up to several thousand network overviews at once (Figure 17). This allows networks to be compared and large networks to be explored using a divide-and-conquer approach. A collection of networks is presented in a table, where each row represents a single network. Columns represent statistics, such as link count, degree distribution or clustering coefficients. Networks

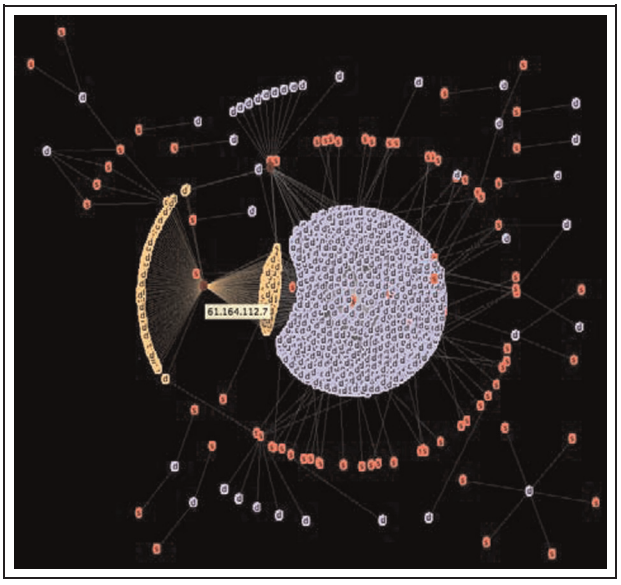

Figure 15. Example of NFlowVis showing communication flows between source and destination hosts.[38]

can also be subdivided and compared based on motifs (small patterns of connectivity), clusters or network-specific attributes.[42]

When it comes to forensic activities, history trees can be useful for providing traceability and the history of the analyst's workflow (Figure 18).

Cyber security is not limited to protecting a network from intrusions and includes other activities, such as malware and vulnerability analysis. DHS listed software assurance as the number one hard problem for cyber security: ''poorly written software is at the root of all of our security problems''.[44] No software security analysis tool is able to detect all the vulnerabilities and a visual analysis environment (Figure 19) was developed to integrate the result from disparate software analysis tools into a unified system.[45] According to the authors, here are the benefits of this approach:

- ''More software analysis tools mean more vulnerabilities will be detected, and vulnerabilities that are detected by multiple tools have a higher confidence that they are accurate;
- Providing interactive information visualization presents results in an understandable format and grants the ability to focus on the most important vulnerabilities; and
- Integration with Systems Development Lifecycle tools provide a streamlined workflow that gives developers more time for coding and less time trying to interpret results''.[45]

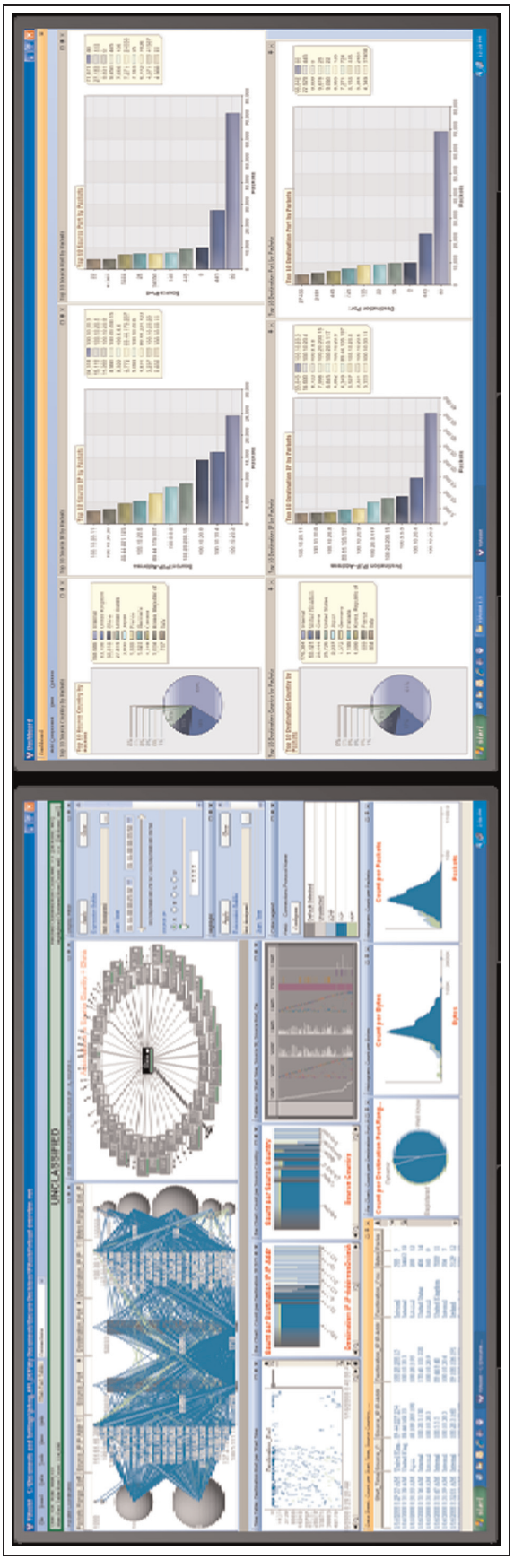

Figure 16. The VIAssist visual analytics platform.[41]

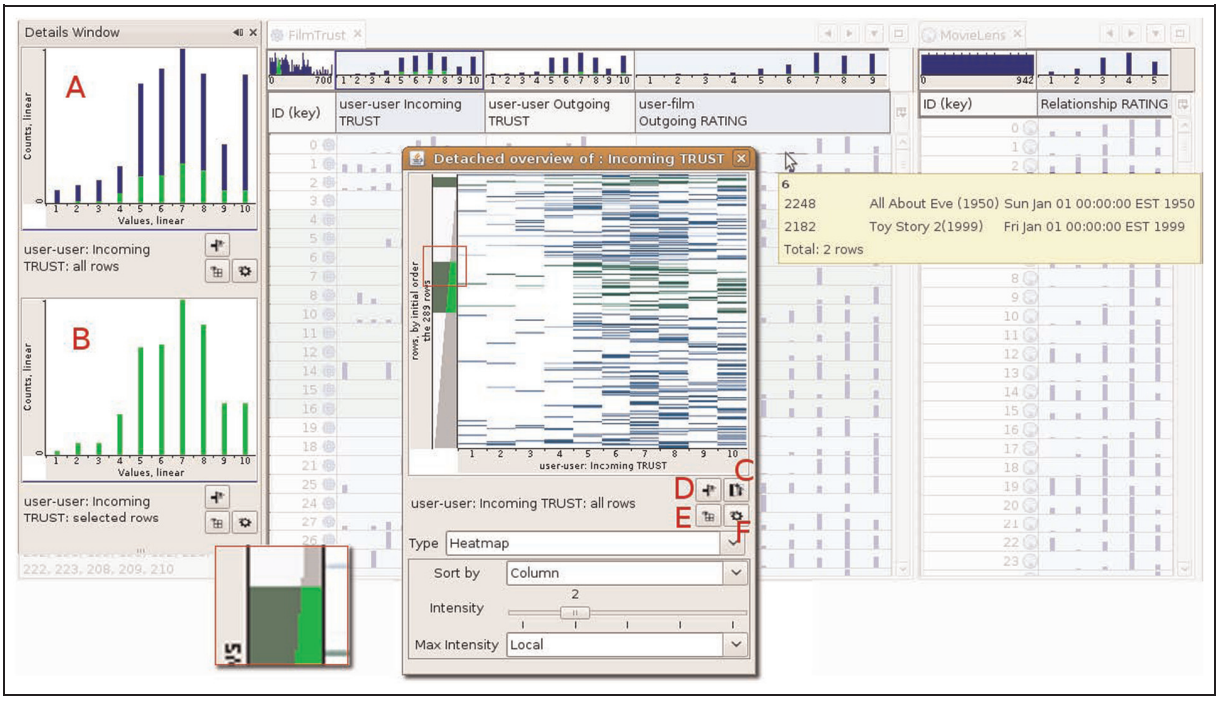

Figure 17. ManyNets is a tool for the simultaneous visualization of many networks.[42]

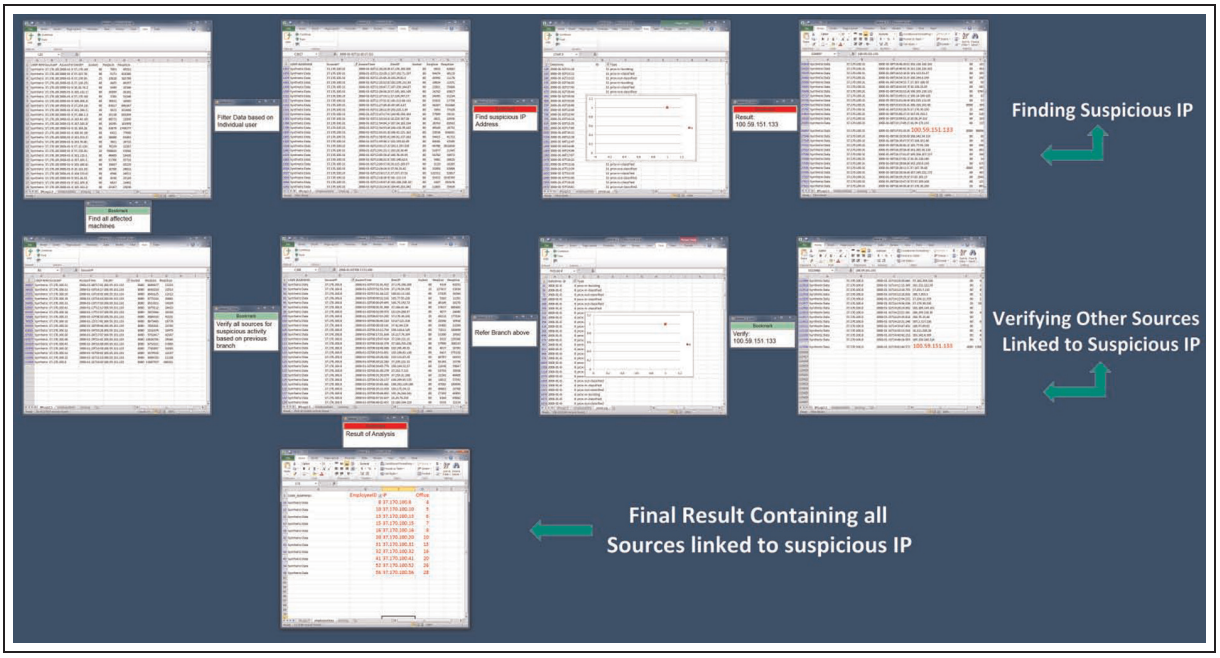

Figure 18. History trees.[43]

Starlight Visual Information System is a VA platform where viewers can interactively move among multiple representations of the information. Starlight can be used for cyber security and computer forensics; the use of Starlight for cyber network analysis is depicted in Figure 20. This commercial software tool can also apply to other

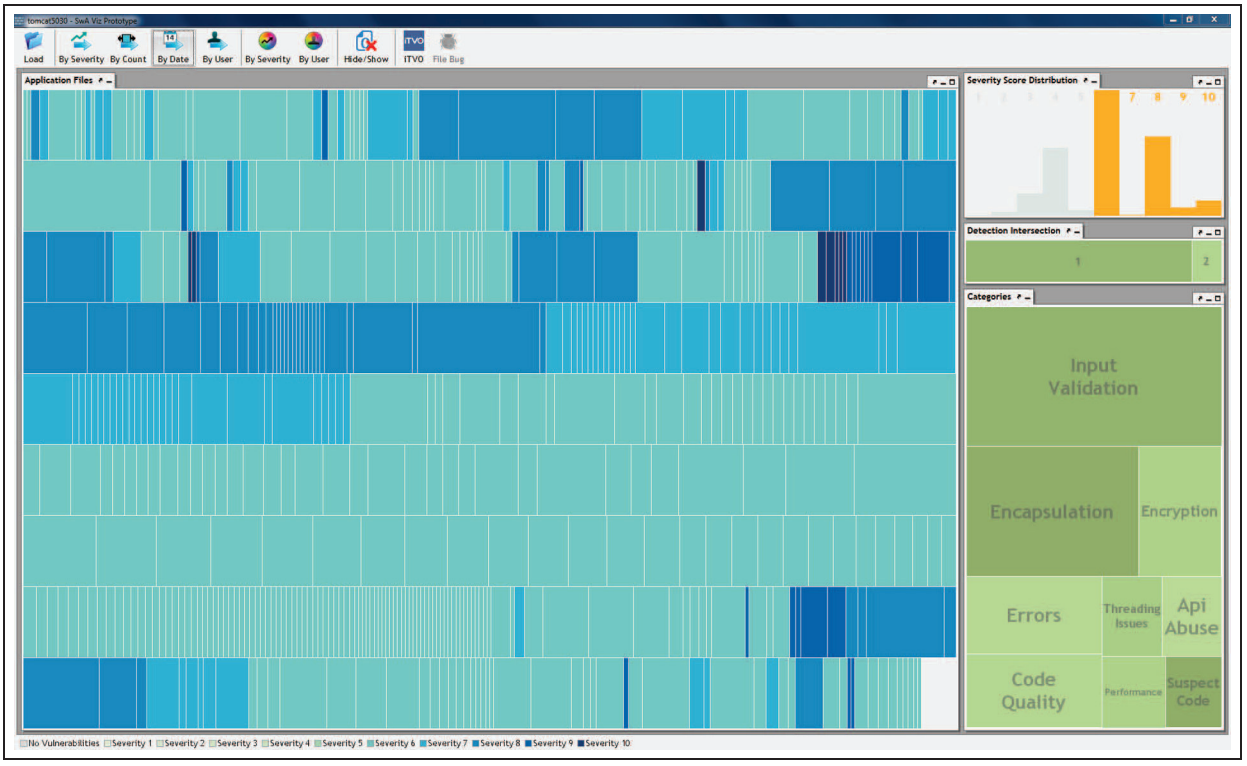

Figure 19. Visualization that shows nearly 34,000 vulnerabilities identified by three software analysis tools.[45]

defense and security domains, such as intelligence analysis, counterterrorism and counter-insurgency (COIN).

5. Intelligence, counterterrorism and counter-insurgency

The purpose of intelligence is to provide commanders and staffs with timely, relevant, accurate, predictive and tailored intelligence about the enemy and other aspects of the area of operations. Intelligence supports the planning, preparing, execution and assessment of operations.[47] In modern conflicts, intelligence activities are more and more concerned with COIN and counterterrorism.

COIN is defined as ''a part of a wider set of irregular activities and threats to a secure and stable environment'', where an irregular activity may be defined as: ''behaviour that attempts to effect or prevent change through the illegal use, or threat, of violence, conducted by ideologically or criminally motivated non-regular forces, groups or individuals, as a challenge to authority''.[48]

> Terrorism is a global threat influencing the attitude and behaviour of a target group by threatening, or carrying out, devastating actions. These actions, as we have seen, can and do include the use of conventional weapons, biological, chemical, or nuclear agents. Today's terrorism also threatens our economic and information resources. Our vulnerability to terrorist attacks expands with our growing reliance on information technologies. Increased access to information and the centralization of vital components of local, national, and global infrastructure threaten both local and national security.[49]

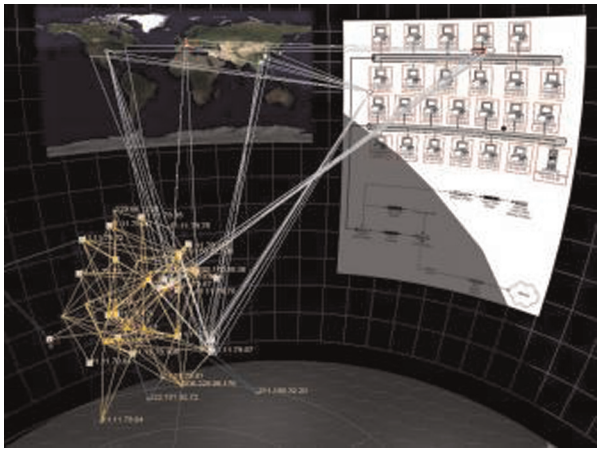

Figure 20. Use of Starlight for cyber analysis.[46]

Inter-agency communication and collaboration is essential to investigate terrorist groups. Law enforcement and intelligence agencies need to work together to collect and analyze data from multiple data sources in order to monitor, penetrate, infiltrate and prevent terrorist activity. A particular attention is needed on preparation activities, such as money transfers, material purchases and personnel movement.

> Actionable intelligence is essential for preventing acts of terrorism. The timely and thorough analysis and dissemination of information about terrorists and their activities will improve the government's ability to disrupt and prevent terrorist acts and to provide useful warning to the private sector and our population.[50]

Intelligence analysis has been a primary target domain for VA applications, but relatively little user task analysis activities were conducted. A study observing experienced intelligence analysts perform a relatively complex intelligence analysis task using a baseline set of querying and browsing tools and concluded that ''baseline tools do not adequately support analysts in meeting the challenges of performing inferential analysis under data overload, leaving them open to making inaccurate statements and missing critical events''.[51] More recently, another study involving intelligence analysis students was performed to characterize their analytical process and methods.[52] The intelligence process they observed appeared to be more parallel and organic than was suggested by the sense-making model from Figure 5. The analysts proceed in parallel steps, building their conceptual model at the same time as they are investigating for new information, producing analysis reports and checking the credibility of previously collected information. From this study, they suggested several design implications for systems supporting intelligence analysis:

- externalize the thinking process: help analysts continuously build a conceptual model;
- support source management: for both pushed and pulled information and organizing sources;
- support analysis with constantly changing information: integrate collection and analysis in a single system and help analysts use structured methods during collection;
- help analysts create convincing production: support insight provenance and sanity checks of analytical products;
- support asynchronous collaboration rather than synchronous collaboration for exploratory analysis;
- unifying the pieces: seamlessly bring together capabilities now scattered into multiple software systems.

> Intelligence is about determining how to answer a question, what to research, what to collect, and what criteria to use. This process becomes part of the analysis - analysis implicitly occurs during the process of the construction. Analysts also explore different sets of analytic techniques to address a problem.[52]

In order to solve complex, multifaceted, real-world problems, intelligence analysts need to develop an understanding of various collections of data and link together information of different types. The use of the previously mentioned Starlight software for intelligence analysis is depicted in Figure 21. This platform enables the visualization of multiple data collections simultaneously in order to uncover correlations that may span multiple relationship types, including networks, geographical data and textual information.

Analysts are also often faced with large collections of unformatted text documents. IN-SPIRE is a text analysis and visualization software that can quickly reveal important information from these datasets and accelerate subsequent investigation and discovery. IN-SPIRE's two main visualizations display representations of the documents in which those with similar or related topics appear closer together (Figure 22). In the Galaxy visualization (upper right), dots represent documents and cluster around center points that represent central topics or themes. In the ThemeView visualization (lower right), users see a relief map where the highest peaks represent the most prevalent topics in the collection.

Oculus nSpace is a web browser-based system of systems for intelligence analysis meant to support multiple analytical styles and workflows.[54] It is the combination of two capabilities called TRIST and Sandbox (see Figure 23). TRIST's multi-dimensional linked views help users find relevant documents (unstructured text, images, videos, etc.) from web services. Queries can be saved and scheduled to be executed repeatedly. The Sandbox is a space where relevant information can be dropped and where analytical sense making happens. Elements can be grouped and collapsed. Graphs and networks are supported and matrices allow analysis of competing hypotheses using groups of evidence. The Sandbox supports flexible visual cognition through spatial arrangement.

As part of the National Consortium for the Study of Terrorism and Responses to Terrorism (START), tools have been developed to consult the Global Terrorism Database. This database holds information about successful and missed terrorism attacks across the world since the 1970s. Figure 24 shows a Theme River representation of terrorism attacks in the world over time. Stripes of different colors show how many terrorist attacks occurred in the corresponding country each year. The number of attacks is represented by the thickness of the stripe at the year on the x-axis.[55]

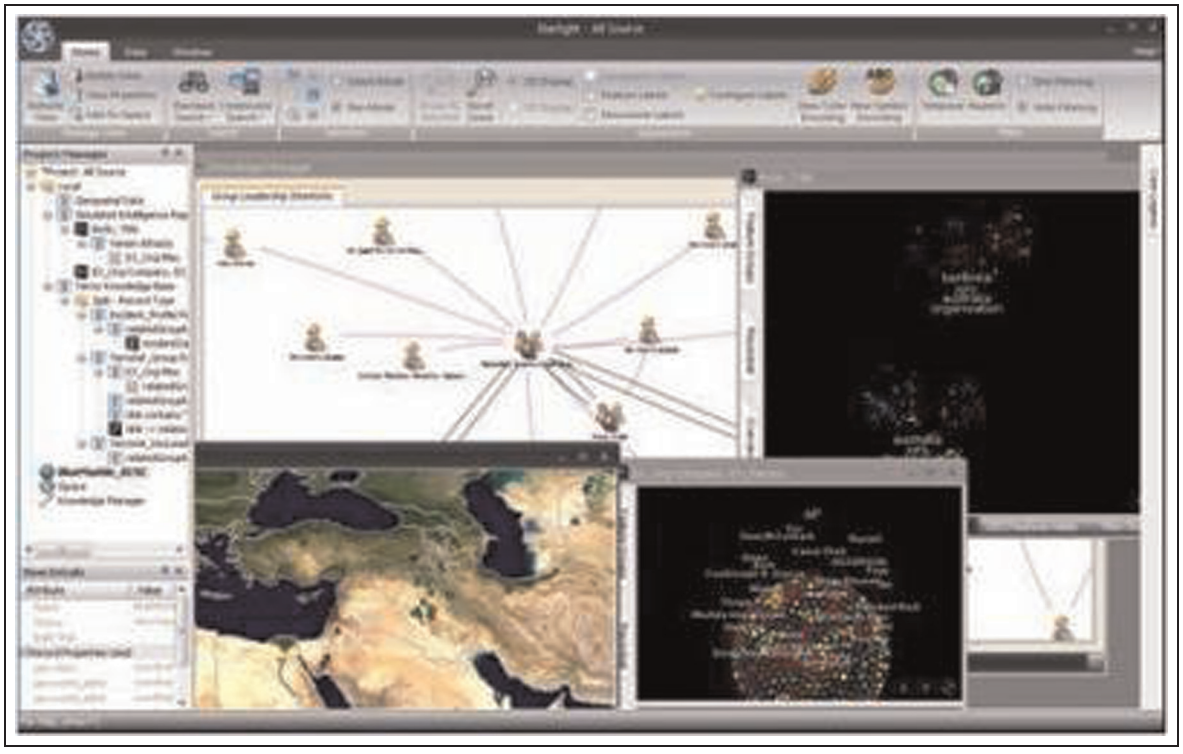

Figure 21. Use of Starlight for intelligence analysis.[46]

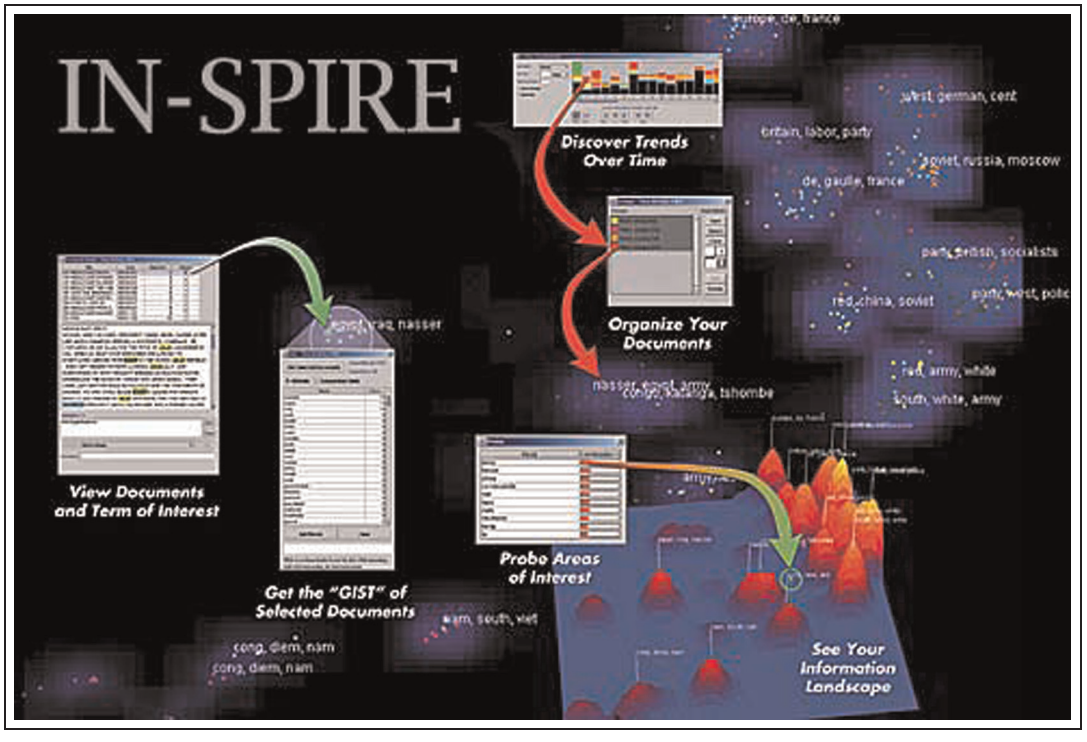

Figure 22. The IN-SPIRE discovery tool integrates information visualization with query and other interactive capabilities.[53]

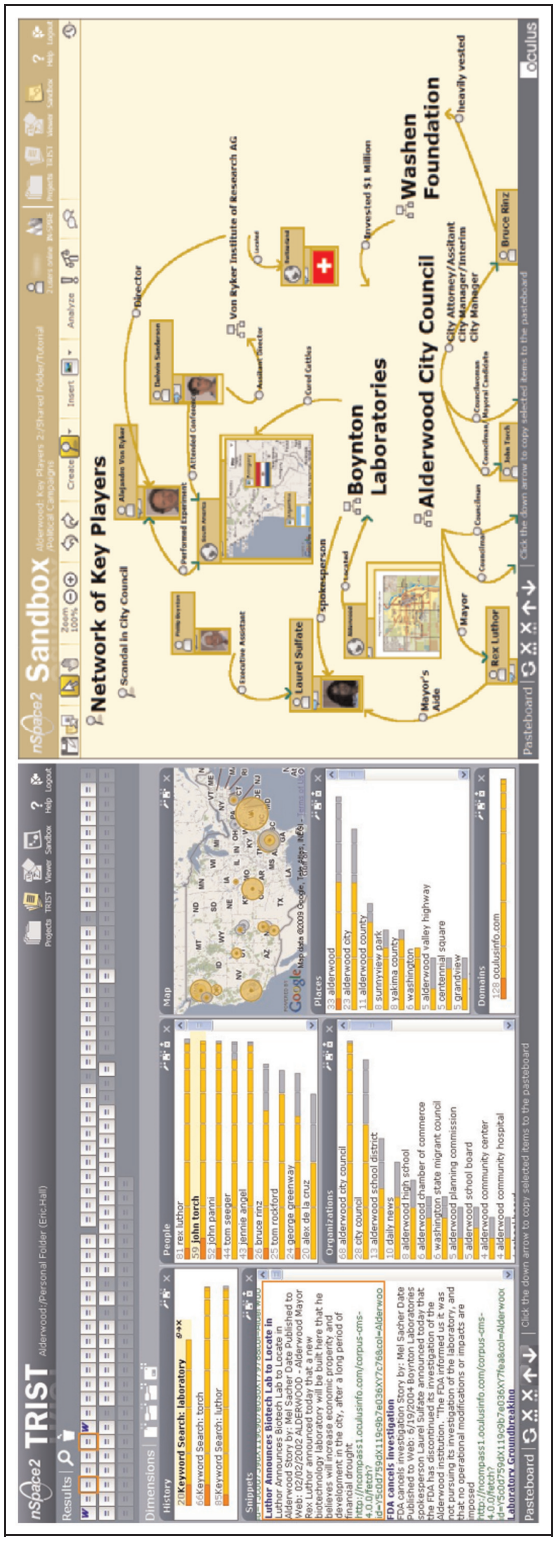

Figure 23. Oculus nSpace TRIST and Sandbox capabilities.[54]

The Basic Ordinance Observational Management System (BOOMsys)[57] is a prototype geovisual application that depicts spatiotemporal data about Improvised Explosive Device (IED) attacks in Iraq (Figure 25). BOOMsys shows the number of IED attacks aggregated by province using circles scaled to the data. Each symbol can be probed to display background information for the province. It also includes tools to explore the data by various time frames, including totals by day, week, month and a composite week by day.

Although not developed specifically for intelligence analysis, Tableau Desktop is a commercial software application that lets users explore their datasets and try out various visualizations easily through simple drag and drop operations. It can be used to graphically analyze virtually any type of structured data, display information in multiple graphic perspectives and produce charts, graphs and reports (see Figure 26). Tableau is also offered publicly for free. The free data visualization tool can help create an interactive visualization and embed it in a website or share it. However, the free version does not allow one to share visualizations in a private network or save it to a personal computer.

GeoTime is another commercially available tool that is well suited for intelligence analysis. It provides the ability to track targets, show communications and relationships, and see behaviors in time and space within a single, interactive 3D display.[59] Events are represented in an *X,Y,T*-coordinate space in which the *X,Y*-plane shows geography and the vertical *T*-axis represents time (see Figure 27). Events animate in time vertically through the 3D space as the time slider bar is moved. GeoTime also provides target timeline comparisons, chart statistics and network analysis tools. Filters and link analysis allow a user to focus on items of interest while automatic pattern analysis tools help quickly identify behavior patterns of moving entities and networks. GeoTime supports time–space annotations of events. Notes, explanations and hypotheses can be captured in a document, including snapshot images. GeoTime has been a winner at the Institute of Electrical and Electronics Engineers (IEEE) Visual Analytics Science and Technology (VAST) conference contest for many years.

6. Moving forward

VA systems prototypes and commercial products for improving the efficiency of cyber security and intelligence analysis are numerous, but few evaluations have been performed to validate this claim. This stems from the inherent difficulty of evaluating user interfaces, especially when it comes to measuring insight gain rather than low-level tasks. Relatively little effort has been devoted to this

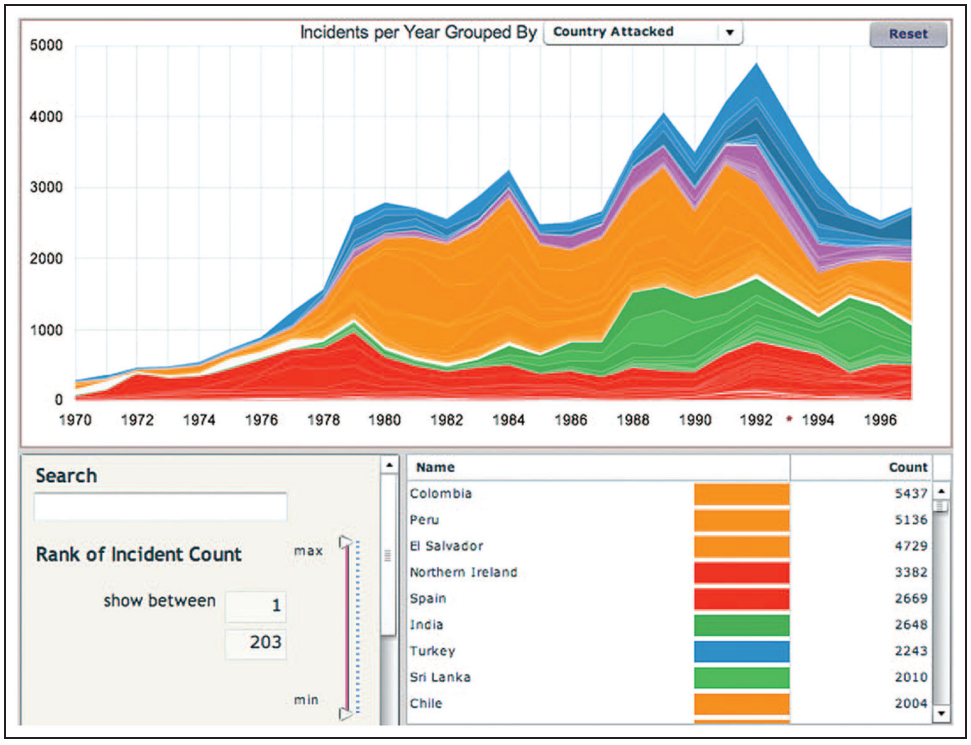

Figure 24. Theme River representation of terrorism attacks in the world over time.[56]

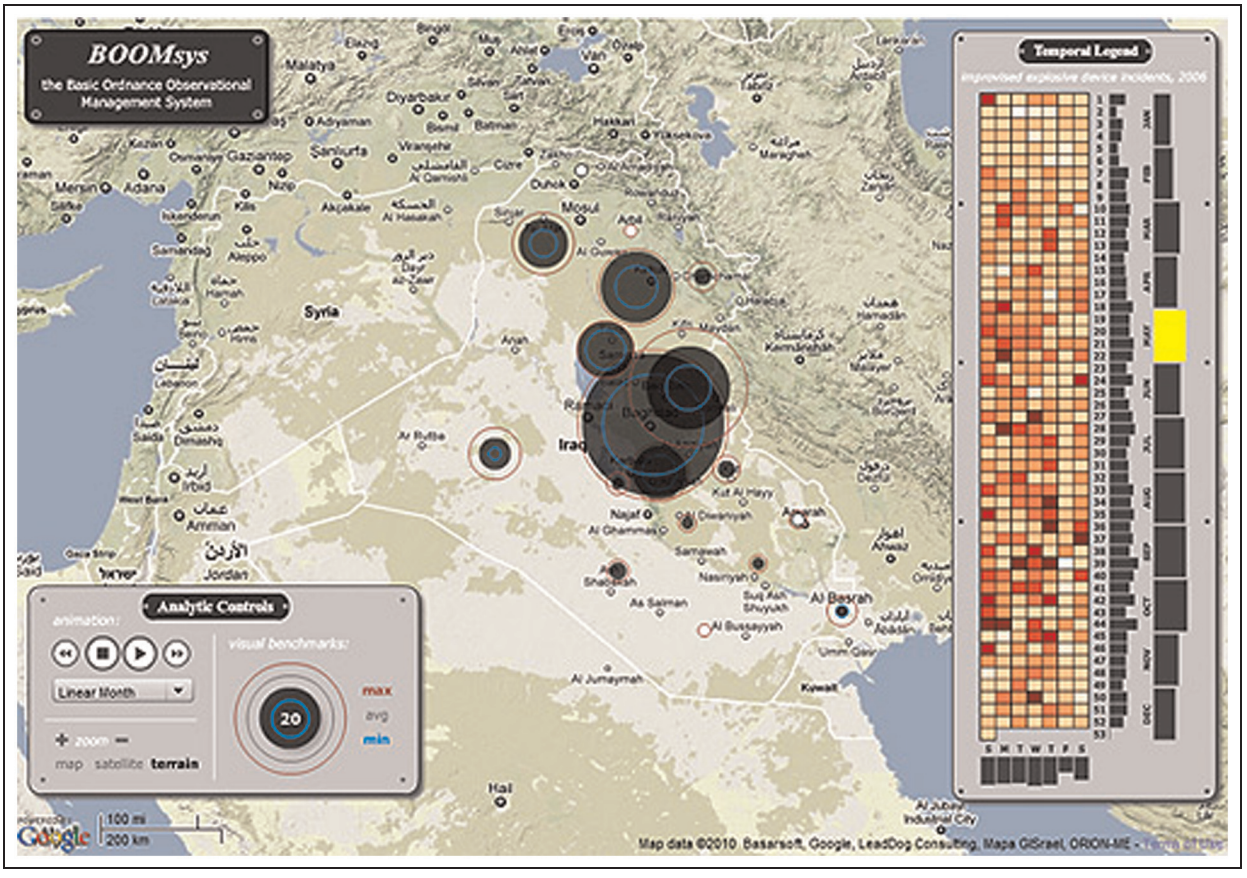

Figure 25. Analysis of Improvised Explosive Device s in Iraq with BOOMsys.[57]

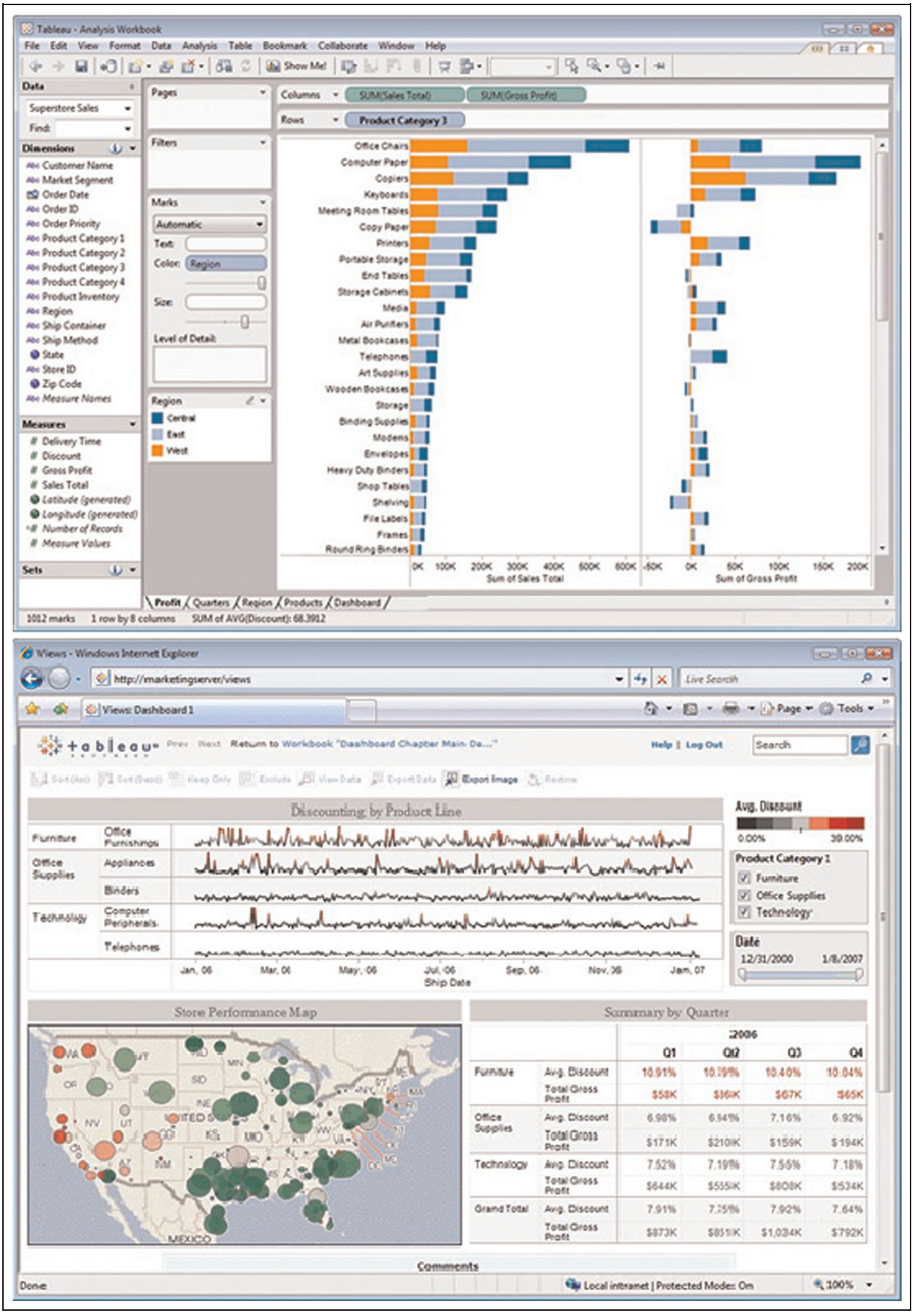

Figure 26. Examples of Tableau Desktop visualizations.[58]

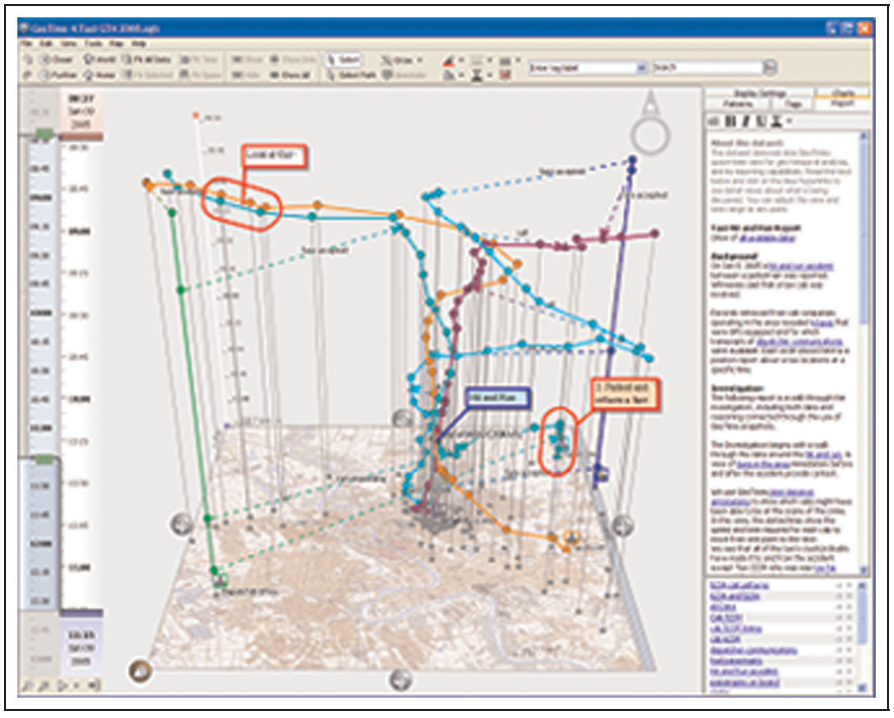

Figure 27. Oculus GeoTime interface.[60]

Table 2. Evaluation approaches.[19]

Method	Most useful for…	Limitations
Observations and interviews	Revealing analytic process	Subjective
Questionnaires and discussion groups	Usability testing – user satisfaction with system	May not reflect true utility/effectiveness
Heuristic evaluation	Usability testing – focus on user interactions/transactions with system	May not reveal deeper insights of cognitive process
Longitudinal studies	In-depth assessment of extent to which tool aligns with process	Tends to use a small sample of participants
Controlled experiments/ performance testing	Comparing alternative VA approaches leading to enduring scientific conclusions	Difficulty in obtaining sufficient number of participants

VA: Visual Analytics.

aspect up to now in comparison to the work devoted to tools development. As more and more VA applications appear, researchers are looking for a better assessment of their effectiveness and utility. The ultimate goal is the creation of a set of metrics that could predict the efficiency of tools for given tasks.

Understanding cognition and creating scientific evaluation methods are the VA fields where research is now most needed. For a number of years, there has been an ongoing discussion related to what is required to make this research domain a truly scientific discipline. There is a need for a coherent theoretical framework.[19] Although basic cognitive processes have been studied by psychologists, there is still much to learn about how a person really makes sense of a situation and gains insight. In turn, this will lead to effective evaluation methods that can measure the effectiveness of tools more accurately and a better informed design methodology.

While waiting for these metrics to be defined, we can use a number of alternative methodologies and approaches among which there are five dominant approaches.[19] They

Table 3. Publicly available VAST Challenge datasets.

Challenge	Type of data
VAST Challenge 2012 (http://www.vacommunity.org/VAST+Challenge+2012)	Big data challenge, large-scale situation analysis and cyber security, global network situation monitoring, unusual malicious computer security events, raw and .cvs processed IDS logs.
VAST Challenge 2011 (http://hcil.cs.umd.edu/localphp/hcil/vast11/)	Geospatial and microblogging for characterization of an epidemic spread, cybersecurity with situational awareness in computer networks and text analytics for investigation into criminal activity.
VAST Challenge 2010 (http://hcil.cs.umd.edu/localphp/hcil/vast10/index.php)	Text records for investigation into arms dealing, hospitalization records for characterization of pandemic spread and genetic sequences for tracing the mutations of a disease.
VAST Challenge 2009 (http://www.cs.umd.edu/hcil/VASTchallenge09/)	Badge and computer network traffic, social network with small geospatial component, video analysis.
VAST Challenge 2008 (http://www.cs.umd.edu/hcil/VASTchallenge08/)	Cell phone social network (phone records), geo-temporal records (building evacuation), unstructured text (wiki edits data and history) and location tracking (boat migrations).
VAST Contest 2007 (http://www.cs.umd.edu/hcil/VASTcontest07/dataset.htm)	Mostly text (news stories and blog entries, along with background information) and some multimedia materials (small maps and data tables) related to a law enforcement/counterterrorism scenario.
VAST Contest 2006 (http://www.cs.umd.edu/hcil/VASTcontest06/dataset.htm)	Mostly text (about 1200 news stories), a few photos, a few maps (in bitmap image form), a few files with other mixed materials (e.g. a spreadsheet with voter registry information or a phone call log) and a couple of pages of background information (in text form).

are summarized in Table 2, along with their advantages and limitations.

The VAST Challenges are a step forward in the right direction. They represent an original mean of gathering VA developers and researchers around shared benchmark datasets and comparing new VA tools. They are proposed every year to the international VA community as part of the IEEE VAST Symposium. Participants can demonstrate the VA capabilities of their tools against an invented scenario and synthetic datasets. The challenge scenario and datasets then remain available to anyone interested in testing their VA tools with representative tasks and datasets. Figure 28 shows visualizations that were submitted to the VAST 2008 Challenge. As shown in Table 3, past VAST Challenges have included a variety of data types related to many defense and security domains, including cyber security and intelligence analysis. These datasets are publicly available online.

7. Conclusion

In the context of modern defense and security operations, analysts are faced with significant data overload problems that prevent them from understanding a situation at hand and anticipating how this situation may develop. Fortunately, VA has emerged as a significant multidisciplinary research field that leverages the human cognitive abilities to comprehend information when presented in a proper way and combined with suitable interaction. To describe VA, this paper has presented a number of information visualization, interaction and analytical reasoning techniques that allow making the relevant information more salient in order to help detect patterns, trends and anomalies.

VA is making its way into defense and security applications, such as cyberspace management and intelligence analysis. Examples of this include performing large-scale network traffic monitoring and intrusion detection in the cyberspace domain, and making sense out of large collections of unformatted text documents for counterterrorism intelligence analysis. VA has a significant momentum and VA research and applications have been growing exponentially over recent years, although more research is required to develop a coherent theoretical framework that will inform evaluation and design of these tools.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful comments. We also thank DRDC CSS (Centre for Security Sciences), in particular Mr Andrew Vallerand and Mr Jack Pagotto, for their indefectible interest in R&D work in VA and for allowing the authors to engage with the VA community.

Funding

This work was done in the course of employment of the authors by the Government of Canada, and received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

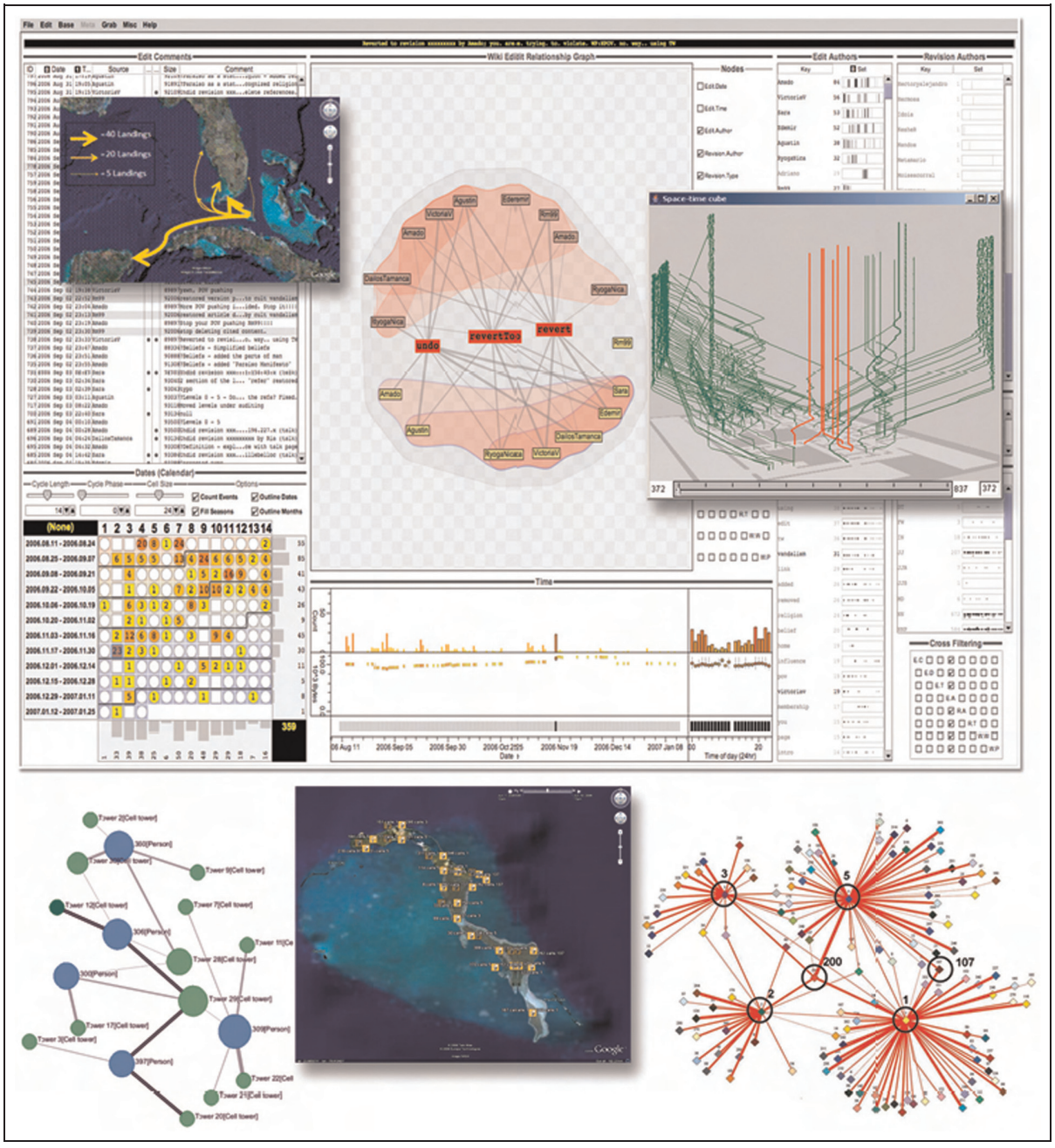

Figure 28. Visualizations submitted to the VAST Challenge 2008, from (clockwise, from upper left): SPANDAC, NEVAC, Fraunhofer Institute, Oculus Info Inc., Palantir, Oculus Info Inc.[61]

References

1. Thomas JJ and Cook KA (eds). *Illuminating the path: the research and development agenda for visual analytics*. IEEE Computer Society Press, 2005. http://www.purdue.edu/discoverypark/vaccine/assets/pdfs/publications/pdf/Illuminating%20the%20Path.pdf
2. Keim D, Kohlhammer J, Ellis G, et al. (eds). *Mastering the information age: solving problems with visual analytics*. Goslar: Eurographics Association, 2010.
3. Thomas JJ. Visual analytics techniques that enable knowledge discovery: detect the expected and discover the unexpected. In: *ACM SIGKDD workshop on visual analytics and knowledge discovery (VAKD '09)*, Paris, France, 28 June 2009.
4. Card SK, Mackinlay JD and Shneiderman B. *Readings in information visualization: using vision to think*. San Francisco, CA: Morgan Kaufmann Publishers, 1999.
5. Healey CG. Perception in visualization, Christopher Haley's web site: http://www.csc.ncsu.edu/faculty/healey/PP/index.html (2009).

6. Kosara R, Miksch S and Hauser H. Semantic depth of field. In: *proceedings of IEEE symposium on information visualization (InfoVis 2001)*, IEEE, 2001, pp.97–104.
7. Stone M. Choosing colors for data visualization, http://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf (2006).
8. Tufte ER. *Envisioning information*. Cheshire, CT: Graphics Press, 1990.
9. Koffka K. *Principles of gestalt psychology*. New York: Harcourt, 1935.
10. Wertheimer M. Laws of organization in perceptual forms. In: Ellis WD (ed.) *A source book of gestalt psychology*. London: Harcourt, 1938, pp.71–88.
11. Simons DJ and Chabris CF. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 1999; 28: 1059–1074.
12. Miller RB. Response time in man-computer conversational transactions. In: *proceedings of the AFIPS fall joint computer conference*, Vol. 33, 1968, pp.267–277.
13. Card SK, Robertson GG and Mackinlay JD. The information visualizer: an information workspace. In: *proceedings of the ACM conference on human factors in computing systems (CHI'91)*, ACM Press, New York, 1991, pp.181–188.
14. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. In: *proceedings of the IEEE symposium on visual languages*, IEEE Computer Society Press, 1996, pp.336–343.
15. Keim DA, Mansmann F, Schneidewind J, et al. *Visual analytics: scope and challenges, visual data mining: theory, techniques and tools for visual analytics*. Lecture Notes in Computer Science (LNCS), Springer, 2008.
16. Heer J and Shneiderman B. Interactive dynamics for visual analysis. *Commun ACM* 2012; 55: 45–54.
17. Patterson ES, Roth EM and Woods DD. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognit Technol Work* 2001; 3: 224–237.
18. Bowden EM, Jung-Beeman M, Fleck J, et al. New approaches to demystifying insight. *Trends Cognit Sci* 2005; 9: 322–328.
19. Greitzer FL, Noonan CF and Franklin L. *Cognitive Foundations for Visual Analytics, PNNL-20207*. Richland, WA: Pacific Northwest National Laboratory, 2011.
20. Ware C. *Information visualization – perception for design*. San Diego, CA: Academic Press, 2000.
21. Pirolli P and Card SK. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: *Proceedings of the International Conference on Intelligence Analysis'05*, 2005, pp.2–4.
22. Pirolli P and Card SK. Information foraging. *Psychol Rev* 1999; 106: 643–675.
23. Russell DM, Stefik MJ, Pirolli P, et al. The cost structure of sensemaking. In: *the INTERCHI '93 conference on human factors in computing systems*, Amsterdam, 1993.
24. Klein G and Moon B. Making sense of sensemaking 2. *IEEE Intell Syst* 2006; 21: 88–92.
25. Henry N, Fedeke J-D and McGuffin MJ. NodeTrix: a hybrid visualization. *IEEE Trans Visual Comput Graphics* 2007; 13: 1302–1309.
26. Wang TD. Interactive visualization techniques for searching temporal categorical data. PhD Dissertation from the Department of Computer Science, University of Maryland, May, 2010.
27. SmartMoney. http://www.smartmoney.com/map-of-the-market/ (2010). Accessed November 23, 2010.
28. d'Ocagne M. *Coordonnées parallèles et axiales: méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles*. Paris: Gauthier-Villars, 1885.
29. Hauser H, Ledermann F and Doleisch H. Angular brushing of extended parallel coordinates. In: *proceedings of the IEEE symposium on information visualization 2002 (InfoVis 2002)*, Boston, MA, 28–29 October 2002.
30. Rübel O, Weber GH, Huang M-Y, et al. PointCloudXplore 2: visual exploration of 3D gene expression. In: Garth C, Hagen H and Hering-Bertram M (eds) *Visualization of large and unstructured data sets*, GI Lecture Notes in Informatics, Vol. S-7. Bonn: Gesellschaft fuer Informatik (GI), 2008, pp.125–137.
31. Soo Yi J, Melton R, Stasko J, et al. Dust & Magnet: multivariate information visualization using a magnet metaphor. *Inform Visual* 2005; 4: 239–256.
32. Green TM, Ribarsky W and Fisher B. Building and applying a human cognition model for visual analytics. *Inform Visual* 2009; 8: 1–13.
33. Collins C and Carpendale S. VisLink: Revealing relationships amongst visualizations. *IEEE Trans Visual Comput Graphics* 2007; 13: 1192–1199; in: *proceedings of the IEEE conference on information visualization (InfoVis '07)*, November–December 2007.
34. D'Amico A and Whitley K. The real work of computer network defense analysts: the analysis roles and processes that transform network data into security situation awareness. In: *proceedings of the workshop on visualization for computer security (VizSec 2007)*, Springer, Berlin, 2008, pp.19–37.
35. D'Amico A, Whitley K, Tesone D, et al. Achieving cyber defense situational awareness: a cognitive task analysis of information assurance analysts. In: *proceedings of the human factors and ergonomics society 49th annual meeting*, 2005, pp.229–233.
36. D'Amico A and Kocka M. Information assurance visualizations for specific stages of situational awareness and intended uses: lessons learned. In: *IEEE workshops on visualization for computer security (VizSec'05)*, 2005, p.13.
37. Wolf J. How to improve IT cyber-security with visual analytics, it security & network security news, eWeek.com, http://www.eweek.com/c/a/Security/How-to-Improve-IT-Cyber-Security-with-Visual-Analytics/ (2009). Accessed April 2, 2013.
38. Mansmann F, Fisher F, Keim DA, et al. Visual support for analyzing network traffic and intrusion detection events using treemap and graph representations. In: *proceedings of the symposium on computer human interaction for the management of information technology (CHiMiT '09)*, 2009.
39. D'Amico A, Goodall JR, Tesone DR, et al. Visual discovery in computer network defense. *IEEE Comput Graphics Appl* 2007; 27: 20–27.

40. O'Brien BF, D'Amico A and Larkin ME. Technology transition of network defense visual analytics: lessons learned from case studies. In: *IEEE international conference on technologies for homeland security (HST2011)*, Boston, MA, 15–17 November 2011.
41. Goodall JR and Tesone DR. Visual analytics for network flow analysis. In: *proceedings of the cybersecurity applications & technology conference for homeland security (CATCH)*, IEEE Press, 2009, pp.199–204.
42. Sopan A, Freire M, Plaisant C, et al. Distribution column overviews in tabular visualizations. HCIL-2010–01, April 2010.
43. Singh A, Endert A, Andrews C, et al. Supporting the cyber analytic process using visual history on large displays. In: *proceedings of the 8th international symposium on visualization for cyber security (VizSec'11)*, 2011, Article 3.
44. Maughan D. The need for a national cybersecurity research and development agenda. *Commun ACM* 2010; 53: 29–31.
45. Goodall JR, Radwan H and Halseth L. Visual analysis of code security. In: *proceedings of the seventh international symposium on visualization for cyber security (VizSec '10)*, 2010, pp.46–51.
46. Future Point Systems, Inc. http://www.futurepointsystems.com/ (2011). Accessed September, 2012.
47. FM 2-0: 2010. Intelligence, Headquarters Department of the Army, Washington, DC, 23 March 2010.
48. DND/CF: 2008. Counter-insurgency operations B-GL-323-004/FP-003, 2008.
49. Visual Analytics Inc. http://www.visualanalytics.com/ (2012).
50. Bush President GW. The Department of Homeland Security, http://www.dhs.gov/xlibrary/assets/book.pdf (2002). Accessed April 2, 2012.
51. Patterson ES, Roth EM and Woods DD. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognit Technol Work* 2001; 3: 224–237.
52. Kang Y and Stasko J. Characterizing the intelligence analysis process: informing visual analytics design through a longitudinal field study. In: *VAST*, 2011.
53. PNNL. http://in-spire.pnnl.gov/ (2012).
54. Wright W, Schroh D, Proulx P, et al. *The sandbox for analysis - concepts and methods*. Montreal, Quebec, Canada: ACM CHI, 2006.
55. Oculus Info Inc. http://www.oculusinfo.com/nspace/ (2012). Accessed September 2012.
56. Lee J. Exploring global terrorism data: a web-based visualization of temporal data. *ACM Crossroads* 2008; 15: 7-16.
57. GeoVISTA. http://www.geovista.psu.edu/BOOMsys/ (2012). Accessed September 2012.
58. Tableau Software. http://www.tableausoftware.com/ (2012).
59. Kapler T, Eccles R, Harper R, et al. Configurable spaces: temporal analysis in diagrammatic contexts. In: *IEEE visual analytics science and technology (VAST 2008)*, 19–24 October 2008, Columbus, OH, IEEE.
60. Oculus Info Inc. http://www.oculusinfo.com/geotime/ (2012).
61. VAC Views. *The VAST 2008 Challenge*, May 2009, pp.14–15. Accessed April 2, 2013.

Author biographies

Valérie Lavigne obtained a master's degree at Université Laval in electrical engineering in 2005. She worked as a consultant in various scientific and defense application domains (optronic surveillance, electro-optical warfare, image processing, hyperspectral imaging, modeling and simulation) before joining Defence R&D Canada – Valcartier in 2008. She is a member of the Future C2 Operations Center Design Group within the Command, Control and Intelligence Section. She is now leading an applied research project on interactive visualization and collaboration for maritime domain analysis. She also contributes to R&D activities concerning social network analysis for COIN operations, intelligence preparation of the battlefield, virtual intelligence analysis capability and future intelligence analysis capability. Her research interests include VA, multimodal human–computer interactions and collaborative technologies.

Denis Gouin holds a bachelor's degree in Computer Sciences. He has been working as a defense scientist with Defence R&D Canada – Valcartier since 1978. Throughout his career, he has been actively engaged in developing novel concepts and prototypes of command and control and intelligence information systems for defense and security applications. His fields of interest include geomatics, human–machine interfaces, information visualization, collaboration environments, enterprise portal technology and open-system architectures. He has led the design and development of the Open Geospatial Datastore Interface (OGDI), which is an open infrastructure that provides an open access to geospatial information. He was the project manager for the COP 21 Technology Demonstration, which has produced and demonstrated a prototype of a Situation Awareness Knowledge Portal. He was the lead of the Situation Awareness sub-project, as part of the Joint Command Decision Support (JCDS) 21 technology Demonstration. Finally, he participates as the Canadian national lead to the TTCP C3I Technical Panel 2 on Command Information Interfaces. He is now the lead of the Future C2 Operations Center Design Group. He is involved in a number of R&D projects that exploit VA and information visualization, applied to maritime domain awareness, intelligence preparation of the battlefield and social network analysis in COIN operations.