Original Article

Scale and complexity in visual analytics

George Robertson[a,*]
David Ebert[b]
Stephen Eick[c]
Daniel Keim[d] and
Ken Joy[e]

[a]Microsoft Research, Redmond, WA 98052, USA.
[b]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA.
[c]SSS Research, VisTracks, Lisle, IL 60532, USA.
[d]Department of Computer and Information Science, University of Konstanz, Konstanz, D-78457, Germany.
[e]Department of Computer Science, University of California, Davis, CA 95616, USA.

*Corresponding author. George Robertson, PO Box 177, Northeast Harbor, ME 04662-0177, USA.
E-mail: ggr@microsoft.com

This article is a product of a workshop on the Future of Visual Analytics, held in Washington, DC on 4 March 2009. Workshop attendees included representatives from the visual analytics research community across government, industry and academia. The goal of the workshop, and the resulting papers, was to reflect on the first 5 years of the visual analytics enterprise and propose research challenges for the next 5 years. The article incorporates input from workshop attendees as well as from its authors.

Received: 27 May 2009
Revised: 7 July 2009
Accepted: 7 July 2009

Abstract The fundamental problem that we face is that a variety of large-scale problems in security, public safety, energy, ecology, health care and basic science all require that we process and understand increasingly vast amounts and variety of data. There is a growing impedance mismatch between data size/complexity and the human ability to understand and interact with data. Visual analytic tools are intended to help reduce that impedance mismatch by using analytic tools to reduce the amount of data that must be viewed, and visualization tools to help understand the patterns and relationships in the reduced data. But visual analytic tools must address a variety of scalability issues if they are to succeed. In this paper, we characterize the scalability and complexity issues in visual analytics. We discuss some highlights on progress that has been made in the past 5 years, as well as key areas where more progress is needed.
Information Visualization (2009) **8,** 247 – 253. doi:10.1057/ivs.2009.23

Keywords: visual analytics; scalability; visualization; analytics

Introduction

In 2004, the US Department of Homeland Security chartered the National Visualization and Analytics Center™ (NVAC™) to lead the research and development of visual analytic techniques for homeland protection. One of the first steps in that process was the development of a long-term research and development (R&D) agenda, which was published in the book *Illuminating the Path*.[1] The R&D agenda focused on developing visual analytic tools to support three primary objectives: preventing terrorist attacks, protecting borders and improving emergency response. One of the grand challenges described in *Illuminating the Path* is the Scalability Challenge. This paper examines and elaborates that challenge.

Our ability to collect data is increasing at a faster rate than our ability to analyze it. EMC[2] reports that the Digital Universe as of May 2009 contained 500 exabytes, and will double every 18 months. They also point out that creation of digital information in 2008 exceeded the total capacity to store it. Analysts, emergency response teams and border protection personnel have massive amounts of information available to them from multiple sources, but the important information may be hidden in a few nuggets. We must create new methods to allow the analyst to examine massive, multi-dimensional, multi-source, time-varying information streams to make effective decisions in time critical situations.

In the 5 years since the NVAC R&D agenda was proposed, some progress has been made toward addressing scalability challenges. However, scalability issues can never be fully resolved as long as the scale of the problems keeps increasing. We need to continually discover ways to handle larger and larger problems. Hence, the basic goals outlined in the agenda remain the same, and much work remains. In addition, visual analytic techniques can be applied to many domains other than homeland security. For example, these techniques can be used for analytic problems in the areas of energy, the environment and basic science,[3] as well as for business intelligence and health care. As we gain experience with visual analytic techniques, some new scalability issues have been observed.

In this paper, we will characterize the scalability and complexity issues in visual analytics. We will discuss some highlights on progress that has been made in the past 5 years, identify key areas where more progress is needed, and describe new scalability issues that must be addressed.

Data Characteristics

The following primitive data types contribute to information overload for analysts.

Textual data. Massive textual data can come from documents, speeches, e-mail messages or web pages. These data are ever increasing in volume. One target reported in 2005[1] was to be able to support analysis of data volumes growing at a rate of one billion new structured messages or transactions per hour, and one million new unstructured messages or documents per hour. These were estimates of what would be required in the intelligence community to detect terrorist threats. Of course, this is a moving target as the total amount of information acquired continues to grow.

Numeric data. The revolution in miniaturization for computer systems has resulted in the production of many types of sensors. The sensors can collect numeric data about their environment (location, proximity, temperature, light, radiation and so on), can analyze these data and can communicate among themselves. Collections of sensors can produce very large streaming sets of data. Methods are needed for analyzing numeric data to efficiently incorporate the data into computerized models.

Image data. Consider the data collected by satellites that image the earth. Commercial satellites can create images at 1-m resolution and collectively create an image of the planet's land surface in a very short time. New methods are needed to permit efficient understanding of image data, especially in the context of other types of data mentioned here.

Video data. Video is often used to enhance the effectiveness of high-risk security and public safety operations. Video recording and content analysis are being used in concert as a powerful tool for improving business processes and customer service. New techniques must be developed to integrate these capabilities for analyzing streaming video data into the analyst's toolbox.

Audio data. Consider the processing of audio from phone calls, 911 calls, radio intercepts, radio traffic during emergency response, and commercial radio and television broadcasts. Techniques exist for word spotting in audio streams. However, that may be insufficient as the volume of audio data increases, because these techniques fail to take context into account.

These primitive data types are organized into collections of various kinds (files, directories, databases and so on). The nature of these organizations and the methods for processing these data are discussed in the companion paper on Data Transformations for Computation and Visualization.[4]

Data present challenges not only because of their diversity, volume and dynamic nature but also because data contain errors and are ambiguous, incomplete, uncertain and potentially intentionally deceptive. Data of multiple types must often be analyzed in concert to gain insight. Important data needed for correct interpretation may be missing, but this may or may not be apparent to the analyst. We must provide mechanisms that help the analyst visually understand the nature of the data being evaluated.

A grand challenge is to support the analyst in distilling the relevant nuggets of information from widely disparate information streams and create an information space containing relevant information that can be used by the analyst in reaching the most timely and well-informed assessment of the situation. We must provide mechanisms that can visually represent the connections between the relevant information in the information streams and allow the analyst to relate concept to data.

A Variety of Scalability Issues

Current technologies cannot support the scale and complexity of the growing analytical challenge. New techniques and underlying scientific foundations are needed to deal with the scale of the problems we are facing in security (threat analysis, emergency management and border protection), global issues of energy, the environment, basic science, health care and business development. Issues of scale cut across every aspect of this challenge.

When considering scalability issues, it is important to understand the context of the development of the computer industry as well as natural human skills and limitations. Moore's Law suggests that basic computer technology performance (processor speed and memory density) will double every 18 months. This trend has continued for 45 years and some projections say it will continue for at least another 5 years before fundamental limitations of physics are encountered.[5] Recently, graphics technology has been improving performance at an even faster rate, doubling every 6 months.[6] Much of the future growth in computational power will come from parallel processing, which is difficult to exploit (this issue is discussed later in the section on computational scalability). All of this added processing power and memory density has enabled the gathering and processing of vast amounts of data.

However, basic human skills and abilities do not change significantly over time. It is true that technology advances, applied carefully, can enable us to use a higher percentage of natural human abilities, but there are basic limits that we are asymptotically approaching. This situation gives rise to the popular notion of *information glut*. That is, we are able to access far more information than we, as humans, can possibly process. The situation also makes scalability issues more difficult to resolve. In addition, analytical challenges often require coping with, sharing, and using information at multiple scales simultaneously. Ultimately, large-scale problems have to be reduced to a scale that humans can comprehend and act on.

Scale may bring opportunities as well. For example, increased scale may help reduce uncertainty of an emerging situation by providing more evidence to either confirm or deny hypotheses. Large data volumes allow analysts to discover more complete information about a situation. As a result, analysts may be able to determine more easily when expected information is missing; sometimes the fact that information is missing offers important clues in the assessment of a situation.

Here, we consider five of the major scale issues that must be addressed: information scalability, visual scalability, display scalability, human scalability and computational scalability.

Information Scalability

Information scalability implies the capability to extract and make sense of relevant information from massive data streams. Methods of information scalability include methods to filter and reduce the amount of data, techniques to represent the data in a multi-resolution manner and methods to abstract the data sets. The companion paper on Data Transformations for Computation and Visualization[4] discusses these methods and techniques, as well as the challenges that must be overcome.

A second form of information scalability has to do with the rate of change of the information. Many existing systems are dynamically updated as data change, but published techniques only deal with modest rates of change. There are two issues that must be addressed with changing data; the new data must be assimilated into the current views the analyst is using, and in some cases, the analyst must be made aware of what has changed. Existing techniques can do both if the rate of change is modest. But, suppose the rate of change radically increased, going from several thousand new data elements per day to several million. Existing techniques would have difficulty keeping up with such a large volume of change, and would also fail to show the analyst what had changed.

Finally, information presentations must be scaled or adapted to the audience. For example, an analyst's presentation to other analysts will contain far more detail than the summary analysis presented to senior management. Current techniques require that this be done manually in an ad hoc fashion. In fact, current practice often involves copying or abstracting parts of an analysis from tools designed for analysis to different tools designed for presentation. One problem with this approach is that during a presentation, there may be limited (or no) tools available to show details of how an analysis was done, or to explore alternatives. An integration of analysis and presentation tools would improve the process. The Scalable Reasoning System[7] is a recent example of a system that integrates analysis and dissemination, as well as provides a means of scaling or adapting the analysis/presentation to the audience.

Relevant information may appear at a variety of scales; the user must be able to change between scales in a way that is easy to understand and track, and must be able to understand cross-scale interactions. We must be able to handle a wide range of dynamic change, and develop systems that semi-automatically scale or adapt information presentations to match a target audience.

One notable recent advance reported by Ingram *et al*.[8] is Glimmer, a multi-level algorithm for multi-dimensional scaling (MDS) designed to run on modern graphics processing unit hardware. MDS is a key technique for reducing high-dimensional data onto a low-dimensional target for presentation. The use of MDS has been somewhat limited because it has been too slow for interactive use when the number of dimensions is scaled up. The Glimmer approach increases speed by a factor of 10–15 for large data sets, making it possible to use MDS interactively on larger data sets. More advances like this are needed.

Visual Scalability

Visual scalability is the capability of visualization representation and visualization tools to display effectively massive data sets, in terms of either the number or the dimension of individual data elements.[9] Factors affecting visual scalability include the quality of visual displays, the visual metaphors used in the display of information, the techniques used to interact with the visual representations, and the perception capabilities of the human cognitive system.

Ware[10] argues that the optimal display is a 4000×4000 pixel resolution monitor, based on human perceptual capabilities such as visual acuity and spatial contrast sensitivity. If each of those 16 million pixels represented one data element, the viewer would see a black screen. If the visual representation of the information requires showing links or labels, as well as separation of the individual objects, then perhaps a few tens of thousands of data elements could be displayed on such a display. For most user tasks, the effective number of data elements that can be displayed is probably much smaller. So, the fundamental problem of visual scalability is how to visually represent a very large number of data elements in a much smaller number of visual display elements, so that the user's task can be performed. Some user tasks can be addressed by filtering the data and showing only the most relevant data; this is information scalability. Other tasks require showing an overview of all of the data (or a very large part of the data) so that large-scale relationships can be seen; this is visual scalability.

Most published techniques in the field of information visualization handle data sets with hundreds to thousands of elements. Some techniques can scale to handle tens of thousands of elements and a very few can handle hundreds of thousands up to one million elements. The InfoVis 2003 Contest focused on the problem of visualizing and comparing large hierarchies. The best technique was TreeJuxtaposer,[11] which could handle two trees of about 100 000 elements and one tree up to 500 000 elements. TreeJuxtaposer used a technique called *Accordion Drawing*. Later work reported by Beerman *et al.*[12] extended these ideas in a system called TJC, to support browsing trees up to 15 million nodes.

However, as described previously, some extreme situations may demand the processing tens of millions of new documents per day, with a total database size of tens of billions of documents. It is reported that at least one existing database has 120 billion documents.[1] It seems likely that these database sizes will increase over time. Clearly the current state of the art is far from being able to visually represent today's data collections, and the need will continue to grow. New techniques are needed to bridge this gap.

One notable recent advance on visual scalability was reported by Chan *et al.*[13] at VAST 2008. The authors describe ATLAS, a visualization tool for temporal data that enables interactive access to a network traffic data set of more than one billion records. They accomplish this by combining high-performance database technology with predictive caching and level of detail management. This approach is limited in the kinds of visual analytic tasks that can be performed rapidly. For example, searching for interesting patterns across time periods cannot be done effectively with predictive caching. Hence more work is required to support the full range of analytic tasks interactively on large data sets.

Another example of recent work in visual scalability is the GreenGrid visualization for electric power grid analytics, reported by Wong *et al.*[14] While the North American power grid has about 50 000 electrical buses, losing 10 of these can generate an enormous number (10^{40}) of scenarios to analyze. GreenGrid uses a weighted force-directed multi-level graph visualization to enable visual analysis of problems at this scale. The system demonstrates how a combination of node and link weighting can make visual analysis significantly easier than the traditional geographic visualization of the power grid.

Display Scalability

Most published visualization techniques are designed for one size display, generally a desktop display (typically 1280×1024 pixels). We need to develop techniques that scale to a variety of display form factors to take advantage of whatever capabilities are available to support analysis and collaboration. Tools should be able to make effective use of everything from a wall-sized display in an emergency response situation room to a personal digital assistant (PDA) or phone-sized display in the hands of a first responder in the field. Studies need to be done to determine how to display information effectively, particularly on small displays.

One recent exploration of display-scale independence was reported by Smith *et al.*[15] in a description of FacetMap, a scalable browser for faceted data. FacetMap is designed to work on any size display, adapting its information layout to reveal more information with larger displays. The same visual representation and interaction techniques are used for all display sizes. This works well until you get down to PDA or phone-sized displays, where the amount of information displayed requires too many interactions to be effective. FaThumb[16] was an alternative facet-based interface for mobile devices with a numeric keypad. It was optimized for the small display. Display scale-independence reduces the need for retraining or learning multiple systems for different-sized displays. However, the experience with FacetMap and FaThumb suggests that effectiveness of systems for very small displays is perhaps more important than display scale-independence.

One thread of recent research has begun to examine the issues of information visualization specifically on large displays. Recent work by Yost *et al.*[17] has shown that displays larger than visual acuity (for example, wall sized), and requiring physical navigation, can be more effective and preferred over smaller displays for some tasks, if the appropriate visualization techniques are used. This is a surprising result, as the larger displays require more complex interaction techniques, including physically moving in order to see and interact with parts of the display.

Another example of recent work on display scalability has to do with table displays (also known as surface computing). Isenberg[18] is exploring the use of table displays for interacting with and sharing information visualizations during collaboration of small collocated teams. As with large displays, surface computing requires exploration of new interaction techniques, including more use of gestures. This is also an example of human scalability work, which is further discussed next.

Human Scalability

Although human skills and abilities do not scale (that is, they are relatively fixed), the number of humans involved in analytical problem-solving activities does scale. Most published techniques for supporting analysis are targeted for a single user at a time. We must develop techniques that gracefully scale from a single user to a collaborative (multi-user) environment. Much of the relevant collaboration research is focused on small groups of collaborators (two or three people). In the scenarios we envision, users may be collaborating from within the same team in an organization, at different levels of an organization or even in different organizations. Each of these cases has its own set of problems that must be solved. One scenario might involve a number of first responders, several regional emergency management centers and a national emergency management center – that is, dozens of users collaborating through the use of shared analytical tools and focusing on different levels of information accessible by everyone involved.

Collaboration issues extend beyond analytical problem-solving activities to decision-making processes. Decision making for an individual or a small team is straight forward compared to the complexity that arises for coordinated decision making in multiple teams, especially if these teams are from different levels of an organization or from different organizations. Visual analytic tools must support the decision-making processes even in these complex situations.

One recent exploration of human scalability is the Many Eyes website,[19] launched in 2007 to enable users to do collaborative analysis by uploading data, creating interactive visualizations and annotating others' work to engage in discussions. Another web-based collaborative visualization and analysis system is Swivel,[20] launched in late 2006. Although Many Eyes provides a much richer set of visualizations than Swivel, it lacks Swivel's ability to mash up parts of multiple data sets. Kosara[21] provides an informal comparison of these two social visualization systems. Both systems seek to make data available to, gain insights from, and share insights with a very large and diverse set of web users. Note that both of these systems involve asynchronous collaboration and analysis. Synchronous collaboration involves many of the same issues, but often requires the addition of some form of turn-taking or floor control to ease the interactions between the people involved.

Computational Scalability

The National Science Foundation has an ongoing 5-year goal for high performance computing to enable petascale computing (10^{15} operations per second) for investigations of computationally challenging problems in science and engineering by the year 2010.[22] In 2007, three Town Hall meetings were held to discuss the challenges of developing exascale computing (10^{18} operations per second) to address global issues of energy, ecological sustainability, basic science and security,[3] The belief is that general-purpose exascale computing will be technologically feasible within the next 15 years. These systems are likely to have 10–100 million processing elements or cores. Adoption of 1000-processor multi-core systems will require a substantial revolution in software and programming techniques for a workforce that has inadequate parallel programming skills. Current code, algorithms, tools and visualization approaches will not work at the exascale level without a paradigm shift.

The report of the exascale computing Town Hall Meetings[3] discusses many of the same topics that we discussed earlier. However, some issues change as we approach petascale and then work towards exascale computing. For example, in reference to handling massive data sets, the report observes, 'Data can be the result of an exascale simulation that must be post-processed for human interpretation, or it can form the input to complex problems via data assimilation. Browsing or looking at data is no longer possible as we near a petabyte. To visualize 1% of 1 petabyte at 10 MB/s takes 35 workdays. There is an enormous need for methods to dynamically analyze, organize, and present data by variability of interest'.

In reference to mathematics for data analysis, the report states, 'A particular gap exists in the mathematics needed to bring analysis and estimation methodology into a data-parallel environment. Parallel linear algebra methods go a long way toward enabling data-parallel analysis, but they do not solve it, just as they would not solve a climate simulation problem. For example, the standard principal component analysis computation does not become data-parallel with a parallel singular value decomposition (SVD) solver, even though the SVD is the core computation in that analysis. Data-parallel solutions for applications on exascale resources will require new mathematics that considers an entire estimation problem for developing scalable data-parallel algorithms in data analysis'.

Exascale applications will generate several terabytes of data per second. Because it is not practical to store raw data generated at such a rate, dynamic reduction of data by incremental summarization, subset selection and other filtering methods will be necessary. For exascale computing, visual analytic methods will be critical for handling the growing impendence mismatch between the size/complexity of data and an analyst's ability to understand and interact with that data.

Other Scalability Issues

In addition to the five major scalability issues just discussed, there are a number of other scalability issues that must ultimately be addressed.

Software scalability. The capability of software systems to be configured to interactively manipulate data sets of various sizes is called *software scalability*. This includes the generation of new algorithms that scale to the ever-increasing information sets that we generate today. We wish to avoid the hidden costs that arise when we build and maintain monolithic, non-interacting, non-scalable software models.

Temporal scalability. Sensemaking often involves temporal reasoning and may require handling data at different time scales. For example, it may be necessary to understand long-term patterns by looking at data over a period of years or even decades and simultaneously understand near-term effects by looking at data over a period of hours or less. Moreover, it may be necessary to integrate and perform correlative analysis on data collected at different temporal scales based on acquisition technology. For instance, in understanding fundamental principles of rain formation in clouds, it may be necessary to integrate data collected 1000 times per second with data collected every several minutes (radar data), and this information may then feed into climate models that work on the scale of years and decades.

Cross-scale issues. Tools are needed that scale to handle 'systems of systems'. These problems are common in science and engineering and may require analysis and combination of data across scales. For example, macrobiology analysis may require understanding the interactions of data simultaneously at the genome, protein, cell, organ, human, country and ecosystem levels. Cancer care treatment requires understanding and integrating data from the biomarker level (for example, integrating metabolics, lipidomics, genomics and proteomics data already at multiple scales), cancer processes at the organ level, environmental exposure, and socioeconomic factors that affect the success and completion of treatment regimens.

Privacy and security issues. Cutting across many scalability issues are concerns with privacy and security, particularly when scaling to multi-user environments. Data privacy and security laws and policies must be adhered to rigorously, which means that software must address challenges such as protecting information from inappropriate access, down to the data item and individual user level. While this may appear to be a universal problem, it is exacerbated by the human scalability problem. For example, when the analytic team involves multiple organizations, some of the data and analysis may not be accessible to everyone involved in the analysis, making the analysis more challenging and potentially making the results less accurate.

Language issues. Scalability issues also arise in dealing with geographically dispersed teams speaking different languages or using different terminology within the same language, and working across teams of people with differing expertise. This is an extension of the human scalability problem. As analytic teams grow in size and become more geographically dispersed, the chances increase that team members will not be using the same terminology or even speaking the same language. Recognition of the potential problem is essential; some form of translation may be needed to resolve the problem.

Conclusion

The fundamental problem that we face is that a variety of large-scale problems in security, public safety, energy, ecology, health care and basic science all require that we process and understand increasingly vast amounts and variety of data. There is a growing impedance mismatch between data size/complexity and human ability to understand and interact with those data. Visual analytic tools are intended to help reduce that impedance mismatch by using analytic tools to reduce the amount of data that must be viewed (for example, by filtering, summarization and abstraction), and visualization tools to help understand the patterns and relationships in the reduced data. But visual analytic tools must address a variety of scalability issues if they are to succeed.

Scalability and complexity issues in visual analytics are themselves quite complex and intertwined. In many cases, what an analyst needs most are simple visualizations of the right subset of the data. This is the information scalability problem; how do you extract the relevant data from a massive stream of data? In other cases, the key insight sought by an analyst requires viewing an overview of the data. This is the visual scalability problem; how do you visualize enormous amounts of data? In either case, the analyst or user of the visual analytic tools may be using those tools on different display devices at different times. The display scalability problem addresses this issue; how do you avoid learning a different system for each size display you work with? As we deal with larger-scale issues, it is often necessary to collaborate on analysis. This leads to the human scalability problem; how do we design visual analytic tools that can gracefully scale from a single user to a collaborative multi-user environment? These tools are all built on a computing infrastructure that is currently approaching petascale capability. Projections indicate that over the next 15 years, exascale computing will become possible. However, it is highly likely that exascale computing will require a paradigm shift in our approach to computing, as it will be highly parallel. This leads to the computational scalability problem; how do we redesign our analysis and visualization tools for exascale computing?

There is no formal research program on scalability issues in visual analytics. Rather, the problems are typically addressed in other visual analytics research and development where the specific scalability issues create a roadblock. Hence, some areas have received more attention than others. The areas that have received the most attention are information scalability for methods to filter data, and visual scalability. However, the solutions to date are point designs that solve specific problems. The areas that have received the least attention are information scalability for large-scale dynamic change, information presentation scaling, display scalability, human scalability and computational scalability. While we must continue to develop point designs that address specific scaling issues that block other visual analytics research and development, we also need to work on systematic solutions to the broader set of scalability issues.

Many of these scalability issues were posed in the NVAC R&D agenda published in *Illuminating the Path*[1] 5 years ago. The basic issues published then remain the same, with the addition of several new issues (computational scalability, temporal scalability and cross-scale problems). Although some progress has been made on many of the goals, dealing with scalability and complexity issues in visual analytic tools will continue to be a challenge as long as the volume of data continues to grow as it has.

References

1 Thomas, J.J. and Cook, K.A. (eds.) (2005) *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society Press.

2 EMC. (2009) IDC: As the economy contracts, the digital universe expands. Multimedia White Paper. http://www.emc.com/collateral/demos/microsites/idc_digital_universe/iview.htm, accessed 12 August 2009.

3 Simon, H., Zacharia, T. and Stevens, R. (2007) Modeling and simulation at the exascale for energy and the environment: Report on the Advanced Scientific Computing Research Town Hall Meetings on Simulation and Modeling at the Exascale for Energy, Ecological Sustainability and Global Security (E3). US Department of Energy, Washington DC, http://www.sc.doe.gov/ascr/ProgramDocuments/ProgDocs.html, accessed 8 April 2009.

4 Kasik, D., Ebert, D., Lebanon, G., Park, H. and Pottenger, W. (2009) Data transformations for computation and visualization. *Information Visualization*, 8(4): 275–285.

5 Crothers, B. (2009) Moore's Law limit hit by 2014? *CNet News*, 16 June. http://news.cnet.com/8301-13924_3-10265373-64.html, accessed 1 July 2009.

6 ExtremeTech. (2009) Moore's Law: Stil for wimps!, http://www.extremetech.com/article2/0,3973,1153970,00.asp, accessed 1 July 2009.

7 Pike, W. *et al*. (2009) The scalable reasoning system: Lightweight visualization for distributed analytics. *Information Visualization* 8(1): 71–84.

8 Ingram, S., Munzner, T. and Olano, M. (2008) Glimmer: Multilevel MDS on the GPU. *IEEE Transactions on Visualization and Computer Graphics* 15(2): 249–261.

9 Eick, S. and Karr, A. (2002) Visual scalability. *Journal of Computational and Graphical Statistics* 11(1): 22–43.

10 Ware, C. (2004) *Information Visualization: Perception for Design*. 2nd edn, San Francisco, CA: Morgan Kaufmann.

11 Munzner, T., Guimbretiere, F., Tasiran, S., Zhang, L. and Zhou, Y. (2003). TreeJuxtaposer: Scalable tree comparison using focus+context with guaranteed visibility. In: *Proceedings of the International Conference on Computer Graphics and Interactive Techniques ACM SIGGRAPH 2003 Papers*; 27–31 July. San Diego, CA. New York: Association for Computing Machinery Press, pp. 453–462.

12 Beerman, D., Munzner, T. and Humphreys, G. (2005). Scalable robust visualization of very large trees. In: K.W. Brodlie, D.J. Duke & K.I. Joy (eds.) Eurographics/IEEE VGTC Symposium on Visualization; 1–3 June. Leeds, UK. Aire-la-Ville, Switzerland: Eurographics Association, pp. 37–44.

13 Chan, S., Xiao, L., Gerth, J. and Hanrahan, P. (2008) Maintaining interactivity while exploring massive time series. In: D. Ebert, & T. Ertl (eds.), IEEE Symposium on Visual Analytics Science and Technology: VAST '08'; 21–23 October, Columbus, OH. Los Alamitos, CA: IEEE Computer Society Press, pp. 59–66.

14 Wong, P.C. *et al*. (2009) A novel visualization technique for electric power grid analytics. *IEEE Transactions on Visualization and Computer Graphics* 15(3): 410–423.

15 Smith, G. *et al*. (2006) FacetMap: A scalable search and browse visualization. *IEEE Transactions on Visualization and Computer Graphics* 12(5): 797–804.

16 Karlson, A., Robertson, G., Robbins, D., Czerwinski, M. and Smith, G. (2006) FaThumb: A facet-based interface for mobile search. *Proceedings of CHI 2006*. NY: Association for Computing Machinery, pp. 711–720.

17 Yost, B., Haciahmetoglu, Y. and North, C. (2007) Beyond visual acuity: The perceptual scalability of information visualizations for large displays. In: R.E. Grinter (ed.) *CHI 2006: Interact, Inform, Inspire: Conference Proceedings: Conference on Human Factors in Computing Systems*; 22–27 April, Montreal, Quebec, Canada. New York: Association for Computing Machinery, pp. 101–110.

18 Isenberg, P. (2007) Information visualization in co-located collaborative environments. *Proceedings of the Grace Hopper Celebration of Women in Computing*, PhD. Forum; 17–20 October. Orlando, FL. Palo Alto, CA: Anita Borg Institute, pp. 223–229.

19 Viegas, F., Wattenberg, M., vanHam, F., Kriss, J. and McKeon, M. (2007) Many eyes: A site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics* 13(6): 1121–1128.

20 Swivel. (2007) Preview: Upload and explore data, http://swivel.com, accessed 8 April 2009.

21 Kosara, R. (2007). Review: Swivel vs. Many Eyes, http://eagereyes.org/VisCrit/Swivel-vs-Many-Eyes.html, accessed 8 April 2009.

22 National Science Foundation (NSF) (2008) High performance computing system acquisition: towards a petascale computing environment for science and engineering. http://www.nsf.gov/pubs/2005/nsf05625/nsf05625.htm, accessed 8 April 2009.