Multi-Focus and Multi-Level Techniques for Visualization and Analysis of Networks with Thematic Data

Michele Cossalter, Ole J. Mengshoel and Ted Selker

Carnegie Mellon University, Moffett Field, CA 94035, U.S.A.

ABSTRACT

Information-rich data sets bring several challenges in the areas of visualization and analysis, even when associated with node-link network visualizations. This paper presents an integration of multi-focus and multi-level that enables an interactive approach to multi-step comparisons across different parts of such data sets. We describe NetEx, a visualization tool providing multi-focus and multi-level techniques that enable users to simultaneously explore different parts of a network and its underlying thematic data, such as time series. NetEx, implemented as a Cytoscape plug-in, has been applied to the analysis of electrical power networks, Bayesian networks, and the Enron e-mail repository. In this paper we focus on data from an electrical power network, and demonstrate how NetEx supports the analytical task of fault diagnosis. Specifically, results from a user study with 25 subjects suggest NetEx enables more accurate isolation of complex faults compared to an especially designed software tool.

Keywords: Network visualization, interactive debugging, fault diagnosis, electrical power network, user study, NetEx, Cytoscape

1. INTRODUCTION

Networks and graph representations are used to model complex systems ranging from engineering artifacts to social relationships. To model and analyze such networks mathematically and computationally, formalisms including graphs (both directed and undirected) and probabilistic graphical models (including Bayesian networks[1]) have been introduced and studied extensively in mathematics, statistics, computer science and artificial intelligence.

The nature of these networks and graphical structures, which are often larger or more complex than a person can easily keep in their mind, makes interactive visualization software a fundamental tool supporting learning and analysis. Consequently, there is a rich literature discussing interactive software tools for network visualization.[2–7] Several techniques and different representations have been developed to improve understandability,[8–13] and enabled an analyst to explore the graph and access low level information.[14–21] Many researchers have investigated approaches for efficient representation of multi-dimensional data.[22–27] Multi-focus interaction[28–30] and the use of multiple coordinated views[31–35] to facilitate comparison have been studied as well.

To show a graph structure on a computer display, a spatial layout of the node-link diagram must be selected. Depending on several graph characteristics (e.g. size, connectivity) and layout criteria, various algorithms can be applied to accomplish this task automatically.[36] Alternatively, the layout can be created manually. For analysis or debugging, analysts often need to compare and contrast the details (such as time-series) of several nodes that may be rendered close or far away in the network layout. In general, locating the right subset of nodes in the network to closely analyze and compare might be non-trivial. We have found this to be true when analyzing even moderately-sized networks (say, with a 100 nodes or more). We denote this the *multi-focus challenge* of visualization. One aspect of the challenge is that for even moderately sized graphs, the complete graph and all its node labels may not be simultaneously visible on the computer screen. Since node labels (or similar annotations to nodes) are no longer visible, an analyst needs to pan and zoom in to understand what a particular node or set of nodes represent. After panning and zooming in, on the other hand, the overall context and other important nodes located far away in the graph layout may be difficult to remember. Another aspect of this challenge is that

Further author information: (Send correspondence to O.J.M.)
M.C.: E-mail: michele.cossalter@gmail.com
O.J.M: E-mail: ole.mengshoel@sv.cmu.edu
T.S.: E-mail: ted.selker@sv.cmu.edu

real-world graph data, may result in spatial layouts that are quite dense and cluttered. This is the case for the Enron e-mail data set,[37] for example.

Another research problem, which we denote the *multi-level challenge*, is the problem of effective and simultaneous visualization of thematic data together with the graph structure. Visualizing a large amount of thematic data embedded within the node-link diagram is generally not feasible. Locating the right portion of the data to analyze in depth in order to solve a problem might be like like finding a needle in a haystack. For example, diagnosis of an electrical power network requires checking and comparing time series data from a large number of sensors to look for abnormal behaviors that could flag possible faults. Moreover, only focusing on local information, such as thematic data related to individual nodes, is typically not enough in that the analyst looses context. Global information, i.e. the structure of the network, plays an important role as well in many problem solving tasks. Considering once again the electrical network example, it is important to understand how power flows from batteries to loads through a distribution network. A key requirement for a network visualization tool is that it should be capable of showing clear connections and relationships between pieces of thematic data and the corresponding graph nodes.

The goal of this research is to develop improved visual analytics techniques that meet the multi-focus and multi-level challenges discussed above. Our emphasis is on interactive visualization and analysis of complex networks where both node-link structure and thematic data are important. Based on the well-understood concept of overview+detail,[38] our NetEx tool enriches the node-link representation of networks, providing direct access to other aspects of data that are not encapsulated in the graph structure. Leveraging existing techniques such as visual style, zooming size capability for problem area and details-on-demand with click-and-drag features,[38–40] and expanding the work of Butkewicz et al. on multi-focused analysis using probes,[28] we develop a visualization tool that enriches the positional relative node-edge representation of networks providing easy access to other aspects of data that can not be directly encapsulated in the graph structure. Combining multiple techniques might help orient a deeper analysis of a network. These techniques allow us to simultaneously analyze different aspects and parts of a data set while preserving the structure of the network. Starting from a network overview, the NetEx tool can be successfully used to simplify complex analysis and fluidly support exploration and comparison of multiple levels of the network and related data.

We hypothesize that the techniques integrated in NetEx, while individually relatively simple, are versatile and can by applied to networks of various sizes and types. We have applied NetEx to Bayesian networks [41,42] in addition to the electrical power network and social networks case studies discussed in this paper. The generality of the simple approach makes it useful in problem solving situations of "needle(s) in a network," for example when an analyst is looking for one or a few "bugs," something that is an anomaly or not intended. In a biological network, an analyst might be hunting for genes that cause disease or extraordinary behavior. In a social network, like the one we created from the Enron emails, one might be trying to identify incriminating evidence of misconduct.

Electrical networks, the focus of this paper's user study, have been interestingly visualized previously.[7,11] Electrical power plays an essential and increasing role in vehicles, including cars, aircraft, and spacecraft. In aerospace, for example, electrical power loads include avionics, propulsion, life support, and thermal management. Unfortunately, electrical power is not without problems. On September 2, 1998, Swissair 111 crashed into the Atlantic Ocean, killing all 229 people on board. It was determined that wires short-circuited and led to an electrical fire. A battery failure occurred on the Mars Global Surveyor on November 2, 2006, ultimately leading to a premature end to its mission. Most likely, a software error caused the battery to overheat due to over-exposure to sunlight. These and similar incidents and accidents emphasize the importance of detecting, isolating, and mitigating faults in electrical power networks. The difficulty of these tasks increases with the size and complexity of the network, the number and complexity of sensors, the sampling rate of sensors, and the number of and degree of interaction and ambiguity between faults.

We first present background and related work in Section 2. Section 3 presents the capabilities of the NetEx software. Starting from a network overview, NetEx analysts can successfully solve complex problems, supported by capabilities for exploration of the network and related data, enabling analysts to compare data values of multiple nodes from various foci in the network. In Section 3 we discuss NetEx in the context of visualization for example networks, with emphasis on a vehicular electrical power network.[43] Here, the analytical task, which can be be considered a knowledge crystallization task,[44] involves examining and comparing time series for a set of sensors (voltage, current, temperature, etc.) in order to diagnose which components of the network broke and at what time it happened. This requires three main steps, namely (i) search for irregularities in sensor time series, (ii) detection of fault candidate nodes in the network, and (iii) isolation of faulty nodes. Section 5 reports on a user study with 25 subjects in which two tools, NetEx and Intelliviz, were compared.

Our results suggest that NetEx improves the accuracy of complex problem solving task compared to Intelliviz, * a custom software tool that has been productively used to visualize several NASA data sets, including data from ADAPT. In Section 6 we conclude and outline future work.

2. BACKGROUND

Appropriate visualization techniques might speed and simplify the analysis and exploration of complex datasets, such as electrical power networks. Some dimensions of heterogeneous data may be naturally represented in the form of a (weighted) graph and visualized as node-link diagrams. A graph $G = (\mathbf{V}, \mathbf{E})$ consists of a set of nodes $\mathbf{V}$ and a collection of pairs of nodes called edges $\mathbf{E} = \{(X, Y) \,|\, X, Y \in \mathbf{V}\}$. A weighted graph (G, W) is obtained by adding a weighting function $W : \mathbf{E} \to \mathbb{R}^+$. Heterogeneous datasets also include other dimensions of data, which in this paper will be referred to as *thematic data D*, that are not directly exposed in the graph structure. The nature of thematic data D varies; examples include time series of physical measurements, as it is for ADAPT, or the textual contents of e-mails or tweets in social networks. The entire network dataset can be described as a tuple $N = (G, W, D)$.

2.1 Graph Navigation and Exploration

Several approaches have been developed to visually explore large graphs. Zooming and panning of the graph representation allow users to switch between overview and detailed representations.[38] More sophisticated interactive graph visualizations provide an integrated representation of details and overview. Research in this area includes the work on subgraph discovery by Faloutsos et al.,[14] the interaction techniques for subgraph selection and manipulation by McGuffin and Jurisica,[16] the work on sigma lenses by Pietriga et al.,[17] and research on fisheye view by Schaffer et al.[18] and Gansner et al.[15] Relevant work on visual exploration of large graphs also includes the combination of semantic and geometric distortions in the visualization system by van Ham and van Wijk,[19] the cross-zooming approach by Wong et al.,[21] and the novel "search, show context, expand on demand" model by van Ham and Perer.[20] Multiple coordinated views, supporting investigation with distinct views showing different aspects of the same conceptual entity, have been used as well.[31,32,35,45]

NetEx integrates some of the above techniques—especially zooming, panning, and multiple coordinated views—with interactive exploration across multiple node foci.

2.2 Multi-level and Multi-dimensional Data

While much information can be encoded in a node-link structure, visualizations have difficulty simultaneously showing underlying thematic data related to nodes in their network relationships. Significant underlying data that exist in a different representation, such as time series, categorization of messages or the main topics discussed in e-mails, is very hard to visualize in a node-link representation. Techniques for multidimensional data analysis and comparison have been investigated, among others, by Lee et al.,[23] Weaver[27] and Lex et al.[24] Also, several multivariate network visualization systems have been developed, such as Wattenberg's PivotGraph,[26] GraphDice by Bezerianos et al.,[22] and FlowVizMenu by Viau et al.[25] The Action Science Explorer[46,47] integrates visualization, statistics, and text analytics; this tool provides a powerful node-arc visualization of scientific citation networks.

Building on work that displays thematic data such as texts, statistics, or time-series, NetEx simultaneously displays such thematic data for interesting nodes in various parts of a network.

2.3 Multi-focus and Multi-point Interaction

Selection of multiple nodes, separated by many links, for detailed comparison might help analysis of graph structures related to large and complex datasets. In this direction, Shoemaker and Gutwin[30] investigated techniques to support multi-point interaction. Also, multi-focus analysis has been studied by Butkiewicz et al.,[28] who describe probe interfaces for geospatial visualization, and Elmqvist et al.,[29] who exploit space folding to guarantee visibility of multiple focus regions. Research has also shown the benefit of linking, as opposed to highlighting, in visual data analysis tasks.[48]

It is the linking of such thematic data across different parts of a network that we think becomes especially interesting for interactive network exploration as provided by NetEx.

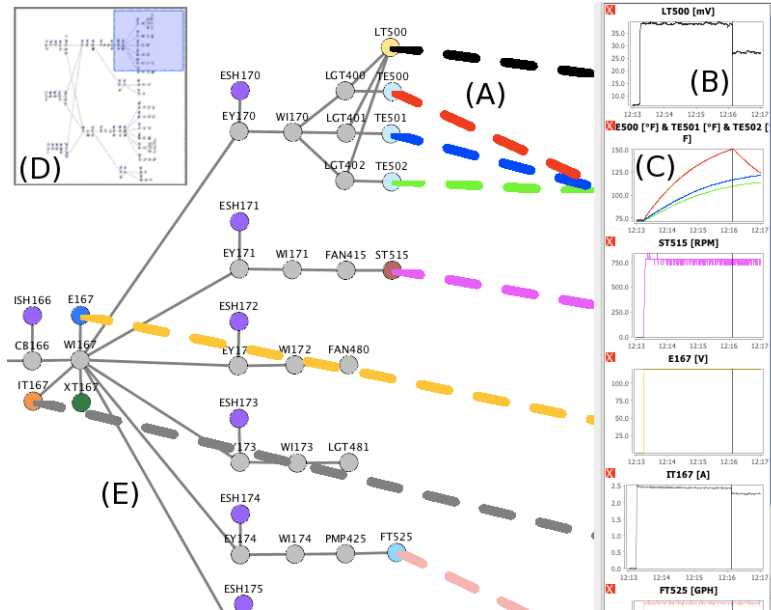*http://www.stottlerhenke.com/projects/intelliviz/

Figure 1. Network visualization with NetEx. Anchoring the thematic data to the network view with dashed bubbles (A) allows low-level focused analysis and comparison. Data boxes (B) can be stacked in a panel alongside of the detailed view (E). A scrollbar allows the panel to host many boxes containing plots. A trace-line is displayed across all plots to facilitate comparisons. Data from different nodes (in this case TE500, TE501, and TE502) can also be merged into a single box just by dragging a box and dropping it over another one (C). The overview window (D) prevents the analyst from losing the complete picture of the network.

2.4 Graph Readability and Filtering

Several techniques have been investigated to address the problems of density and clutter sometimes associated with node-link visualizations, e.g. the hyperbolic browser by Lamping et al.,[10] the flow map layout by Phan et al.,[12] EdgeLens by Wong et al.,[13] and Holten's hierarchical edge bundling.[9] Dynamic filtering to hide the least interesting edges, as well as using opacity to emphasize important edges while keeping all edges visible are other typical solutions to clutter.[36] Overbye and Weber[11] have investigated different methods for visualization of electric power systems, including animation, dynamic device sizing, contouring, data aggregation and 3D views. Also, a very compact representation of graphs can be obtained by replacing traditional node-link diagrams with adjacency matrices or hybrid representations, e.g. NodeTrix by Henry et al.[8] However, in an adjacency matrix one may loose some structure that can be easily seen in node-link diagrams.

NetEx could be extended with the above readability, filtering, and clutter-reduction techniques; adding to them the idea of interactively evaluating thematic data for interesting nodes in various parts of a network.

2.5 Summary

While other techniques and tools discussed above are good at focusing and browsing, our goal is to support visualization for the purpose of interactive problem solving in networks with thematic data. We now turn to a more detailed presentation of the visualization techniques that we are investigating in this paper.

3. NETWORK VISUALIZATION AND ANALYSIS WITH NETEX

We describe a multi-focus multi-level visualization approach, implemented in a software tool NetEx that supports integrated network analysis and interactive visualization. NetEx gives the analyst a familiar graph view with modified layout to focus on multiple parts simultaneously using a distinct bubble attachment to other visually associated views. We integrate existing techniques[28, 31, 38] to improve multi-concept combination comparisons. The visualization technique considers overview+detail as well as focus+context, in a seamless fashion. The visual representations can be zoomed and panned at all times for overview+detail and provide enhanced interactive focus+context techniques on demand.

NetEx[41,42] is an interactive visualization tool for integrated network analysis, implemented in the form of a Cytoscape[49] plug-in. It gives the analyst a familiar graph view with modified layout to focus on multiple parts of a network simultaneously, distinguishing data at different representation levels with a stylized "thinking bubble" connection, taken from the thought bubbles found in cartoon strips. Each bubble link connects a node in the graph and a pop-up data box showing its thematic data, a time-series plot in the case of electrical power networks.[†] This facilitates the comparison of thematic data, such as time-series, for multiple nodes while at the same time connecting the thematic time-series to their respective locations in the graph structure. The bubble links support both multi-level and multi-focus analysis, which we now discuss.

3.1 Multi-level Analysis

Figure 1 shows a screen-shot of NetEx. The network *overview* is always visible (Figure 1(D)). The *detail view* allows comparison of semantics not brought out in the overview. Zooming and panning allow the analyst to restrict attention to a particular area of the graph, which is visualized in the center of the display (Figure 1(E)). NetEx enables the analyst to explore data at the desired level of detail (i.e., *multi-level analysis*) without losing the complete picture of the network. Thus, the tool is preserving the capability of being orienting even when a very small area is selected for focus in the detail view (*overview+detail*[38]).

NetEx promotes interactive visual exploration, allowing users to select a particular element of the graph, such as a single node or a set of nodes, and to access underlying information about that element not directly exposed in the network structure (thematic data). In the ADAPT network example, thematic data consists of line chart views of time series data corresponding to sensor readings. The thematic data is attached to the network view with a line of bubbles which anchors the data box to the node it is related to (Figure 1(A)). The use of bubbles instead of solid lines distinguishes between graph edges (solid lines) and anchors (bubble lines).

3.2 Multi-focus Analysis

By supporting the simultaneous opening and comparison of thematic data, such as sensor reading time series, for nodes from different parts of a network, NetEx enables what we call *multi-focus analysis*. This feature, which can be seen as an extension of the probes proposed by Butkiewicz et al.[28] to graphs, should be useful to any system that has more than one level of representation. Many analysis tasks—including fault diagnosis—require users to inspect and compare multiple nodes at the same time in order to make sense of what is happening in the system under study.

NetEx simplifies comparison of different nodes, along with their thematic data, in several ways. First, the analyst can "merge" thematic data from different nodes into a single box by dragging a box and dropping it over another one (*Drag & Merge*), as illustrated in Figure 2. This is particularly useful when the number of compared nodes is large. A different color is associated with each of the nodes to be compared. Both the line in the plots and the bubbles are painted with this color to allow for immediate association of nodes with corresponding data. Original colors are preserved in the merged box to help the analyst maintaining association between nodes and data, even when the latter is moved into a different box. Moreover, plots are lined up vertically in the side panel and a trace-line is displayed across all of them to help identifying events occurring at the same time for different nodes (Figure 1(B)). NetEx also allows the analyst to change the axis scale of data boxes so that a common scale is adopted.

4. VISUALIZATION AND ANALYSIS OF NETWORKS: CASE STUDIES

One of the major drivers of the progress of humanity is the enhancing of our abilities through external aids.[50] There is an increasing gap between, on the one hand, the rich world emerging through the steadily increasing capabilities of sensors, networks, and computers, and the capability of humans to absorb, analyze, and act accordingly on the other hand.

Software tools for visualization and analysis of networks, such as NetEx, make up one avenue of attack to bridge this gap. We have applied NetEx in several case studies. In a first case study, discussed in Section 4.1, we use our visualization tool to diagnose faults in an electrical power network. Here, we use data from ADAPT,[43] a real-world electrical power network. We compare NetEx with a baseline tool; this comparison is the basis for the formal user study in Section 5. In a second case study, NetEx is applied to a social network built from the Enron e-mail data set.[37] Section 4.2 shows how NetEx helps detect two key players in the company, namely Jeff Dasovich and Bill Williams, that were not previously

---

[†]NetEx provides, for the data boxes, both a completely separate visualization as well as floating diagrams. We experimented and found pros and cons of both approaches, and therefore ended up providing both options in NetEx.
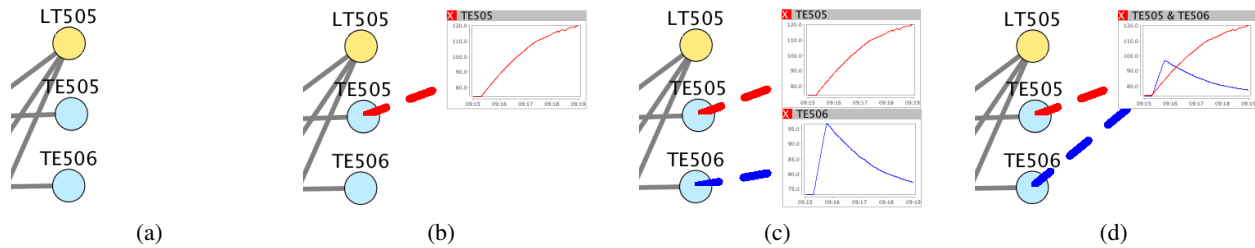
Figure 2. Alternative NetEx displays of a node-link graph in combination with detailed data shown in data boxes, illustrated using the ADAPT electrical power network dataset: (a) no data box is shown; (b) the box of TE505 is shown; (c) the box of TE506 is also shown; (d) the two boxes are merged. Original colors are preserved in the common box to help the analyst maintaining association between nodes and data, even when the latter is moved into a different box.

known to the analysts using NetEx. A third case study shows how the techniques described in this paper allow better interactive analysis of Bayesian networks,[1] especially comparison of conditional probability tables within them.[41, 42]

4.1 Diagnosing Faults in Electrical Networks

This section describes the use of NetEx to diagnose faults in a real-world electrical power network, known as ADAPT,[43] which has capabilities for power storage, distribution, and consumption. ADAPT contains batteries, electromechanical relays, circuit breakers, and different kinds of loads, such as pumps, fans, and light bulbs. Several sensors are available, measuring voltage, current, relay position, temperature, light intensity, and liquid flow. We consider scenarios taken from the ADAPT testbed, more specifically Tier 2 of DX 2009 competition dataset.[‡] These scenarios consist of nominal runs, where faults were not injected into ADAPT, as well as faulty runs involving one or more faults in components or sensors.

4.1.1 Fault Diagnosis Examples

We are interested in diagnosing which component(s) broke and at what time it (they) broke. By examining and comparing time series for a selected set of sensors (voltage, current, temperature, etc.), it is possible to diagnose a faulty scenario. For example, consider a case in which light bulb LGT400 breaks, stopping emitting light and heat and drawing no current. Current sensor readings upstream of the broken component will show an abrupt drop, and the light (LT500) and temperature (TE500) sensors attached to the bulb will stand clearly out and show abnormal behavior (Figures 1 and 3). Specifically, a sudden drop of light intensity can be seen for LT500, while a sharp inversion of trend (from slowly increasing to decreasing) is observed for TE500. In contrast, other sensors on the same path, such as TE501 and TE502, show the expected slowly increasing behavior. This suggests that power is correctly supplied to most loads, and the problem must be related to one load, LGT400.

The fault isolation task in an interactive visualization involves several steps such as scrolling and zooming, as well as multi-step comparisons. Our hypothesis is that NetEx, by supporting users in comparing readings from multiple sensors and quickly identifying interesting parts of the network, provides strong support for fault diagnosis processes.

4.1.2 Baseline Visualization Tool: Intelliviz

For the purpose of comparison, we used *Intelliviz*,[§] an existing visualization tailor-made for the ADAPT electrical power network. The tool provides two windows: a *network window* (Figure 3(B)), and a *sensor window* showing sensor readings (Figure 3(A)). There is no link between the windows, so that a plot can only be associated with the corresponding network node based on the common name. Plots of sensor readings are organized alphabetically. Plots corresponding to nodes which are very close in the network and interacting with each other may be very far apart in the sensor window and hard to compare. Since the ADAPT network contains about 170 nodes, half of which are sensors, the sensor window must be scrolled down several times in order to see all the plots. Finally, no overview of the network nor zooming are provided.

In order to perform the fault isolation task with this tool, the following steps are necessary:

1. Search for irregularities in sensor time series. Since the number and position of components and sensors involved in faults are unknown, all plots must be analyzed, but due to limited screen space it is not possible to look at all plots at the

---

[‡]http://www.dx-competition.org/
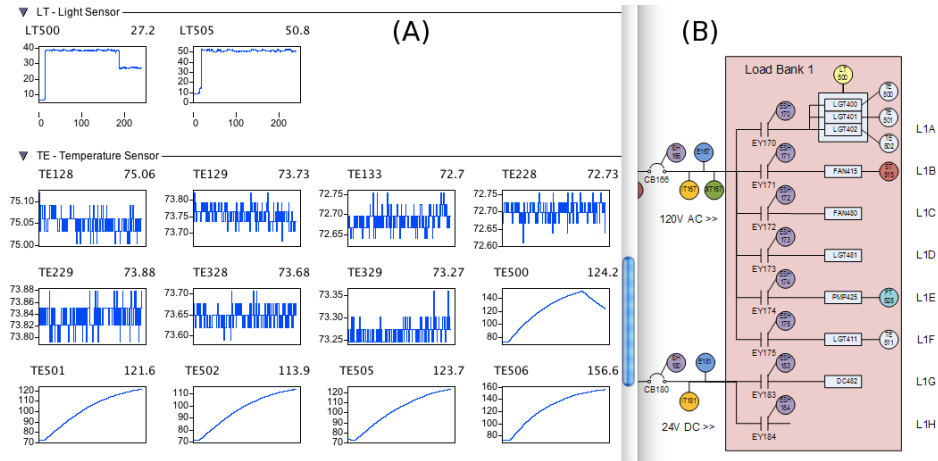[§]http://www.stottlerhenke.com/projects/intelliviz/dxcompetition/index.htm

Figure 3. Intelliviz, a visualization tool for the ADAPT electrical power network. Two windows are provided: sensor readings window, with time series plots, shown to the left in the foreground (A) and network window shown to the right in the background (B).

Diagnosis Step	Intelliviz	NetEx
1) Search for irregularities in sensor time series	• Cannot mark interesting plots • All plots automatically shown	• Interesting plots can be kept • Plot creation takes time and effort
2) Detect fault candidate nodes	• Plots potentially hard to compare due to ordering according to type and name • No visual link between plots and nodes	• Plots easy to compare thanks to box stacking, *Drag & Merge*, and common scale • Bubbles allow for immediate matching of plots and nodes
3) Isolate faulty nodes	• Requires switching between sensor and network windows • Harder due to lack of preparation in search and detect steps	• No need to switch between different windows • Easier thanks to preparations made in search and detect steps

Table 1. Comparison between Intelliviz and NetEx with regards to each step of a diagnosis task. NetEx allows for plots of interest to be selected and shown together in the same view, anchoring data to respective nodes through bubbles. During the first step, NetEx sets an analyst up for the second and third steps, which can consequently be performed faster than using the baseline tool, Intelliviz.

same time. Also, there is no way to mark interesting plots, such as LT500 and TE500 in Figure 3(A). The analyst is thus required to remember (or manually write on the side) the names of the fault candidate nodes.

2. Detect fault candidate nodes in the network. Since there is no link between the plots and the network view, fault detection requires the analyst to match sensor plots (Figure 3(A)) with network nodes (Figure 3(B)) based on component names only. Due to lack of explicit links (such as the bubble anchors of NetEx), this may imply visual inspection of all nodes in the network. Many panning operations could be necessary, besides continuous switching between network window and sensor window (or list of candidate node names, either written in a separate text editor or on a sheet of paper).

3. Isolate faulty nodes. The location of faulty components and sensors can be inferred by carefully analyzing the detected irregularities (sudden current drop, slowly decreasing temperature, etc.), taking into account the relationships between candidate nodes (current nodes on the same path as faulty light bulb LGT400, sensors LT500 and TE500 connected to LGT400). This step also requires panning and switching between the network and sensor windows (Figure 3).

4.1.3 Novel Visualization Tool: NetEx

Assuming that the side panel is used to display thematic data, NetEx enables an analyst to perform the same steps as Intelliviz, but in a different way:

1. Search for irregularities in sensor time series. The analyst needs to first select a group of nodes (Figure 1(D)) with one cursor drag, and then display the corresponding sensor plots in the side panel (Figure 1(B)) with one key stroke. Plots corresponding to candidate nodes can be kept in the side panel, making the following steps straightforward. For example,
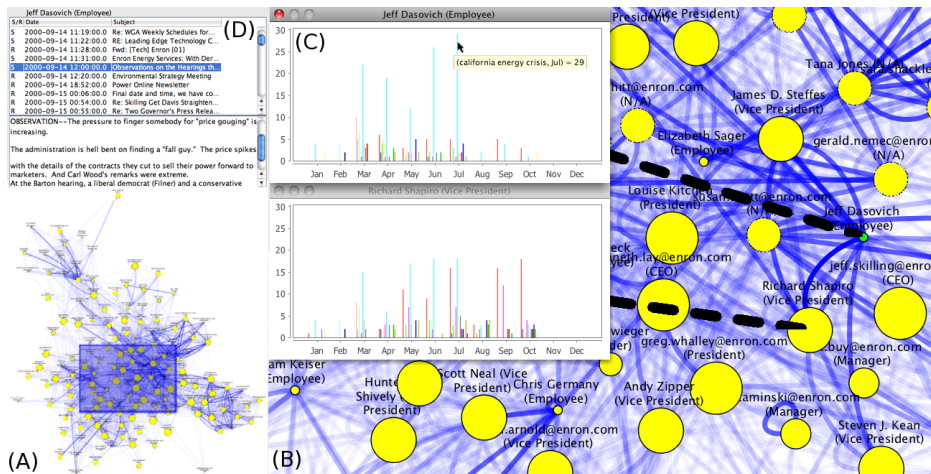
Figure 4. Multi-focus and multi-level representation, in NetEx, for Enron social network: (A) overview level, (B) detail level, (C) data level, and (D) datum level. Anchoring the data level to the network view with large dashed bubbles allows low-level focused analysis and comparison while preserving the structure of the network.

let us consider Figure 1, where the analyst has selected and expanded a number of nodes. After visual inspection of node data boxes, an analyst may realize that only LT500, TE500 and IT167 are abnormal. The analyst might thus remove from the side panel the node boxes related to all other sensors, as they are no longer relevant to fault diagnosis.

2. Detect fault candidate nodes in the network. Bubble links enable immediate matching of candidate plots in data boxes with the corresponding nodes (Figure 1(A)). No visual search of node labels in the network, nor switching between different windows, is needed.

3. Isolate faulty nodes. Comparison of data is easier in this case because (i) only interesting plots are shown in the side panel, and (ii) data from different nodes can be merged in the same box (Figure 2). There is no need to switch between different windows thanks to the visual link between nodes and data (bubbles). Consider again Figure 1: after irrelevant boxes are removed, bubble lines enable an analyst to quickly realize that all interesting sensors (LT500, TE500 and IT167) are located on the same path.

4.1.4 Comparison of NetEx and Intelliviz

Both Intelliviz and NetEx support electrical power system analysis and problem solving. Specifically, we discussed above how three problem solving steps of diagnosis—search, detect, and isolate—are supported by both these software tools. The different capabilities of each tool, with respect to the three problem solving steps, are summarized in Table 1.

4.2 Social Network Visualization

NetEx has been applied to visualize and analyze the well-explored Enron data set.[52] Consisting of $250,000$ e-mail messages exchanged between 150 employees over a period of about 3 years, this corpus has been widely used for demonstrating knowledge discovery and visualization techniques. NetEx enabled the analysts to detect two key players in the company, namely Jeff Dasovich and Bill Williams, that they had not previously identified using other visualization tools.

Focusing on Dasovich, NetEx shows interesting areas of the graph that are very clearly distinguishable at a glance, even if the whole network is displayed (Figure 4(A)). By using the overview+detail capability of NetEx, the analyst's attention was captured by a very small node (i.e. marginal role in the organization) with (i) an impressive number of exchanged messages (i.e. very dark edges), (ii) a significant number of edges, and (iii) connections with large nodes (i.e. prominent roles). Zooming and panning gives a more detailed view of this node, corresponding to the employee Jeff Dasovich, and its neighbors (Figure 4(B)). Now it is quite easy to realize that there is a significant interaction between Dasovich and two vice presidents, namely James Steffes and Richard Shapiro. These relations, which NetEx enabled the analysts to discover quite naturally, had not been noted previously in other visualizations of the data. As pointed out by Pathak and Srivastava,[53] on one hand the frequency of interaction between these people is relatively high with respect to the average of the network. On

the other hand, this also are concealed relations, meaning that only the involved individuals are privy to these relationships, while the rest of the network is not aware of them. Concealed relations such these are of great interest to analysts since they may hide malicious activities.

We can get more information about the topics discussed by Dasovich in his e-mails by activating the thematic data box, another multi-level capability of NetEx, for his node. As Figure 4(C) shows, *California Energy Crisis* is the dominant topic of Dasovich's messages, especially in the summer of 2001, just before the Enron scandal broke. This information might be very relevant for analysts interested in identifying the main actors involved. The concealed nature of these relationships can be investigated in detail using the NetEx message viewer (Figure 4(D)).

5. USER STUDY: DIAGNOSING FAULTS IN ELECTRICAL NETWORKS

In order to test our hypotheses about the benefits of NetEx compared to Intelliviz, we conducted a user study in the area of debugging electrical networks (see Section 4.1). In this section we report on the design and results of this user study, including both objective and subjective results.

5.1 Subjects, Tutorials, and Tasks

We asked 25 subjects to use the two tools to diagnose faults in the ADAPT electrical power network. Subjects were graduate students at the CMU Silicon Valley campus, with at least some knowledge about electrical circuits. Subjects (23 males and 2 females) were in the age range 20–30 years and volunteered to participate in the study with no cash compensation. Before an experiment, we guided the subject through a tutorial presenting the ADAPT electrical power network and the visualization tools. We also guided subjects through one tutorial task per tool, similar to the discussion in Section 4.1, to show how Intelliviz and NetEx can be used for fault diagnosis.

During training, each subject was shown how to use the tool going through two instructive tutorials which took them approximately 30 minutes. Tutorial 1 is an example of a fault occurring to a load (FAN415). Tutorial 2 is an example of a fault occurring to a connecting component (CB136).[¶] Each of the tutorial tasks aligns better to one of the experimental tasks, but since all users used both tools (as further discussed below), we believe there is no or little bias. Subjects were instructed to look at the time series plots showing sensor readings to find out which component(s) or sensor(s) failed, and at what time it happened. Subjects were not given any particular indication about preferring speed or accuracy. They were asked to analyze data and look for faults, with a hard time limit of 20 minutes per scenario.

We considered two different scenarios (*Task 1* and *Task 2*), each with two concurrent failures. Subjects were randomly placed into two groups (13 people in *Group 1* and 12 in *Group 2*). Each subject used both Intelliviz and NetEx. Subjects in *Group 1* first used Intelliviz for *Task 1*, and then NetEx for *Task 2*. Subjects in *Group 2*, instead, first used NetEx for *Task 1*, and then Intelliviz for *Task 2*.

Task 1 consisted of a scenario with two faulty loads (light bulbs LGT406 and LGT407). The faults can be detected, see Figure 5(a), by noticing the drop of temperature reported by the sensors attached to the loads (TE506 and TE507), the drop of light intensity reported by the sensor LT505, and the drop of current in the path from the battery to the loads (sensors IT261, IT267, and IT340). This task was intended to be relatively easy, as there was very little ambiguity in the evidence and a very similar example had been shown during the tutorial.

Task 2 consisted of a scenario with a faulty circuit breaker (CB180) and a faulty sensor (ESH160A). As illustrated in Figure 5(b), the circuit breaker fault can be detected by noticing the change of state, from closed to open, reported by the sensor attached to the circuit breaker (ISH180) and noticing the drop of current and power downstream of the circuit breaker (IT181 and E181). The issue with the sensor could be identified by noticing that all the other components downstream of the relay were working fine. In other words, the problem was the sensor itself, not the relay. This task was intended to be harder than *Task 1*, due to the possible confusion from having two concurrent faults of different nature on the same path.

[¶]Tutorial 1 (Exp_792_pb_t2), Tutorial 2 (Exp_758_pb_t2), Task 1 (Exp_814_pb_t2), and Task2 (Exp_766_pb_t2ff) are available on-line at `http://www.stottlerhenke.com/projects/intelliviz/dxcompetition/datamontage/competition/index_t2.htm`.
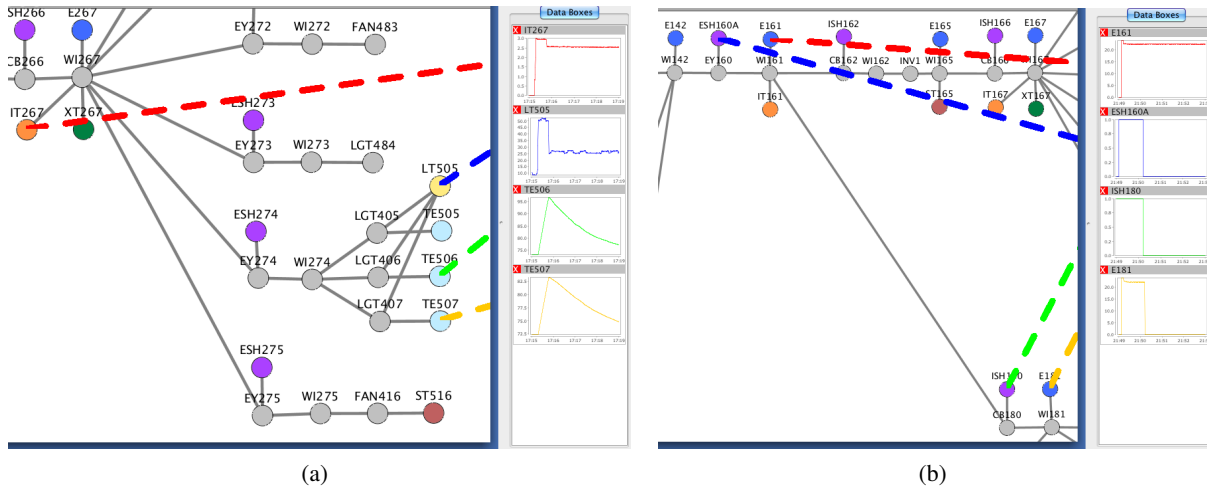
Figure 5. Key NetEx plots for experimental tasks. (a) *Task 1*: the drop of temperature reported by sensors TE506 and TE507, the drop of light intensity reported by sensor LT505, and the drop of current reported by sensor IT267 indicate that light bulbs LGT406 and LGT407 are very likely to be faulty. (b) *Task 2*: the change of state reported by sensor ISH180 and the drop of power reported by sensor E181 indicate that circuit breaker CB180 is very likely to be faulty. The change of state reported by sensor ESH160A is not supported by a drop of power in sensor E161, suggesting that the sensor itself might be faulty in this case.

5.2 Answers and Analysis

Answers from subjects were recorded both during the experiments with Intelliviz and NetEx as well as afterwards. During the experiment, for each detected fault, subjects were asked to report the faulty component or sensor, along with the time when the problem happened, and a short explanation of their answer. We also recorded the time when they detected a fault; no feedback was provided to the subject as to whether the fault they had found was correct or not. The answers were then analyzed as discussed below; our results are presented in Section 5.3. After the experiment, subjects were asked to answer a post-task questionnaire, providing their experience around the tasks and tools. These answers were then analyzed as discussed below, leading to the results presented in Section 5.4 and Section 5.5.

We used two objective metrics to analyze users' performance on each task: the F_1 score to measure accuracy and the diagnosis time to measure speed. The F_1 score is defined as:

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN},$$

where TP is the number of correctly identified faults (*true positives*), FP is the number of false alarms (*false positives*), and FN is the number of missed faults (*false negatives*). The F_1 score measures the accuracy of a subject's answer, taking into account both correct diagnoses and mistakes. The best F_1-value is 1, meaning no mistakes ($FP + FN = 0$), while the worst F_1-value is 0, meaning no correct diagnoses ($TP = 0$).

In five cases, the subject's brief explanation for a diagnosis did not agree with the reported fault. We determined that subjects accidentally confused the name of a component with the name of the attached sensor, and gave preference to the explanation. Moreover, subjects who obtained a $F_1 = 0$ score for a task were not considered in the following analysis for that task.[||] This happened for one subject in *Task 1* and six subjects in *Task 2*.

5.3 Objective Results

Figure 6(a) summarizes the results of subjects using Intelliviz and NetEx on Task 1 (simple task) and Task 2 (complex task). The time subjects used to finish a task is summarized on the *x*-axis, while the accuracy is summarized on the *y*-axis. Averages and standard deviations for both dimensions are shown. Indeed, the task difficulty was about right, as most of

---

[||]Note that no subjects abandoned the study. However, some subjects obtained a $F_1 = 0$ score, which we took to mean that they either did not understand the task or ran out of time before finding a meaningful answer to the task.

Task	Tool	F_1 **Accuracy**	**Diagnosis Time**
Task 1	IntelliViz	0.92 ± 0.16	8.08 ± 4.31
Task 1	NetEx	1.00 ± 0.00	9.18 ± 5.72
Task 2	IntelliViz	0.73 ± 0.22	9.44 ± 3.17
Task 2	NetEx	0.87 ± 0.17	12.90 ± 3.41

Table 2. Mean and standard deviation of the F_1 score and diagnosis time for each of the two groups on each of the two tasks. See Figure 6(a) for an overview.
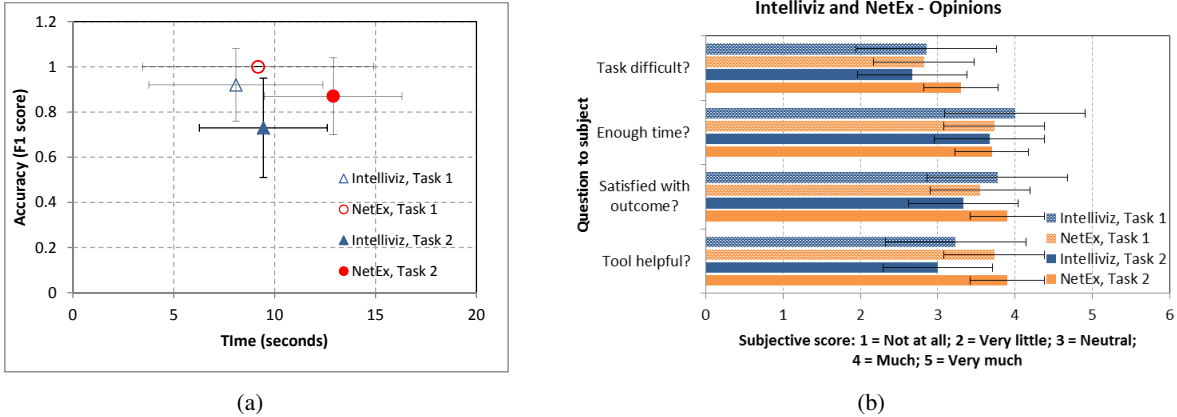

(a)   (b)

Figure 6. Comparison of Intelliviz and NetEx. (a) Performance in terms of time (*x*-axis) and accuracy (*y*-axis). Subjects were faster but less accurate in their diagnostic problem solving when they used Intelliviz compared to when they used NetEx. (b) Answers summarized from post-questionnaire. When using NetEx, subjects appropriately thought that Task 2 was more difficult than Task 1; there is no such difference for Intelliviz.

the subjects took much of the allocated time to finish *Task 2*. For both Task 1 and Task 2, the results suggest that NetEx is more accurate while Intelliviz is faster. This is especially prominent for Task 2 (the complex task).

Table 2 reports the mean and the standard deviation of the F_1 score for each of the two groups on each of the two tasks. For both tasks, subjects who used NetEx were on average more accurate than subjects who used Intelliviz. To analyze this statistically, a *t*-tests was used to test whether the mean for one group is greater than for the other group. For Task 1, NetEx was more accurate than Intelliviz with $p = 0.061$;** there was no significant difference between the completion times. For Task 2, Intelliviz was faster than NetEx with $p = 0.018$; however NetEx was more accurate than Intelliviz with $p = 0.075$.

One interpretation of the above result is that NetEx enables more accurate results thanks to its key multi-level and multi-focus capabilities: the bubble anchors, the side panel, and the capability of showing and hiding plots on demand. The combination of these features, in fact, allows for immediate matching of candidate plots with the corresponding nodes in the network. This is particularly crucial for *Task 2*, where the interaction between the faulty component requires more than one step of the analysis.

Table 2 reports the mean and the standard deviation of the diagnosis time, in minutes, for each of the two groups on each of the two tasks. Subjects who used Intelliviz were on average significantly faster than subjects who used NetEx on the more difficult *Task 2* ($p = 0.018$); this is not the case for *Task 1* ($p = 0.298$).

We have two hypotheses for why Intelliviz users were faster, on average, than NetEx users. First, some subjects using Intelliviz quit very quickly (perhaps thinking, typically incorrectly, that they had solved the diagnosis problem and were done) but in fact did not perform very accurately compared to subjects using NetEx. This hypothesis is supported by the lower accuracy score for Intelliviz, see Table 2 and Figure 6(a). Second, Intelliviz shows all plots simultaneously in the *sensor window*, and in addition this Intelliviz window can host a larger number of plots than the NetEx side panel in a single view (about 20 as opposed to 5). This second point suggests that the search step will generally be faster in Intelliviz than in NetEx. However, NetEx users are supported by the more interactive and scalable approach of being able to select a

---

**We realize that the use and interpretation of *p*-values is somewhat controversial and emphasize that when $0.1 \geq p > 0.05$ this is often considered "weakly significant" or a "trend." In this paper, we include the specific *p*-value for $p < 0.1$, or show error bars, and avoid the use of descriptive terms (like "strongly significant" and "weakly significant").

What do you think of these NetEx features?

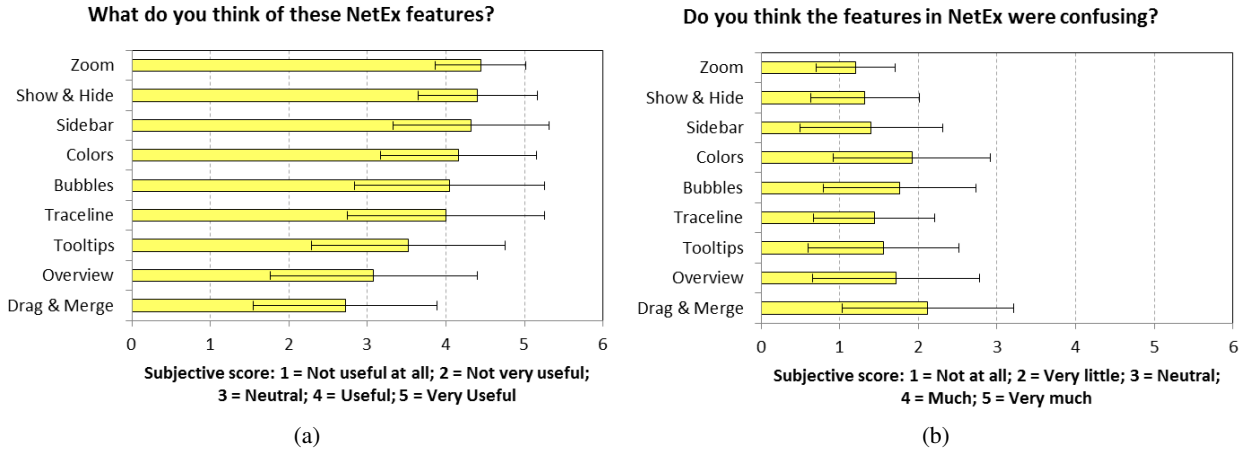**Do you think the features in NetEx were confusing?**

Figure 7. Opinions about different features of NetEx, summarized from post-questionnaires. Only the features Drag & Merge and possibly Overview were seen as not useful–see panel (a); no features were seen as confusing–see panel (b).

group of nodes they want to inspect and to execute another action (key stroke) to see the plots. The selection is performed by clicking and dragging the mouse over the desired nodes.

5.4 Subjective Results: Intelliviz versus NetEx

In one part of the post-questionnaire, we asked the same questions about the experimental system NetEx and the control Intelliviz. Figure 6(b) summarizes the results, in terms of averages and standard deviations (error bars) of four post-experiment questions asked about Intelliviz and NetEx. Note that exactly the same questions were asked about Intelliviz and NetEx. While there is substantial overlap between the error bars and means in many cases, several interesting trends emerge. For *Task difficult?*, only when NetEx was used is there a (correct) indication that Task 2 (the complex task) was harder. Using a *t*-test, we found in Task 2 that the *Task difficult?* score for NetEx is higher than that for Intelliviz (with $p = 0.060$). For *Satisfied with outcome?*, there appears to be a difference in outcome satisfaction in favor of NetEx only for the complex task. Using a *t*-test, the score for NetEx is for Task 2 greater than that for Intelliviz (with $p = 0.053$). For *Tool helpful?*, there appears to be a bigger gap between NetEx and Intelliviz for Task 2, in favor of NetEx. Using a *t*-test, in Task 2 the score on the *Tool helpful?* question for NetEx is greater (better) than the score of Intelliviz, with $p = 0.019$.

5.5 Subjective Results: Opinions about NetEx

Figure 7(a) and Figure 7(b) show results, in terms of averages and standard deviations (error bars), for two post-experiment questions asked about nine different features of NetEx. The nine features were: Bubbles, Colors, Drag & Merge, Overview, Show & Hide, Sidebar, Tooltips, Traceline, and Zoom. Figure 7(a), showing results for the question *What do you think of these NetEx features?*, is ordered according to average usefulness, with the most useful feature at the top. The only NetEx feature with an average score worse than (i.e., less than) Neutral is Drag & Merge. We ran a statistical hypothesis test, specifically a one-population *t*-test, with the null hypothesis that $\mu = \mu_0 = 3.0$ (since Neutral = 3.0). All results are (in terms of usefulness of NetEx features) significantly away from Neutral = 3.0, except for Drag & Merge and Overview.[††]

We hypothesize that the reasons for these NetEx features being less useful include the following. Subjects did not use Drag & Merge very much; they may have been able to perform the tasks just by comparing plots (taking advantage of the Trace-line). Perhaps this feature was not required to solve the experimental tasks, or perhaps learnability was a reason this feature was not used much. Drag & Merge may not have been emphasized enough during the tutorial compared to other features such as Bubbles. Also, it was not possible to split plots that had been merged together by means of Drag & Merge, and this might have prevented subjects from using this feature as much as we had hoped for. As for Overview, we now think that the experimental ADAPT graph was not large enough to really test the usefulness of this feature.

[††]Using a one-population *t*-test, the level of significance was better than $\alpha = 0.025$ for Tooltips and $\alpha = 0.005$ or better for the remaining NetEx features (excluding Drag & Merge and Overview).

Figure 7(b) is structured essentially the same way as Figure 7(a), except we here focus on results for the question *Do you think the features in NetEx were confusing?* Interestingly, all averages are better than (less than) Neutral and statistically significantly away from Neutral = 3.0.[‡‡] This suggests that no NetEx features were, in general, confusing to users.

5.6 Discussion

Our results suggest that NetEx users perceived the task to be harder compared to Intelliviz users ($p = 0.060$); they were nevertheless (and correctly) more confident about their performance on the task ($p = 0.053$). Moreover, NetEx users were more satisfied with the tool ($p = 0.019$).

For the relatively small ADAPT graph, our results suggest that Intelliviz is faster than NetEx on the search step, because it automatically shows thematic data for all nodes. All the time series plots must be examined and, at least for a relatively small network such as ADAPT, this can be done quicker when all the plots are available in the same window, as it is in Intelliviz. Sadly, Intelliviz shows thematic data for all nodes and this approach does not scale. Also, for more complex tasks looking at all thematic data might be daunting, even for quite small networks. Indeed, for specific comparisons and for networks that are larger than ADAPT, NetEx may have an advantage as it allows for plots that need to be compared to be selected and shown together in the same view. More importantly, during the search step NetEx sets an analyst up for the detect and isolate steps (see summary in Table 1). Our results suggest that these steps can then be performed faster and more accurately using NetEx compared to Intelliviz.

6. CONCLUSION AND FUTURE WORK

The multi-focus and multi-level analysis techniques discussed in this paper enable an analyst to identify and compare multiple nodes of interest. While many visualizations focus on analysis, NetEx focuses on showing a visually recognizable views of the node-link structured, coupled with multi-focus display of thematic data for important nodes. This paper demonstrates how comparing multiple views in a visual interactive analysis tool can reduce errors in comparing data, for the purpose of fault diagnosis, in different parts of a network. The multi-focus and multi-level techniques succeed by supporting simultaneous analysis of multiple parts of a dataset while giving the analyst different views of and easy access to the network's graph structure. Using data from an electrical power network, we showed in a user study how NetEx may improve the analytical task of diagnosis. Our results suggest that, compared to a baseline tool, NetEx enables more accurate isolation of faulty nodes in complex situations involving multiple faults.

In future work, we would like to borrow from Intelliviz by showing, in NetEx, all the plots in a larger side panel at the beginning of the analysis. The analyst could then remove all plots showing a nominal behavior as in the current implementation, ending up with a relatively small number of candidate plots (typically less than 10 for tasks like the ones in this study), and leaving space for the overview and detail window, as well as the bubble connections. NetEx could also be strengthened by drawing parts of large networks together with multiple fisheye foci. We also expect to further evaluate these visual interfaces and interaction techniques with additional graph exploration tasks.

The size and complexity of network-related data continue to expand. The power and value of this work is that it is part of the quest for scalable interactive computing systems that keep an analyst aware of the whole while exploring the details. We call for creating tools that aid people's limited memory while simultaneously exploring and comparing multiple nodes or areas of interest in networks. We encourage other researchers to work, as we are, on improving interactive techniques for exploring large and complex problems.

REFERENCES

[1] Pearl, J., [*Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*], Morgan Kaufmann, San Mateo, CA, USA (1988).

[2] Abello, J. and van Ham, F., "Matrix Zoom: A Visual Interface to Semi-External Graphs," in [*Proc. of InfoVis'04*], 183–190 (Oct 2004).

[3] Abello, J., van Ham, F., and Krishnan, N., "ASK-GraphView: A Large Scale Graph Visualization System," *IEEE Transactions on Visualization and Computer Graphics* **12**(5), 669–676 (2006).

[4] Adar, E., "GUESS: A Language and Interface for Graph Exploration," in [*Proc. of CHI'06*], 791–800 (Apr 2006).

[‡‡]Using a one-population t-test, the level of significance was $\alpha = 0.005$ or better for all NetEx features.

[5] Auber, D., "Tulip : A Huge Graph Visualisation Framework," in [*Graph Drawing Softwares*], Mutzel, P. and Jünger, M., eds., *Mathematics and Visualization*, 105–126, Springer-Verlag (2003).

[6] Perer, A. and Shneiderman, B., "Integrating Statistics and Visualization for Exploratory Power: From Long-Term Case Studies to Design Guidelines," *IEEE Computer Graphics and Applications* **29**(3), 39–51 (2009).

[7] Wong, P. C., Schneider, K., Mackey, P., Foote, H., Jr., G. C., Guttromson, R., and Thomas, J., "A Novel Visualization Technique for Electric Power Grid Analytics," *IEEE Transactions on Visualization and Computer Graphics* **15**(3), 410–423 (2009).

[8] Henry, N., Fekete, J.-D., and Mcguffin, M. J., "Nodetrix: a Hybrid Visualization of Social Networks," *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1302–1309 (2007).

[9] Holten, D., "Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data," *IEEE Transactions on Visualization and Computer Graphics* **12**(5), 741–748 (2006).

[10] Lamping, J., Rao, R., and Pirolli, P., "A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies," in [*Proc. of CHI'95*], 401–408 (May 1995).

[11] Overbye, T. J. and Weber, J. D., "New Methods for the Visualization of Electric Power System Information," in [*Proc. of InfoVis'00*], 131–136 (Oct 2000).

[12] Phan, D., Xiao, L., Yeh, R., Hanrahan, P., and Winograd, T., "Flow Map Layout," in [*Proc. of InfoVis'05*], 219–224 (Oct 2005).

[13] Wong, N., "EdgeLens: An Interactive Method for Managing Edge Congestion in Graphs," in [*Proc. of InfoVis'03*], 51–58 (Oct 2003).

[14] Faloutsos, C., McCurley, K. S., and Tomkins, A., "Fast Discovery of Connection Subgraphs," in [*Proc. of KDD'04*], *KDD '04*, 118–127 (Aug 2004).

[15] Gansner, E., Koren, Y., and North, S., "Topological Fisheye Views for Visualizing Large Graphs," in [*Proc. of InfoVis'04*], 175–182 (Oct 2004).

[16] McGuffin, M. J. and Jurisica, I., "Interaction Techniques for Selecting and Manipulating Subgraphs in Network Visualizations," *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 937–944 (2009).

[17] Pietriga, E., Bau, O., and Appert, C., "Representation-Independent In-Place Magnification with Sigma Lenses," *IEEE Transactions on Visualization and Computer Graphics* **16**(3), 455–467 (2010).

[18] Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., and Roseman, M., "Navigating Hierarchically Clustered Networks Through Fisheye and Full-Zoom Methods," *ACM Transactions on Computer-Human Interaction* **3**(2), 162–188 (1998).

[19] van Ham, F. and van Wijk, J. J., "Interactive Visualization of Small World Graphs," in [*Proc. of InfoVis'04*], 199–206 (Oct 2004).

[20] van Ham, F. and Perer, A., ""Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest," *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 953–960 (2009).

[21] Wong, P. C., Mackey, P., Cook, K. A., Rohrer, R. M., Foote, H., and Whiting, M. A., "A Multi-Level Middle-Out Cross-Zooming Approach for Large Graph Analytics," in [*Proc. of VAST'09*], 147–154 (Oct 2009).

[22] Bezerianos, A., Chevalier, F., Dragicevic, P., Elmqvist, N., and Fekete, J.-D., "GraphDice: A System for Exploring Multivariate Social Networks," in [*Proc. of EuroVis'10*], 863–872 (Jun 2010).

[23] Lee, B., Smith, G., Robertson, G. G., Czerwinski, M., and Tan, D. S., "FacetLens: Exposing Trends and Relationships to Support Sensemaking within Faceted Datasets," in [*Proc. of CHI'09*], 1293–1302 (Apr 2009).

[24] Lex, A., Streit, M., Partl, C., Kashofer, K., and Schmalstieg, D., "Comparative Analysis of Multidimensional, Quantitative Data," *IEEE Transactions on Visualization and Computer Graphics* **16**(6), 1027–1035 (2010).

[25] Viau, C., McGuffin, M. J., Chiricota, Y., and Jurisica, I., "The FlowVizMenu and Parallel Scatterplot Matrix: Hybrid Multidimensional Visualizations for Network Exploration," *IEEE Transactions on Visualization and Computer Graphics* **16**(6), 1100–1108 (2010).

[26] Wattenberg, M., "Visual Exploration of Multivariate Graphs," in [*Proc. of CHI'06*], 811–819 (Apr 2006).

[27] Weaver, C., "Multidimensional Data Dissection Using Attribute Relationship Graphs," in [*Proc. of VAST'10*], 75–82 (Oct 2010).

[28] Butkiewicz, T., Dou, W., Wartell, Z., Ribarsky, W., and Chang, R., "Multi-Focused Geospatial Analysis Using Probes," *IEEE Transactions on Visualization and Computer Graphics* **14**(6), 1165–1172 (2008).

[29] Elmqvist, N., Henry, N., Riche, Y., and Fekete, J.-D., "Mélange: Space Folding for Visual Exploration," *IEEE Transactions on Visualization and Computer Graphics* **16**(3), 468–483 (2010).

[30] Shoemaker, G. and Gutwin, C., "Supporting Multipoint Interaction in Visual Workspaces," in [*Proc. of CHI'07*], 999–1008 (Apr 2007).

[31] Baldonado, M. Q. W., Woodruff, A., and Kuchinsky, A., "Guidelines for Using Multiple Views in Information Visualization," in [*Proc. of AVI'00*], 110–119 (May 2000).

[32] Collins, C. and Carpendale, S., "VisLink: Revealing Relationships Amongst Visualizations," *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1192–1199 (2007).

[33] Heer, J., "Exploring enron." http://hci.stanford.edu/jheer/projects/enron (2005).

[34] Stasko, J., Gorg, C., and Liu, Z., "Jigsaw: Supporting Investigative Analysis through Interactive Visualization," *Information Visualization* **7**(2), 118–132 (2008).

[35] Weaver, C., "Building Highly-Coordinated Visualizations in Improvise," in [*Proc. of InfoVis'04*], 159–166 (Oct 2004).

[36] von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J. J., Fekete, J.-D., and Fellner, D. W., "Visual Analysis of Large Graphs," in [*Proc. of EG'10 - STARs*], 37–60 (May 2010).

[37] Shetty, J. and Adibi, J., "The Enron Email Dataset Database Schema and Brief Statistical Report," tech. rep., University of Southern California (2004).

[38] Cockburn, A., Karlson, A., and Bederson, B. B., "A Review of Overview+Detail, Zooming, and Focus+Context Interfaces," *ACM Computing Surveys* **14**(1), 1–31 (2008).

[39] Herman, I., Melancon, G., and Marshall, M. S., "Graph Visualization and Navigation in Information Visualization: A Survey," *IEEE Transactions on Visualization and Computer Graphics* **6**(1), 24–43 (2000).

[40] Selker, T., "Style and Function of Graphic Tools," in [*Proc. of GI'99*], 123–131 (Jun 1999).

[41] Cossalter, M., *Multi-focus Multi-level Network Visualization*, Master's thesis, Carnegie Mellon University (Aug 2011).

[42] Cossalter, M., Mengshoel, O., and Selker, T., "Visualizing and Understanding Large-Scale Bayesian Networks," in [*Proc. of the AAAI'11 Workshop on Scalable Integration of Analytics and Visualization*], 12–20 (Aug 2011).

[43] Poll, S., Patterson-Hine, A., Camisa, J., Garcia, D., Hall, D., Lee, C., Mengshoel, O. J., Neukom, C., Nishikawa, D., Ossenfort, J., Sweet, A., Yentus, S., Roychoudhury, I., Daigle, M., Biswas, G., and Koutsoukos, X., "Advanced Diagnostics and Prognostics Testbed," in [*Proc. of DX'07*], 178–185 (May 2007).

[44] Russell, D. M., Stefik, M. J., Pirolli, P., and Card, S. K., "The cost structure of sensemaking," in [*Proc. of INTER-CHI'93, ACM Conference on Human factors in Computing Systems*], 269–276 (1993).

[45] Heer, J. and Boyd, D., "Vizster: Visualizing Online Social Networks," in [*Proc. of InfoVis'05*], 32–39 (Oct 2005).

[46] Dunne, C., Shneiderman, B., Gove, R., Klavans, J., and Dorr, B., "Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization," *JASIST: Journal of the American Society for Information Science and Technology* (2012).

[47] Gove, R., Dunne, C., Shneiderman, B., Klavans, J., and Dorr, B., "Evaluating visual and statistical exploration of scientific literature networks," in [*VL/HCC '11: Proc. 2011 IEEE Symposium on Visual Languages and Human-Centric Computing*], 217–224 (2011).

[48] Steinberger, M., Waldner, M., Streit, M., Lex, A., and Schmalstieg, D., "Context-preserving visual links," *IEEE Transactions on Visualization and Computer Graphics* **17**, 2249–2258 (2011).

[49] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T., "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Research* **13**(11), 2498–2504 (2003).

[50] Norman, D., [*Things that make us smart: Defending human attributes in the age of the machine*], Perseus Books (1993).

[51] Shetty, J. and Adibi, J., "The enron email dataset database schema and brief statistical report," tech. rep., University of Southern California (2004).

[52] Pathak, N. and Srivastava, J., "Automatic extraction of concealed relations from email logs," in [*International Workshop and Conference on Network Science*], (May 2006).