Line Graph Explorer: Scalable Display of Line Graphs Using Focus+Context

Robert Kincaid
Agilent Laboratories
5301 Stevens Creek Blvd., MS 54U-SC
Santa Clara, CA 95051
robert_kincaid@agilent.com

Heidi Lam
University of British Columbia
201-2366 Main Mall
Vancouver BC V6T 1Z4
hllam@cs.ubc.ca

ABSTRACT
Scientific measurements are often depicted as line graphs. State-of-the-art high throughput systems in life sciences, telemetry and electronics measurement rapidly generate hundreds to thousands of such graphs. Despite the increasing volume and ubiquity of such data, few software systems provide efficient interactive management, navigation and exploratory analysis of large line graph collections. To address these issues, we have developed Line Graph Explorer (LGE). LGE is a novel and visually scalable line graph management system that supports facile navigation and interactive visual analysis. LGE provides a compact overview of the entire collection by encoding the y-dimension of individual line graphs with color instead of space, thus enabling the analyst to see major common features and alignments of the data. Using Focus+Context techniques, LGE provides interactions for viewing selected compressed graphs in detail as standard line graphs without losing a sense of the general pattern and major features of the collection. To further enhance visualization and pattern discovery, LGE provides sorting and clustering of line graphs based on similarity of selected graph features. Sequential sorting by associated line graph metadata is also supported. We illustrate the features and use of LGE with examples from meteorology and biology.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces-Graphical User Interfaces; I.3.5. Computer Graphics: Methodology and Techniques – Interaction Techniques; J.3[Life and Medical Sciences]:Biology and genetics

General Terms
Measurement, Design, Human Factors

Keywords
Focus+Context, Line Graph

1. INTRODUCTION
Line graphs may date back over 4000 years to ancient Egypt [7] while a more "modern" instance is attributed to Robert Plot [19] in 1684. Since that time, our ability to collect and store such data has grown at an incredible rate. While we have access to more information than ever before, we also have the additional burden to make sense of these massive data collections.

This paper introduces an interactive visualization for large collections of line graphs. Our motivating application and inspiration is in the domain of molecular biology, more specifically in protein and DNA analysis using gel electrophoresis. This process uses an electric field to separate the components of a sample migrating through a porous gel medium. The observed separation is generally proportional to the size or molecular weights of the components.

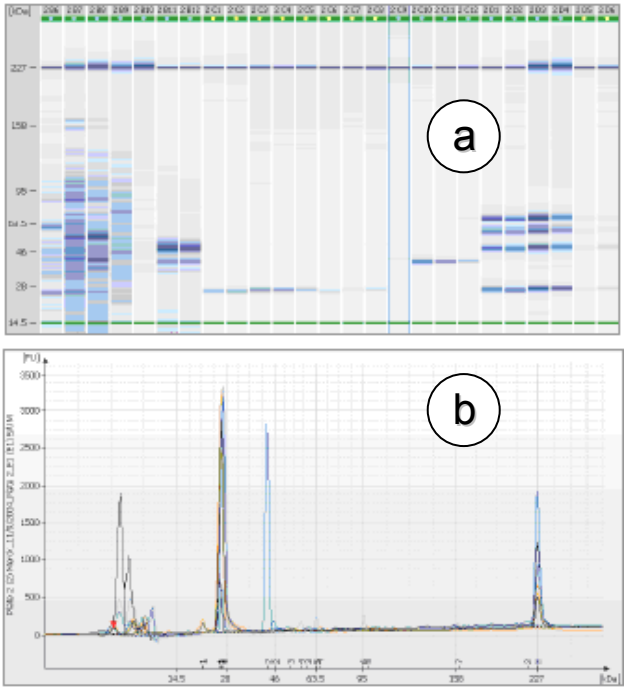

Figure 1. Typical outputs from an electrophoresis instrument. (a) A traditional vertically oriented gel image where each column represents a sample and the vertical dimension depicts the separation criterion used in the experiment (in this case, the molecular weight). The intensity of the signal is shown by the intensity of the color of the horizontal bands. (b) Overlaid plots where each line graph represents the detailed electropherogram of a single sample. Here the vertical dimension is the intensity of the signal, while the horizontal dimension is the separation criterion (molecular weight).

AVI '06, May 23-26, 2006, Venezia, Italy.

The data produced by these experiments can be viewed as 2D line graphs (Figure 1(b)), with the x-dimension being the separation criterion (for example molecular weight), and the y-dimension being signal intensity. Alternatively, the same data can be visualized in a manner similar to traditional 1D gel images (Figure 1(a)) where samples can be stacked side-by-side as columns, thereby allowing better sample comparison. For each sample, the separated sample components are shown as horizontal bands, with the intensity of the band representing the signal intensity. The y-dimension represents the separation criterion. Modern high throughput instruments can process thousands of samples per day, resulting in large collections of line graphs.

Frequently collections of line graphs are either overlaid as in Figure 1(b), or stacked. Overlaying many graphs has clear issues with occlusion. On typical computer displays, stacking graphs only permits visualizing a small number of graphs at one time. For large collections, some method of scrolling the collection or selecting relevant subsets is typically necessary, thus losing overall context of the data set. While not generally applied, the 1D view shown in Figure 1(a) does allow more graphs to be stacked, but precise visualization of graph detail is lost in the color encoding, and only qualitative comparisons can be made.

To address these issues we developed Line Graph Explorer (LGE), a visualization system that provides an overview similar to the gel-like view in Figure 1(a), along with the details available in the line graph view in Figure 1(b). Our goal is to design a visually scalable display that effectively enables data exploration and visual analysis of these large collections, thereby allowing scientists to explore and discover underlying trends and patterns in such data. In addition, to enhance visual pattern discovery, LGE allows sorting and clustering of line graph features. Sequential sorting by associated line graph metadata is also supported.

2. PROBLEM DESCRIPTION

Before introducing the features of LGE, we will start by describing the data and tasks it supports.

2.1 Data

While Line Graph Explorer is generically applicable to a wide range of data, we are initially motivated by data derived from high throughput life science instruments. The data can typically be reduced to a collection of related line graphs, each representing a biological sample. Each graph consists of linearly ordered data. For example, the process of electrophoresis usually produces a component migration time series called electropherograms (Figure 1, Figure 5, Section 5.2). In genetic analysis, microarray data [14] can be ordered by genomic location (Figures 6 and 7, Section 5.3). Before presenting several examples, it is useful to note the following common characteristics of such data:

- The individual line graphs are based on linearly ordered data, in which case the ordered axis (usually the x-axis) remains fixed and is not generally reorderable.
- The collection of graphs *may* consist of independent samples stacked vertically, whose ordering *can* be changed to juxtapose and compare graphs.
- Graphs consist mostly of a baseline with fluctuations from the base line as *peaks* or *valleys* (for example, spectra).
- For biological or chemical samples these peaks and valleys *are* the regions of interest, and we refer to them as *features*.
- The data consists of (or can be reduced to) a modest enough number of features that visual analysis is feasible.
- The line graphs are appropriately aligned *prior* to visualization to permit valid comparison of feature alignment.
- The analysis (whether computational or visual) is to compare the size, location and alignment of these features.
- We expect that some line graphs will show similar features and others will not. This similarity or dissimilarity is what we seek in the analysis, and represents a corresponding biological or chemical similarity or dissimilarity.
- The number of line graphs in a given study (i.e. the subject of analysis) can be quite large, with hundreds to thousands of samples.

Thus, we need a visually scalable and efficient means to inspect, compare and manage large collections of related line graphs. We chose a Focus+Context approach to provide both a scalable overview and detailed line graphs within a single integrated view. The interactions and visualization are optimized assuming the data properties listed above. For graphs with little or no correlated behavior, we would expect the visualization to be less useful, although it would still allow some degree of useful graph management and inspection. While our motivation is high throughput instrument data, any linearly ordered data sharing most of the characteristics listed above should be amenable to the visualization we propose, particularly correlated time series data.

2.2 Tasks

The high-level task of LGE is comparison between large collections of line graphs. We describe the lower-level tasks for which it was designed in terms of the task taxonomy of Amar et al. [2]:

- Extrema, Range and Distribution: These tasks are used to characterize the overall properties of a graph and allow quick identification or comparison to other graphs in the collection.
- Correlation and Cluster: It is useful to organize the graph collection to make similar graphs adjacent. Not only is this useful in it's own right, but it increases the likelihood of finding and assessing the extent of visual correlations.
- Anomalies: The analyst is often interested in finding missing, unexpected and/or shifted peaks in the line graph collection. This search can be supported directly by visual analysis. In addition, sorting or clustering can bring similar anomalies together or help identify an anomaly existing within an otherwise similar group of graphs.

As we will describe later, LGE provides a number of interactions and computational methods intended to support these tasks.

3. RELATED WORK

The solution we propose was actually hinted at in Bertin's synoptic of graphical constructions [5]. He refers to such collections of line graphs as an "array of curves." In his graphical representation of the synoptic, he even indicates these curves are reorderable. However, the application of this construction is never fully developed. Further, Bertin does not indicate depicting these curves as anything other than 2D line graphs.

To fully develop this concept we considered several other research topics that are related to LGE.

Color Encodings

Using color to encode numerical data is a widely used technique discussed in detail by Bertin [5]. Such encodings are typically implemented in the form of heatmap displays [6]. Bioinformaticians frequently use a red/green [8] and sometimes yellow/blue color gradient to represent gene expression ratios. Spotfire [26] provides a generic mechanism for coloring points in 2D and 3D graphs by mapping numeric attributes to a color gradient. Recently, Saito et al. [23] described a compact color encoding for one-dimensional data that attempts to preserve some aspects of the underlying 2D line graph. Their scheme affords visual scalability, but does not directly support visualizing full graph details.

Focus+Context

Many information visualization systems employ Focus+Context techniques, where some kind of visual compression is used to view the entire data set within the application window [6]. Additionally, a distortion or lens effect is employed to facilitate viewing data of interest in detail while maintaining the context of the entire dataset. Table Lens [21] is a well known example that applies Focus+Context to tabular spreadsheet-like data. LGE can be considered an extension of Table Lens to include line graph data in addition to simple tabular data.

Reorderable Matrix

The *reorderable matrix* was first introduced by Bertin [5]. The rows and columns of a table of graphical representations are permuted to reveal correlated features. The best known implementation of this concept is again Table Lens. Siirtola [25] has performed empirical user studies of the reorderable matrix with positive results. More recently VistaClara [13] used a fully permutable matrix to analyze microarray data.

Linearly Ordered Data

One of the most widely investigated and relevant cases of *linearly ordered data* is time series data. Müller and Schumann [17] provide a review of techniques for time-dependent data. TimeSearcher [12] displays overlaid line graph data and allows filtering based on graph features. QueryLines [22] provides a mechanism for approximate queries of time series based on visually specified criteria. BinX [4] displays a single line graph that focuses scalability on the x-dimension. VistaChrom [14] analyzes microarray data as a form of linear data, addressing scalability primarily through data aggregation as opposed to directly visualizing the complete data set. Hao et al. [11] developed an importance-driven, space-filling layout for time series data, but does not allow for easily aligned comparisons.

2D Line Graphs

This basic plotting style is ubiquitous, with more examples than can be enumerated here. Spotfire [26] may be the best known system associated specifically with exploratory information visualization that also supports a variety of line graph displays. These systems typically employ traditional views of overlaid or stacked line graphs. We believe none have adequately addressed the visual scalability issues of large collections of such plots.

4. DESIGN

LGE can be viewed as an extension of Table Lens [18], where one of the columns consists of a collection of line graphs as seen in the right panel of Figure 2(b). A table of associated metadata is also provided and allows ad hoc sorting of the rows in the collection. This table is shown in the left panel of Figure 2(a). By reordering the table, patterns may emerge either from the line graph views or from the metadata table itself, or both.

4.1 Viewing Global Context

A collection of line graphs actually consists of *three* dimensions: the ordered independent x-axis, the dependent y-axis, and the additional dimension consisting of the reorderable list of graphs. Such collections can be rendered in a 3D view, but there are serious occlusion and perspective distortion problems with this approach. To provide a compact overview we reduce the rendered dimensionality to 2D by initially displaying them as thin 1D ribbons with the y-dimension of the data encoded by color saturation and luminance instead of vertical position. Higher y-values are represented by more saturated and brighter colors. LGE offers three different forms of color mapping:

Linear, where the normalized y-dimension is used directly for saturation and brightness;

Sigmoidal, with the normalized y-dimension value x mapped to saturation and brightness level i, with a user-defined constant s using the function:

$$i = \frac{2}{1 + e^{-sx}} - 1$$

Logarithmic, with the normalized y-dimension value x mapped to saturation and brightness level i, using the function:

$$i = \begin{cases} \log|x| + 1 & ; x > 0 \\ -\log|x| + 1 & ; x < 0 \\ 0 & ; x = 0 \end{cases}$$

Using this compressed encoding of the line graph data, we avoid overlaying graphs, which is unsatisfactory as occlusion becomes a serious problem with large numbers of graphs. We also avoid 2D stacking of the graphs, which suffers from limited visual scalability, and we avoid 3D stacking, which carries the same occlusion and navigation issues as most 3D visualizations. Further, since the thickness of a given strip can be quite narrow and juxtaposed with other graphs in a dense display, the eye can perceive more of the overall correlated patterns in the data. For example, the graph features align in closer proximity in our overview than they would with stacked 2D plots

4.2 Viewing Detail

By encoding the y-dimension of the line graph using color instead of vertical height, we intentionally sacrifice the level of perceivable detail in the graph in order to improve visual scalability and to allow comparison of line graphs across the entire collection. To compensate for this perceptual loss, LGE uses a simple form of semantic zooming to allow viewing the full line graph on demand. There are two forms of display for the full 2D line graphs. The user can either select individual graphs, as shown in Figure 2(c), or create a collection of graphs in the form of a *lens*, as shown in Figure 2(d). The lens is designed to display a series of neighboring graphs, with the focal graph in full magnification, and the upper and lower graphs in increasingly lower magnifications of the y-scale. The style is very similar to the parallel projection chosen by Baudisch et al. [3] for text magnification. The lens can be moved to show a different set of adjacent line graphs, and it can be resized to contain more or less graphs. Since the line graphs are initially in the dense overview state, they are closely packed together, which sometimes makes

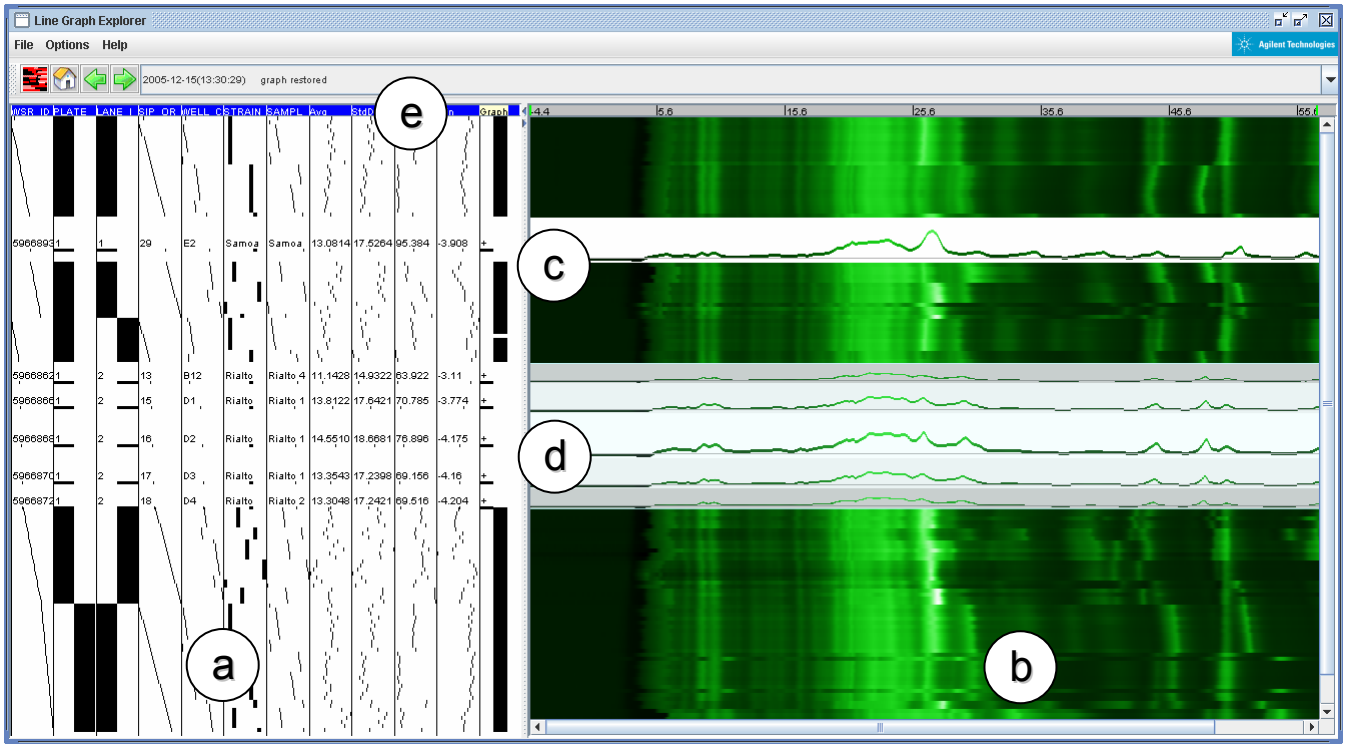

Figure 2. The Line Graph Explorer interface: LGE has two main panels: (a) the metadata panel and (b) the line graph panel. The metadata panel contains attributes of the line graph data displayed. In this example, it shows the sample name, and other specifics of the electrophoresis examples including plate and well information. The line graphs show the electropherogram data. The line graphs can be viewed either individually, as in (c), or under a lens, as in (d). LGE offers sorting by metadata and sorting and clustering based on line graph features. Sorted results are stored as history (e) and can be retrieved.

selection of individual graphs more difficult. The lens facilitates graph selection by enlarging the line graphs. This enlargement not only aids visual inspection, but also provides a larger screen target for selection operations. Additionally, the lens provides a convenient way to scan the data when the lens is moved across the display.

Since the lens displays different progressive magnifications of the y-, but not the x-scale, lens graphs are displayed at slightly different aspect ratios. This difference in aspect ratios might be confusing to users since scales of adjacent graphs within the lens are not directly comparable. LGE attempts to minimize possible confusion in two ways. First, LGE uses different background colors for the different magnification levels of the plots, with darker colors for lower magnifications. This scheme attempts to create an illusion of depth where plots of lower magnifications will be perceived to be on a lower visual layer and farther away from the user. Second, LGE doubly encodes the y-dimension of the line graph with space and color as a visual reminder of the y-scale, and to visually link the two visual representations of the line graph data. The reader should note that the intent of lens graphs surrounding the focus graph is to aid navigation more than to provide analytical comparison. For analysis purposes, the lens can be enlarged or the graphs opened directly to provide more directly comparable aspect ratios.

4.3 Basic Interactions

Users can view the 2D plots of individual line graphs by a single left-mouse click on the overview (Figure 2(b)). The result is shown in Figure 2(c). Clicking on an open 2D plot will return it to the closed overview state. Dragging the left-mouse button opens a range of graphs generating a series of adjacent 2D line graphs. Dragging upwards will close previously opened line graphs. Similarly, a moveable and resizable lens can be created by an alt-left-mouse click, as shown in Figure 2(d). In contrast to a simple 2D graph, users can *move* the lens up and down through the collection with an alt-left-mouse drag or by using the arrow keys. One resizes the lens by placing the mouse on the upper or lower edges of the lens and performing an alt-left-mouse drag operation. All of these actions are synchronized with the metadata panel to maintain horizontal data alignment.

4.4 Metadata Panel

Metadata can be associated with each line graph and displayed in tabular form in the left panel, as shown in Figure 2(a). Similar to the handling of line graphs, such data can also be displayed in either the overview or the detail state. For the overview state, LGE displays graphical bars representing either categories or numerical data. When a line graph is opened to the detailed 2D state, the available cell height is sufficient to show the values of metadata as text strings in addition to the graphical representation. The position and state of the metadata for each sample are linked to those in the line graph panel allowing the user to inspect details of the metadata at the same time as inspecting details of the line graphs. This behavior can be seen in Figure 2.

Several statistical values are computed for the line graphs and provided as default metadata. These values include the mean, standard deviation, minimum and maximum of each line graph, and can be used as attributes for reordering graphs. LGE also provides a column of metadata called *Graph State* which is + for open graphs and – for closed graphs, and adjusts dynamically as graphs are opened and closed. This attribute is contained in the far

right column of metadata in Figure 2. The Graph State attribute permits the user to rapidly collate all open graphs together for quick comparison.

Similar to the implementation of Table Lens, LGE allows sequential sorting of metadata for ad hoc data exploration. Once sorted, both the metadata and the line graph panels will display the samples in their new row order.

4.5 Computationally Assisted Interactions

LGE also offers sorting and clustering based on individual line graph features. LGE currently supports Euclidean distance and Pearson correlation as distance measures. Users can select a particular line graph, or optionally select a part of the graph, and use either the selected portion or the entire graph as a basis for sorting the remaining graphs by similarity. This operation results in the selected graph as the topmost graph in the panel with the remaining graphs ordered by decreasing similarity to the selected graph.

Users can similarly cluster all line graphs based on the entire or a selected portion in the x-dimension. We currently use standard agglomerative clustering based on simple distance measures, including Euclidian distance and Pearson correlation. To simplify the display and since we are primarily interested in grouping similar graphs together for comparison, we currently do not display the tree structure of the cluster hierarchy. Also, our current implementation assumes aligned graph data where corresponding measurements exist at each time point across all graphs.

4.6 Sort History

Any operation which reorders the data is stored in a history queue and is accessible on the toolbar shown in Figure 2(e). Browse buttons allow sequential navigation through the history and a combo box allows a direct selection. The basic model is similar to a web browser history mechanism.

5. RESULTS

In this section, we illustrate the features of LGE using three sets of real measurement data: time series climate data, electropherograms and microarray data. For brevity, we largely omit details of the various experimental platforms and the specific data types. The cited references can help clarify some of these points.

5.1 Sorting Climate Data

As a simple initial example, we consider a relatively large collection of time series climate data. Figure 3 shows a sorted display of daily mean temperature from 324 international cities for the past 10 years, adapted from the data provided by the University of Dayton [27]. There are over one million data points rendered. Even at a very high-level overview, it is easy to see the periodicity of the data with warm summer months in red alternating with cold winter months in green. In this example we have first sorted by standard deviation of the location's temperature and then by hemisphere (North/South). In Figure 3 we have also opened a lens so we can probe for a pleasant climate. We use the lens to scan quickly for a location that exhibits a low standard deviation (consistent temperature) and a mild mean temperature. Since in this example, we consider a low variance a desirable trait, we can quickly determine either visually (for example, the line graph for Honolulu is relatively featureless) or from the pre-computed metadata. We quickly find Honolulu as a candidate destination.

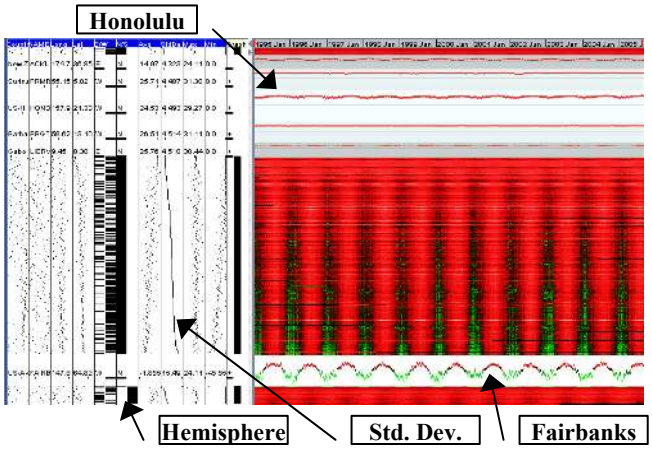

Figure 3. Daily mean temperature from 1995-2005 for 324 international cities. Over one million data points are shown in this overview, encoded with a bi-color scheme where red represents positive (warm) values and green represents negative (cold) values. The annual periodicity of the data is clearly shown in this overview. Graphs are sorted by standard deviation and then hemisphere. A lens is opened around Honolulu Hawaii. The open graph in the lower portion is Fairbanks Alaska for comparison.

It is interesting to consider Fairbanks Alaska, also shown in Figure 3. The increase in seasonal temperature fluctuations is readily apparent. Alaska clearly has a more extreme climate when compared to Hawaii.

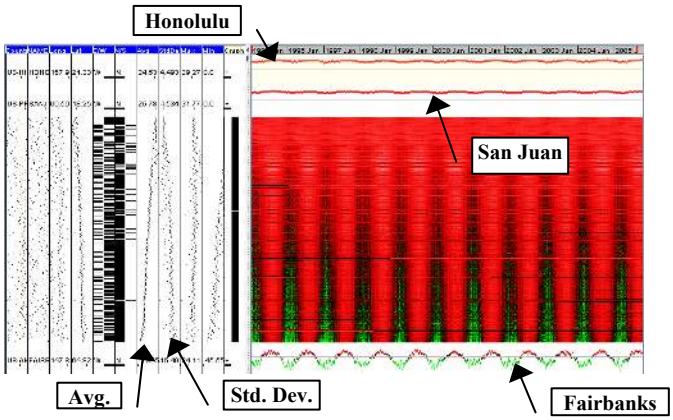

Figure 4. Daily mean temperature data sorted by similarity to Honolulu Hawaii. San Juan Puerto Rico has the most similar temperature profile to Honolulu. Fairbanks occurs at the end of the sorted list as it is the most extreme opposite to Honolulu in the data set.

Once we find a location of interest we might want to consider alternatives that have similar climates. We can sort the data set by similarity to Honolulu using Euclidean distance as the similarity measure. The result of sorting is shown in Figure 4.

We can open up the graph for Honolulu and compare it to its nearest similar neighbor, which is San Juan Puerto Rico. Based on climate, San Juan might make a good alternative to Hawaii. It is also interesting to note that based on the similarity measure, Fairbanks Alaska is the extreme opposite of Honolulu.

Sorting by similarity makes a more complex comparison of the graphs versus just by comparing a single attribute. Careful viewing of the metadata shows the similarity sorted ordering is strongly correlated with average temperature, but not precisely the same order obtained by simply sorting by this attribute. The annual variance in the data also contributes to the similarity measure in this case, and is evident in the table by observing the weaker but obvious correlation with standard deviation.

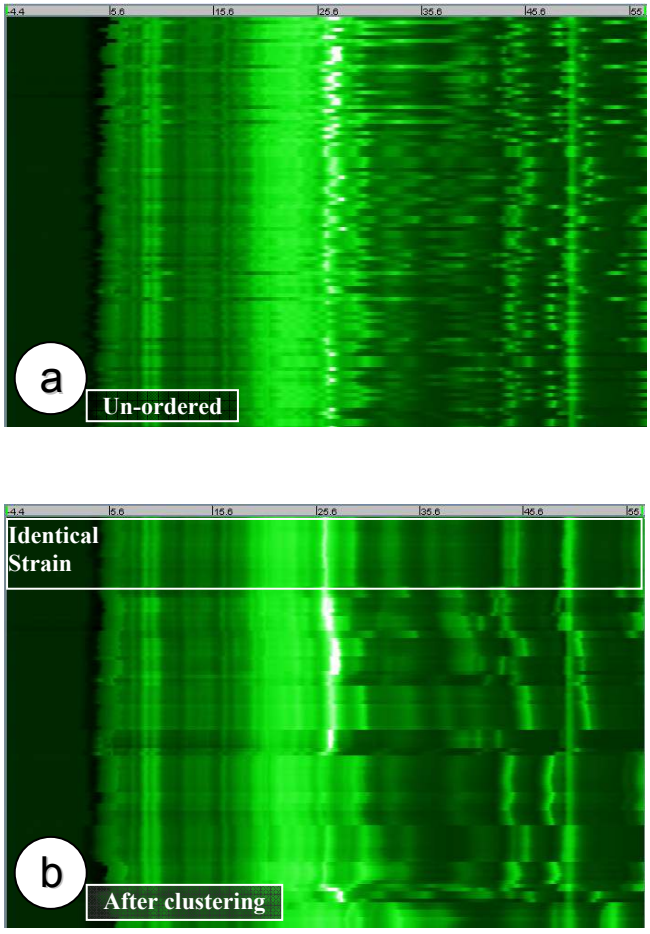

Figure 5. Clustering wheat strains: (a) The original un-ordered list of electropherograms. (b) The result of clustering. It is clearly visible that similar electropherogram profiles are now grouped together to enable quick comparison between similar graphs, as well as across related groups of graphs. One group of identical strain is outlined for illustration, but other similar groupings are also visible.

5.2 Clustering Wheat Strains

The dataset for this example is produced by a high throughput electrophoresis instrument [1] and consists of the electropherograms of 112 wheat strain samples. For each electropherogram, the x-axis is migration time, which is proportional to molecular weight. Here we are essentially looking at the molecular weight distribution of proteins in the sample, which can provide a kind of line graph signature for each wheat strain.

The unordered collection is shown in Figure 5(a). At this preliminary stage we can see that the overview is effective at summarizing the entire data and shows clearly correlated peaks between some samples, and also obvious differences. There are also hints in the patterns that there might be some systematic differences due to underlying biology.

We can easily group similar graphs with the built-in facility to hierarchically cluster using Euclidean distance between graphs. The effect of this operation is shown in Figure 5(b). It is immediately obvious that samples with similar electropherogram profiles are grouped near each other indicated by large regions with essentially identical aligned profiles. In most cases these matches correspond to identical strains derived from different flour samples. In addition related groups of different strains are clustered together.

5.3 Exploring Graph Features in Breast Cancer Data

Cancer cells often have increased or decreased copies of genomic DNA. Recently, an area of considerable interest is how DNA copy number changes correlate with gene expression (the amount of mRNA being transcribed from DNA in protein synthesis). Both DNA copy number and gene expression can be measured using microarrays, often in the form of ratios between a test sample and a normal non-diseased reference sample.

In this scenario, we use LGE to jointly analyze both DNA copy number and gene expression data. The data is from 41 breast tumor cell lines as measured by Pollack et al. [20]. To keep our example simple, we consider only chromosome 17. Gene expression data is mapped to the genomic location of each gene so that it is comparable to the DNA copy number data. The dataset is smoothed as a 10-point moving average of the original log ratio data. For the purposes of this example, we simply consider the line graphs as abstract representations of increased/decreased DNA copy number, or gene over/under expression.

We encode the data with a typical two-color scheme where red represents high ratios (log ratio > 0) and green represents low ratios (log ratio < 0). This view is shown in Figure 6 where we have segregated the DNA copy number and gene expression data for comparison. A quick glance at the overview reveals that while the measurements fluctuate in both DNA copy number and gene expression data, the fluctuations seem to be more extreme and frequent in the gene expression data. This observation is not surprising since gene expression is more dynamic than the static property of DNA copy number.

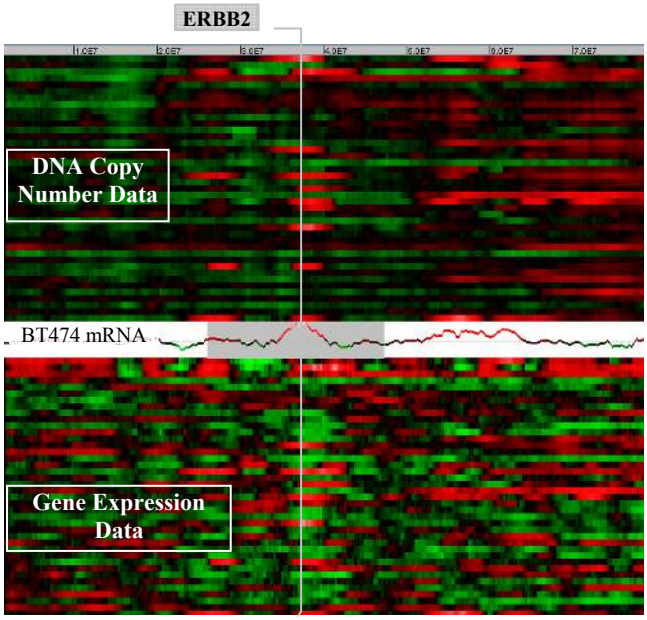

Figure 6. DNA copy number and gene expression data, sorted by measurement class. BT474 mRNA is opened to select a region that encompasses the gene ERBB2. Red indicates increases in DNA copy number or gene expression. Green indicates decreases. The gray region in the open graph indicates the selected region.

We also note that there are a number of similar intense red peaks that occur near the middle of chromosome 17. These peaks correspond to significantly increased copy number or a similar increase in gene expression. One of the strongest is highlighted as a gray rectangle in the gene expression profile of cell line BT474 (Figure 6). We note that the top of this peak is in the vicinity of the gene ERBB2, also known as HER2. This gene is strongly associated with some types of breast cancer. The drug Herceptin (trastuzumab) specifically targets cells with increased expression of this gene [16]. Hence these peaks represent important and interesting features both graphically *and* biologically and worthy of further exploration. With LGE, we can select just the relevant region (indicted by the gray rectangle in Figure 6), and can sort by similarity to *only this region*, ignoring other graph features. The result is shown in Figure 7.

We immediately see that we have isolated a series of profiles that have similar peaks in the vicinity of ERBB2 indicating increased DNA copy number or gene over-expression in this important region. What is more striking is that frequently the DNA copy number profile and gene expression profile for the same cell line occur adjacent to each other, indicating a strong degree of similarity between the increases in gene expression and DNA copy number. This observation suggests that perhaps the over-expression of genes found in cancer cell lines may often be simply due to increases in DNA copy number, instead of more complicated mechanisms involving gene expression regulation.

Beyond the specifics of these three examples, an important observation is that relatively large collections of related line graphs can be interactively explored with LGE to provide insights about the correlated behavior of the data. Further, this exploration can be quickly accomplished with a small number of simple user interactions.

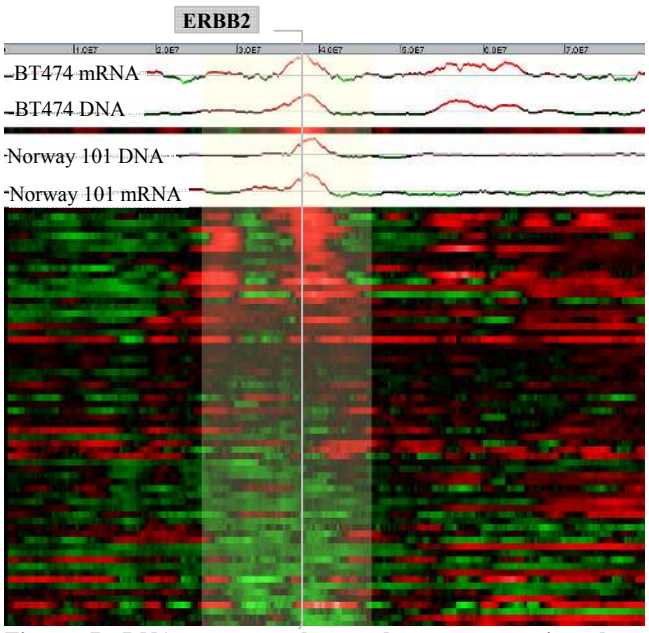

Figure 7. DNA copy number and gene expression data, *reordered* by clustering Euclidean distance of the ERBB2 Region. Once clustered, we find that cell line-matched DNA copy number and gene expression measurements are often next to each other or very close together, indicating a very high correlation between gene copy number and over/under expression of mRNA. Two examples are shown in the figure: the BT474 and the Norway 101 cell lines.

6. CONCLUSION

This paper presents Line Graph Explorer, a visually scalable technique for visualizing large collections of line graph data. The combination of a Table Lens-like interface for reordering line graphs based on associated metadata and a Focus+Context approach for inspecting line graph details affords a powerful and facile interaction technique for exploratory analysis of large collections of line graph data. Our preliminary evaluation of the prototype with data from a number of different domains suggests LGE's potential as a general tool for line graph data visualization. Viewed as an extension of Table Lens, it seems reasonable that LGE would likely inherit many of the same interaction and visualization benefits.

While our current work focuses on the combination of reorderable metadata and compressed line graphs, there remains an open question in the current design as to the most appropriate techniques to provide detail and context simultaneously. Studies of the effectiveness of Focus+Context techniques have yielded mixed results. While such techniques have been found to be beneficial for some navigation tasks [10, 24], their benefit is more questionable for interactive layout [9] and visual scanning [15]. We are currently initiating a user study comparing the use of Focus+Context, Overview+Detail and Details-on-Demand approaches for exploring line graph data.

We are also interested in examining methods to magnify all graphs at a specific point on the x-axis (a vertical lens). Such methods, when coupled with user-variable magnification, would

allow the user to drill to an arbitrary level of graph detail in either dimension. In addition, we are interested in expanding LGE's data analysis tool set by adding interactive data filtering as well as more powerful methods of feature comparison and search.

7. Acknowledgements

We would like to thank Stuart Card for helpful discussions and encouragement and Tamara Munzner for guidance and manuscript review.

8. References

[1] Agilent Technologies, Agilent 5100 Automated Lab-on-a-Chip Platform http://www.chem.agilent.com, 2005.

[2] Amar, R., Eagan, J. and Stasko, J., Low-Level Components of Analytic Activity in Information Visualization. in *Proc. of the IEEE Symp. on Information Visualization (INFOVIS '05)*, 2005, 111-117.

[3] Baudisch, P., Lee, B. and Hanna, L., Fishnet, a fisheye web browser with search term popouts: a comparative evaluation with overview and linear view. in *Proc. of the Working Conf. on Advanced Visual Interfaces (AVI'04)*, 2004, 133-140.

[4] Berry, L. and Munzner, T., BinX: Dynamic Exploration of Time Series Datasets Across Aggregation Levels. in *Posters Compendium of the IEEE Symp. on Information Visualization (INFOVIS'04)*, 2004, 5-6.

[5] Bertin, J. *Graphics and Graphic Information-Processing*. Walter de Gruyter, 1981.

[6] Card, S.K., Mackinlay, J.D. and Shneiderman, B. *Readings in information visualization : using vision to think*. Morgan Kaufmann Publishers, San Francisco, Calif., 1999.

[7] Clarke, S. and Engelbach, R. *Ancient Egyptian construction and architecture*. Dover Publications, New York, 1990.

[8] Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U S A*, *95*, 25 (1998), 14863-14868.

[9] Gutwin, C. and Fedak, C., A comparison of fisheye lenses for interactive layout tasks. in *Proc. of Graphics Interface*, 2004, 213-220.

[10] Gutwin, C. and Skopik, A., Fisheye View are Good for Large Steering Tasks. in *Proc. of the Conf. on Human Factors in Computing Systems (CHI'03)*, 2003, 201-208.

[11] Hao, M.C., Dayal, U., Keim, D.A. and Schreck, T., Importance-driven visualization layouts for large time series data. in *Proc. of the IEEE Symp. on Information Visualization (INFOVIS'05)*, 2005, 203-210.

[12] Hochheiser, H. and Shneiderman, B., A dynamic query interface for finding patterns in time series data. in *Proc. of the Conf. on Human Factors in Computing Systems (CHI'02)*, 2002, 522-523.

[13] Kincaid, R., VistaClara: an interactive visualization for exploratory analysis of DNA microarrays. in *Proc. of the ACM Symp. on Applied computing (SAC'04)*, 2004, 167-174.

[14] Kincaid, R., Ben-Dor, A. and Yakhini, Z. Exploratory visualization of array-based comparative genomic hybridization. *Information Visualization*, *4*, 3 (2005), 176-190.

[15] Kobsa, A. User Experiments with Tree Visualization Systems. *Proc. of the IEEE Symp. on Information Visualization (INFOVIS'04)* (2004), 9-16.

[16] Molina, M.A., et al. Trastuzumab (Herceptin), a humanized anti-HER2 receptor monoclonal antibody, inhibits basal and activated HER2 ectodomain cleavage in breast cancer cells. *Cancer Research*, *61*, 12 (2001), 4744-4749.

[17] Müller, W. and Schumann, H., Visualization for modeling and simulation: visualization methods for time-dependent data - an overview. in *Proc. of the Winter Conf. on Simulation*, 2003, 737-745.

[18] Pirolli, P. and Rao, R. Table lens as a tool for making sense of data. in *Proc. of the Workshop on Advanced Visual Interfaces (AVI'96)*, ACM Press, Gubbio, Italy, 1996, 67-80.

[19] Plot, R. *Philosophical Transactions of the Royal Society of London*, *169* (1685), 930-931.

[20] Pollack, J.R., et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. U S A*, *99*, 20 (2002), 12963-12968.

[21] Rao, R. and Card, S.K., The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. in *Proc. of the Conf. on Human Factors in Computing Systems (CHI'94)*, 1994, 318-322.

[22] Ryall, K., et al. QueryLines: approximate query for visual browsing. in *Extended Abstracts of the Conf. on Human Factors in Computing Systems (CHI '05)*, ACM Press, Portland, OR, USA, 2005, 1765-1768.

[23] Saito, T., et al., Two-tone pseudo coloring: compact visualization for one-dimensional data. in *Proc. of the IEEE Symp. on Information Visualization (INFOVIS'05)*, 2005, 173-180.

[24] Schaffer, D., et al. Navigating Clustered Networks through Fisheye and Full-Zoom Methods. *ACM Trans. Comput.-Hum. Interact.*, *3*, 2 (1996), 162-188.

[25] Siirtola, H., Interaction with the Reorderable Matrix. in *Proc. of the International Conf. on Information Visualisation (IV'99)*, 1999, 272.

[26] Spotfire, Inc., Spotfire http://www.spotfire.com, 2005.

[27] University of Dayton, Temperature Data Archive http://www.engr.udayton.edu/weather, 2005.