Visualizing very large layered graphs with quilts

Benjamin Watson, David Brink, Matthias Stallmann, Ravi Devarajan, Matthew Rakow, Theresa-Marie Rhyne and Himesh Patel

SAS Institute Inc. and North Carolina State University

ABSTRACT

Traditional node-link depictions of layered graphs such as flow charts and process or genealogy diagrams are in widespread use. Layers emerge from applied context (e.g., process stages or familial generations), or are inserted to improve visual clarity. However, for many applications these diagrams quickly lose their utility as graph complexity grows. We introduce quilting, an interactive, matrix-based depiction for very large layered graphs that remains useful even when optimized node and link depictions have become unintelligible. We demonstrate quilting using an activity-based management (ABM) application that must depict layered graphs with thousands or even hundreds of thousands of nodes. Unlike node-link depictions, quilts depict 500-node graphs quite clearly. On typical desktop displays, quilts depicting larger graphs must be summarized.


Additional Keywords: graph drawing, layered graphs, crossing minimization, matrix depiction.

1 INTRODUCTION

Layered graphs such as structure charts, process diagrams, and flow charts have wide-ranging applications. In these graphs, nodes are grouped into layers defined either by the application context, or introduced to increase visual clarity.

Traditional node-link depictions of layered graphs arrange members of a layer into a line [1][4]. Proper links connect nodes on adjacent layers; we call remaining links skip links. Crossing minimization algorithms reduce the intersection of proper links, and can improve legibility. Nevertheless, as the number of links grows, these depictions can become quite muddled, with viewers having trouble understanding graph connectivity (Figure 3).

To address the scalability problem for layered graphs, we introduce quilts (Figure 1), a new depiction that uses matrices to visualize layered graphs. Below, we describe the meaning and construction of quilts and show how quilts might be used in one application.

2 THEORY

The quilt depiction of a layered graph is a simple adaptation of the matrix depiction for unlayered graphs [2]. We represent proper links with dots in an achromatic matrix, and chain these matrices together with additional colored levels (rows or columns of cells) representing layers (Figure 1). Each level cell corresponds to an individual layer node. To distinguish levels from one another and from matrices, we assign a unique chroma and saturation to each level. We assign each level cell a unique luminance, effectively making the color of every graph node unique.

Each level acts as the source level for the following matrix, and the destination level for the previous matrix. Nodes on odd-numbered levels are lined up horizontally while those on even-numbered levels appear vertically. A proper link from an odd to even level makes a left turn whose corner is the cell representing that link. Similarly, a simple link from an even to odd level makes a right turn through the cell for that link.

We depict skip links with colored cells at the end of the row or below the column that follows level nodes. The color of each skip link is the same as its destination level node (recall that the color of each node is unique). To improve clarity, we sort the skip links using level number as the primary key and cell number (within the level) as secondary key.

2.1 Interactivity

To ensure good scalability, quilt cells are typically too small to support display of application-assigned node properties, especially textual labels. Viewers can reveal a node or link’s properties by hovering over its corresponding cell (Figure 3).

As noted by Ghoniem et al. [3], one of the weaknesses of a matrix depiction is path following. We address this shortcoming by allowing viewers to highlight graph paths in the quilt by clicking on node or link cells (Figure 3). The first such click highlights the cell itself, as well as immediately adjacent nodes. Each additional click highlights nodes that are one more link distant. We call this click-through. A backward click-through, removing highlights from the nodes last reached, is available by clicking on the node while the control key is pressed. Pressing the shift key while clicking on a node highlights all the nodes reachable from the clicked cell. Clicking on any portion of the quilt not containing a node or link removes the highlight.
To provide viewers with the strengths of both quilts and the node-link depiction, we couple these depictions in an interactively linked view (Figure 3). Clicking in either depiction will highlight the appropriate graph path in the other depiction.

2.2 Summarization

When graphs contain thousands of nodes or more, the corresponding quilts can require summarization, since they will not fit in a typical display. To summarize quilts, we cluster nodes. Each summarized level cell except the last represents $s$ unsummarized nodes. Each summarized matrix cell represents up to $s \times r$ links. In applied settings, we find that summarized matrix cells represent far fewer links. We map the number of links in summarized matrix cells to cell luminance, with darker cells indicating more links (Figure 2).

Hovering over a summarized level cell indicates the number of cells being summarized, as well as statistical summarizations of any application properties in the summarized set of cells. Hovering over summarized link and skip link cells has similar functionality.

Selecting a summarized cell highlights summarized level and matrix cells in a manner much like that in unsummarized quilts.

![Figure 2. Summarized quilt, 16 nodes per summarized node.](image)

3 DISCUSSION

Scalability. Quilts remain legible even as graphs grow to contain hundreds of nodes. In contrast, node-link depictions with only a few dozen nodes can be difficult to understand.

Property visibility. Quilt scalability comes at the price of some property visibility, with node and link properties and labels hidden inside small cells until the user hovers over the cell. This can make finding nodes or links difficult. However, property visibility in node-link depictions also suffers as graphs grow in size.

Link visibility. Following a series of links in small node-link depictions is simple, since all links begin and end directly at the linked nodes. Links in similarly sized quilts require the viewer to trace the matching row and cell across the matrix that joins them or examine the color of the skip links. In larger graphs this is less of a disadvantage, as following links in node-link depictions becomes more difficult.

Skip links. Node-link and quilt legibility suffers as the ratio of skip to proper links grows, since layers become a less useful visual grouping. To aid in skip link legibility, hovering over a skip link draws a line to its corresponding destination node. Perhaps the most effective depiction of such irregularly connected graphs is a single matrix: because it does not depict layers, it does not require special cells to depict skip links.

Crossing minimization. In node-link depictions, preserving legibility as graph size moves from small to moderate requires crossing minimization – an NP-hard problem. Quilt depictions remain legible even without crossing minimization, though such minimization (which would concentrate link dots along matrix diagonals) may still prove beneficial.

3.1 An Applied Example

We developed the quilt in the applied context of SAS’s Activity-Based Management application (ABM). ABM is an analytical application that models an organization's processes to determine accurately the cost and profitability of products and customers. ABM models the interactions between corporate groups and assigns revenue to those responsible for the products or services.

SAS’s ABM system uses a directed graph to model the interactions within an organization. Large organizations and their processes yield complex models often containing hundreds of thousands of vertices and millions of links. While answering specific cost or revenue questions using these models is fairly simple, realizing the data’s true wealth requires visualizing them to understand hidden trends. Unfortunately, most of our ABM graphs overwhelm traditional node-link depictions.

Figure 3 shows a small ABM graph depicted using linked node-link and quilt views. The graph models the cost processes involved in an airlines’ flight catering service. Here we have highlighted the food prep group in our linked views. By following the incoming/outgoing links we can see the objects that contribute to/receive contributions from this group. Even with highlighting, in the node-link depiction it is difficult to discern exactly how highlighted nodes are connected to the food prep group; in the quilt this connectivity is much clearer. On the other hand, in the quilt we cannot see the labels or attributes of any nodes except the one selected.

![Figure 3. Example with ABM data.](image)

4 CONCLUSION

We have presented a new depiction of layered graphs, which we believe remains legible even when the depicted graphs have several hundred nodes. We plan to investigate layer and crossing optimization, clustering and zooming in summarization view, interactive editing of the quilt, new ways to represent OLAP data, and user interface improvements. We would also like to conduct a user study to compare the effectiveness of quilts in comparison to node-link depictions of layered graphs.

REFERENCES


