Dynamic Graphics for Network Visualization

Richard A. Becker Stephen G. Eick Eileen O. Miller Allan R. Wilks

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Abstract

Network data involves statistics that are associated with the nodes or links in a network. We describe several dynamic graphics tools for visualizing such data.

1. Introduction

Along with the explosive growth of computer networks in recent years has been a concomitant explosion in the volume of network data, measuring such quantities as capacity, flow, blocking and delay. Analyzing such data is made difficult because of the need to incorporate into the analysis the often complicated topological structure of the network. Because of this, little attention has been paid to the graphical display of network data, beyond simply displaying the topological information. In light of the growing importance of network management and analysis, new visualization tools for network data are needed.

Network data arises in contexts other than computers, of course. Examples are telecommunications networks, financial flows and migration flows. For an interesting discussion of visualizing migration flows, see Tobler [3]. In all of these applications, exploring and modeling the data can be enhanced through the use of graphical tools.

In the next section we suggest a number of ideas for the static display of network data, while motivating the need for interaction through dynamic graphics. Section 3 has a brief discussion of dynamic graphics in general and then in section 4 we specialize this to the case of network data. We conclude with an indication of the major application area that motivated this work and in which we have been able effectively to use our tools.

2. Static Displays for Networks

Broadly speaking there are two sorts of network data—those associated with the links of the network and those associated with the nodes. Node-oriented data may be, for example, the aggregation of link-oriented data around each node. In what follows we assume that network data will be displayed on an underlying map, which will be either a real geographical representation of the spatial node locations, or a schematic representation of the network layout.

A common technique for displaying link data on a map is to draw line segments between each pair of nodes for which there is data. Each segment may be colored to encode the value of the statistic. Alternatively, its thickness might encode the value. Since such data is often directional, it is possible to show data in both directions by bisecting the segments and always associating the half connected to a node to show the statistic with that node as the originating node. Figure 1 illustrates some of these ideas (the parts of the display below and to the left of the map will be described in section 4). The display shows blocking data from the AT&T long distance network on October 17, 1989, the day of the California earthquake.

For all but the simplest networks these link data displays may have many intersecting lines and are difficult to interpret. There are several possibilities for reducing clutter. One is to shorten the line segments, that is, instead of drawing the line segments 50% of the way between nodes, draw them 30% or 10%, say. Another is to draw only lines whose corresponding statistic falls above or below some threshold. The difficulty with these ideas is that it is quite hard to come up with a good heuristic for setting these thresholds or line lengths (or overall line thicknesses, for that matter) before making the display. This difficulty may be overcome spectacularly by dynamic manipulation of the graphical display as described below.

Node-oriented data may be displayed by showing a symbol such as a circle at each node on the map, with the symbol scaled to represent the value of the statistic. More complex symbols can be used to represent more than one statistic simultaneously. For example, if flow is the statistic under consideration, there will typically be an inflow and an outflow at each node; these

CH2913-2/90/0000/0093/$01.00 – 1990 IEEE

can be shown by using rectangles for the symbols, with the width and height proportional to the two statistics. As with link data displays, it is difficult to determine *a priori* what the overall size of the symbols should be, and yet this decision has a large effect on the interpretability of the resulting display. Figure 2 shows data on call attempts for the same event as Figure 1.

Sometimes it is desirable to study the evolution of a node or link statistic through time. One possibility is to make a number of displays, one for each time point, and then scan these visually. The difficulty here is that there are often many time points so that it is impractical to produce all the displays and it can be difficult to interpret them by the scanning process. Here again direct manipulation of the display can come to our aid.

3. Dynamic Graphics

We have used dynamic graphics techniques in the development of several visualization tools for network data. By dynamic graphics we mean direct manipulation, in which the user interacts directly with the graphics by use of a mouse [2]. The mouse is used to adjust, among other things, various sliders that control the parameters mentioned previously: overall line thickness, overall symbol size, line length, and so on. The net effect is the ability to quickly and naturally filter through many static pictures that may be uninteresting and to focus on the ones that are meaningful. This could mean focusing on data of interest, such as only looking at large flows, or it could mean finding unexpected anomalies.

While the static displays mentioned earlier have merit in the context of final data presentation, we believe that dynamic graphics is critical for the prior process of understanding and finding patterns and anomalies in network data. Thinking of the possible displays of a particular set of data as being parameterized by symbol size, line length, time, and so on, the space of all possible parameter values is large and most combinations of the parameters do not lead to useful or even interpretable displays. In fact, our experience is that interpretability is quite sensitive to the values of the parameters. It is the *dynamic* modification of the parameters, coupled with visual feedback, that allows the analyst to adjust the parameters efficiently to produce informative displays.

4. Dynamic Displays for Networks

One of our tools (see Figure 1) is specifically tailored to looking at link-oriented data. It takes as input several link statistics (possibly the same statistic at several different times), together with information on where to place the nodes on the display. A list of the input statistics is shown on the screen and the user can move continuously through the displays for each statistic by "brushing" on this list [1], that is, by moving the cursor across the list with a mouse button held down. The tool displays one of these statistics using the line segment display described earlier. By moving sliders, the user may continuously adjust the length and thickness of the line segments. The color legend that describes the color coding of the lines has two slider bars superimposed on it, representing upper and lower thresholds. These are initially placed at the two ends of the legend; by moving them the user can restrict which lines are shown to those whose values lie between the two thresholds.

There are also several operations available on the link display. When the cursor is near a node, that node becomes *current*, and some node-specific identification text is displayed. By clicking a mouse button near a node, that node becomes the *anchor* node. Subsequently, whenever another node is current the actual values for all the statistics between the current and anchor nodes appear in a fixed place on the display. Using brushing, nodes can be *deactivated*: all lines into and out of them are turned off. Brushing with a different button held down reactivates such nodes. In addition to these capabilities there are a number of other mouse activated screen "buttons", including one that allows zooming in on a piece of the network.

A second tool (see Figure 2) is oriented towards the display of node data. Its input is one or more node statistics for a number of time periods. The user controls which of these statistics are selected for display; if one is selected it is encoded in the size of circles centered at the nodes and if two are selected they are encoded by the width and height of rectangles at the nodes. For example, the statistics might be inflow and outflow. It is often important to identify sources or sinks in the network. To aid in this, rectangles that are short and fat or tall and thin are colored red or green, respectively, and all other rectangles are colored white. The amount by which the aspect ratio of a rectangle must deviate from 1 to cause it to be colored is controlled through a slider by the user. A second slider controls the overall size of the symbols.

If the node data is given for more than one time period, a third slider may be manipulated to select the time period for which data is currently being displayed. The tool also has the capability of playing automatically through all the time periods, and another slider is provided to control the playback speed. Our experience has been that using this automatic playback feature allows us to understand patterns in quite large volumes of data as well as to rapidly detect unusual patterns or anomalous events.

5. Application

Our motivation for this work came from analyzing large collections of data involving AT&T's long distance network, although our tools have wide applicability to the analysis of other network data. In the AT&T network there are over 100 nodes and these are essentially fully interconnected. This means that link displays can have over 10,000 line segments on them.

Using our tools we have been able successfully to view this data and make interesting discoveries. We have produced a video tape illustrating the use of our tools to analyze the calling patterns after the 1989 California earthquake.

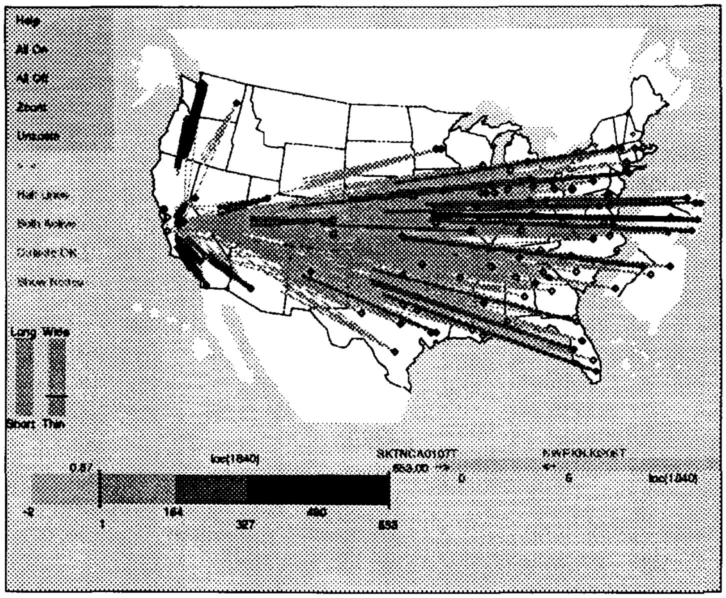
References

[1] Richard A. Becker and William S. Cleveland, "Brushing Scatterplots", *Technometrics*, Vol. 29, No. 2, pp. 127-142, May, 1987.

[2] Richard A. Becker, William S. Cleveland and Allan R. Wilks, "Dynamic Graphics for Data Analysis", *Statistical Science*, Vol. 2, No. 4, pp. 355-395, November 1987.

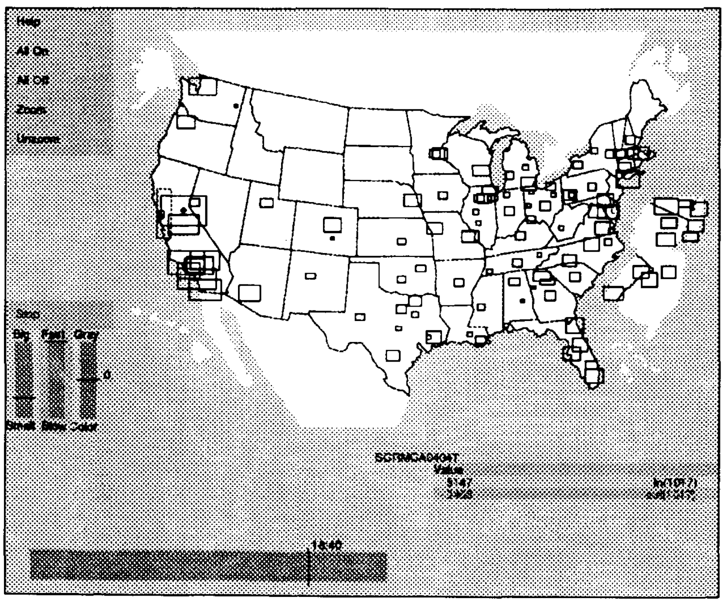
[3] Waldo R. Tobler, "Experiments In Migration Mapping By Computer", *The American Cartographer*, Vol. 14, No. 2, pp. 155-163, 1987.

Figure 1. Overflowing Calls Following the October 17, 1989 Earthquake ***(Color Plate 45, page 467)***



This figure shows the overflowing calls (non-completed due to congestion) in the AT&T network on a node-to-node basis. Overflowing calls are represented by a line extending from the originating switch half-way to the terminating switch; the line color encodes the number of calls according to the color scale. Notice how the number of overflowing calls *into* the San Francisco Bay area, the site of the earthquake, is much greater than out of the Bay area. This was due to manually instituted network management controls that made it possible for callers to get information out of the disaster site. Here, Newark is an anchor node and Stockton is the current node (described in section 4). The big "island" off the East Coast is a blowup of the New York City/New Jersey area. Note: data values have been altered.

Figure 2. Call Attempts Following Earthquake ***(Color Plate 46, page 467)***



This figure shows the number of calls into and out of each AT&T network switch at 18:40 network time (17:40 Pacific Daylight Time), just after the earthquake. Each rectangle is centered about a switch; its horizontal dimension is proportional to the number of incoming calls in the preceding 5-minute period and the vertical dimension encodes the outgoing calls. It is easy to see the abnormally large numbers of calls processed by the Bay Area switches. This effect is even more pronounced when the day's data is animated (see section 4). Sacramento is the "current" node, and its statistics are shown in the region below the screen.