Graph Data in R, a User's Perspective

Chris Volinsky

February 14, 2004

Data which can be represented by a graph (ie. as a collection of nodes and edges between those nodes) have been generating a lot of interest lately. Graphs have become an increasingly popular way of representing data in many different domains, ranging from web connectivity to disease transmission to relationships between monks in closed communities, and many more. Traditionally, the tasks of graph creation, layout, manipulation, modelling and rendering might be handled by several different tools. In the last few years the R community has developed packages for integrating many of these tools into one cohesive environment. They include:

- **graph**: The R **graph** library is a toolbox of graph creation and manipulation tools that allow a graph to be created from already existing R structures. The representation of the graph can be as a list of nodes and edges, or as a distance matrix, depending on the nature and size of the graph. **graph** contains many useful tools for graph manipulation and modelling, including boundary calculations, graph distance, and random graph probability measures.
- **Rgraphviz**: The **graphviz** library is a commonly used graph layout tool. The **Rgraphviz** package supports two major graph layout algorithms, dot (a hierarchical layout) and neato (a radial layout).
- **Rggobi**: **Ggobi** is a powerful interactive data visualization tool which has developed some nice features for graph rendering and manipulation. **Rggobi** was developed as an interface for R to interact with Ggobi in a seamless manner, and allows the user to do brushing and spinning and many more interactive data tasks.

Using these three tools together provides a powerful toolbox for data analysis with graph data. An R object can be turned into a graph object using the **graph** library, **Rgraphviz** is then used to generate a snazzy layout, which is then passed to **Rggobi** tools to render the plot so that the user can interact with the data.

I have used these tools to help in analyzing telecommunications data at AT&T. I look at "Communities of Interest", or COI, which are small subgraphs of our massive callgraph, which we use as a network signature for the purposes of fraud detection. In this talk, I will present some COI data in the context of a fraud application, and show how I have used the graph tools above to learn about, analyze and model these graphs.