Visualizing Augsburg Traffic Data with VanGogh

Rene Keller [1]
Universitätsstr. 14, 86135 Augsburg, Germany
E-mail: Rene.Keller@student.Uni-Augsburg.de

Keywords: Networks; Exploratory Data Analysis; Traffic Data

Abstract

Visualizing networks is a common task in statistics. As networks can be very large, new approaches must be tested. Techniques that just concentrate on the visualization of the network itself can give valuable information, but other graphics can help as well. This makes interactive features such as linking and selection necessary. VanGogh is a software solution that deals with networks in an interactive way to find new and interesting properties of a given dataset. This article will give a brief introduction to VanGogh and will show how such an analysis of network data works in general.

1 Introduction

Many datasets contain network data, but there are hardly any software solutions to work with networks interactively. daVinci[2] and netdraw[3] are free software for dealing with networks, but they lack interactive features. For this purpose VanGogh (**V**isualizing **a** **n**etwork, **G**raphics **o**verviews **g**raphics **h**ighlighting) was developed, to test how interactive tools can be applied to networks.
This article will introduce VanGogh and show how VanGogh can be used. For that purpose I will give an example for a typical analysis that can be performed with it.

[1] Computeroriented Statistics and Data Analysis, Augsburg University, Germany
[2] Download at: www.b-novative.com/products/daVinci/daVinci.html
[3] Download at: www.analytictech.com/downloadnd.htm

"Augsburg Traffic Data" is a dataset provided by the city of Augsburg to measure traffic densities downtown. In a very constructive cooporation with the city government several datasets were provided.
A network $\mathcal{N} = \{N, F\}$ is a set of nodes $N = \{1, ..., n\}$ and flows F with $F \ni f = \{i, j\}$ and $i, j \in N$. Nodes typically stand for static objects such as websites, streets or countries, while flows represent connections, trades and links between these nodes. Flows and nodes may have additional attributes a_{ijk} $(i, j \in N)$ and b_{il} $(i \in N)$ such as weights, lengths or coordinates. Networks are typically stored in adjacency matrices if there is only one additional attribute for the flows. If there are more attributes, it is useful to store this kind of data in a data matrix with one flow in each row (see Appendix A). Transformation of these two formats can be done easily.

2 Augsburg Traffic Data

"Augsburg Traffic Data" comes in two different datasets for the years 2001 and 2002. The datasets were provided by the city of Augsburg and are the result of studies initiated by the city council. The question was, which streets suffer from very high traffic density and how these problems can be solved. Another problem was, how the results of the analysis can be presented to politicians in an appropriate way.
The dataset is real. It was obtained by observing the registration numbers of all cars driving into and out of the city during a certain period of time. Only through traffic was observed, that means that only cars were observed, that left the city again within 30 minutes after entering it. In total 24261 cars were observed in 2001 and 7046 in 2002. This resulted in 1104 flows in 2001 and 168 in 2002. In 2002 only the southern half of the city was surveyed, as at the end of 2001 a new bypass was completed there. The files `Flows2001.txt` and `Flows2002.txt` contain the data. The files `CityImage.txt` and `City.jpg` are for importing a background image into VanGogh.
The variables of the datafile stand for different aspects of the resulting networks:

- von: The names of streets where cars enter the city of Augsburg.
- nach: The names of streets where cars leave the city of Augsburg.
- x1: The corresponding x coordinate of the von street in the imagefile.
- y1: The corresponding y coordinate of the von street in the imagefile.
- x2: The corresponding x coordinate of the nach street in the imagefile.
- y2: The corresponding x coordinate of the nach street in the imagefile.
- anz: The number of cars that entered the city at von and left the city at nach.
- wann(only in `Flows2001.txt`): The time of the observation. 0: morning and 1: afternoon.

The variable names “von” (from), “nach” (to), “anz” (count) and “wann” (when) are used, because this dataset is normally presented to Germans. The exact categories of von and nach differ in the two datafiles, because in 2002 only traffic in the south of the city was surveyed (some streets were not included, others were added).
This kind of dataset requires two very important concepts. One is weight variables. Weight variables contain weights for every case (flow) of the dataset. This is very important here, because cars with same entering and leaving point have been aggregated. The weight variable (here anz) specifies the number of cars driving the same route. In Flow Diagrams weights are used to specify the size of the resulting flow representations. If not used, all flows will have the same size. If weight variables are used in Histograms, Barcharts and Fluctuation Diagrams, the size of the objects are not proportional to the number of cases any more, but to the sum of the weights.
The second concept is aggregation. If one takes a closer look to the `Flows2001.txt` dataset, it is clear that there are always two flows from node A to node B, one flow for the morning and one for the afternoon (in variable wann). In a Flow Diagram these two flows will always overlap. Aggregation means, that all flows with same source and target nodes are summed up to one flow, adding the weights specified by a weight variable. Aggregation is an important part of the construction of Barcharts, Fluctuation Diagrams, Histograms, SymmDiagrams and Map Diagrams, and can be applied to Flow Diagrams as well.

3 VanGogh

VanGogh is an interactive software for visualizing networks. It is written in JAVA, so it is possible to run VanGogh on any platform. It only requires JAVA 1.3. I am developing VanGogh for my Diploma Thesis. In this article I will concentrate on the Windows version of VanGogh, the Macintosh version differs in some aspects from the descriptions below. VanGogh can be downloaded at

www1.math.Uni-Augsburg.de/VanGogh.

3.1 Getting started

The first Impression

VanGogh is started with a doubleclick on `VanGogh.jar`. The main window will open (see Fig. 1). It is divided into the menubar and the toolbar at the top and a virtual desktop. The two rightmost buttons in the toolbar are for specifying user preferences and for online-help. The online-help provides guides to most of the features of VanGogh. It is recommended that the online-help is searched if anything is unclear. The four buttons labelled I, II, III and IV allow the user to switch between four different virtual desktops, so it is possible to have 4 different datasets or collections of views of the same dataset open at once. The active buttons on the left are for opening a dataset and for quitting VanGogh.

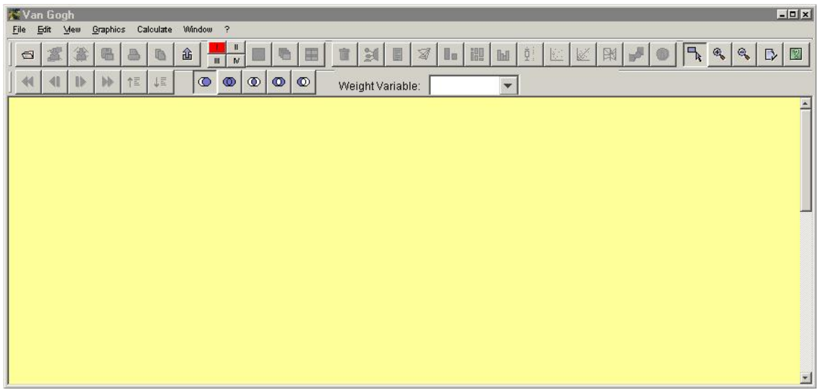

Figure 1: The Main Window of VanGogh

Opening a Dataset

A dataset can be opened with the corresponding button in the toolbar or the menu. With a standard file dialog, the dataset file can be chosen. The datafile must contain ASCII text, separated by tabs or spaces, and must include a single row of variable names (see Appendix A). VanGogh uses a standard data matrix with one row for each entry. Data provided in adjacency matrices must be converted into this form by the user.
After selecting a datafile and loading it into VanGogh, two windows will be shown in the currently selected virtual desktop (see Fig. 2). The one on the left (Dataset Frame) contains a list of all variables contained in the currently open dataset as well as information about the dataset itself (number of variables, cases and currently selected cases). Doubleclicking on a variable opens a list of all entries of this variable.

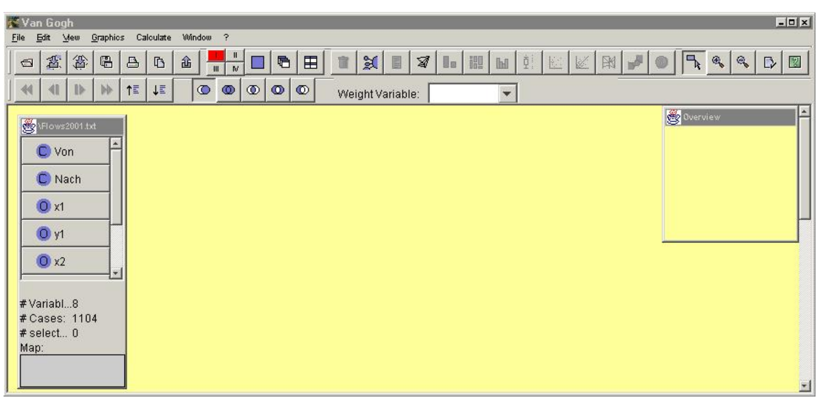

Figure 2: The Main Window of VanGogh after opening a dataset

The window topright is the Overview and is for zooming into graphical displays. It will give information about the currently visible area of the active graphical display. More buttons become available after opening a dataset. The buttons next to the Open button are for importing a map (see Appendix C) and image files (see Appendix B) for better visualization of the resulting networks. Buttons for rearrangements of open windows are enabled now as well.

3.2 Graphical Displays

Several graphical displays are available in VanGogh. Each of them can be opened by selecting variables and then choosing the corresponding button in the toolbar or the Graphics menu.

Plot Capabilities

As per default there is no weight variable set, weight variables can be added to all graphical displays available in VanGogh (except Dotplots, Scatterplots and Parallel Coordinates) using ALT-W. It is possible to drag a weight variable directly from the Dataset Frame into the desired graphic. Choosing a weight variable in the selection box in the middle of the toolbar is also possible. All plots that are opened from then on will have this variable as a weight variable per default.
Each graphic can be copied to the clipboard, saved as a `.jpg` or printed out. Zooming, panning and selection are possible as well. Available options for a plot can be displayed by pressing ALT-O or can be found in the online-help.
In the next part of this paper I will introduce the graphical displays available in VanGogh and describe how they work.

Flow Diagrams

Flow Diagrams are available via the corresponding menu item in the Graphics menu. A dialog box appears in which the variables for source and target node variables and coordinate variables are set. Confirming these settings opens a Flow Diagram.
If an image or map file has been imported before, the plot will show this map or image in the background, if not, the plot will be empty. This is, because only selected flows are visible per default. To show all flows, press CTRL-A. Now the flows will be visible (see Fig. 3 left).

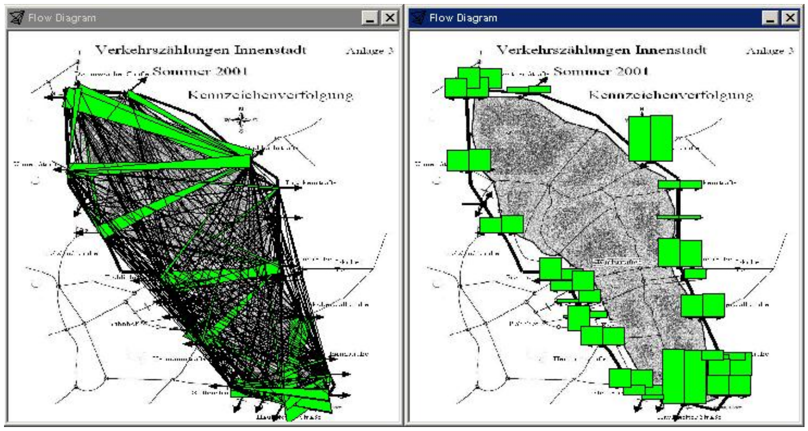

Figure 3: Flow Diagram with flow displays on the left, node displays on the right

The Flow diagram offers a wide range of different views of the dataset. It is possible to use Object and Censored Zooming (described below) to increase or decrease the sizes of the arrows representing the flows.
Pressing ALT-O brings up the Options menu for this plot. It is divided into four parts. The last part is for selecting a weight variable (see above). The third part specifies, whether images or maps should be visible in the background. Every image or map can be used as a background for a Flow Diagram, but it is clear, that the coordinates have to match!
The first part contains all options available for visualizing flows:

- Show Flows: Switch on/off whether flows should be visible at all.
- Show all Flows: Show all flows, including not-selected ones.
- Show not selected with Alpha Blending: Show selected flows as normal. Not-selected flows will be drawn transparently.
- Show only selected Flows: Show only selected flows.
- Arrows, Representations, Lines, Bars: Choose different representations of the flows.
- Aggregate Flows: Sum up all flows with same source and target nodes.
- Transfer Points: Select x and y coordinate variables to replace the straight flows by Bèzier curves. **Attention:** This will only show a result, if Arrows were chosen for the flows.

The second part of the Options menu concerns options for visualizing nodes (see Fig. 3 right):

- Show Nodes: Switch on/off whether nodes should be visible at all.

- Show only selected Nodes: Show only selected nodes.
- Bars: Bars are drawn as visualizations of the nodes (see Barchart below).
- Circles: Circles and squares are drawn. Circles represent targets, squares represent sources.

Selection and zooming apply to both nodes and flows. If there is any ambiguity, only nodes are selected (for instance if a selection is made at a city entry point). Only displayed objects can be selected.

Barcharts

Barcharts are opened by selecting at least one categorial variable and then choosing Create Barcharts (Example see Fig. 4). Barcharts are capable of Censored Zooming described below. Pressing ALT-O brings up the Options menu for this plot.

- Barchart, Spine Plot: Switch between a Barchart and a Spineplot. Note: Censored Zooming is not available in Spine Plots.
- Add Weight Variable: Adds a weight variable, described above.
- Add Confidence Intervals: Choose whether confidence intervals are shown or not. For selected bars a binomial model is used, and for not selected ones a Poisson model. Confidence intervals can only be added to Barcharts.

Selecting, zooming and querying works in the standard way, but there is more possible with a Barchart. By dragging a label of a bar, the order of the categories can be changed. Sorting is possible as well. CTRL-UP and CTRL-DOWN sorts the categories of a Barchart by their absolute selected values. If a Spine Plot is visible the categories are sorted by the relative selected values. The results of such operations are also applied to all other open plots that inherit the same variable as this Barchart (Barcharts and Fluctuation Diagrams).
Another possibility in Barcharts is grouping. Selecting the corresponding button from the toolbar creates a new variable with a new category holding all selected cases even if only parts of bars are selected.

Fluctuation Diagrams

A Fluctuation Diagram is special case of a Mosaicplot (see [2]) for a set of categorial variables. The first variable is split up into its categories horizontally as in a Spineplot. This is repeated with all other variables, by alternating between the horizontal and vertical directions. The result will be a Mosaicplot with the area of a mosaic equal to the size of the cases falling into the same categories. In a Fluctuation Diagram

the mosaics are shifted to a linear grid and adapted in size so that the mosaics are proportional but do not overlap. If a weight variable was selected, the sizes of the mosaic are not proportional to the number of cases falling into the same categories, but to the cumulated weights of these cases.
Fluctuation Digarams can be opened via the corresponding button in the Graphics menu. If one or more categorial variables are selected, a Fluctuation Diagram shows up (see Fig. 5).
Fluctuation Diagrams allow Censored Zooming via the cursor panels. The Options menu allows switching between a Fluctuation Diagram and a Mosaic Plot and to specify a weight variable.
Dragging of variable names changes the order of the variables for the construction of the Fluctuation Diagram.
Note: Sorting in Barcharts may have an effect on Fluctuation Diagrams! If a variable that is contained in the Fluctuation Diagram is sorted in a barchart, the order of categories will be changed in the Fluctuation Diagram as well, leading to a new view to the data.

Histograms

A Histogram is opened by selecting a variable and choosing Create Histograms (Example see Fig. 6). Histograms are useful to get an overview of continuous variables. As the anchor point and the binwidth have a large effect on the resulting Histogram, these values can be changed in different ways. The easiest one is to drag the corresponding line to a new value. There are situations where this is not precise enough. By querying these lines a menu will appear, allowing the setting of specific values for the anchor point and binwidth or to select certain binwidth rules. The rules for the binwidth are taken from [4]. Per default Scotts rule is applied.
In the Options menu you can specify a weight variable and confidence intervals. Confidence intervals use a Poisson model for not selected cases and a binomial model for selected ones (see Barcharts above).

Figure 4: Barchart

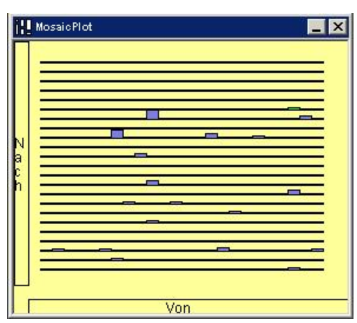

Figure 5: Fluctuation Diagram

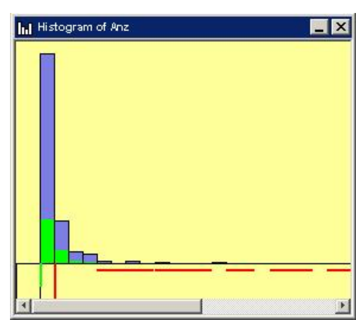

Figure 6: Histogram

Dotplots and Scatterplots

Dotplots and Scatterplots are opened via the Graphics menu. One continuous variable must be selected for Dotplots and at least two for Scatterplots. If there are more than two variables selected and Create Scatterplots was chosen, a Scatterplot matrix will appear.
Dotplots (see Fig. 7) and Scatterplots (see Fig. 8) both support Object Zooming and Alpha-Blending per default. CTRL-S will change a horizontal Dotplot into a vertical one and in the case of a Scatterplot the axes are switched. In Scatterplots some density estimators have been implemented, because especially in Scatterplots overlapping of points can be a big problem. Hexagonal and rectangular binning and a gaussian kernel estimation are available. The binwidth and the windowwidth can be controlled using the cursor keys with ALT. For more information about density estimation in Scatterplots see [4] and [1].

SymmDiagrams

SymmDiagrams visualize flow balance in opposite directions. In the "Augsburg Traffic Dataset" it helps to detect whether there are differences between the number of cars driving from street a to street b and from street b to street a.
SymmDiagrams are a special form of Scatterplots. They are made for two variables and a weight variable. By selecting the corresponding button in the Graphics menu a SymmDiagram can be drawn.
For each pair of nodes a and b the number of flows n_{ab} from a to b and n_{ba} from b to a is calculated. In the diagonal view n_{ab} is the x coordinate and n_{ba} the y coordinate of the point representing the nodes a and b. As there is no natural order in the nodes, each pair of nodes appears twice in the SymmDiagram. Because this view is symmetric to the first main diagonal, this diagram is called a SymmDiagram. The Tukey view of the SymmDiagram, which is the default, takes $\frac{n_{ab}+n_{ba}}{2}$ in the x direction and $\frac{n_{ab}-n_{ba}}{2}$ in the y direction. This view is symmetric to the zero line in x direction. The closer a point is to the symmetry line, the more balanced are the flows in opposite directions. If a weight variable was chosen, replace n_{ab} by the sum of weights x_{ab} of flows from a to b.
Figure 9 shows the resulting SymmDiagram of "Augsburg Traffic Data". The isolated point in the north of the diagram has the largest difference of flows in opposite directions. This example will be discussed later in section 4.2

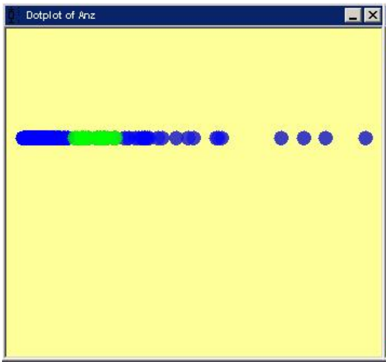

Figure 7: Dotplot

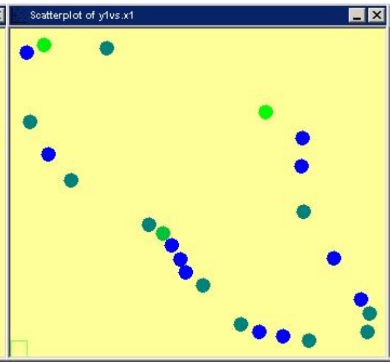

Figure 8: Scatterplot

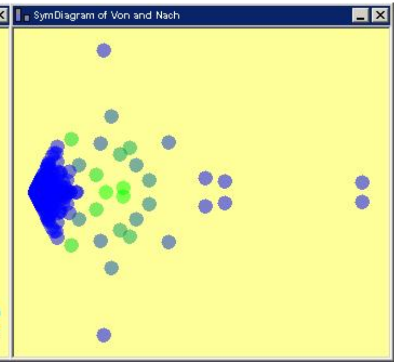

Figure 9: SymmDiagram

Parallel Coordinates

Parallel Coordinates are discussed in [3]. Selecting a number of variables and then choosing Create Parallel Coordinates opens this display (see Fig. 10). In connection with networks, Parallel Coordinates can be useful if there is more than one weight variable available (e.g. traffic at different times or for different kinds of traffic). Then the different weights can be compared easily.
Parallel Coordinates support Object Zooming and Alpha-Blending. Per default only selected cases are displayed, though this can be changed using the Options menu. All interactive possibilities work in the standard way, but it is also possible to change the order of the variables contained in the plot by dragging them into a new position. It is also possible to sort the variables according to mean, median or standard error and to draw the variables all on the same scale or on individual scales.

Map Diagrams

Map Diagrams are only available, if a polygon map has been imported. To open a Map Diagram select a categorial variable that matches the names of the polygons of the map file (see Appendix C) and then choose the corresponding button from the Graphics menu. Figure 11 shows a Map Diagram taken from the analysis of a dataset concerning traffic between districts of Augsburg.
The Options menu allows you to specify a weight variable and to select how different values of polygons are drawn. As Augsburg Traffic Data does not come with a polygon map, this diagram has not been used for this dataset. Figure 11 is a result of another dataset about traffic in Augsburg region.

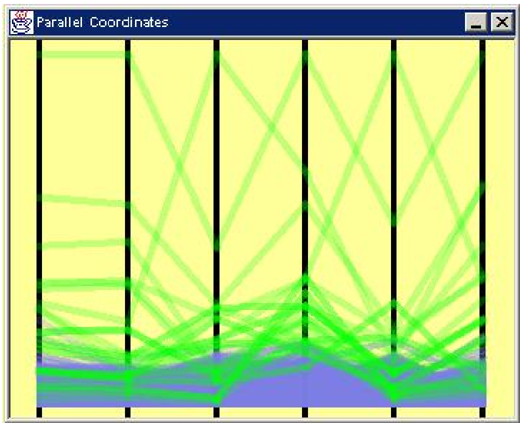

Figure 10: Parallel Coordinates

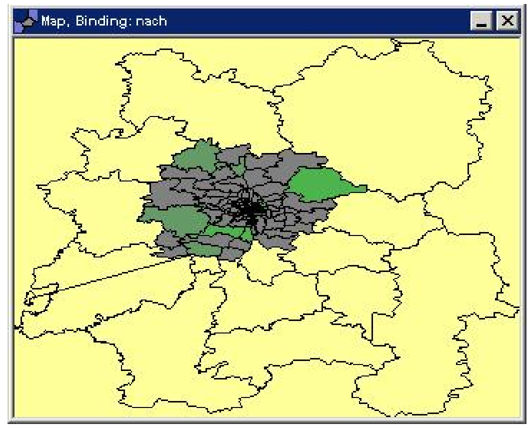

Figure 11: Map Diagram

3.3 Interactive Capabilities

Interactivity is essential for Exploratory Data Analysis, because it allows you to switch between different views quickly. VanGogh supports a number of interactive features described in the literature (see [6]). Most of these features work in all plots (Selection, Linking, Zooming, Panning and Querying), some of them can only be applied to some displays (Alpha Blending, Object and Censored Zooming). Features not described here may be found in the descriptions of the graphical displays themselves in section 3.2.

Selection and Linking

Selection is made just by dragging a rectangle in the selected window. All elements that intersect with this rectangle will be selected. For selection there are Undo and Redo capabilities implemented in VanGogh. The corresponding buttons can be found on the left of the toolbar. Selection can be performed with different modi. The corresponding buttons can be found next to the Undo and Redo buttons.
Selection is linked in all plots. this means that in every open plot there will be the same cases selected at any time.

Zooming and Panning

Zooming is possible with a mouseclick into the desired plot when zooming mode has been previously activated. Switching between zooming and selection mode can be performed using the buttons to the left of the Preferences button. Every plot can be enlarged, but Barcharts, Histograms, Dotplots and Parallel Coordinates are not

enlarged actually, they are only stretched horizontally. The result of the zooming can be observed in the Overview window.
Panning is possible with the scrollbars next to the graphical display and in the Overview, by dragging the rectangle showing the currently visible area of the plot. See Fig. 12 for an example of zooming.

Querying

As VanGogh is an interactive software, it supports querying. Therefore full labelling of graphical displays is not needed, which is rather a topic for static displays. Querying can be performed using the right mousebutton (CTRL-Click on Mac). Every plot supports querying and every object that is displayed in a graphic may be queried. If several objects overlap, a summary of all underlying objects will be made except in the Flow Diagram when nodes and flows overlap. In this case only the node will be queried. For a typical query result, see Fig. 12.

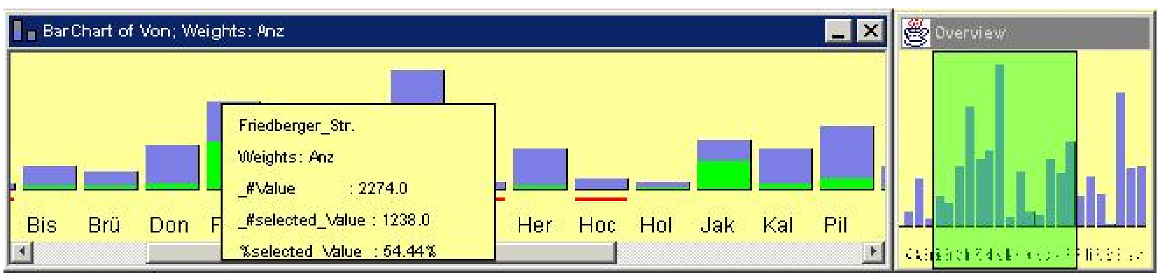

Figure 12: Zooming and Querying a Barchart, Overview on the right

Object Zooming

Object Zooming means the enlargment of objects in a plot. The plot is not enlarged by itself, but only the objects in it. In a Scatterplot this means the control over the point size (see Fig. 13). With this (in addition to alpha blending capabilities) it is possible to focus on certain aspects of the plots by enlarging them. Object Zooming is supported by all "point plots" Scatterplots, SymmDiagrams and Parallel Coordinates and the Flow Diagram. Object Zooming can be controlled using the UP and DOWN cursor keys with ALT.

Figure 13: Original Scatterplot on the left, Object Zooming result on the right

Censored Zooming and Alpha Blending

Censored Zooming is related to Object Zooming. Censored Zooming does not magnify every object in the plot, but only the small ones, while the large objects do not grow. This is very useful for comparison of small objects without creating overlapping. The problem is that it is not possible any more to compare all objects, because Censored Zooming does not maintain proportionality. If this happens all objects that cannot be compared to others any more get a red border. Figure 14 shows the result of a Censored Zooming operation. In the left Fluctuation Diagram it is not clear whether there are any non empty bars except the few visible ones. One might guess that in the north of the diagram there are no cases at all. The right diagram shows (after Censored Zooming) that there are actually cases in the north, especially in the north east. Censored Zooming is supported by Flow Diagrams, Barcharts and Fluctuation Diagrams.
Alpha blending works the other way round to Censored Zooming. Every object in the plot will be drawn transparently. So only large clusters of objects can be seen at all after Alpha Blending. Alpha Blending can be performed in all “point plots”.
Both, Censored Zooming and Alpha Blending can be controlled using LEFT and RIGHT cursor keys with ALT.

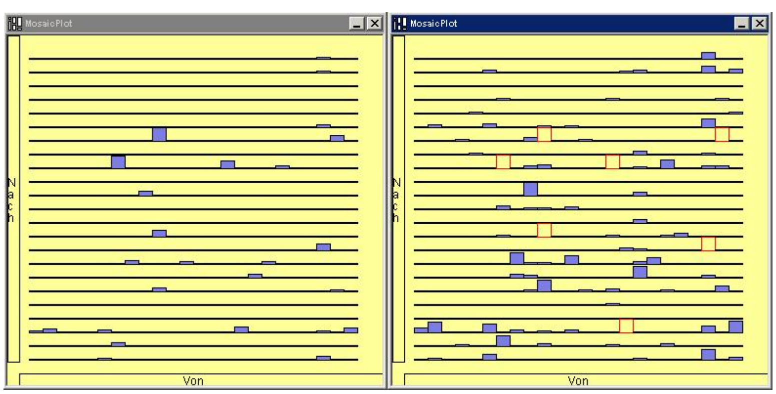

Figure 14: Original Fluctuation Diagram on the left, Censored Zooming result on the right

3.4 Conclusion

VanGogh offers different graphical displays which are useful for exploratory data analysis. The usage of VanGogh is very intuitive and user friendly. Together with its interactive features and plots specially designed for visualizing networks, VanGogh is a powerful tool for analysing network data.

4 Visualizing and Analysing Traffic Data

To analyse the Augsburg Traffic Data, choose Traffic Wizard from the File menu. This wizard will guide you through the typical process of opening a dataset containing traffic data. In the first step select `Flows2001.txt` as the Dataset to be opened. Choose Next, and take the defaults in the nodes and coordinates. As a weight variable, choose Anz and as an image file `CityImage.txt`. In the next step select all possible graphical displays that are available in this dialog. Then choose finish, and VanGogh should look like Fig. 15. The Flow Diagram on the right is empty, as no cases are selected. The Barcharts, the Fluctuation Diagram and the Flow Diagram include the weight variable specified in the wizard.

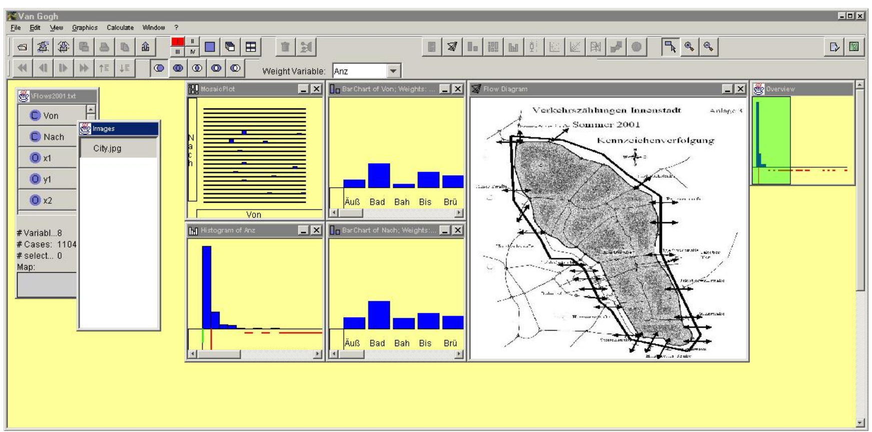

Figure 15: VanGogh after the Traffic Wizard

4.1 Overview

The first thing to do, is to get an overview of the network itself. Go to the Flow Diagram on the right and choose Show all Flows. Now all flows are shown in the Flow Diagram (see Fig. 16).
It can be easily seen, that some large flows go directly through the city, while most of the flows (and the larger ones) are located either in the north or the south of the city. The large almost horizontal flows in the middle of Augsburg are streets that

come from the east of Augsburg and head towards the main station. It is known, that these streets are always very busy. If you knew Augsburg in 2001 also the large amount of traffic in the south is no surprise. There is one place named “Rotes Tor” (see Fig. 17) a very old city gate. The street next to it is a very busy one with a high risk of traffic jam.

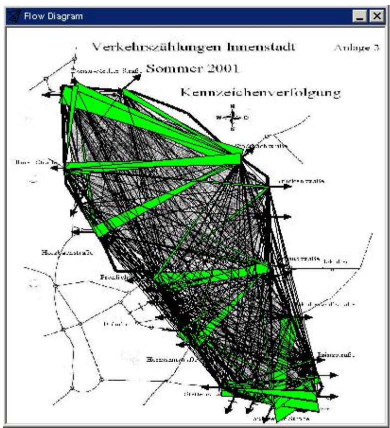

Figure 16: All flows visible in a Flow Diagram

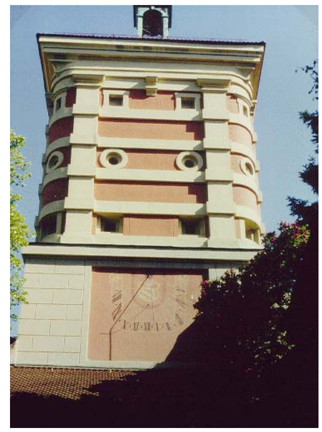

Figure 17: “Rotes Tor” in Augsburg

4.2 Questions

In this part I will give typical questions that were asked by the dataset owners, the administration of the City of Augsburg. I will give answers to these questions and I will show, which graphical displays available in VanGogh are useful to visualize these results.

In which Street do the largest amount of Cars enter/ leave?

This question can be answered in different ways. I will concentrate on the question, where most cars enter.
The easiest thing to do, is to look at Barcharts. Select the Barchart of Von. Press CTRL-A to select all cases and then sort the categories in this Barchart. The resulting image can be seen in Fig. 18. It can be easily seen by querying the largest bar, that at Haunstetter_Str. most of the cars are entering Augsburg. The second and third largest flows of cars entering Augsburg are Stadtbachstr. and Friedberger_Str.. Selecting these tree bars leads to the Flow Diagram shown in Fig. 19.
With Barcharts it is also possible to answer questions like: Where are all the cars heading to, which enter the city at Haunstetter_Str.. Again, go to the Barchart of Von, select the bar of Haunstetter_Str., go to the Barchart of Nach and sort the bars according to the abolute selected values. As you see, most of the cars entering the city at Haunstetter_Str., leave the city at Friedberger_Str.. That is no surprise, as many people working at “Siemens” in Augsburg have to take this route to drive home.

The main question of this section can also be answered, by just looking at a Flow Diagram alone. Select Show Nodes, compare the size of all bars and query the largest one (see Fig. 20). As you will see (with zooming), it will again be Haunstetter_Str..

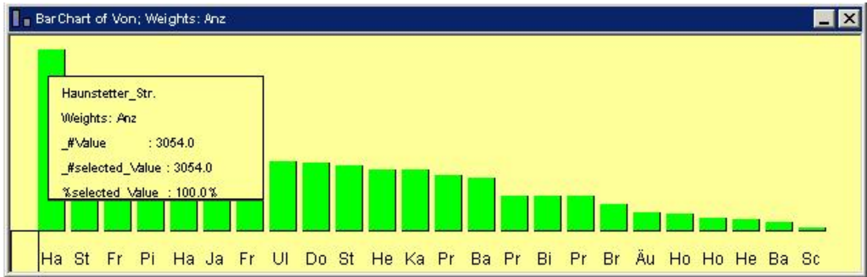

Figure 18: Barchart of Von

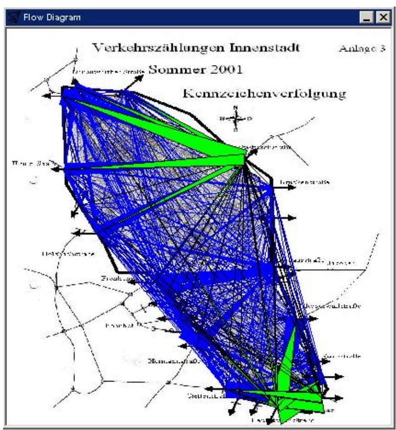

Figure 19: Flow Diagram with largest entering streets selected

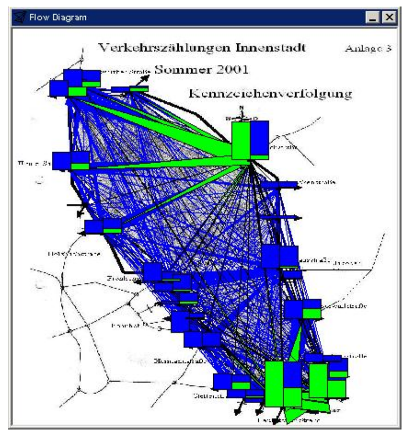

Figure 20: Flow Diagram with visible bars

Which Route has the highest Traffic Density? Are there any Outliers?

This question cannot be answered by a Flow Diagram, as there is too much overlapping. You might guess, that probably one of the flows in the south may be the largest one.
For answering this question, we move our attention to the Fluctuation Diagram (see Fig. 21). We see that there are a lot of relatively small bars and only two really large ones. Querying these two bars shows, that the largest flow is that from Haunstetter_Str. to Friedberger_Str.. Selecting this bar and looking at the Flow Diagram confirms, what we guessed beforehand.
The same as above can be done by using a histogram. Scroll to the largest values in the histogram and select them. The resulting Flow Diagram will be the same.
A closer look to the histogram shows that in about $\frac{3}{4}$ (exact: 796) of the flows less than 18 cars were observed (see Fig. 22).

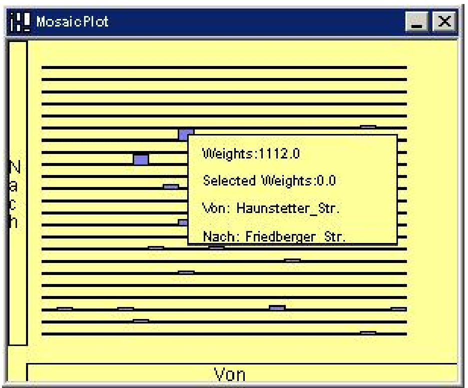

Figure 21: Fluctuation Diagram

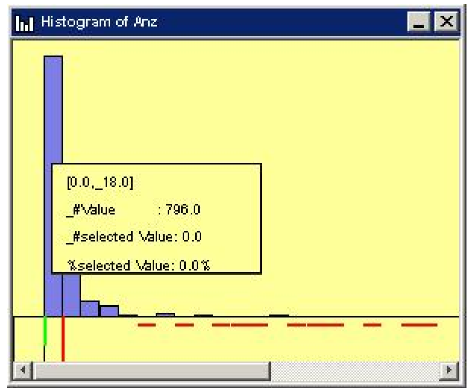

Figure 22: Histogram of Anz

Are there preferred Directions of Flows? Are there any One Way Streets?

One way streets can be detected easily with Barcharts. Every category appearing in a Barchart of Von or of Nach and not appearing in the other one is a street, where either cars only enter or cars only leave the city. Three One Way Streets can be detected: Halderstr., Hochfeldstr. (only in Von) and Jakoberstr. (only in Nach).
It is much more difficult to detect preferred directions of flows with standard plots. For this purpose VanGogh provides the SymmDiagram. In a SymmDiagramm, this question can be answered easily (See Fig 23). The further the distance of a point is to the zero line, the more different are the weights of the flow directions. Points lying on the left boundary envelope of this plot show the One Way streets. Selecting the flows with the highest differences gives the Flow Diagram in Fig. 24. The flow with the largest difference is the flow from Halderstr. to Hermannstr. with a difference of 459 cars.

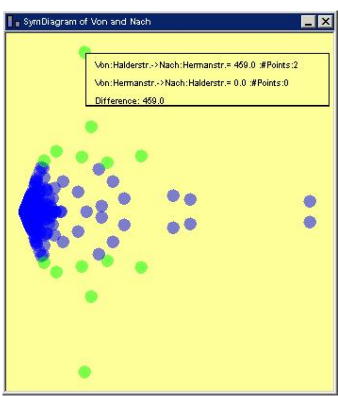

Figure 23: SymmDiagramm with greatest differences selected

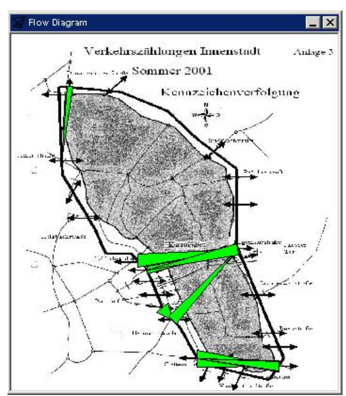

Figure 24: Flow Diagram showing preferred flow directions

Are there Differences between morning and afternoon?

This can be answered by creating a Barchart of Wann(when) and either selecting 0 or 1 and comparing all different plots.
About half of the cars were observed in the morning, the other half in the afternoon. Comparing the plots with 0 selected, shows that in Barcharts of Von and Nach almost in every bar half of the points are selected. This can be seen more easily, using Spineplots (see Fig. 25). A Mosaicplot of Von and Nach comes to the same conclusion. Comparing Histograms and Flow Diagrams shows as well, that there is only a subtle difference between morning and afternoon.

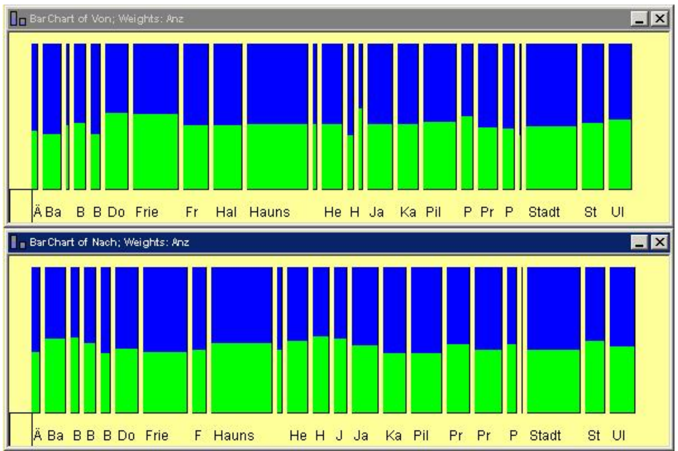

Figure 25: Spineplots of Von and Nach; 0 selected

Are there Differences between 2001 and 2002?

This question cannot be answered with the datafile `Flows2001.txt` alone. Go to another virtual desktop and open the dataset `Flows2002.txt` with the traffic wizard described in section 4. This is necessary, because in this dataset only the south of the city was observed and the observation points differ in some cases.
The resulting Flow Diagram is shown in Fig. 26. As you see, only streets in the south of Augsburg were observed. One obvious difference is visible at once: The flows from Haunstetter_Str. to Friedberger_Str. are not the largest flows any more, they are dominated by flows more in the western part of the city.The largest flow is now Hermannstr. to Fuggerstr. (see Fig. 27) with 740 cars, while Haunstetter_Str. to Friedberger_Str. only has a weight of 107. That Hermannstr. to Fuggerstr. is so large, is no surprise, as all cars driving to the north of the city are counted as leaving the city in Fuggerstr..
The flow from Haunstetter_Str. to Friedberger_Str. has decreased because a bypass around “Rotes Tor” has been completed at the end of 2001 to reduce the traffic density in that specific area. The observation of cars in 2002 measures the effect of that bypass.

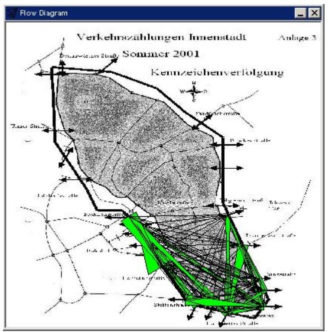

Figure 26: Flow Diagram in 2002

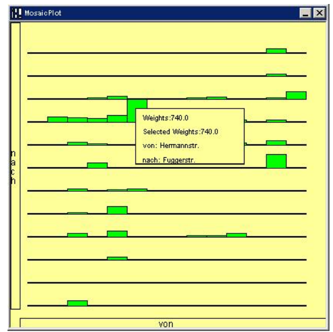

Figure 27: Fluctuation Diagram in 2002

4.3 Results

There are a lot of characteristics in the dataset that can be easily visualized with VanGogh. The owners of the dataset were very happy, that features of the dataset which were known beforehand, could be visualized and presented to people not knowing the dataset as well. Especially the Flow Diagram is a very powerful tool for presenting this kind of data, as it shows the characteristics in a way that everyone (at least everyone knowing Augsburg) can understand them.

5 Conclusion

This article showed, how interactive features can be applied to networks, using VanGogh. I described the analysis of a typical dataset.
There are other datasets, that can be analysed with VanGogh. The city of Augsburg again provided traffic data for urban city districts calculated with a mathematical model. This dataset is much bigger than the one described above (22000 flows). In this dataset interactive selection and hiding of not selected values is even more important, because hardly anything can be seen in the Flow Diagrams, due to overlapping.
Another kind of dataset is for crossroads. In this kind of dataset one is interested in how many cars drive through the crossroad in a certain direction and at different times. For that purpose, the city needs a layout of graphical displays. With VanGogh it is possible to visualize this kind of data, by modifying the data files in an appropriate way.
Another kind of data are trade flows. I analysed the “Oil” Dataset, introduced in [7] as well. This one is very interesting, as it contains a polygon world map that can be used as an additional visualization of the flows.
Very new is the application in sports, especially in Volleyball (other sports as tennis can be analysed with VanGogh as well). Transition matrices that appear in the modelling of a game can be visualized using a network and VanGogh. It is even possible to show the game field as an underlying map or image behind the game moves that are represented by flows.

The datasets described above are networks with natural x and y coordinates. Therefore VanGogh does not need to lay the network nodes using layout algorithms described in the literature (see [8]). These kind of networks and more complex ones will be a topic of further research.

References

[1] Godtliebsen, F., Marron J. S, Chaudhuri P.(2002). Significance in Scale Space for Bivariate Density Estimation. Journal of Computational and Graphical Statistics, 11, 1-22.

[2] Hofmann, H. (2000). Exploring categorical data: interactive mosaic plots. Metrika 51(1): 11-26.

[3] Inselberg A. (1999). Dont panic ... just do it in parallel. Computional Statistics, 14, pp. 53-77.

[4] Scott D. (1992). Multivariate Density Estimation. Theory, Practice and Visualization. John Wiley & Sons, Inc

[5] Ullenbohm, C. (2003) . Java ist auch nur eine Insel, Galileo Press GmbH, Bonn 2003.

[6] Unwin A. (1999). Requirements for interactive graphics software for exploratory data analysis. Computational Statistics 14:7-22

[7] Unwin, A., Sloan, B., Wills, G. (1989), Interactive Graphical Methods for Trade Flows. in New Techniques and Technologies for Statistics. 1992. Bonn: Office for Official Publications of the EC.

[8] Wills, G. (1999) . NicheWorks - Interactive Visualization of very large Graphs. Journal of Computational and Graphical Statistics, 8(2):190-212.

A Example Dataset Format

The dataset must be a text file, cases are to be separated by tabs. VanGogh does not support missing values and spacings in variable names or category names are not supported. The following extract is the first part of “Augsburg Traffic Data” in file `Flows2001.txt`.

Von	Nach	x1	y1	x2	y2	Anz	Wann
Aeussere_Uferstr	Aeussere_Uferstr	220	873	220	873	12	0
Aeussere_Uferstr	Aeussere_Uferstr	220	873	220	873	12	0
Badstr.	Aeussere_Uferstr	169	501	220	873	0	0
Badstr.	Aeussere_Uferstr	169	501	220	873	13	1
Bahnhofstr.	Aeussere_Uferstr	324	277	220	873	0	0
Bahnhofstr.	Aeussere_Uferstr	324	277	220	873	0	1
Bismarckstr.	Aeussere_Uferstr	436	71	220	873	5	0
Bismarckstr.	Aeussere_Uferstr	436	71	220	873	4	1
Brueckenstrstr.	Aeussere_Uferstr	499	619	220	873	2	0
...	...	...	...	...	...	...	...

B Example Image File Format

The Image File must be a standard text document, with tab as separator. An image file consists five entries. The first one is a pointer to an existing `.jpg` file, the last four entries give the coordinates of the upper left and lower right points (normally, 0,0,imagewidth, imageheight are good values). These values are essential, if the coordinate variables in the dataset are coordinates that do not match with the pixel size of the image.
The following file is `CityImage.txt` from the “Augsburg Traffic Data”.

City.jpg
0
0
720
1037

C Example Map File Format

Maps come in different file formats. The maps that can be imported into VanGogh contain several polygons, separated by empty lines. A polygon consists of a header with its name, a label "Polygon" and the number of points that define the polygon. Again the file must be simple text, separated by tabs.
The following extract is from the "Oil Dataset", file `Poly.txt`.

1000	Polygon	8
7.7567978	6.8837285	
7.5755286	6.6183047	
7.5302114	6.4855928	
7.3942599	6.3528810	
7.5755286	6.3528810	
7.9833837	6.6625423	
7.9833837	6.8394914	
7.7567978	6.8837285	
2000	Polygon	10
7.3942599	6.3528810	
7.5755286	6.3528810	
7.5755286	6.2201691	
...	...	