DOI: http://dx.doi.org/10.7551/978-0-262-33027-5-ch068

Omnigram Explorer: A Simple Interactive Tool for the Initial Exploration of Complex Systems

Tim Taylor, Alan Dorin and Kevin Korb

Faculty of Information Technology, Monash University, VIC 3800, Australia
tim@tim-taylor.com alan.dorin@monash.edu kbkorb@gmail.com

Abstract

We describe the design of Omnigram Explorer (OMG), an open-source tool for the interactive exploration of relationships between variables in a complex system. OMG is designed to help researchers gain a holistic, qualitative understanding of the relationships between variables in their data at a preliminary stage of analysis; such exploration might highlight interactions that warrant further quantitative investigation using other tools. We illustrate OMG's use on real-world data, and also describe its potential as a tool for communication to non-specialists.

Introduction

When working with models of complex systems, visualisation tools can be invaluable in helping researchers gain an understanding of the system's behaviour. Visualisation can be used at all stages of the research process, from the earliest stages of design through to the eventual communication of results to a variety of different audiences (Dorin and Geard, 2014; Grimm, 2002).

The work reported in this paper was developed in the context of a collaboration between the authors and a group of epidemiologists at the University of Melbourne.[1] The epidemiologists use a variety of different modelling techniques in their work, ranging from agent-based models and other simulations to more traditional mathematical approaches. Their models often involve several dozen independent and dependent variables.

We wished to develop an interactive tool that would allow our colleagues to gain a quick qualitative understanding of their models at the initial stages of analysis, and highlight features to be investigated in more detail. The resulting tool, named Omnigram Explorer (which we abbreviate as OMG), is described in this paper. In the following sections we review relevant previous work, describe the principles that drove OMG's design, describe the tool's main features, and give examples of using OMG on real-world data.[2]

[1] See Acknowledgements section for details.

[2] Various supplementary materials are indicated in the paper. These are available at `http://www.tim-taylor.com/omnigram/ecal2015/`.

Omnigram Explorer is a free, open-source tool developed in Processing.[3] The source code, binary executables (for Windows, Mac and Linux), documentation and related materials are available at `http://www.tim-taylor.com/omnigram`.

Previous work

Traditional approaches to visualising the relationships among multiple variables include the Scatter Plot Matrix (SPLOM) and Parallel Coordinates (Heer et al., 2010). These are widely used, although neither is without problems. For example, SPLOM visualisations focus on pairwise relationships between variables (Kosara et al., 2003), and the effectiveness of Parallel Coordinates can greatly depend on factors such as the linear order in which data dimensions are plotted (Zhang et al., 2012).

In addition to representation, *interaction* is an increasingly important aspect of information visualisation systems (Ward and Yang, 2004; Yi et al., 2007). For the kinds of visualisation problems we wished to address in our project, relevant and interesting early work was produced by Spence and Tweedie (1998). They developed a tool called *Attribute Explorer* in which multivariate data was presented as a set of histograms, one for each variable. Each histogram displayed the distribution of values in the data-set for its associated variable. The user could select subsets of data by adjusting a slider under a histogram. The subset of data points so selected was represented in the other histograms using a specific highlight colour (a technique known as *linking and brushing*).

The tool described in the current paper, Omnigram Explorer, took inspiration from Spence and Tweedie's work as a starting point. We added a variety of novel and principled extensions (as described in the following sections), and have made it available as a free, open-source tool.

Design principles

An effective data visualisation should provide an intuitive way for the user to gain insight into the organisation of that

[3] `http://processing.org`

Tim Taylor, Alan Dorin, Kevin Korb (2015) Omnigram Explorer: A Simple Interactive Tool for the Initial Exploration of Complex Systems. Proceedings of the European Conference on Artificial Life 2015, pp. 381-388

data. Effective interaction in an information visualisation should allow the user to develop a mental model of correlations and relationships in the data (Kosara et al., 2003).

A guiding principle behind the development of Omnigram Explorer was to leverage properties of the human visual system to allow complex information to be easily processed by the user. In particular, we used several Gestalt grouping principles[4] in the design of the system to allow it to convey a large amount of information about correlations in the data in a deceptively simple way, and to allow users to focus on some variables and not on others. We will expand upon these features in the following sections.

Features

General Usage

OMG is a tool for visualising pre-recorded data of parameter settings and output values from a set of runs of a target model.

At start-up, the user is prompted to select a model definition file that describes the system's variables and also points to a data file. The data file contains the samples of values for each variable. OMG loads this data (or a random subset of it if so instructed) and presents it to the user in graphical form for interactive exploration. OMG is agnostic about the particular form of the system and its variables: variables may be designated as inputs or outputs if desired, but this is not required. The model definition file may also contain information about causal relationships between variables, which can then be visualised to provide hints to the OMG user of interesting relationships to explore. Information about causal relationships might have been generated from existing knowledge of the system or from tools such as Bayesian causal network discovery software.[5]

As described in the following sections, OMG provides various modes of interaction that allow the user to explore the relationships in the system in different ways.

User Interface and Nodes

A general view of the OMG interface is shown in Figure 1.[6]

The most important component of the user interface is the *node*, which is an interactive, graphical representation of the data associated with a particular variable in the system. The detailed user interface of a single node is shown in Figure 2. A node displays a histogram showing the distribution of samples of a particular variable observed in the data read in from the data file.

The node's histogram always shows the distribution of *all* samples read in from the file. However, a subset of these samples may be highlighted in different colours, either (for focus nodes) to indicate that they lie within a range of values selected with the range selector widget by the user, or (for non-focus nodes) to represent brushing in response to focus nodes (see later for further details of focus nodes and brushing). A small circle is drawn below the histogram to indicate the position of the median (or mean) value of the selected samples.

By showing the selected samples highlighted against non-selected samples, the node therefore provides *context* and *guidance* for later exploration; this was one of the guiding principles of *Attribute Explorer* (Spence and Tweedie, 1998). We use the term *omnigram* to refer to the simultaneous visualisation of histograms of all of a system's parameters; an omnigram extends these principles of context and guidance from a single parameter to the whole system.

Modes of Interaction

There are four basic modes of interaction in OMG: *Single Node Brushing*, *Multi Node Brushing*, *Omnibrushing* and *Sample View*.

Single Node Brushing In this mode, only one node can have the focus at any one time. Clicking on the histogram area of a node gives it the focus (shown by the red focus indicator in the top left of the node), and removes the focus from any other node.

The range selector of the focus node can be adjusted to select a subset of samples from the overall distribution (i.e. a subset of bins from the histogram). The handles at each end of the selector can be dragged left or right to change the upper and lower limits of the selected range, and the main selector bar can be dragged left or right to shift the whole range up or down. The bins within the selected range are shown in dark red, and the other bins in the focus node are shown in white.

When a range of samples has been selected in the focus node in this way, all of the other nodes are updated to show where the same samples lie in the distributions of the other variables in the model. The matching samples are shown in dark blue. This is OMG's implementation of the *linking and brushing* technique mentioned earlier.

The real power of linking and brushing in OMG becomes apparent when you interactively change the range selection in the focus node, and watch the resulting changes in the other nodes. In particular, selecting a fairly small subset of values with the range selector in the focus node (i.e. having a short range selector bar), dragging the bar from left to right and back, and watching how the change in values of the focus node is associated with changes in the other nodes, is a very effective technique.

The human visual system is very attuned to noticing multiple objects moving in the same direction (the Gestalt principles of Continuity and Common Fate). OMG makes use of

[4] See, e.g., (Wolfe et al., 2009).

[5] E.g. CaMML (Korb and Nicholson, 2011).

[6] The screen-shots in this section show OMG using data from the standard Auto MPG data-set downloaded from the UCI Machine Learning Repository (Lichman, 2013) and with causal links generated by CaMML (Korb and Nicholson, 2011).

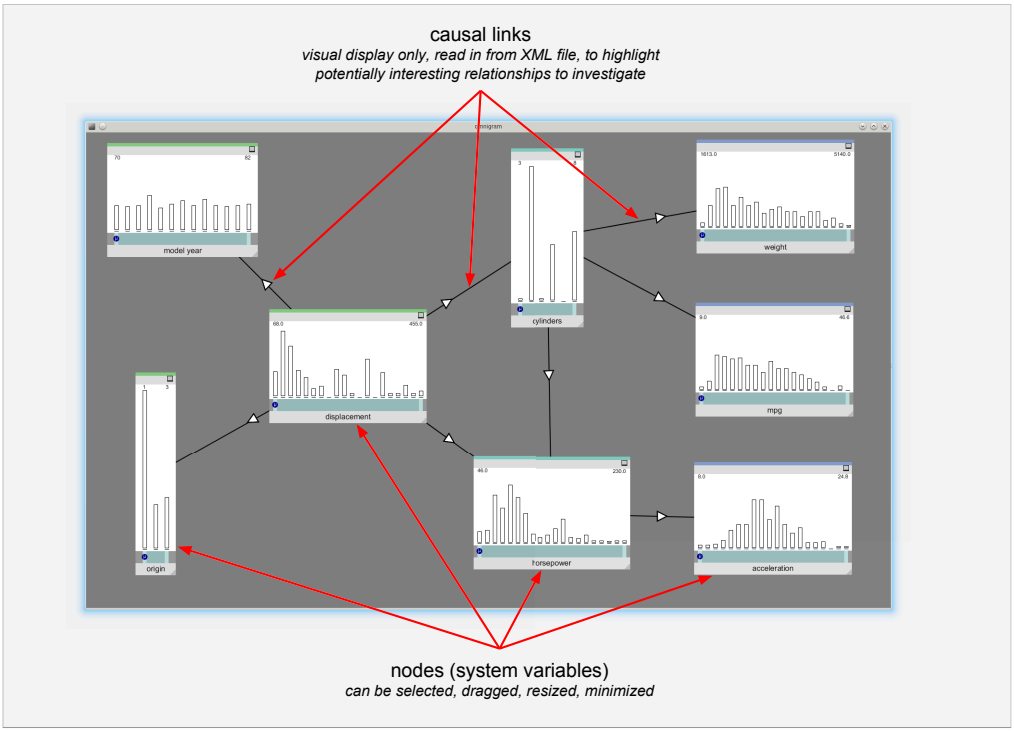

Figure 1: The Omnigram Explorer user interface.

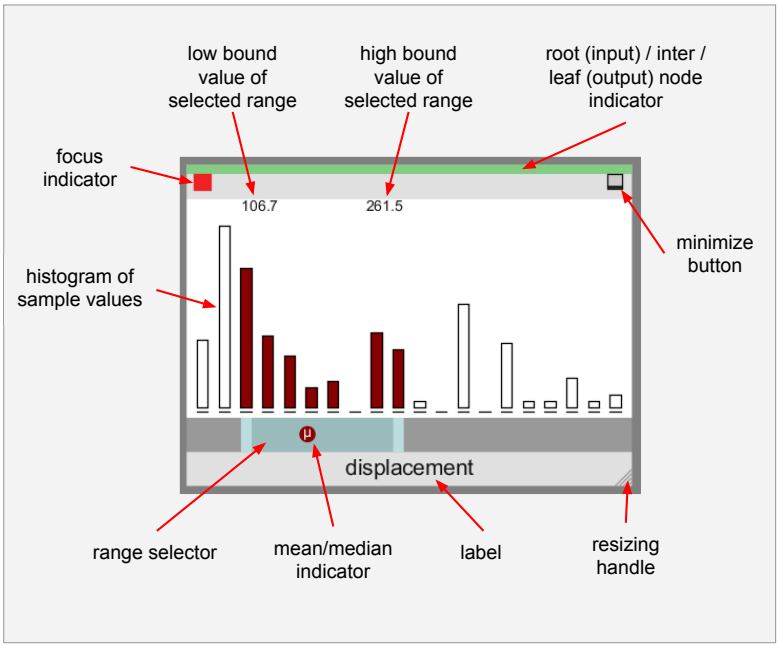

Figure 2: Detail of user interface for a Node.

Tim Taylor, Alan Dorin, Kevin Korb (2015) Omnigram Explorer: A Simple Interactive Tool for the Initial Exploration of Complex Systems. Proceedings of the European Conference on Artificial Life 2015, pp. 381-388

this fact to allow the user to easily spot correlations between many different variables in *Single Node Brushing* mode. Because these principles rely on movement, it is difficult to convey the power of this technique in writing or still images. By interacting with the system (moving the range selector bar) and watching how the selected set of samples change in each histogram as a result, it is very easy to see how changes in one variable are correlated to changes in all of the other variables in the system. Feedback from informal focus group meetings[7] suggests that this is effective even with 30–40 nodes on-screen at once. A demonstration video is provided in the supplementary materials.

Multi Node Brushing *Multi Node Brushing* is an extension to *Single Node Brushing* that provides information about sensitivity of the model. An example screen-shot of *Multi Node Brushing* mode is shown in Figure 3.

In this mode, more than one node can have the focus at the same time. In the example shown in the figure, the three leftmost nodes all have the focus (indicated by the red squares at the top left of each node). The range selectors in each of the focus nodes can be adjusted just like in *Single Node Brushing* mode, to select a subset of samples.

The information displayed in non-focus nodes is more detailed than in *Single Node Brushing* mode. As before, dark blue is used to indicate the location of samples that have been selected by the range selection in the focus nodes. In *Multi Node Brushing*, OMG looks at the conjunction of the range selections made on the focus nodes: to take an example from Figure 3, a blue patch on a histogram bin in the *horsepower* node indicates that there is a sample that has that horsepower value that also lies within the selected range of *model year* and lies within the selected range of *displacement* and lies within the selected range of *origin*.

You will notice from the figure that non-focus nodes also show other colours. A light green patch indicates that there is a sample that has that value that also lies within the selected range of values for *all but one* of the focus nodes. So, in this example, a light green patch on a histogram bin in the *horsepower* node might indicate that there is a sample that has that horsepower value that also lies within the selected range of *model year* and lies within the selected range of *displacement*, but not within the selected range of *origin*. It might also indicate that the origin and model year ranges were respected, but not the displacement, etc. That is, the light green colour indicates that one of the focus node ranges was not respected, but it does not tell you which one.

Similarly, a light red patch indicates that two of the focus node ranges were not respected, and white indicates that three or more were not respected.

The colour therefore gives some indication of the sensitivity of the model to the value ranges chosen. For example, the presence of a light green patch in a non-focus node indicates that by increasing the selected range of just one of the focus nodes, the patch can be turned blue, i.e. it can be made to satisfy the constraints on all of the focus nodes by loosening a single constraint.

Omnibrushing *Omnibrushing* mode works in a somewhat different way to *Single Node Brushing* and *Multi Node Brushing*. Like in *Single Node Brushing*, only one node can have the focus at a time in *Omnibrushing* mode. When a node receives the focus, a different colour is assigned to each of that node's bins. The hue of each bin changes steadily from the leftmost bin to the rightmost bin, with reds on the left changing to greens and blues on the right.

The colours in each of the non-focus nodes are updated to reflect those of the focus node. For each bin in a histogram of a non-focus node, a fraction of the area of the bin is given the same colour as each of the bins in the focus node according to the fraction of samples in that non-focus bin that belong to the bin in the focus node corresponding to that colour. An example of *Omnibrushing* mode is shown in Figure 4; for further examples, see the Example Usage section and Figure 6 below.

Thus, *Omnibrushing* mode can be thought of as a composite *Single Node Brushing* mode, where each bin in the focus node has been given a different colour, and the corresponding brushing of all other nodes has then been superimposed into a single representation.

Sample View *Sample View* mode provides a different way to visualise the data. The overall distribution of values for each node is shown, as with the other modes, but these are shown partially faded out in the background of each node, and are provided just as a reminder of the overall shape of the observed data. The real point of interest in this mode is the display of individual samples from the data file. Each sample is shown as a small coloured circle placed in the position of the histogram bin corresponding to the sample value of that variable. An example screen from *Sample View* mode is shown in Figure 5.

At any one time, a fixed number of samples is shown. By default, OMG will automatically cycle at a moderate speed through different selections of samples from the data. At each step in the cycle, one sample (the one that has been displayed for the longest time) is removed from the currently displayed set, and a new sample replaces it. The user can interactively increase or decrease both the number of samples shown at a time and the speed of cycling. The user may also choose to step forwards and backwards through the different sample selections manually rather than having them updated automatically.

In *Sample View* mode, only a single node can have the focus at any one time. By default, the samples are coloured according to which bin they belong to in the focus node (for

[7] These meetings comprised eight participants (including the software author). More rigorous tests are planned in future work.

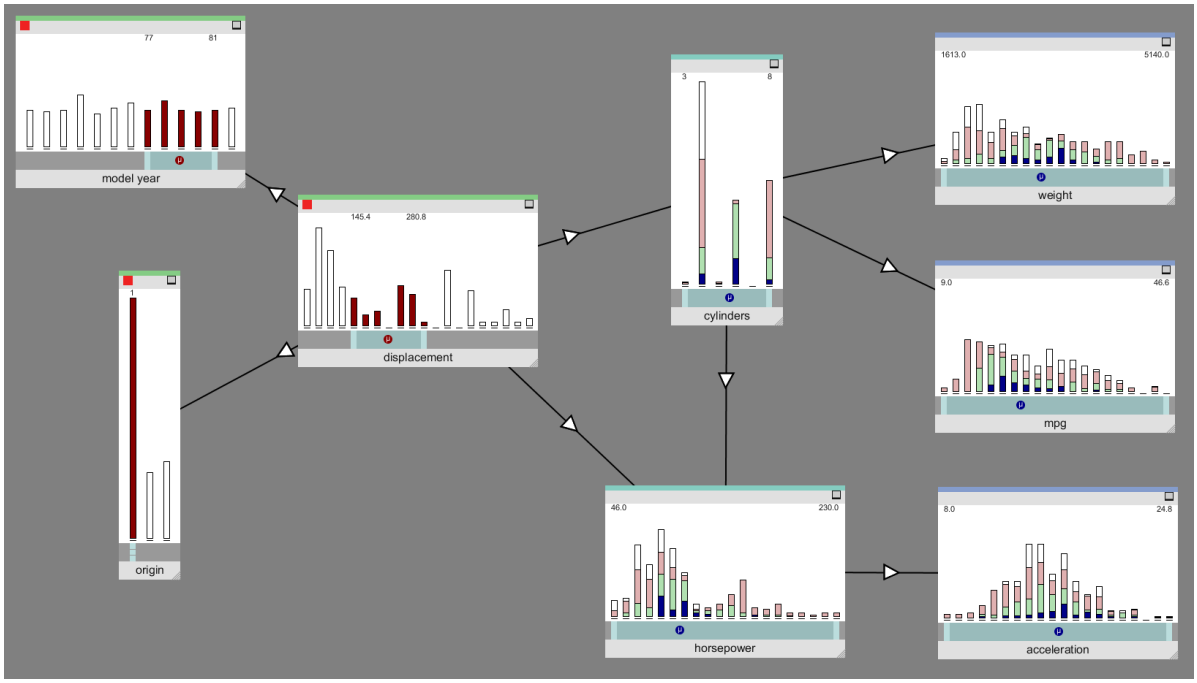

Figure 3: Multi Node Brushing mode (focus nodes: model year, origin, displacement)

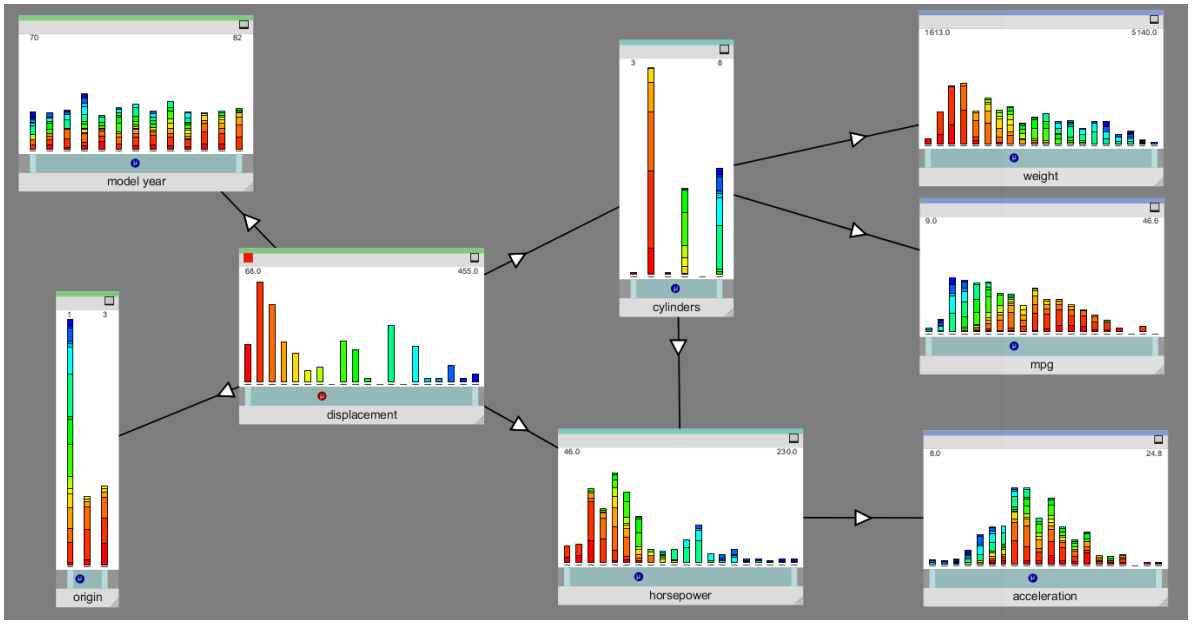

Figure 4: Omnibrushing mode (focus node: displacement)

Tim Taylor, Alan Dorin, Kevin Korb (2015) Omnigram Explorer: A Simple Interactive Tool for the Initial Exploration of Complex Systems. Proceedings of the European Conference on Artificial Life 2015, pp. 381-388

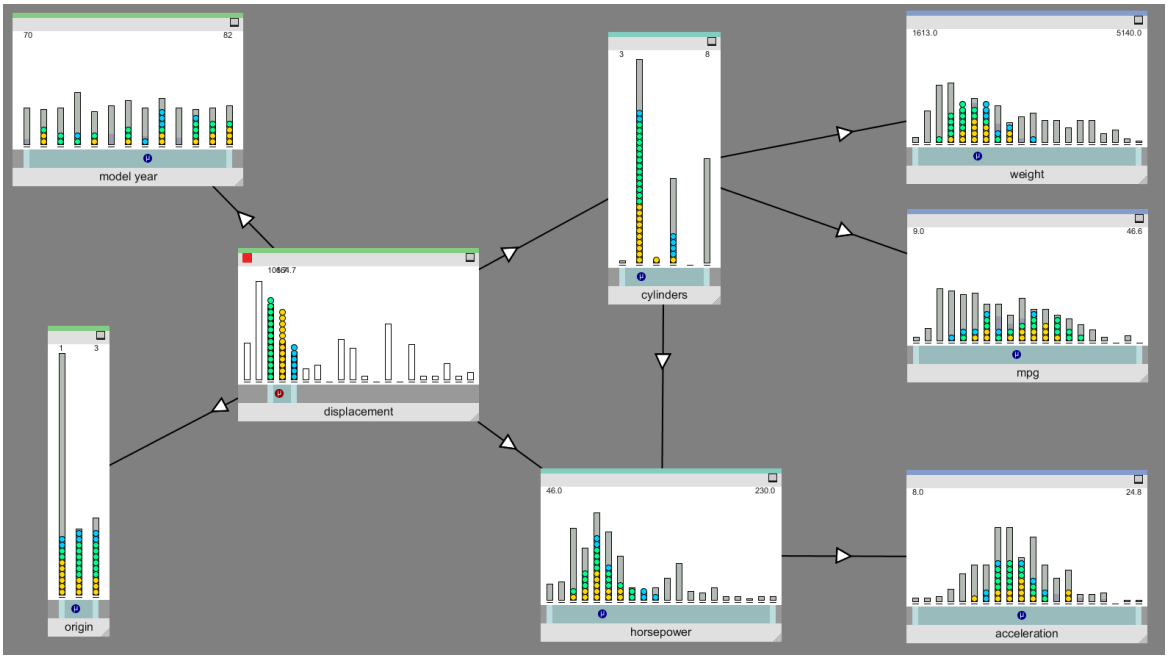

Figure 5: Sample View mode (focus node: displacement)

an example, see Figure 5, where the focus node is the *displacement* node, indicated by the red focus indicator in the top left of the node). The user can quickly see how these samples are distributed in the other nodes by looking at the distribution of colours.

If preferred, users can select a colouring scheme in which samples are coloured at random, irrespective of their bin in the focus mode. This scheme is most useful for studying a small number of samples. The colours allow the user to determine, for each coloured sample in the focus node, the exact bins into which they fall for all other nodes.

Other features

OMG has various other features that we will not describe in detail here. These include the ability to interactively create *brush links* between nodes in *Multi Node Brushing* mode, so that when the range selector on one focus node is moved, the range selector on another linked focus node also moves in a specified way. This allows the user to interactively explore complex relationships between multiple variables.

As mentioned before, the model definition file may also contain information about causal links between nodes, and OMG can display these as arrows between nodes to provide hints about relationships that might be worth exploring.

The initial placement of nodes on the screen is done automatically, and the input file may contain hints about which nodes should be placed where. But regardless of initial placement, all nodes can be dragged to new positions interactively by the user. One situation in which this can be particularly useful is as follows: if nodes have been identified that behave in a similar way (e.g. when dragging the range selector of the focus node), those nodes can be grouped together by dragging them to one part of the screen. The user can then start to partition a complex model into subsets of similarly performing variables by spatially grouping the nodes on screen. Placing similarly-behaving nodes in close spatial proximity allows the user to perceive them as a single group (the Gestalt principle of Proximity), thereby simplifying the process of mental model formation. Other methods to increase or decrease the focus on specific nodes include the ability to interactively resize a node, and to minimise it so that the histogram is completely hidden from view.

Example Usage

As an introductory example of using OMG on real data, we will briefly discuss its use on the standard Heart Disease data-set available from the UCI Machine Learning Repository. We use the reduced Cleveland subset of the data, comprising 297 complete samples, each with 13 input attributes and one disease diagnosis attribute to be predicted. See Table 1 for a summary and (Lichman, 2013) for full details.

The data was loaded into OMG. In *Single Node Brushing* mode, the diagnosis node (*num*) was given the focus, and the range selector bar reduced so that a single bin in the *num* histogram was selected. This narrow range selector bar was then dragged left and right to select different bins in the histogram in rapid succession. The resulting patterns of brushing in the other nodes gave a clear indication of which

Label	Description
age	Age in years
sex	Sex (0=female, 1=male)
cp	Chest pain type (1-4)
trestbps	Resting blood pressure (mmHg)
chol	Serum cholesterol (mg/dl)
fbs	Fasting blood sugar > 120 mg/dl (0=no, 1=yes)
restecg	Resting ECG results (0-2)
thalach	Maximum heart rate achieved (bpm)
exang	Exercise induced angina (0=no, 1=yes)
oldpeak	ST depression induced by exercise
slope	Slope of peak exercise ST segment (1-3)
ca	Num major vessels coloured by fluoroscopy
thal	Summary of heart condition (3=normal)
num	Disease diagnosis (0=no presence)

Table 1: Heart Disease Data-set attributes

Attribute	Apparent Correlation	Pearson Correlation
age	Medium	0.222
sex	Low	0.227
cp	High	0.404
trestbps	Medium	0.160
chol	None	0.066
fbs	None	0.049
restecg	Low	0.184
thalach	-Medium	-0.421
exang	Low	0.392
oldpeak	High	0.501
slope	Medium	0.375
ca	High	0.521
thal	High	0.513

Table 2: Qualitative judgement of apparent correlations of attributes with disease diagnosis using OMG in *Single Node Brushing* mode (column 2) and calculated Pearson correlation coefficients (column 3). See text for details.

nodes were correlated with *num* (and the movement of the "median selected value" indicator under the histograms provided an additional indication). This process is illustrated in a video in the supplementary materials.

To quantify the effectiveness of this procedure, after performing this brushing technique for a few seconds, each node was given a qualitative rating of the degree to which it appeared to be correlated with the *num* node. Correlations were rated as *None*, *Low*, *Medium* and *High*. The results are shown in Table 2, along with the Pearson correlation coefficient of each attribute with *num*, calculated from the raw data. The qualitative assessments of correlation from OMG were then converted to numeric scores under the mapping $\{$*None*=0, *Low*=1, *Medium*=2, *High*=3$\}$, and the Pearson correlation coefficient between these scores and the calculated correlations was calculated. This came to 0.84, which demonstrates that the qualitative impression of the 13 correlations that can be gleaned from a few seconds of using OMG corresponds very well to the correlations calculated using Pearson's correlation coefficient.

A useful next step is to use *Omnibrushing* mode to gain another perspective on these relationships. A screen-shot of using OMG on this data in *Omnibrushing* mode, with the diagnosis node (*num*) selected as the focus node, is shown in Figure 6. Because *num* (the rightmost node in the figure) has the focus, each bin in that node is drawn using a different colour, ranging from reds on the left to blues on the right. The distributions of these samples in each of the other nodes is plotted using matching colours.

A wealth of information is revealed in Figure 6, but just a few points will be highlighted here. It is immediately apparent that this data shows that heart disease (indicated by *num*$>$0) is much more prevalent in males (*sex*=1) than females: look at the relative fraction of each bin in the *sex* node filled by healthy (red) samples. The prevalence of heart disease appears to be considerably higher in older people: the bins on the right-hand side of the *age* node, starting from the tallest bin (which corresponds to age 55), each show around 50% or more of samples with a disease diagnosis (non-red). On the other hand, the serum cholesterol level (*chol* node) does not appear to be strongly correlated, as each bin has roughly the same fraction of samples from each diagnosis.

These two initial investigations have suggested some relationships that might be worth exploring further. *Multi Node Brushing* mode can be used to begin to explore more complex relationships between multiple nodes. And *Sample View* mode can provide more fine-level investigation of patterns from individual samples in the data. Taken together, these investigations allow one to build up a useful qualitative overview of the data. These initial investigations can then suggest interesting questions for further quantitative analysis using other tools.

Other uses of OMG

Our colleagues at the University of Melbourne have started to use OMG to explore some of their data-sets. In addition to the kind of exploratory analysis described above, they have found it useful in other ways. Firstly, it is useful for quickly checking the independence of factors in a model. Secondly, it is suitable as a tool for communicating the behaviour of complex models to non-specialists such as government policy makers. Features that make it well suited for this task include the fact that data is presented as bar charts, which are easily understood. In addition, the scientists can deliver the OMG tool and allow non-specialists to interact with it using a previously generated data-set, thereby providing some constraints on the interactions (i.e. the non-specialists cannot break the model by doing something unexpected).

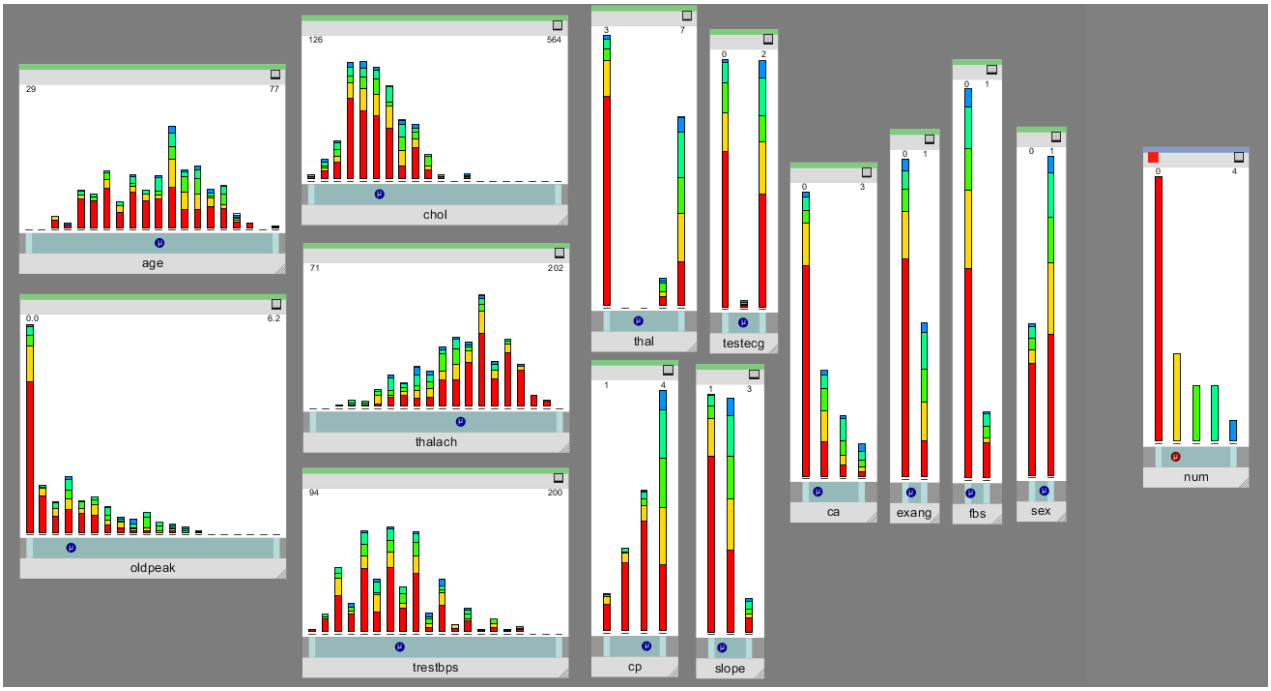

Figure 6: Example using Cleveland Heart Disease Data-set in *Omnibrushing* mode (focus node: num)

In future work we plan to extend OMG's capabilities in other ways. In particular, we are investigating connecting OMG to live processes in the form of simulations and Bayesian nets for live interaction and guided exploration.

Conclusion

We have described Omnigram Explorer, a tool for early-stage, qualitative exploration of data from simulations and other models of complex systems. The tool leverages properties of the human visual system to allow users to quickly spot relationships in the data through interactive manipulation. It has also been suggested that OMG has desirable properties for a tool for the communication of model dynamics to non-specialists. The tool is open-source, and we encourage readers to use it, and to extend it.

Acknowledgements

Omnigram Explorer was developed by the authors in collaboration with Jodie McVernon, James McCaw, Nicholas Geard and Patricia Campbell at the Melbourne School of Population and Global Health at the University of Melbourne. The project was funded by ARC grant number DP110101758.

References

Dorin, A. and Geard, N. (2014). The practice of agent-based model visualisation. *Artificial Life*, 20(2):271–289.

Grimm, V. (2002). Visual debugging: A way of analyzing, understanding and communicating bottom-up simulation models in ecology. *Natural Resource Modeling*, 15(1):23–38.

Heer, J., Bostock, M., and Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6):59–67.

Korb, K. B. and Nicholson, A. E. (2011). *Bayesian Artificial Intelligence*. CRC Press, 2nd edition.

Kosara, R., Hauser, H., and Gresh, D. L. (2003). An interaction view on information visualization. In *State-of-the-Art Proceedings of EUROGRAPHICS*, pages 123–137.

Lichman, M. (2013). UCI machine learning repository. http://archive.ics.uci.edu/ml. University of California, Irvine, School of Information and Computer Sciences.

Spence, R. and Tweedie, L. (1998). The attribute explorer: information sythesis via exploration. *Interacting with Computers*, 11:137–146.

Ward, M. and Yang, J. (2004). Interaction Spaces in Data and Information Visualization. In Deussen, O., Hansen, C., Keim, D., and Saupe, D., editors, *Eurographics / IEEE VGTC Symposium on Visualization*. The Eurographics Association.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzy, R. L., and Lederman, S. J. (2009). *Sensation and Perception*. Sinauer Associates, 2nd edition.

Yi, J. S., ah Kang, Y., Stasko, J. T., and Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224–1231.

Zhang, Z., McDonnell, K. T., and Mueller, K. (2012). A network-based interface for the exploration of high-dimensional data spaces. In *IEEE Pacific Visualization Symposium*, pages 17–24. IEEE.

Tim Taylor, Alan Dorin, Kevin Korb (2015) Omnigram Explorer: A Simple Interactive Tool for the Initial Exploration of Complex Systems. Proceedings of the European Conference on Artificial Life 2015, pp. 381-388