



Cognitive measurements of graph aesthetics

Colin Ware¹
 Helen Purchase²
 Linda Colpoys²
 Matthew McGill²

*Data Visualization Research Lab, Durham, U.S.A.;
 School of Computer Science and Electrical
 Engineering, The University of Queensland, St
 Lucia, 4072, Brisbane, Australia*

Correspondence:
 Colin Ware, Data Visualization Research Lab,
 Center for Coastal and Ocean Mapping,
 University of New Hampshire, 24 Colovos
 Road, Durham, NH 03824, U.S.A.
 Tel: 603 862 1138; Fax: 603 862 0839;
 E-mail: colinw@cisunix.unh.edu

Abstract

A large class of diagrams can be informally characterized as node–link diagrams. Typically nodes represent entities, and links represent relationships between them. The discipline of graph drawing is concerned with methods for drawing abstract versions of such diagrams. At the foundation of the discipline are a set of graph aesthetics (rules for graph layout) that, it is assumed, will produce graphs that can be clearly understood. Examples of aesthetics include minimizing edge crossings and minimizing the sum of the lengths of the edges. However, with a few notable exceptions, these aesthetics are taken as axiomatic, and have not been empirically tested. We argue that human pattern perception can tell us much that is relevant to the study of graph aesthetics including providing a more detailed understanding of aesthetics and suggesting new ones. In particular, we find the importance of good continuity (ie keeping multi-edge paths as straight as possible) has been neglected. We introduce a methodology for evaluating the cognitive cost of graph aesthetics and we apply it to the task of finding the shortest paths in spring layout graphs. The results suggest that after the length of the path the two most important factors are continuity and edge crossings, and we provide cognitive cost estimates for these parameters. Another important factor is the number of branches emanating from nodes on the path.

Information Visualization (2002) 00, 00–00. doi:10.1057/palgrave.ivs.9500013

Keywords: Graph layout; aesthetics; cognitive modeling; network visualization

Introduction

Many diagrams use boxes or circles to show entities and linking lines drawn between them to represent relationships between the entities. An example is a software structure chart where subroutines are shown as boxes and the linking lines show which subroutines call each other. Other examples of node–link diagrams include organization charts, data flow diagrams, flow charts and the set of diagrams encompassed by the unified modeling language (UML).¹

Graphs are abstractions of node–link diagrams studied by mathematicians, and the discipline of graph drawing has developed to investigate methods for laying out graphs according to a set of aesthetic principles that are assumed to improve readability.² Many of the principles and methods of graph drawing can be applied to the broad area of drawing effective node–link diagrams and thus have the potential for wide ranging applicability.

Some commonly applied graph layout aesthetics are the minimization of edge crossings, displaying symmetry of graph structure, minimizing bends in edges. However, despite the considerable effort that has gone into constructing algorithms to optimize according to these aesthetics, surprisingly little work has gone into the empirical validation of these aesthetic principles. A notable exception is the pioneering work of Purchase³ which compared task performance on five pairs of graphs that were designed to differ according to the aesthetic principles of edge bends, edge crosses, maximizing the minimum angle, orthogonality and symmetry. This study

concluded that edge crossings was 'by far the most important aesthetic' when compared with the other four aesthetic criteria. However, the crossings conclusion itself was based on only two hand drawn graphs, one with many crossings and one with few. A visible inspection suggests a number of confounding factors: in the version with more crossings the geometric path lengths appear to have been longer, there is a less even distribution of nodes, and the paths appear to be less continuous than the alternative (ie: their angular deviation from a straight line is greater than the version with less crossings).

The results of the initial Purchase study suggest that there is scope for more experimental work in this area. Experiments that are more controlled, and which consider a greater range of aesthetics *within the same graph drawings* (rather than merely between two drawings representing the extremes of each aesthetic) are more likely to indicate the relative importance of the differing aesthetics.

Perceptual and cognitive issues in graph aesthetics

Much of what we know about human pattern perception originates with the early work of the Gestalt psychologists.^{2,4} They produced a set of gestalt laws that determine whether we see something as a 'figure' as opposed to 'ground'. One gestalt law that is especially relevant to graph drawing is the principle of good continuation as illustrated in Figure 1. This suggests that we will more easily see smooth continuous contours than jagged ones. It also suggests that we will be able to interpret Manhattan layout graphs (with only vertical and horizontal lines) less easily than graphs that use smoothly curved lines, because the continuous lines are more likely to 'pop out' as perceptually complete objects.⁵

More recent results from neurophysiology bear on the issue of edge crossings in graphs. Rapid early-stage neural processing is thought to underlie the tendency of certain simple forms to 'pop out' from their surroundings.⁶ At the early stage of visual processing every part of the visual field is processed in parallel with a set of orientation tuned neurons that respond preferentially to bars with particular orientations. These detectors are only coarsely tuned, roughly within $\pm 30^\circ$.⁷ This suggests that when

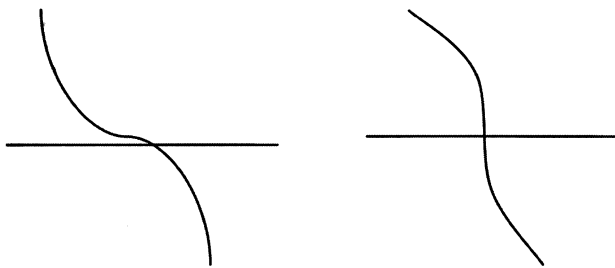


Figure 1 The pattern on the left (a) is perceived as a curved line overlapping a rectangle (b) rather than as shown in (c).

edges cross at acute angles, they will be more likely to cause visual confusion when rapid interpretation is important, than when they cross close to 90° . See Figure 2.

Cell assemblies are responsible for capturing entire contours, and recent research has cast light on how they may function, thereby giving the principle of good continuation a more robust empirical and theoretical foundation. A set of experiments by Field *et al*⁸ quantified the good continuation principle, and their stimulus design is illustrated in Figure 3. These studies showed that when contour segments are aligned along a smooth curve they are easier to see than when they are not, and the ease with which the contour can be seen is a direct function of the continuity of the path. According to the emerging theory, points detected along a curved edge are linked together by a neural mechanism that allows edges as a whole to be perceived through a set of local rules that link the output of independent feature detectors.⁸ Whether individual cell tuning properties or cell assemblies can best account for how easily people can perceive paths in graphs is beyond current theory, but it is likely that both levels of analysis are important.

Good continuation has two direct applications to graph drawing. First, it suggests that a path will be more readily perceived if the nodes are not in a zigzag pattern, but form a smooth continuous sequence. This point is illustrated in Figure 4. Good continuation also suggests that curved lines can be used to make certain paths more apparent (as illustrated in Figure 5) although this is not the focus of our present study.

Our present study was motivated by two primary concerns: we wished to follow up the earlier work of Purchase using a more refined procedure to determine the relative importance of different aesthetics. In particular we were interested in the problem of path continuity (Figures 4 and 5). In many cases edge crossings can only be reduced at the cost of constructing a more indirect or less continuous path. If this were the case then algorithms should be constructed that take both factors into account.

The second goal was to place the value of graph aesthetics on a firmer empirical foundation, by developing and applying a methodology for measuring the cognitive cost of graph aesthetics. By using a large number of trials and varying a set of interesting parameters we hoped to be able to deduce the relative cognitive cost of different factors in graph layout (eg edge crossings, path continuity, geometric line length).

Perceptual and cognitive modeling The discipline of human-computer interaction (HCI) has developed simple cognitive models of common computer interaction tasks, such as selection of objects using a mouse.¹⁰ For example, Fitt's law¹¹ has been adopted to describe how long it takes to make a visually guided hand movement (using a computer mouse) as a function of how small the target is and how far the distance to be moved.

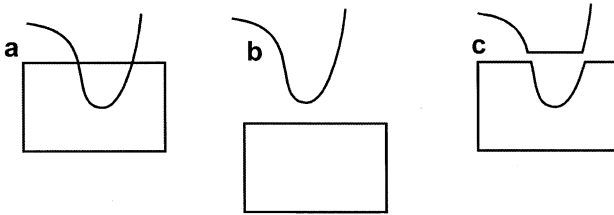


Figure 2 The coarse orientation tuning of edge detectors in the brain suggests that lines that cross at an acute angle as shown on the left are more likely to be confusion than lines that cross nearly at 90° as shown on the right.

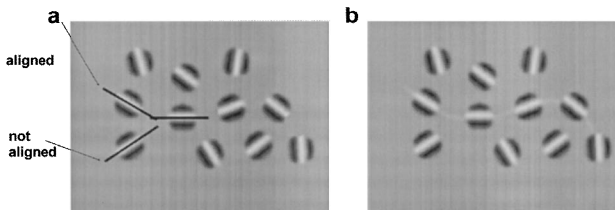


Figure 3 A schematic diagram illustrating the experiments conducted by Field *et al.*⁸ If the elements are aligned as shown in (a) so that a smooth curve can be drawn between them, then the curve shown in (b) is perceived. In the actual experiment Gabor patches were used.

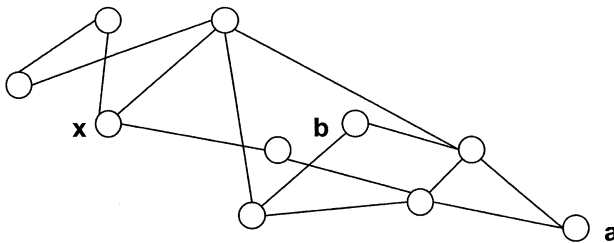


Figure 4 The principle of good continuation suggests that it should be easier to see the shortest path from x to a than from x to b, despite the fact that in both cases, the shortest path length is 3.

The general strategy for building simple perceptual or cognitive models is applicable to the problems of graph aesthetics. If we consider a task such as determining the shortest path between two nodes, we can measure how the time to perform such a task depends on various factors, such as the length of the path, the continuity of the path and the number of edge crossings on the path. If such an approach can produce robust and reproducible results, we can estimate the cognitive cost of, for example, the number of edges leaving intermediate nodes, each crossing, and the angle of the edge crossings. The result could then be used to develop optimal layouts to support a set of tasks that are anticipated in the use of a node-line diagram.

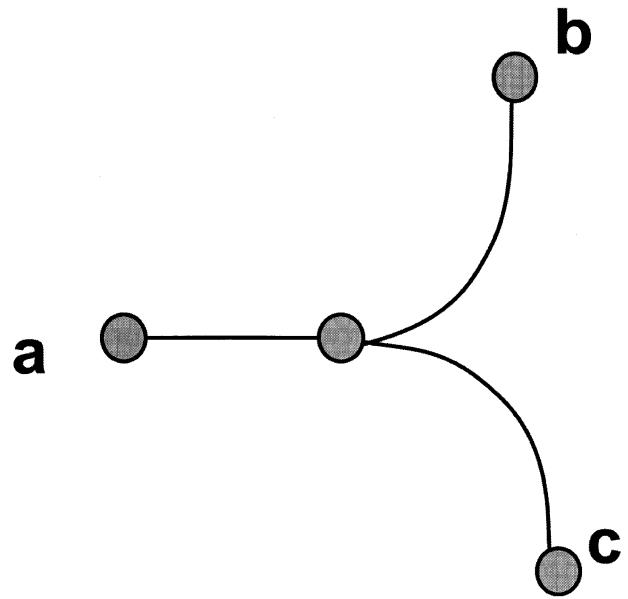


Figure 5 Good continuation suggests that the path from a to b should be easier to perceive than that from b to c.

Experimental determination of perceptual and cognitive costs

In order to estimate the cognitive cost of various graph layout aesthetics we developed an experiment to measure the time to perceive the shortest path between two specified nodes in a spring layout graph. The method involved generating a large number of graphs each with a unique shortest path between two specified nodes, in which the following factors varied:

Continuity ('path bendiness') (con) We defined continuity at a node as the angular deviation from a straight line of the two edges on the shortest path which emanate from the node. To get the total path continuity the angular deviations were simply summed at all nodes on the path. By this metric a straight path has a low value.

Number of crossings (cr) and average crossing angles (aca) We recorded the number of edge crossings on the shortest path, as well as the angle of each edge crossing (from which we derived the average cosine crossing angle). We expected that acute angles would be more disruptive than more perpendicular angles.

Number of branches (br) For each of the intermediate nodes on the shortest path, we recorded the number of edges leaving the node which were not part of the shortest path itself (ie the degree of the node -2). Every branch on a path represents a possible alternative path that might be considered in determining the shortest path. Thus, as the total number of branches on the shortest path increases, we can expect the task to become more difficult.

Shortest path length (*spl*) We were not interested in the number of edges in the shortest path as such; however, in order to generalize the results we thought it desirable to evaluate the other factors in the context of the length of the shortest path.

In addition to the primary factors we were also interested in:

Total geometric line length (*tll*) The spring layout algorithm we used is designed to produce edges of approximately constant length. Nevertheless, combined spring forces can cause edges to be shorter or longer than the designed length. Therefore we were interested in the actual total geometric length of the shortest path.

Total crossings in the graph (*tcr*) We recorded the total number of crossings in the entire graph drawing. We did not expect this measurement to have any bearing on the result, but included it as this was the measurement used by Purchase in the earlier study with which we wished to compare our results.

Experimental task The subjects were asked to determine the shortest path length of each graph drawing presented to them, as produced by the spring layout algorithm. For any given graph drawing, the shortest path between the two highlighted nodes was between three and five. The start and end nodes were both highlighted.

The diagrams Each graph drawing used in the experiment was a simple non-meaningful node-link diagram, which did not describe a specific domain. The parameters for all 180 experimental graphs were determined prior to the experiment, and stored in a file. The drawings themselves generated during the experiment itself, using the stored parameter information. Each trial entailed generating a graph drawing using spring forces and simulated annealing.¹² There were 42 nodes in the graph, and the number of edges on each node was randomly varied between one and five. The size of the window was 700 pixels square, and physically measured 19.4 by 19.4 cm. The subjects sat directly in front of the computer screen, at a distance of 40–50 cm. The spring constant was set so that the mean of the edge lengths was approximately 50 pixels (this converts to approximately 1.4 cm).

For each trial, the algorithm randomly determined a start node that had at least two incident edges. It then used a breadth first search to find a path to another node such that there was a unique shortest path between the two nodes of length 3, 4 or 5.

Examples of typical diagrams used in the experiment can be seen in Figure 6.

Experimental methodology

Experimental documents Subjects were given a set of experimental materials to familiarize themselves with

the task and the online system. These materials consisted of a consent form, instructions for the online system, and a tutorial sheet.

The instruction sheet described the task and the online system that would be used by the subject. The tutorial explained graph drawings and the concept of the shortest path between two nodes, as well as presenting six example graph drawings. These drawings had two nodes highlighted, and the correct answer for the shortest path question was indicated for each.

The subjects were given 5 min to sign the consent form, reach through and understand the materials, ask questions, take notes, or draw diagrams as necessary.

Online task Following the preparation time, subjects were told to start the custom-built online experimental system. The graph drawings were presented individually on the screen and subjects were required to respond by indicating the length of the shortest path between the two highlighted nodes. Three adjacent keys on the keyboard were used for this purpose: they were labeled 3, 4, and 5, the complete range of possible answers.

The subjects were shown 10 blocks of 20 graph drawings each, with the first two blocks being identical to the final two sets. These first two blocks of diagrams were used as practice diagrams and were excluded from the data, as were the first two diagrams in each consecutive block. These practice diagrams were used to accommodate for the learning effect (whereby the subjects' performance on the task may improve over time, as they become more competent in the task).

The eight different experimental blocks of graph drawings were presented in random order for each subject, though the order of the diagrams within the blocks remained the same. This meant the two practice blocks at the beginning of the experiment differed for each subject, as they were always the same as the last two blocks of diagrams in the experiment.

Each diagram was displayed until the subject either pressed the space bar in order to see the following diagram, or pressed one of the labeled numeric keys (3, 4 or 5). There was no time limit on the display of the diagrams and only once the subjects pressed the space bar would the next diagram appear. A computer beep indicated an incorrect response, to encourage the subjects to perform as best they could, and to identify subjects who were not taking the task seriously.

In addition, there was a rest period between each block of graph drawings, allowing the subjects to recover their concentration before continuing with the experiment. The length of this rest break was controlled by the subjects.

Data collection The experimental data collected was the response time and accuracy of the subjects' responses to the experimental diagrams: this information was recorded by the online system during the experiments. Elapsed times were measured through system calls with millisecond nominal resolution. However, temporal reso-

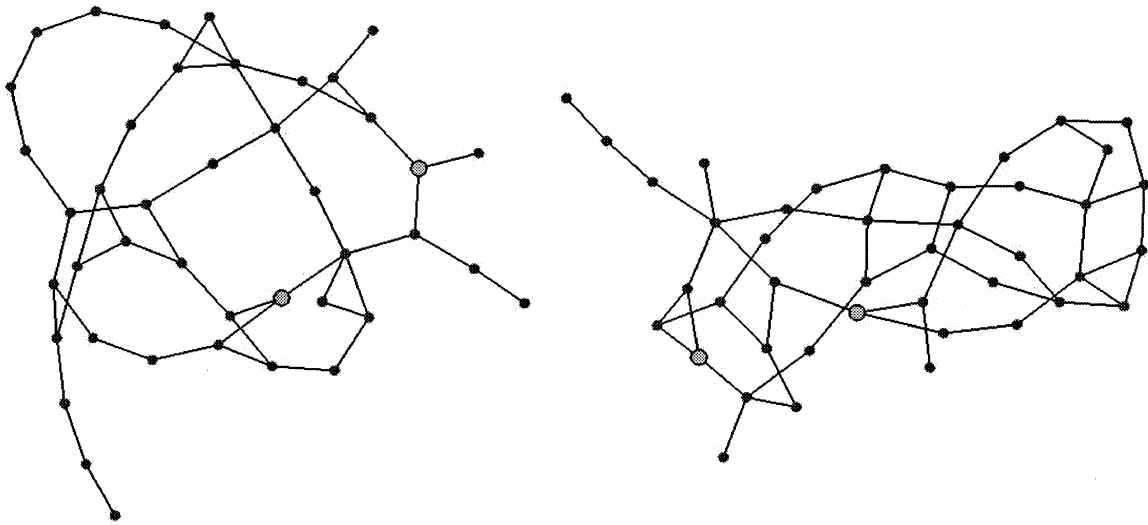


Figure 6 Two of the experimental graph drawings, as viewed by the subjects. In the originals the small nodes were red while the larger end-point nodes were light blue.

lution could be no better than the 16 millisecond granularity imposed by the monitor refresh rate.

Independently of the experiments, graphical measurements were calculated and recorded for each of the 180 graph drawings used. These included:

- the correct value for the shortest path length (**spl**): 3, 4 or 5
- the continuity ('bendiness') of the shortest path (**con**): measured in degrees.
- the number of crossing edges on the shortest path (**cr**)
- the cosine of the angle at which each crossing edge crossed the shortest path. The purpose of this was to allow us to weight shallow angle crossings higher than orthogonal crossings. From this we computed the average cosine crossing angle (**aca**)
- the total number of crossed edges in the graph drawing (**tcr**)
- the number of edges branching from nodes along the shortest path (**br**)
- average geometric line length along the shortest path, per edge (**all**). This was computed in arbitrary units. To get centimeters it is necessary to multiply by 1.4.
- total geometric line length of the shortest path (**tll**): same units as **all**

Subjects The 43 subjects were second and third year Computer Science and Information Systems students at the University of Queensland. The subjects were paid \$15 for their time, and, as an incentive for them to take the experiment seriously, the best performer was given a CD voucher.

Results

Ninety-three per cent of the subjects' responses were correct. The response time data was analyzed using only

the correct trials. Subjects saw 10 blocks of 20 drawings each. Seven graph drawings were eliminated from the analysis because of an unexpected perceptual error, where an edge passing under a node, looked as if it were two edges attached to that node. Eliminating the practice graphs (the first two blocks, 40 drawings), and the first two drawings of each subsequent set ($2 \times 8 = 16$ drawings), left us with 137 drawings in our experimental set.

Our independent variable was the average response time over all subjects for each graph drawing (*rt*), and our dependent variables were the measured geometric features of each drawing: *spl*, *con*, *cr*, *aca*, *trc*, *br*, *all*, *tll*.

Table 1 shows the linear correlations between the independent variables and the average response time. Figure 7 shows selected scatter graphs to demonstrate some of the most important correlations identified.

Data analysis: linear correlations

The results presented above (Table 1, Figure 7) use linear correlations, showing the relationships between each independent variable and the dependent variable of response time. From these linear correlations we observe that:

- The two key variables of crossings (*cr*) and continuity (*con*) are almost independent.
- The variability of the data increases with increase in 'bendiness': straight paths were generally responded to between 2 and 5 s while paths with a high measure of 'bendiness' were responded to in between 2 and 14 s.
- A similar effect can be seen with the total geometric line length of the shortest path: responses to short paths only varied between 2 and 4 s whereas for long path the range was between 4 s and about 14 s.

Table 1 Linear correlation coefficients between the measured graph drawing properties, including the shortest path length for the task and the average response time over all subjects. Shaded cells indicate results that are statistically significant

	spl	con	cr	aca	tcr	br	all	tll	rt
spl	1	0.484	0.191	0.134	0.059	0.379	0.064	0.930	0.736
con		1	0.019	0.082	0.125	0.119	-70.294	0.331	0.633
cr			1	0.141	0.347	0.267	0.428	0.332	0.449
aca				1	0.064	0.208	0.099	0.167	0.148
tcr					1	0.116	0.011	0.066	0.216
br						1	0.353	0.475	0.462
all							1	0.419	0.050
tll								1	0.623
rt									1

Data analysis: multiple regression

However, linear correlations are insufficient for appropriate interpretation, as there are many significant correlations between the independent variables themselves (Table 1). We need to be sure that the internal relationships between the individual independent variables are 'factored out', so that we can identify the relative contribution of each variable, independent of its relationship to the other variables.

We used stepwise multiple regression in two steps. In the first step, we included only the shortest path length (spl) into the equation, as we knew that it would have the greatest effect, and we first wished to remove the variance in the data that it caused. We did not force the number of edge crossings (cr) into the equation at this point (despite the prior work by Purchase which concluded that crossings was the most important factor) as we wished to determine whether the number of edge crossings really was more dominant than the other factors.

In the second step, we put all the remaining independent variables into the equation and performed a stepwise analysis. Thus, the most significant independent variable (ie the variable that explains the most of the variance that has not already been explained by the previously entered variables) would be used first, then the next most significant independent variable, etc., with the relative order of importance indicated by the left-to-right ordering of the terms in the equation. Any independent variables that were not significant at a 0.05 significance level would not appear in the final multiple regression equation.

Performing a stepwise multiple regression analysis on the 137 graph drawings, with response time as the dependent variable, the following equation relates the response time to the shortest path length, the continuity of the shortest path, the number of crosses on the shortest path and the number of additional branches from the intermediate nodes on the shortest path. The other independent variables were not significant.

$$rt = -4.970 + 1.390 \text{ spl} + 0.01699 \text{ con} + 0.654 \text{ cr} + 0.295 \text{ br} \quad (1)$$

When the data has been normalized to eliminate the constant, the following equation better allows us to see the relative contributions of the independent variables:

$$rt = -0.414 \text{ spl} + 0.406 \text{ con} + 0.317 \text{ cr} + 0.172 \text{ br} \quad (2)$$

An R^2 value can indicate the extent to which the dependent variable correlates with the independent variables on the left hand side of the equation. The change in the R^2 value after each independent variable is included in the stepwise regression, indicates the relative effect of each of the variables. R^2 after shortest path length (spl) is included: 0.542. R^2 after continuity (con) is included: 0.642 (change=0.100). R^2 after crossings (cr) is included: 0.760 (change=0.118). R^2 after branches (br) is included: 0.784 (change=0.024).

Discussion

As a contribution to graph layout research perhaps our most significant result is the finding that path continuity is an important factor in perceiving shortest paths. Considering that path length is an intrinsic property of a graph, this makes continuity the most important aesthetic property in the set of graphs that we generated. Simply put, the results show that 100° of continuity (bendiness) on a path contribute 1.7 s to finding the shortest path while each edge crossing contribute 0.65 s. A more general way of stating this is to say that the cognitive cost of a single crossing is approximately the same as 38° of continuity. The practical consequence is that it may be worth allowing for an occasional crossing in a graph layout if it reduces the bendiness of paths. Using this information it should be possible to construct graph layout algorithms that are more effective for analytic tasks where path finding is important. More studies are required to determine the robustness of our findings of this study for other than spring layout graphs and for different graph sizes, but the theoretical arguments given in the introduction would lead us to expect the results to be generalized.

In comparing these results with those of Purchase³, it was found that her measurement of the total number of edge crossings in the graph drawing is not a significant indicator of response time. For shortest path tasks, we

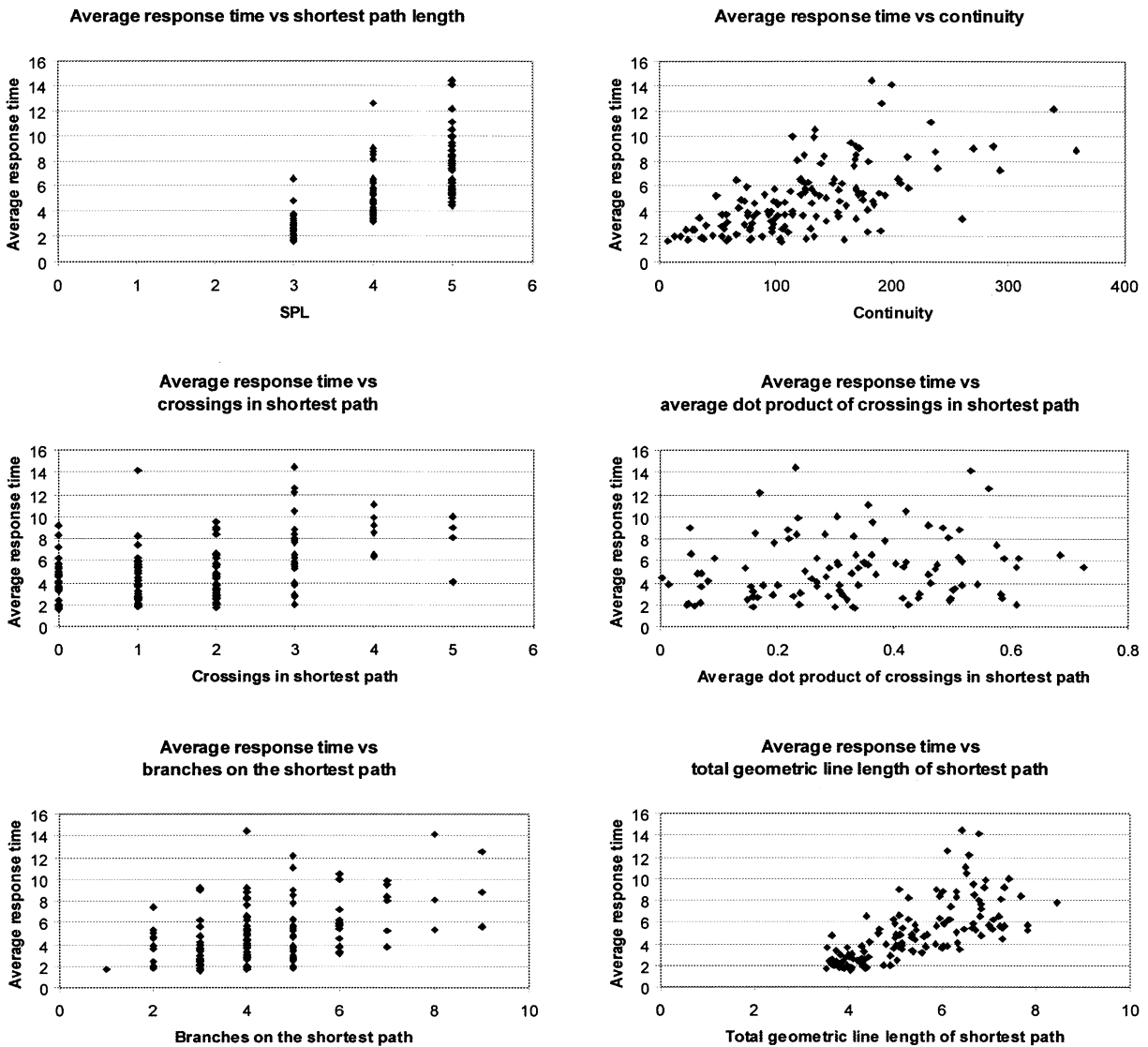


Figure 7 Scatter plots showing the linear correlations between some of the measured graph drawing properties, and the average response time for each graph drawing over all subjects (time is measured in seconds).

have shown here that it is the number of edges that cross the shortest path itself that is important, rather than the total number of edges crossing in the drawing. The results of the Purchase experiment, were, however, not purely based on a shortest path task: the task of identifying the nodes and edges which, when removed, would cause two nodes to be disconnected, were also included in her performance data. The effect of the total number of crossings in the whole graph drawing may have been more important in these latter two tasks than in the shortest path task.

Perhaps the result that is most difficult to explain is the negative intercept for the regression equation (1). A simple extrapolation from this would imply that with very short straight paths the task can be accomplished

in negative time. Of course, all of our results were positive. A possible explanation is that short straight paths can be perceived in roughly constant time and that it is only when the path length exceeds 3 that a significant cognitive cost is incurred. This is made more plausible by research that has shown that we can count visual objects up to three ‘at a glance’, that is, in constant time.¹³ Thus the cognitive cost of counting nodes is incurred only after the length of a path exceeds three.

We believe that a major part of their contribution in this study is the methodology itself. In many visualization problems there are similar tradeoffs between different optimization criteria. The same experimental methodology described here can be applied to any display problem that has similar characteristics and a



well-defined elementary task. For example, it is common practice to display vector fields using particle traces, but the optimal way of doing this is unknown. As with graph layout there are many free display parameters to be optimized: the length of the particle traces, their density, and their sizes can all be manipulated. Several tasks are available for optimization: one is the perception of vector field strength, another is the prediction of advection (ie where in the field a dropped particle would end up.)

Two areas of future work can stem from this research: firstly the data presented here needs validating to determine whether it truly can be used as a predictive model for shortest path tasks in graph drawing. Secondly, the methodology can be applied to a variety of other simple

visual tasks for which the cognitive cost of differing visual features would be of interest.

Acknowledgments

We are grateful to the students from the School of Information Technology and Electrical Engineering at the University of Queensland, who willingly took part in the experiments, and to Julie McCreddon and Maureen Tingley for advice on the statistical analysis. Discussions with Sue Whitesides at McGill University were instrumental in shaping this study. The Australian Research Council partly funded this research, and partial funding for Colin Ware came from the U.S. National Science Foundation. Ethical clearance for this study was granted by the University of Queensland, 2001.

References

- 1 Rumbauch J, Jacobson I, Booch G. *The Unified Modeling Language Reference Manual*. Addison Wesley Longman, Inc.: Reading, Mass. 1999.
- 2 Battista GD, Eades P, Tamassia R, Tollis IG. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, 1999.
- 3 Purchase H. Which aesthetic has the greatest effect on human understanding. In: Battista G (Ed). *Graph Drawing 97*, 1353 of Lecture Notes in Computer Science. Springer Verlag: 1997 284–290.
- 4 Koffka K. *Principles of Gestalt Psychology*. Harcourt Brace: New York. 1935.
- 5 Ware C. *Information Visualization: Perception for Design*. Morgan Kaufman. 1999.
- 6 Triesman A, Gormican S. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 1988; **95**: 15–48.
- 7 Blake R, Holopogan K. Orientation selectivity in cats and humans assessed by masking. *Vision Research* 1985; **23**: 1459–1467.
- 8 Field D, Hayes A, Hess FR. Contour integration by the human visual system: Evidence for a local “association field”. *Vision Research* 1993; **33**: 173–193.
- 9 Parent P, Zucker S. Trace interference, curvature consistency and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1989; **6**: 661–675.
- 10 Card S, Moran TP, Newell A. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates: New Jersey, 1983.
- 11 Fitts PM. The information capacity of the human motor system in controlling the amplitude of movements. *Journal of Experimental Psychology* 1954; **47**: 381–391.
- 12 Davidson R, Harel D. Drawing graphs nicely using simulated annealing. *ACM Trans. on Graphics* 1996; **15**: 301–331.
- 13 Dehaene S. *The Number Sense: How the Mind Creates Mathematics*. Oxford Press: New York, 1997.

IVS

Manuscript No.	011_02	
Author		
Editor		
Master		
Publisher		

Information Visualization
Typeset by Elite Typesetting
for Palgrave 



QUERIES: to be answered by AUTHOR/EDITOR

AUTHOR: The following queries have arisen during the editing of your manuscript.
Please answer the queries by marking the requisite corrections at the appropriate positions
in the text.

QUERY NO.	QUERY DETAILS	QUERY ANSWERED
	The references within the manuscript have been renumbered to run numerically	