An Experimental Study of the Basis for Graph Drawing Algorithms

H.C.Purchase
The University of Queensland,
R.F.Cohen
The University of Newcastle, Australia
and
M.I.James
The University of Queensland

Designers of graph drawing algorithms and systems claim to illuminate application data by producing layouts that optimise measurable aesthetic qualities. Examples of these aesthetics include *symmetry* (where possible, a symmetrical view of the graph should be displayed), *minimise arc crossings* (the number of arc crossings in the display should be minimised), and *minimise bends* (the total number of bends in polyline arcs should be minimised).

The aim of this paper is to describe our work to validate these claims by performing empirical studies of human understanding of graphs drawn using various layout aesthetics. This work is important since it helps indicate to algorithm and system designers what are the aesthetic qualities most important to aid understanding, and consequently to build more effective systems.

Categories and Subject Descriptors: I.3 [**Computer Graphics**]: Picture/Image Generation—*Display algorithms*; H.5 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation*

General Terms: Graph Drawing Aesthetics

Additional Key Words and Phrases: Human-computer interaction

This work is funded by an Australian Research Council grant.
An abbreviated version of this article appeared in Lecture Notes in Computer Science (no. 1027), Springer, 1995.
Name: H.C.Purchase
Affiliation: School of Information Technology, The University of Queensland
Address: St Lucia, Queensland 4072, Australia
Name: R.F.Cohen
Affiliation: Computer Science Department, The University of Newcastle
Address: New South Wales 2308, Australia
Name: M.I.James
Affiliation: School of Information Technology, The University of Queensland
Address: St Lucia, Queensland 4072, Australia

1. INTRODUCTION

The visualisation produced by a graph drawing subsystem should *illuminate* application data. That is, it should help the user to understand and remember the information being visualised. A good layout can be a picture worth a thousand words; a poor layout can confuse or mislead. Designers of graph drawing algorithms and systems claim to illuminate application data by producing layouts that optimise measurable aesthetic qualities, on the assumption that considering these aesthetics makes the drawing 'nice' and easier to read. Examples of these aesthetics include:[Eades 1984; Ferrari and Mezzalira 1969; Tamassia 1987]

—*bends*: the total number of bends in polyline arcs should be minimised [Tamassia 1987; Trickey 1988]

—*crossings*: number of arc crossings in the display should be minimised [Ferrari and Mezzalira 1969]

—*symmetry*: where possible, a symmetric view of the graph should be displayed [Eades 1984; Lipton et al. 1985]

—*orthogonality*: arcs or arc segments should be parallel to the coordinate axes [Tamassia 1987]

—*minimum angle*: the minimum angle formed by the drawing of consecutive arcs around a node should be maximised [Vijayan 1986; Formann et al. 1993; Garg and Tamassia 1994]

The aim of this paper is to describe our work to validate these claims by performing empirical studies of human understanding of graphs drawn using various layout aesthetics. This work is important since it helps indicate to algorithm and system designers what are the aesthetic qualities most important to aid understanding, and consequently to build more effective systems.

There has been very little work in this area. This is the first study to address the effect of aesthetics on general understanding of graphs. Prior work has focused on specific applications [Batini et al. 1985; Ding and Mateti 1990] and comparisons of two- and three-dimensional layouts [Ware et al. 1993].

Although many algorithms consider more than one aesthetic in their attempt to create an illuminating graph display, this initial study focuses on individual aesthetic principles, leaving the testing of the effectiveness of the various algorithms which embody them for a further project. We aim to determine empirically the validity of these aesthetics, and place a priority ordering on their criteria in the ease of reading of a graph. Graph layout algorithms often compromise between more than one aesthetic: by providing a priority listing, we aim to assist algorithm developers in determining which aesthetics to emphasise.

We study the following popular graph drawing aesthetics: minimise arc crossings, minimise bends, and maximise symmetries. Our initial results are encouraging, indicating that minimising arc bends and minimising arc crossings are both important aids to human understanding of graph drawings. However, our study is inconclusive as to the importance of maximising symmetry. This final non-result is surprising to us, since we believed that maximising symmetry would have a high impact on the understandability of drawings.

The rest of the paper is organised as follows. Section 2 describes the scope of our experiment, section 3 explains our methodology, section 4 presents our results, and section 5.1 describes our analysis. Finally, in section 6 we give conclusions and directions for future work.

2. SCOPE

2.1 Hypotheses

Three common graph-drawing aesthetics were chosen for this study: *symmetry* (where possible, a symmetrical view of the graph should be displayed), *minimise arc crossings* (the number of arc crossings in the display should be minimised), and *minimise bends* (the total number of bends in polyline arcs should be minimised). We restricted our study to undirected graphs.[1]

The hypotheses for our experiment were therefore:

—Increasing the number of arc bends in a graph drawing decreases the understandability of the graph

—Increasing the number of arc crossings in a graph drawing decreases the understandability of the graph

—Increasing the local symmetry displayed in a graph drawing increases the understandability of the graph

There are two ways in which "understandability" may be measured. A purely relational ("syntactic") method measures the efficiency and accuracy with which people can *read* a graph structure and answer questions about it. Such graph-theoretic questions need to be generic and application-independent, and include questions of the form "What is the shortest path from node A to node B?" A more application-specific ("semantic") method rather considers a graph *interpretation* task: in this case it is more appropriate that the effectiveness of the graph drawing is measured within the context in which the application-specific graph is usually used. Thus, instead of eliciting answers to specific questions asked about the graph itself, it is more suitable to look at whether the graph has assisted the user in accomplishing a particular application task. Suitable questions for this approach would include (in the area of software engineering) "What object classes would be affected by changing the external interface to class X?"

In this study, we concentrate on the syntactic perspective, leaving the semantic approach for a later study. Three graph-theoretic questions were therefore chosen as our means for measuring the understandability of a graph drawing, and the performance of a person in answering these questions about a graph drawing was considered an appropriate indicator of their understanding of the drawing. The three questions used were:

[1]As this initial study, as well as producing its own results, also served as a pilot study of a novel experimental method, we chose to restrict the number of aesthetics considered. The study of other aesthetics (the direction of flow for directed graphs, maximising the minimum angle between arcs from one node, and orthogonality) has been left for a later project, when this experimental method may be refined in the light of this initial experience.

(1) How long is the shortest path between two given nodes?
(2) What is the minimum number of nodes that must be removed in order to disconnect two given nodes such that there is no path between them?
(3) What is the minimum number of arcs that must be removed in order to disconnect two given nodes such that there is no path between them?

2.2 The Graphs

A dense graph and a sparse graph were defined. The sparse graph has 16 nodes and 18 arcs. The dense graph was created by adding 10 extra arcs onto the sparse graph. They were designed so that for each of the three questions above, pairs of nodes could be identified that would give a range of answers. For the first question four node-pairs could give answers of 2,3,4,5; for the second question, three node-pairs gave answers of 1,2, and for the third question, the possible answers were 1,2,3. Care was taken to ensure that in defining the node pairs corresponding to the three answers for question 3 (the number of arcs to be removed to disconnect two nodes), the answers to the questions were not easily determined by the degree of either of the nodes.[2]

2.3 The Graph Drawings

Nine drawings of each graph were created, with the number of bends, crossings and the amount of perceived symmetry varied appropriately (Figures 1 and 2). Metrics for the number of crossings and bends were easy: we merely counted them; determining a metric to measure the amount of symmetry in a graph drawing was more difficult. We needed to define a metric that could give a numerical value to perceived symmetry. The method used in this study is described in Appendix A (see also [Bhanji et al. 1995]).

The graphs were drawn with the following constraints:

—no two bent arcs were parallel (as this would introduce regularity into the drawings which may confound the results)
—the angle of the bends was less than 150 degrees (so that the bend was obvious)
—symmetry, bends and crossings were spatially distributed throughout the graph drawings
—when drawing a graph with one aesthetic variation in it, the values of the other two aesthetics were held at zero

To ensure that the subjects would not recognise that the same graphs were being used for all the drawings, each drawing had its nodes relabeled randomly: the same

[2]Initially, we included a sub-graph of 4 nodes in both the sparse and dense graphs which was not attached in any way to the main graph. The purpose of this was to enable us to include questions that had the answer 0 for the second and third question. Comments from subjects during a pilot test indicated that this was one of the main aspects of the drawings that they found confusing. More seriously, if the answer to the second question was 0, and one of the nodes in the third question also referred to the sub-graph, the answer (0) to this third question was obvious. As we did not wish any extraneous features independent of the aesthetics to have any confounding effect, and as it was not essential that there be questions with the answer 0 for any of the questions, the sub-graph was removed from both the dense and sparse graph drawings in the main experiment.

Table 1. The aesthetic values for the 18 graph drawings

graph drawing		bends	ratio	crossings	ratio	symmetry	ratio
sparse-bends-few	(**sbf**)	5	(1)	0	–	0	–
sparse-bends-some	(**sbs**)	15	(3)	0	–	0	–
sparse-bends-many	(**sbm**)	26	(5.2)	0	–	0	–
sparse-crossings-few	(**scf**)	0	–	4	(1)	0	–
sparse-crossings-some	(**scs**)	0	–	16	(4)	0	–
sparse-crossings-many	(**scm**)	0	–	28	(7)	0	–
sparse-symmetry-few	(**ssf**)	0	–	0	–	4	(1)
sparse-symmetry-some	(**sss**)	0	–	0	–	17.5	(4.4)
sparse-symmetry-many	(**ssm**)	0	–	0	–	34	(8.5)
dense-bends-few	(**dbf**)	6	(1)	0	–	0	–
dense-bends-some	(**dbs**)	18	(3)	0	–	0	–
dense-bends-many	(**dbm**)	30	(5)	0	–	0	–
dense-crossings-few	(**dcf**)	0	–	6	(1)	0	–
dense-crossings-some	(**dcs**)	0	–	24	(4)	0	–
dense-crossings-many	(**dcm**)	0	–	42	(7)	0	–
dense-symmetry-few	(**dsf**)	0	–	0	–	4.6	(1)
dense-symmetry-some	(**dss**)	0	–	0	–	25.7	(5.6)
dense-symmetry-many	(**dsm**)	0	–	0	–	51	(11.1)

lettering (A – T) was used in each, but the relationships between the nodes was different. The corresponding node-pairs associated with the questions were also relabeled accordingly.

For each aesthetic, three drawings of each graph were produced, one with a small aesthetic measurement (*few*), one with an interim aesthetic measurement (*some*), and one with a large aesthetic measurement (*many*). The values of these measurements for all the graph drawings are shown in Table 1.

Note that every attempt was made to maintain the same *few*:*some*:*many* ratios for each aesthetic between both graphs, to enable comparisons to be made between the results for each graph.

3. METHODOLOGY

3.1 Subjects

The subjects in the experiment were all second year computer science students from the University of Queensland. 49 subjects worked on the dense graph, and 35 of these same subjects worked on the sparse graph a week later. Participation in the experiments was voluntary.

3.2 Experimental booklet

Each subject was given an experimental booklet of the following form:

(1) An introduction which included a brief description of graphs, and definitions of the terms *node*, *arc*, *path*, and *path length*.

(2) Instructions to the subjects, and explanations of the three graph-theoretic questions they would be required to answer about the experimental graph drawings.

(3) A simple example graph drawing (including some bends, crossings and symmetry), with explanations of how the correct answers for all three questions were achieved.

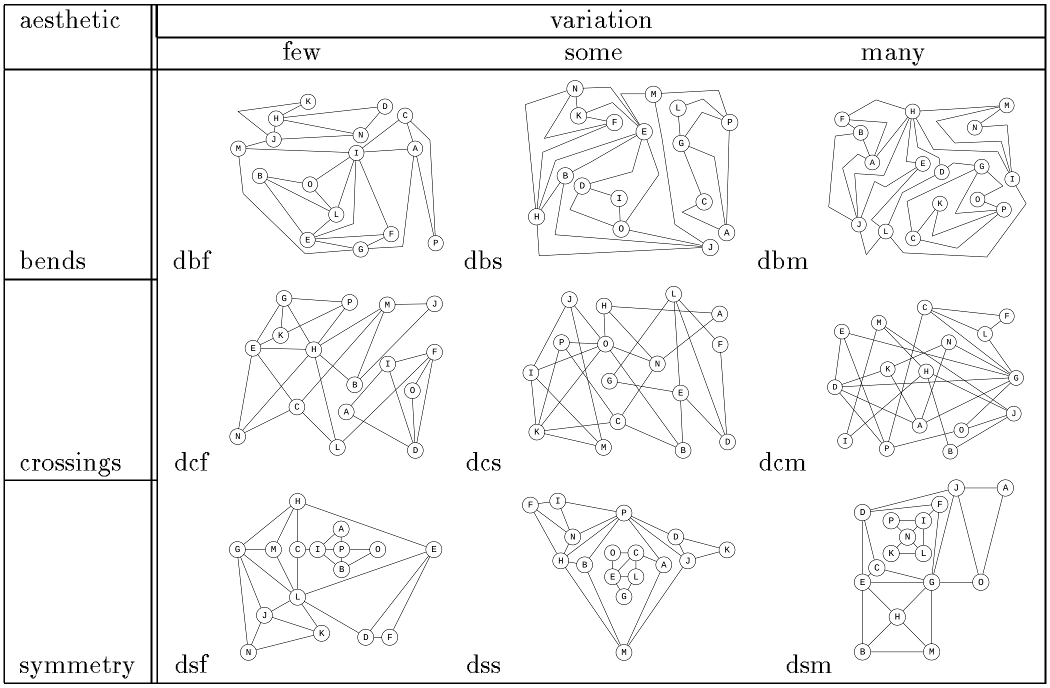


Fig. 1. The Experimental Graph Drawings: dense graph

At this stage in the experiment, the subjects were asked if they had any questions about graphs in general, or about the experiment. It was important to ensure that all the subjects knew what was going to be expected of them.

(4) Three drawings of two "practice" graphs (sparse: 20 nodes and 21 arcs; dense: 20 nodes and 32 arcs). Each graph was drawn three times: once with crossings, once with bends and once with some symmetry. The three graph theoretic questions were asked of each of these drawings. The purpose of this section was to familiarise the subjects with the nature of graph drawings, and to ensure they were comfortable with the task and the questions. The subjects were unaware that these graph drawings were not experimental.

(5) A "filler" task which engaged the subjects' mind on a small problem unrelated to graphs. This was to ensure that their performance on the subsequent experimental graphs was not affected by any follow-on effect from the practice graphs. We used a simple word puzzle.

(6) The nine experimental graph drawings (all sparse, or all dense), in a random order for each subject.
The questions themselves were randomised too: although the same three graph-theoretic questions were asked of each drawing, the pair of nodes chosen for each question was randomly selected from a list of node-pairs. This list of node-pairs was defined so that there was a range of possible correct answers to each question for each graph drawing (see section 2.2). This randomisation was to ensure that any variability in our data could not be explained away by

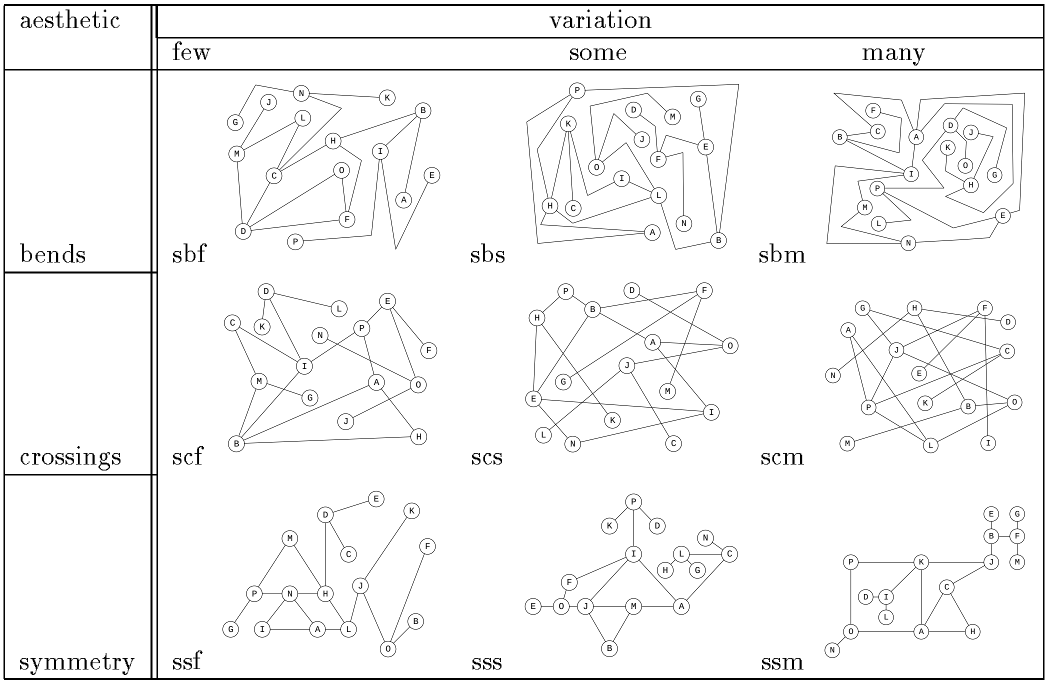


Fig. 2. The Experimental Graph Drawings: sparse graph

the varying difficulty of the questions, which may have been the case if each graph drawing had had a fixed set of question-answer pairs.

(7) A questionnaire, asking the subjects about their current degree, how far they were into their degree, any previous experience of reading graphs, and if there were any particular features about the graphs that they could identify as assisting or hindering their performance (this information was for our own interest only: it was not used in the quantitative analysis of the results).

The time allowed for each component of the experiment is shown in Table 2. Therefore, the variables associated with the experimental methodology were:

Table 2. The controlled timing for the sections of the experimental booklet

Task	Time: dense graph	Time: sparse graph
Introduction and example	As necessary	As necessary
Practice graph 1	60 seconds	45 seconds
Practice graph 2	60 seconds	45 seconds
Practice graph 3	55 seconds	40 seconds
Practice graph 4	55 seconds	40 seconds
Practice graph 5	50 seconds	35 seconds
Practice graph 6	50 seconds	35 seconds
Filler task	60 seconds	60 seconds
Experimental graphs 1–9	45 seconds	30 seconds

—*control variables:* the time allowed for answering the three questions for each graph drawing; the graph structure; the questions

—*independent variables:* the number of bends, the number of crossings and the value of the symmetry metric

—*dependent variable:* the number of errors made in answering the questions for each graph drawing

4. RESULTS

A within-subject analysis was chosen for this experiment to reduce any variability that may be attributable to the difference between the subjects (age, experience *etc*). The main disadvantage with within-subject analysis is the 'learning effect', where the subjects' performance on the later questions may be better, not because of the nature of the graphs or questions, but merely because they have had more practice. In this experiment, any learning effect was minimised by the large number of graph drawings used, the use of practice graphs at the beginning, and the random order of the drawings for each subject.

We chose the Friedman two-way analysis of variance by ranks, used when a number of matched ordinal samples are taken from the same population. Each aesthetic metric was varied in a ordinal manner (i.e.*few–some–many*). Additionally the samples for each metric were matched because all values of each metric were tested on every subject. The Friedman test is a nonparametric statistical test, so small populations are suitable for analysis: we chose $N = 25$ as the floor on our sample size. (Appendix B gives a brief description of the Friedman two-way analysis of variance.)

A requirement of the Friedman test is that the data points be ranked. For each graph, the results for each aesthetic were collated and ranked. The number of correct answers on the three questions for each subject was ranked from 1 to 3: the highest score obtained by a subject was given a rank of 1 and the lowest score was ranked 3. If any scores were tied, their rankings were averaged. As an example, Table 3 shows the results and rankings for the *dense-bends* drawings.

This Friedman test was used to determine the χ^2 value for each aesthetic, and the probability that the ranks were produced by chance (p). The results are shown in Table 4. With a level of significance set at $\alpha = 0.05$, these results are significant for bends and crossings in both the dense and sparse graphs, as $p < \alpha$ in both these cases.

Figures 3 and 4 show the trends in the average number of errors made by each subject, when plotted against the variation in each aesthetic.[3]

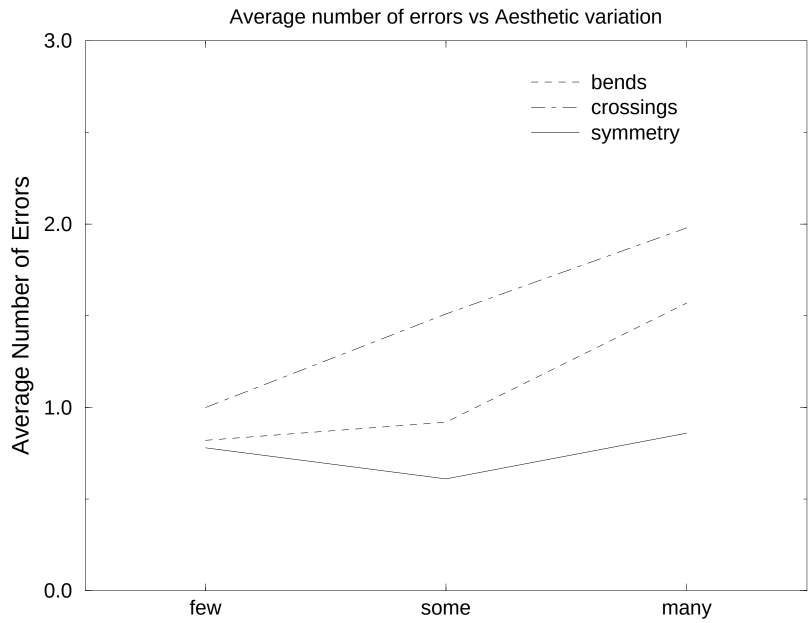
[3]Note that even though the average number of errors for the *bends-some* drawing is slightly higher than that for the *bends-many* drawing, the important feature to note for the data analysis is the fact that the overall trend for all three data points is upwards.

Table 3. Subject scores and rankings for *dense-bends* drawings

Subject	Scores			Rankings		
	dbf	dbs	dbm	dbf	dbs	dbm
1	3.0	3.0	3.0	2.0	2.0	2.0
2	3.0	3.0	0.0	1.5	1.5	3.0
3	3.0	1.0	1.0	1.0	2.5	2.5
4	2.0	2.0	1.0	1.5	1.5	3.0
5	0.0	0.0	0.0	2.0	2.0	2.0
6	2.0	2.0	1.0	1.5	1.5	3.0
7	1.0	0.0	1.0	1.5	3.0	1.5
8	2.0	1.0	1.0	1.0	2.5	2.5
9	0.0	1.0	0.0	2.5	1.0	2.5
10	0.0	1.0	1.0	3.0	1.5	1.5
11	3.0	3.0	1.0	1.5	1.5	3.0
12	3.0	3.0	3.0	2.0	2.0	2.0
13	3.0	3.0	1.0	1.5	1.5	3.0
14	3.0	2.0	1.0	1.0	2.0	3.0
15	2.0	2.0	1.0	1.5	1.5	3.0
16	2.0	3.0	0.0	2.0	1.0	3.0
17	2.0	1.0	1.0	1.0	2.5	2.5
18	1.0	2.0	2.0	3.0	1.5	1.5
19	1.0	1.0	0.0	1.5	1.5	3.0
10	3.0	2.0	2.0	1.0	2.5	2.5
21	2.0	1.0	2.0	1.5	3.0	1.5
22	3.0	3.0	2.0	1.5	1.5	3.0
23	2.0	2.0	1.0	1.5	1.5	3.0
24	3.0	2.0	1.0	1.0	2.0	3.0
25	3.0	0.0	1.0	1.0	3.0	2.0
26	3.0	2.0	2.0	1.0	2.5	2.5
27	3.0	3.0	2.0	1.5	1.5	3.0
28	2.0	2.0	2.0	2.0	2.0	2.0
29	2.0	1.0	2.0	1.5	3.0	1.5
30	3.0	3.0	1.0	1.5	1.5	3.0
31	2.0	2.0	1.0	1.5	1.5	3.0
32	2.0	2.0	0.0	1.5	1.5	3.0
33	2.0	3.0	3.0	1.0	2.5	2.5
34	3.0	3.0	3.0	2.0	2.0	2.0
35	3.0	3.0	3.0	2.0	2.0	2.0
36	2.0	2.0	2.0	2.0	2.0	2.0
37	2.0	3.0	2.0	2.5	1.0	2.5
38	2.0	3.0	1.0	2.0	1.0	3.0
39	3.0	3.0	2.0	1.5	1.5	3.0
40	3.0	2.0	1.0	1.0	2.0	3.0
41	3.0	2.0	2.0	1.0	2.5	2.5
42	2.0	2.0	1.0	1.5	1.5	3.0
43	1.0	2.0	2.0	3.0	1.5	1.5
44	2.0	1.0	1.0	1.0	2.5	2.5
45	1.0	3.0	2.0	3.0	1.0	2.0
46	1.0	2.0	2.0	3.0	1.5	1.5
47	3.0	3.0	3.0	2.0	2.0	2.0
48	3.0	3.0	2.0	1.5	1.5	3.0
49	2.0	3.0	0.0	2.0	1.0	3.0

Table 4. The χ^2 value for each aesthetic, and the associated p (the probability that the ranks were produced by chance)

graph	aesthetic	χ^2	p
dense	bends	17.48	< 0.001
	crossings	23.09	< 0.001
	symmetry	2.3	< 0.3
sparse	bends	10.99	< 0.01
	crossings	10.56	< 0.01
	symmetry	0.27	< 0.9



	few	some	many
bends	6	18	30
crossings	6	24	42
symmetry	4.6	25.7	51

Aesthetic Variation

Fig. 3. Results for the dense graph

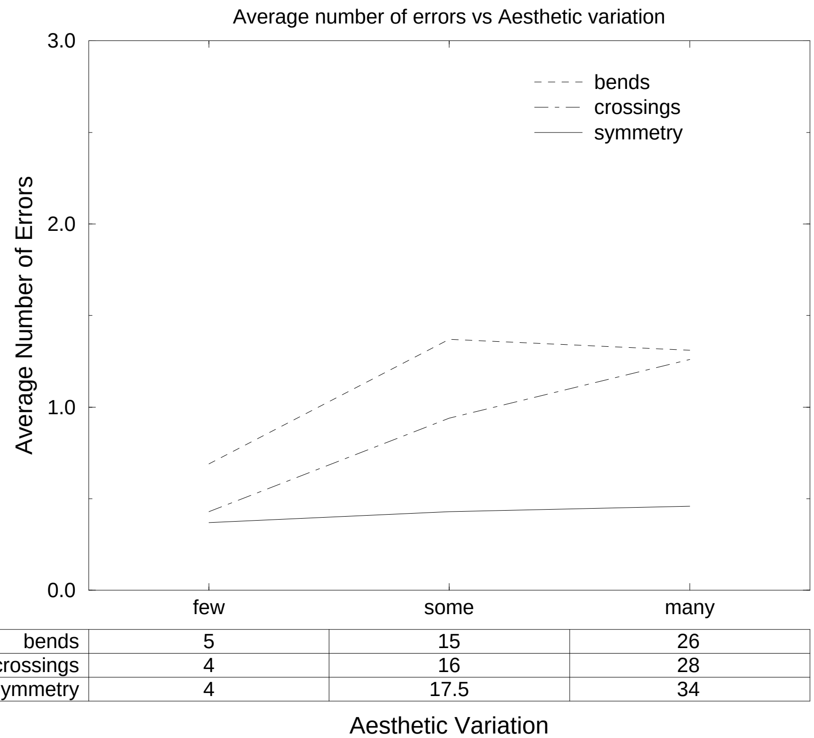
One of the questions we asked the subjects at the end of the questionnaire was: "Can you identify any characteristics of the graph drawings that made them more difficult to read?" Responses to this question included:

—"Many interwoven arcs like spiderwebs"
—"Lines crossing over"
—"Crooked lines"
—"The zig-zag lines and lack of efficient structure"
—"Disorganised, unstructured"

5. ANALYSIS

5.1 Analysis of Results

Two of our hypotheses are confirmed in both sparse and dense graphs (crossings and bends), but the symmetry hypothesis is inconclusive in both graphs. We believe that this is due to a 'ceiling effect'. The length of time for the subjects to answer the questions (45 seconds for dense, 30 seconds for sparse), was generous: most of the subjects managed to get all the questions right for the symmetrical graphs (even those which had a low measure of symmetry). This meant that there was no variability in the symmetry data, and thus the results of this analysis are not significant. In pilot tests, when we used times as high as 75 seconds for dense graphs and 60 seconds for sparse graphs, we observed a similar ceiling effect on the graph drawings for all three aesthetics.



bends	5	15	26
crossings	4	16	28
symmetry	4	17.5	34

Fig. 4. Results for the sparse graph

Additionally, symmetry is very different in nature to the other two aesthetics considered here. It is identified by humans more by pre-attentive perception than by an attentive computational quality. Other significant points which may explain why the symmetry hypothesis was not as conclusive as the other two hypotheses are:

—The symmetry metric we defined may not adequately quantify perceptual symmetry. An alternative metric could examine weighting the metric by the proportion of nodes involved in symmetrical relations as well as the proportion of arcs. We also only considered local symmetry: it is possible that running similar experiments with global symmetries might produce better results.[4]

—Secondly, these experiments tested the ability of the subjects to understand the structure of graph drawings. It is possible that maximising the amount of perceptual symmetry in a graph drawing is of more benefit when the symmetry is used for the perceptual organisation of semantic concepts instead of merely the organisation of the graph structure. Future experiments will test subjects' ability

[4]Note that by only considering local symmetries, our metric produces results that may appear anomalous: thus, a *symmetry-some* graph drawing with a few large global symmetries may appear to have as much (or more) symmetry as a *symmetry-many* drawing which has many small local symmetries. In the time since this initial project was performed, a new metric for symmetry has been defined which weights symmetric sub-graphs by their area [Purchase and Leonard 1996]: this new metric will be used in all future studies.

to understand the meaning of graph drawings from specific application areas.

—In controlling the experiment, we avoided introducing potential interaction between the aesthetics by having no variation of two aesthetics while varying the third. For example, all the graph drawings that varied the number of bends had no crossings and a zero value for symmetry. There is a potential conflict here between the nature of our three hypotheses: with bends and crossings, the hypothesis was that *increasing* their number would result in increased errors; with symmetry, the hypothesis was that *decreasing* the value of the metric would result in increased errors. By keeping the number of crossings and bends at zero for the symmetrical graph drawings, the drawings were being made simpler. This is a contrast to the bends and crossings diagrams, where keeping the symmetry metric at zero made the drawings more complex. Therefore, perhaps it is no surprise that a ceiling effect was observed with the reading of the symmetry drawings, when the same time period was used for reading all the graphs of the three aesthetics.

5.2 Discussion of Methodology

While this experiment has produced some interesting results, it has also served as a trial of a novel experimental methodology for investigating the quality of graph drawings with respect to human understandability. Every experiment has its limitations and can be improved upon, as no experiment can be perfect [Gottsdanker 1978]. The results reported here need to be considered in terms of the limitations of this particular experiment and its methodology.

By performing experiments in an area where there were no similar prior studies for us to refer to, there were a number of issues for which we needed to propose new methods, some of which may be seen as potentially limiting the application of these results.

—**A metric for symmetry:** We needed to define a computational method for measuring perceptual symmetry in a graph drawing, as this had not, to our knowledge, been done before. This definition of symmetry may not be an appropriate measure of perceptual symmetry, focusing as it does on local rather than global symmetries. Empirical tests which require that subjects produce an ordering on a series of drawings, indicating relative perceived symmetry, may assist with the production of a more appropriate measurement by identifying the specific features that contribute most to perceptual symmetry.

—**Measurement of understanding:** We needed to define how we could measure understanding. Having chosen to concentrate on a relational (or "syntactic") understanding, we needed a measurement for this understanding. While reaction time may be the most appropriate measurement, the use of paper-based experimental booklets for a large number of subjects meant that recording reaction time was impractical, and we chose to control time, and measure errors instead. Using an on-line system would enable the reaction time of each subject to each question to be measured.

—**Experimental methodology:** We needed to define the aesthetic variations, the control and independent variables, the nature of the task, and support for the experimental method. The choice of the three questions, and the introductory information and filler task all fulfilled their purpose well. The practisc drawings, and the randomisation of the order of the drawings and the choice of nodes used in the questions, ensured that the experiment was appropriately controlled for the learning effect and for the difficulty of the questions.

—**Graphs and graph drawings:** We needed to define the two graphs, identify suitable node-pairs for the three questions, and create the graph drawings by hand, carefully varying the aesthetics as needed. There was a potential conflict between the maximising and minimising hypotheses, as all three aesthetics were controlled at a value of zero. This conflict could be avoided in future versions of the experimental drawings by controlling the aesthetics at an intermediate level.

6. CONCLUSIONS

In this paper, we have described our study to validate important graph drawing aesthetics. Considering our original individual hypotheses, this study confirms two of them:

—Increasing the number of arc bends in a graph decreases the understandability of the graph.

—Increasing the number of arc crossings in a graph decreases the understandability of the graph.

These results confirm the general intuition that in order to draw a graph to maximise human understanding, arc bends and arc crossings should be minimised.

Our third hypothesis remains unconfirmed:

—Increasing the local symmetry displayed in a graph increases the understandability of the graph.

If we consider the three hypotheses together, although we cannot conclusively state that the symmetry hypothesis is confirmed, we can conclude that increasing the number of crossings and bends is more detrimental to the understandability of the graph than a reduction in symmetry. The similarity in the results for crossings and bends makes it difficult to place any priority of one over the other. The comparisons between the aesthetics are made possible because aesthetic values were kept constant between the graph drawings so that there were no confounding factors. For example, all the drawings which were related to crossings and all those relating to symmetry had no bends.

This result is important in the design of graph-drawing algorithms which intend to produce drawings that maximise understandability. If, for example, a compromise needs to be made between the competing aesthetics of crossings and symmetry, the study reported here clearly indicates which aesthetic is more important. By taking this information into account, the algorithm designer can produce more understandable drawings.

6.1 Further Work

This initial study was performed using paper-based experimental booklets, as a means of getting data, as well as of trialling the experimental method. In future studies, we intend to conduct the experiments on-line, using an interactive experimental system. This will allow us to:

- —measure the time taken to answer the questions correctly as the dependent variable, rather than the number of errors made within a fixed time. This may be a more appropriate measure of understanding.[5]
- —study the effect of aesthetics on human understanding of graphs drawn on a display device, rather than on paper.

Future work also includes the following:

- —studying other graph drawing aesthetics (for example, orthogonality, maximising the minimum angle, etc).
- —studying the effectiveness of various graph drawing algorithms.
- —studying the effectiveness of graph drawing aesthetics when used in application domains (for example, software engineering design diagrams).

ACKNOWLEDGMENTS

The authors would like to thank the following people and organisations for their helpful advice and participation: Shelina Bhanji, Julie McCreddon, Tim Mansfield, and the staff and students of the University of Queensland Computer Science Department. This work is funded by the Australian Research Council.

APPENDIX

A. A METRIC FOR SYMMETRY

A measurement of the amount of symmetry in a graph drawing was an essential part of the study: we needed to be able to vary the symmetry displayed in a graph drawing, and to have some metric by which we could measure it. For the purposes of this experiment, symmetry is quantified by [Bhanji et al. 1995]:

(1) counting all the symmetrical relationships about each axis of symmetry.
(2) adding the symmetry values from all the axes together.
(3) multiplying this total by the proportion of the graph drawings' arcs which are symmetrical objects.

This resulted in the following formula for measuring symmetry:

$$\sigma = \left(\frac{A_o}{A_{total}}\right) \times \sum_{i=1}^{M} \left(\frac{A_{di}}{2} + \frac{N_i}{2} + A_{si}\right)$$

where:

[5]In this initial study, it was impossible to measure time as the dependent variable on a large number of subjects using paper-based experimental booklets.

—σ is the measure of symmetry; the higher the value of σ, the more symmetrical the drawing
—A_o is the number of arcs in the graph which are symmetrical objects
—A_{total} is the total number of arcs in the graph
—M is number of axes of symmetry
—A_{di} is the number of double symmetrical arcs about axis i
—N_i is number of symmetrical nodes about axis i
—A_{si} is the number of single symmetrical arcs about axis i

It appears that this metric gives an appropriate measure of the perceived symmetry of a graph drawing [Bhanji et al. 1995].

Definitions

An *axis of symmetry* is defined as a straight line that has a minimum of three symmetrical objects (nodes (N), singly symmetrical arcs (A_s) and doubly symmetrical arcs (A_d)) mirrored on either side. The minimum of three symmetrical objects prevents an axis of symmetry being drawn at right angles, midway through every arc in a graph. Thus preventing every arc and pair of directly connected nodes contributing to symmetry, regardless of the overall appearance of the graph.

Around any axis of symmetry is a group of symmetrical objects. Collectively these objects make a *locality of symmetry*. While a *locality of symmetry* is defined by one axis of symmetry, the same group of symmetrical objects can have numerous axes. There is no restriction on the direction of these axes, but for each *locality of symmetry*, no two axes can be parallel. This prevents the same symmetrical relationship being counted over multiple axes.

The following objects are considered *symmetrical objects*:

—A double symmetrical arc (A_d) is an arc that is mirrored by another arc about the axis of symmetry.
—A single symmetrical arc (A_s) is a single arc that is bisected at right angles by the axis of symmetry.
—An axially symmetrical arc (A_a) is an arc that runs along the axis of symmetry.
—A symmetrical node (N) is a node that is mirrored by another node about the axis of symmetry.

The following three relationships are considered *symmetrical relationships*:

—Two mirrored double symmetrical arcs (A_d).
—Both halves of a single symmetrical arc (A_s).
—Two mirrored symmetrical nodes (N).

The final metric value is obtained by multiplying the number of symmetrical relationships in the graph by the *proportion of arcs which are symmetrical*. Thus, axially symmetrical arcs (A_a) contribute to this proportion even though they are not involved in a symmetrical relationship. If all the arcs in a graph are symmetrical objects then σ is simply the total number of symmetrical relationships over all axes of symmetry.

B. THE FRIEDMAN ANALYSIS OF VARIANCE BY RANKS

In this experiment, the error scores were ranked within the subject. Thus if a subject got 0 errors for *bends-few*, 1 error on *bends-some*, and 3 errors on *bends-many*, the ordinal nature of these error values was coded by ranking the scores as *bends-few*=3, *bends-some*=2, *bends-many*=1 (see Table 3). As the number of errors may not be an interval measure of ease of comprehension and since the small range of the error data possibly restricted it from being normally distributed, a non-parametric analysis of variance was used.

The Friedman Analysis of Variance test was performed on the ranked data, where matching across conditions was fulfilled by using a within subjects design. If there were no differential effect of a given aesthetic on the level of difficulty of the drawings, we would expect the ranks for that aesthetic to be spread evenly across the three drawings. However, if there were a differential effect on the number of errors, we would expect the ranks to be clustered in specific conditions.

The χ^2 value was calculated on this ranked data as follows:

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1)$$

where

—N = number of subjects (49 for the dense graph; 35 for the sparse graph)

—k = number of conditions = 3

—R_j = sum of ranks for the jth condition

By comparing the χ^2 value obtained by this formula with a table of critical values for χ^2, the probability (p) that the ranks were produced by chance can be determined (see Table 4). Those aesthetics for which $p < 0.05$ were deemed to have a significant effect on the number of errors.

REFERENCES

BATINI, C., FURLANI, L., AND NARDELLI, E. 1985. What is a good diagram? a pragmatic approach. In *Proc. 4th Int. Conf. on the Entity Relationship Approach* (1985).

BHANJI, S., PURCHASE, H., COHEN, R., AND JAMES, M. 1995. Validating graph drawing aesthetics: A pilot study. Technical Report 336, University of Queensland Department of Computer Science.

DING, C. AND MATETI, P. 1990. A framework for the automated drawing of data structure diagrams. *IEEE Transactions on Software Engineering SE-16*, 5, 543–557.

EADES, P. 1984. A heuristic for graph drawing. *Congressus Numerantium 42*, 149–160.

FERRARI, D. AND MEZZALIRA, L. 1969. On drawing a graph with the minimum number of crossings. Technical Report 69-11, Istituto di Elettrotecnica ed Elettronica, Politecnico di Milano.

FORMANN, M., HAGERUP, T., HARALAMBIDES, J., KAUFMANN, M., LEIGHTON, F., SIMVONIS, A., WELZL, E., AND WOEGINGER, G. 1993. Drawing graphs in the plane with high resolution. *SIAM Journal of Computing 22*, 5, 1035–1052.

GARG, A. AND TAMASSIA, R. 1994. Planar drawings and angular resolution: Algorithms and bounds. In J. VAN LEEUWEN Ed., *Proceedings of the Second Annual European Symposium on Algorithms, ESA94*. Utecht, The Netherlands: Springer-Verlag. Lecture Notes in Computer Science, 855.

GOTTSDANKER, R. 1978. *Experimenting in Psychology.* Prentice-Hall.

LIPTON, R., NORTH, S., AND SANDBERG, J. 1985. A method for drawing graphs. In *Proc. ACM Symp. on Computational Geometry* (1985), pp. 153–160.

LOHSE, G., BIOLSI, K., WALKER, N., AND RUETER, H. 1994. A classification of visual representations. *Communications of the ACM 37*, 12 (December), 36–49.

PURCHASE, H. AND LEONARD, D. 1996. Graph drawing aesthetic metrics. Technical Report 361, University of Queensland Department of Computer Science.

SIEGEL, S. 1956. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill.

TAMASSIA, R. 1987. On embedding a graph in the grid with the minimum number of bends. *SIAM J. Computing 16*, 3, 421–444.

TRICKEY, H. 1988. Drag: A graph drawing system. In *Proc. Int. Conf. on Electronic Publishing* (1988), pp. 171–182. Cambridge University Press.

VIJAYAN, G. 1986. Geometry of planar graphs with angles. In *Proc. ACM Symp. on Computational Geometry* (1986), pp. 116–124.

WARE, C., HUI, D., AND FRANCK, G. 1993. Visualizing object oriented software in three dimensions. In *CASCON 1993 Proceedings* (1993).