Graph Drawing Aesthetics—Created by Users, Not Algorithms

Helen C. Purchase, Christopher Pilcher, and Beryl Plimmer

Abstract—Prior empirical work on layout aesthetics for graph drawing algorithms has concentrated on the interpretation of existing graph drawings. We report on experiments which focus on the creation and layout of graph drawings: participants were asked to draw graphs based on adjacency lists, and to lay them out "nicely." Two interaction methods were used for creating the drawings: a sketch interface which allows for easy, natural hand movements, and a formal point-and-click interface similar to a typical graph editing system. We find, in common with many other studies, that removing edge crossings is the most significant aesthetic, but also discover that aligning nodes and edges to an underlying grid is important. We observe that the aesthetics favored by participants during creation of a graph drawing are often not evident in the final product and that the participants did not make a clear distinction between the processes of creation and layout. Our results suggest that graph drawing systems should integrate automatic layout with the user's manual editing process, and provide facilities to support grid-based graph creation.

Index Terms—Evaluation/methodology, graphs and networks, user interfaces.

1 Introduction

Typically, the aesthetic criteria for graph layout algorithms have been based on the intuition of the algorithm designers. Some empirical work has been done in an attempt to verify the usefulness of these criteria in reading and understanding graphs (as measured by graph tasks, often shortest-path tasks) [1]. More recently, work has been done on the process of graph comprehension, using eye-tracking data [2]. These empirical studies have all related to the reading and understanding of graphs, rather than their creation.

Two recent empirical studies have investigated a more hands-on approach to users' perception of graph drawings. van Ham and Rogowitz [3] considered the creation of graph layouts, looking at how people prefer to layout graphs when they have the opportunity to move nodes and edges around. This is a very different empirical task to that of reading and interpreting a graph, and van Ham and Rogowitz [3] investigated, in particular, the depiction of clusters, the presence of edge crossings, edge length distribution, and orientation. Dwyer et al. [4] performed a similar experiment using larger graphs, and two different interfaces.

The work presented here builds on the work of van Ham and Rogowitz [3] and Dwyer et al. [4], but has an important difference: our participants drew the graph from scratch, rather than simply moving nodes in a predrawn graph—this removed any layout bias present in the original drawing. As in Dwyer et al. [4], our experiment was run twice, with two different interfaces: sketch and formal.

- *H.C. Purchase is with the School of Computing Science, University of Glasgow, Glasgow G12 8QQ, United Kingdom. E-mail: helen.purchase@gla.ac.uk.*
- *C. Pilcher and B. Plimmer are with the Computer Science Department, University of Auckland, Science Centre Building 303, 38 Princes St, Auckland 1010, New Zealand. E-mail: cpil016@aucklanduni.ac.nz, beryl@cs.auckland.ac.nz.*

Manuscript received 5 May 2010; revised 20 Sept. 2010; accepted 20 Nov. 2010; published online 16 Dec. 2010.
Recommended for acceptance by R. Balakrishnan.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-2010-05-0095.
Digital Object Identifier no. 10.1109/TVCG.2010.269.

This paper describes previous graph drawing empirical work and discusses the work and conclusions of van Ham and Rogowitz [3] and Dwyer et al. [4]. It describes the experiment, and presents the results. We have already reported on a preliminary exploratory study of the sketched drawings [5]; here, we present the results of the formal experiment, together with more extensive sketch experiment data, and statistical analysis of data comparing the two modes of interaction.

Three aspects of the human graph drawing process are discussed: the product, the process, and the preferences.

2 Related Work

Many graph layout algorithms that have been devised over several decades [6] have typically been designed in accordance with the intuitions of the algorithm designers. Over the years, a set of assumed "graph drawing aesthetics" has emerged, defining the criteria by which the "goodness" of the graph drawing produced by a layout algorithm can be assessed [7], [8]. Such aesthetics include, for example, a minimum number of edge crossings, as few edge bends as possible, a display of symmetric substructures, and large angles between edges incident at a node. Graph layout algorithms therefore tend to be valued for the extent to which their output drawings conform to these aesthetic criteria.

Graph drawings should also be assessed according to the extent to which they assist human comprehension of the relational information represented in the graph. Some empirical work has been done to this end, investigating whether the aesthetic criteria used by algorithm designers do indeed assist with comprehension. Findings include the overwhelming evidence for the reduction of edge crossings [1], [9], [10], some evidence for the reduction of bends and depiction of symmetry [1], placement of important nodes at the top of the graph [11], and large angles between incident edges [2]. All these studies have been conducted by asking participants to answer graph-based questions on a variety of presented graph drawings, each carefully controlled for the aesthetic criteria.

The recent publication by van Ham and Rogowitz [3] has taken a different empirical approach to determining the best graph layout for human use. They asked participants to manually adjust the layout of existing graph drawings: "rearrange the nodes in the network in a way that you think best reflects their interconnections." They used four graphs of 16 nodes, each with two clusters: these clusters were separated by one, two, three, and four edges, respectively. These were both presented in a circular and a spring layout [12], giving a total of eight starting diagrams, which were presented in random order. Users of the Many Eyes visualization service [13] were invited to take part in the experiment by adapting the layout of the drawings. They collected 73 unique drawings, which they visually analyzed according to the number of edge crossings and evidence of clustering, as well as other computational measures such as edge length distribution and cluster distance. They found that most participants separated the two node clusters, the human drawings contained 60 percent fewer edge crossings than the automatically produced drawings, and that humans did not value uniform edge length as much as the spring algorithm did.

van Ham and Rogowitz [3] acknowledge the limitations of their work. In particular, using a web-based experiment means that they have no information on their participants, apart from the fact that they had "some sophistication in data visualization." In addition, the fact that the graphs were presented with an initial layout may have biased the resultant drawings, and the names used as node labels may have had an effect (for example, when participants attempted to avoid label overlaps).

Dwyer et al. [4] used two interface conditions: surface and mouse. Participants were asked to layout two social networks of 50 nodes and 74 and 77 edges, respectively, each with a circular initial arrangement. The focus of the experiment was for the participants to lay the graphs out in a way that would best support the identification of cliques, chains, cut nodes, and leaf nodes. Dwyer et al. [4] focused their analysis on the process of layout with the two interfaces rather than on the product, and the only observation that they make about the graphs produced is that users removed edge crossings. The interface method appeared to have no effect on the number of crossings.

The experiment reported here improves on the methodology of van Ham and Rogowitz [3] and Dwyer et al. [4] in several important ways:

- The participants had to both draw the graph, as well as lay it out, a more complex task than both van Ham and Rogowitz [3] and Dwyer et al. [4].
- The experiment was conducted face to face, so demographic information about the participants is available and we know that all participants did all drawings (this was not the case for van Ham and Rogowitz [3]).
- The participants drew the graphs from scratch, so were not biased by any initial layout (both van Ham and Rogowitz [3] and Dwyer et al. [4] used an initial configuration).
- A sketching tool was used, so the physical drawing process was unhindered by a clumsy editing process and participants could draw curved or bent lines if they wished (this interface is similar in interaction method to the surface interface used by Dwyer et al. [4]).
- We collected video data, so were able to analyze both the process and product of creation (this was done by Dwyer et al. [4], but not van Ham and Rogowitz [3]).
- We discussed layout preferences with the participants in a post-experiment interview, so were able to find out more about their thoughts on the process (this was done by Dwyer et al. [4], but not van Ham and Rogowitz [3]).
- Our node labels were simple letters, enclosed within the node boundary (van Ham and Rogowitz [3] used name labels, and reported problems with doing so; Dwyer et al. [4] used smiley faces).

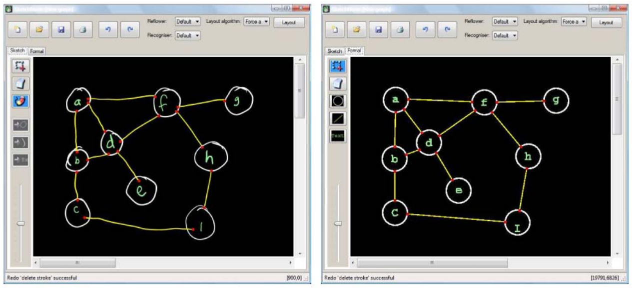


Fig. 1. SketchNode in sketch mode and formal mode.

Including the task of graph creation as well as layout and using a comprehensive, face-to-face methodology meant that the graphs in our experiment are not as large as those in either van Ham and Rogowitz [3] or Dwyer et al. [4], so as to make the duration of the experiment acceptable to participants. We used four graphs (two practice graphs and two experimental graphs), and the graphs were limited to 10 nodes each.

Our experiment has produced extensive and rich data, in terms of product, process, and preferences. The results can inform the design of automatic graph drawing algorithms by highlighting those features that users consider important when they are unconstrained in their own drawing of graphs and are not subject to any layout bias.

3 Graph Drawing Experiment

Our primary research question is: "*Which graph drawing aesthetics do people favor when creating their own drawings of graphs?*" This is an exploratory question: we asked participants to draw graphs and analyzed their drawings for evidence of the common aesthetics. Although we had not originally thought that the actual process of drawing the graphs would be of interest, analysis of the video screen casts revealed a tendency for some aesthetics (not present in the final product) to be followed during the drawing process. Preference data collected by interview were analyzed to support the findings.

Our secondary question is: "*Do the aesthetics favored differ between two different interface styles?*" In this case, a between-participants experimental design was followed, with some participants using a sketch interface, and some using a formal interface. Data representing both the extent of conformance to aesthetics in the drawings as well as the drawing process were statistically analyzed.

3.1 Equipment

A graph drawing sketch tool, SketchNode [14] (Fig. 1) was used on a tablet PC. The tool has two drawing modes: *sketch* and *formal*. The sketch mode allows nodes, edges, and node labels to be drawn with a stylus on the tablet screen, laid flat, thus allowing the same hand movements as pen and

Graph A	(A,D) (A,C) (B,D) (C,D) (B,C) (B.E) (C,E) (E,J) (F,G) (J,F) (F,I) (G,I) (J,H) (I,H)
Graph B	(J,F) (J,I) (G,I) (H,I) (G,H) (F,H) (G,J) (F,A) (F,G) (G,E) (A,E) (D,E) (D,C) (D,B) (C,B) (A,B) (A,C) (B,E)

Fig. 2. Adjacency lists for the experimental graphs.

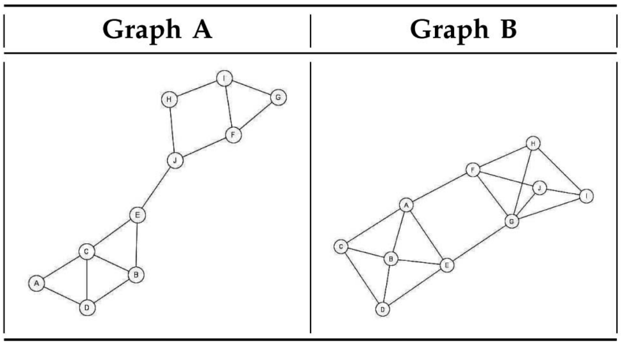


Fig. 3. Graphs A and B.

paper, and giving a more natural interaction than using an editing tool. Edges can be created at any time. Unlike pen and paper, however, the SketchNode interface allows nodes (or groups of nodes) to be selected and relocated (with corresponding movement of attached edges, which are straightened when moved). Nodes and edges can be erased. It thus has the advantages of pen and paper, as well as the advantages of a graph drawing editing tool.

The formal mode is similar to the many currently available graph drawing systems (for example, yEd). While the stylus can still be used on the tablet interface, nodes are created by a single tap on the panel, labels are entered using an on-screen keyboard, and straight edges are created by dragging the stylus between nodes. Unlike sketching, an edge cannot be created unless it connects two existing nodes.

3.2 Task

The experiment was run twice, once in sketch mode and once in formal mode. The experiment was exactly the same for both modes. Participants were given an adjacency list of edges (Fig. 2) and asked to draw the graph in SketchNode, with the instruction to *Please draw this graph as best as you can so to make it "easy to understand."* They were deliberately not given any further instruction as to what "*easy to understand*" means. In particular, they were not primed with any information about common graph layout aesthetics, for example, minimizing edge crossings, use of straight lines, etc. They were given as long as they liked to draw and adjust the layout of the graphs.

3.3 Graphs

We designed two experimental graphs: graph A had 10 nodes and 14 edges; graph B had 10 nodes and 18 edges. We were unable to use graphs as big as those used by van Ham and Rogowitz [3] and Dwyer et al. [4], as we were required to keep the duration of the experiment to a reasonable time, and the screen size of the tablet PC was limited. However, these graphs were still designed with the aims of van Ham and Rogowitz [3] in mind, as they both had identifiable clusters: graph A had two clusters separated by one cut edge; graph B had two clusters separated by two cut edges. Both graphs are planar. Fig. 3 shows these two graphs produced by yEd's spring layout algorithm [15], clearly showing the two clusters.

3.4 Experimental Process

After reading the information sheet and signing the consent form, participants answered a pre-experiment questionnaire which asked for some demographic information, including information about their experience with maths, theoretical computer science (CS), graphs, and pen-based technology.

The participants were then given a demonstration of the SketchNode system, including all the relevant interface features for the mode they were working in: node and edge creation using the stylus, bends, and curved lines, selecting and moving nodes and edges, selecting and moving subgraphs, labeling nodes, erasing, undoing, and redoing actions, zooming, and scrolling. They were also shown how to represent a simple four-element adjacency list as a four-node and four-edge graph drawing. Participants were given ample chance to ask questions about the SketchNode system and the process of representing an adjacency list as a graph drawing.

Besides the two experimental graphs, A and B, two practice graphs were defined, $\mathrm{P1}(n = 5, e = 6)$ and $\mathrm{P2}(n = 8, e = 8)$, and these were presented first. Participants were not aware that these were practice graphs—their use ensured that the participants were comfortable with the task and with the system before they drew the two experimental graphs that we were interested in. Exactly the same instructions were given to the participants for the practice graphs as for the two experimental graphs: *Please draw this graph as best as you can so to make it "easy to understand."*

The two experimental graphs A and B, were then presented to the participants, with the edges listed in a different random order for each participant. Each experiment was conducted individually, with only the experimenter and participant present. The order of presentation of the two graphs was switched between participants so as to control for any possible ordering effects.

3.5 Participants

The data from 34 participants were collected for analysis: 17 each for the sketch and formal experiments. The participants are numbered 3-19 for the sketch experiment and 2-18 for the formal experiment, as the initial experiments were pilots in both cases. Participants were friends, family, and classmates of the student experimenters, and were students and non-students, of both genders. Only some of the student participants were studying Computer Science (Table 1).

Seven of the participants took part in both the sketch and the formal experiments; as there was a period of over six months between the two experiments, and the interaction method used for drawing the graphs was very different, it is unlikely that participants would have remembered the graphs well enough for performance in the former experiment to affect that in the latter one.

3.6 Data Collection

All interactions with the tablet were recorded using Morae software [16], producing video screen casts of all the interactions with the interface, as well as a corresponding audio track. The time taken for the drawing of the graphs was recorded.

Our notational convention is: graph A drawn by participant 3 is **3A**; graph B drawn by participant 8 is **8B**. The suffix **s** or **f** indicates sketch or formal mode.

TABLE 1
Demographic Information about the Participants

	Sketch 17 participants	Formal 17 participants
Gender	7F, 10M	6F, 11M
Occupation	10 students	7 non-students
Age	max 49, min 20, median 24, mean 27.65	max 57, min 19, median 24, mean 28.9
Education	13 have some university-level education	10 have some university-level education
Experience of graphs	8 yes, 9 no	8 yes, 9 no
Experience of university maths/CS	8 yes, 9 no	8 yes, 9 no
Pen-based technology experience	4 none, 9 minimal, 3 moderate, 1 extensive	3 none, 8 minimal, 5 moderate, 1 extensive

At the end of the experiment, the participants were asked "*Why did you arrange the graphs in the way you did?*" in a recorded interview.

4 Results

Analyzing the sketched graphs produced and the graph drawing process captured by the screen cast videos necessarily required some subjective judgements to be made. In all cases, the coding was done first by one author, verified and reverified by another author, and then confirmed by the third author.

4.1 The Product: What Do the Drawings Look Like?

Of the 34 sketch drawings (Fig. 4), three were incorrect. 15As had an additional edge (G,H), 19As represents node H twice, and 16Bs is missing the (D,C) edge : this had been drawn by the participant at the start, but had been lost in the later editing process.

Of the 34 formal drawings (Fig. 4), three were incorrect. 8Af was missing (A,F), 8Bf was missing (F,A), and 9Bf was missing (A,E). As the focus of the experiment was on how participants represented graphs (and not on whether they drew them correctly or not), incorrect graphs were not removed from the analysis.

The graph drawings were analyzed for the following layout features used by van Ham and Rogowitz [3] (Table 2):

- *Number of edge crossings.* Points outside of the node boundaries where one or more edges cross.
- *Representation of clusters.* True, if a straight line can be drawn through the graph drawing so that the two predefined clusters in the graph can be visually separated.
- *Number of hulls.* Clusters that are bounded by edges. This feature is either 1 or 2, and is not applicable if the "Representation of clusters" feature is false.
- *Cluster extraction.* The ratio of the average inter-node distances within clusters to the average inter-node distances over the whole drawing; the smaller the value, the more the clusters have been separated.

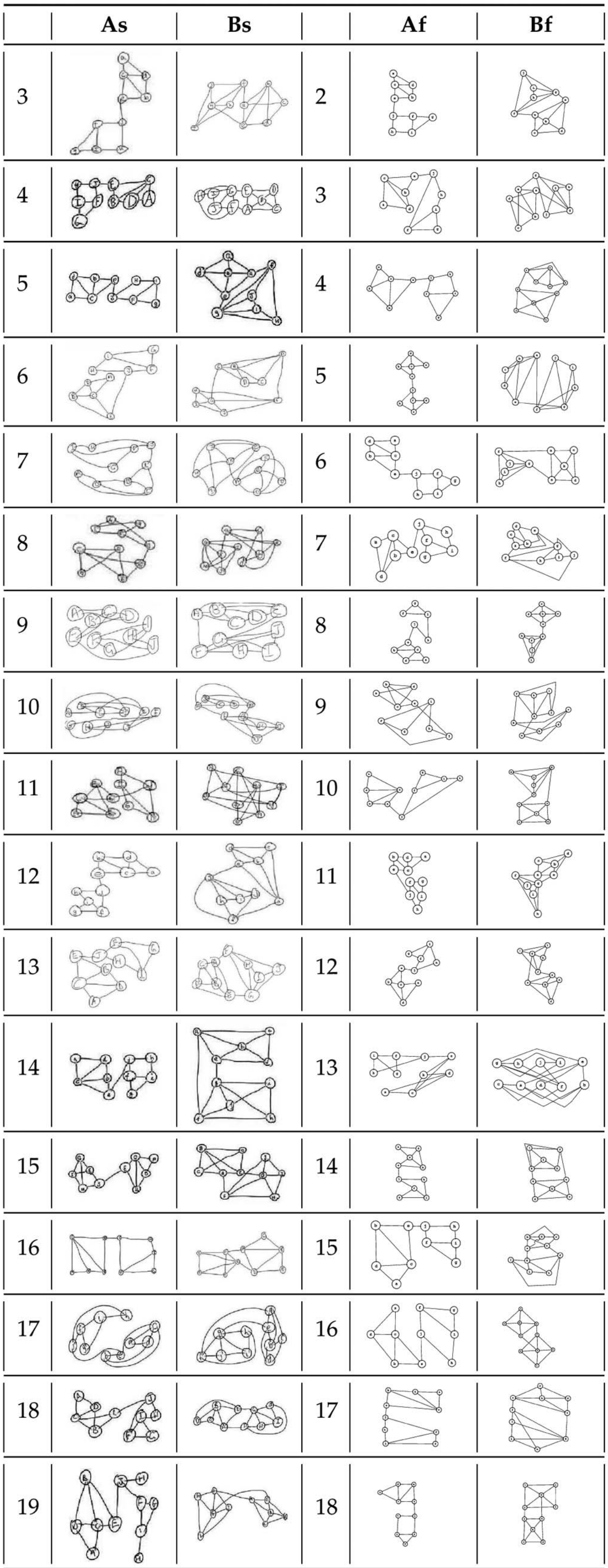


Fig. 4. The graph drawings in sketch and formal modes.

- *Cluster distance.* The minimum and average inter-node distance between nodes in the two clusters, giving an estimate of the relative positioning of the clusters.

TABLE 2
Features of the Drawings

		GA $n = 10, e = 14$		GB $n = 10, e = 18$		p-value when comparing sketch and formal modes
		sketch	formal	sketch	formal	
Edge crossings	Total number of edge crossings	23	11	51	39	0.457
	Mean	1.35	0.65	3	2.29	
	Median	0	0	2	1	
	Max	7	6	22	18	
	Mean percentage edge crossings	1.6%	0.8%	2.1%	1.6%	0.367
	Drawings with zero crossings	11	16	6	8	0.143
	Total edge crossings excluding obvious outliers (11Bs and 13Bf)	23	11	29	21	0.268
	Mean percentage of edge crossings excluding outliers	1.6%	0.8%	1.3%	0.9%	0.250
Clusters and Hulls	Drawings with two clusters	10	12	11	10	0.805
	Mean percentage visible clusters represented as hulls	65.0%	92.3%	100%	100%	0.172
Cluster extraction ratio	Mean internode distance within clusters/mean internode distance in the whole graph	0.75	0.73	0.78	0.77	0.602
Cluster distance	Min	0.23	0.21	0.22	0.21	
	Mean	0.60	0.62	0.57	0.61	0.036*
Vertical/horizontal edges	Mean percentage edges horizontal or vertical	38.2%	42.0%	29.1%	27.1%	0.874
	Graphs clearly drawn with a grid arrangement in mind	6	9	4	4	0.449
Straight edges	Drawings with all edges straight	12	16	6	11	0.021*
	Mean percentage of edges straight	89.1%	99.6%	87.3%	93.5%	0.068*
	Drawings with no edges straight	1	0	1	0	
Edge length distribution	Mean variance of edge lengths within the drawings (ELD)	0.243	0.224	0.240	0.241	0.250

In all four cases, the number of graphs drawn was 17—a total of 68 drawings.

- *Edge length distribution (ELD).* The standard deviation of the normalized edge lengths within each graph drawing: a low standard deviation indicates that edge lengths are uniform in length.

We also included additional measures in our analysis:

- *Number of straight lines.* In sketching mode, the lines are unlikely to be geometrically straight, so a vertical assessment as to whether the edge was intended to be straight was made; in formal mode, it was possible for edges to be drawn with bends.
- *Number of vertical or horizontal edges.* A visual assessment was made as to whether the edge was intended to be horizontal or visual for both sketch and formal modes.

4.1.1 The Nature of the Drawings

Analysis of the 68 drawings revealed that even though some graphs were drawn with several edge crossings (e.g., 8Bs, 11Bs, and 13Bf), most drawings had been drawn so as to minimize crossings: 40 drawings had zero crossings, while eight had one crossing and six had two: thus, only 14/68 (20 percent) drawings had more than two crossings. It is clear that this was the most important layout aesthetic used by the participants, and this result concurs with other empirical studies (e.g., [1]) and with the findings of van Ham and Rogowitz [3]. 15Bf shows an interesting method of avoiding a crossed edge.

Most participants recognized the presence of the two clusters, and separated them appropriately. van Ham and Rogowitz [3] explicitly asked participants to layout the graphs so that they best *...reflect[ed] their interconnections*; in our experiment, we were not so explicit in our instructions regarding the connectedness of the graph—despite this, our participants still performed well when it came to grouping clustered nodes together. This can be seen visually from the diagrams, although the relatively high cluster extraction ratios (0.73-0.78) do not support this computationally; the ratio reported in van Ham and Rogowitz [3] for clusters separated by one cut edge is approximately 0.6. The between cluster distances show (relatively) how far apart the clusters were drawn; it is unsurprising that the clusters in graph B are closer together (0.62, 0.61) than those of graph A (0.60, 0.57), as they have two connecting edges (a result concurring with van Ham and Rogowitz [3]). All graph B clusters were surrounded by hulls; the sketch version of graph A only has 65 percent of visible clusters surrounded by hulls.

Approximately one third of all edges drawn were aligned along the horizontal or vertical axes, and 23 of the 68 drawings were clearly drawn with a grid-like formation in mind (i.e., most edges are horizontal or vertical, and most nodes are aligned vertically and horizontally, e.g., 5As, 14Bs, and 2Af).[1] van Ham and Rogowitz [3] do not analyze their drawings with respect to horizontal or vertical edges, and their published examples show little evidence of this feature. It is possible that their starting configurations of circular and spring layouts, neither of which favors the

1. The drawings determined to be "grid-like" are 3As, 4As, 4Bs, 5As, 12As, 14As, 14Bs, 15Bs, 16As, 16Bs, 2Af, 5Af, 6Af, 6Bf, 11Af, 14Af, 14Bf, 15Af, 16Af, 16Bf, 17Af, 18Af, and 18Bf.

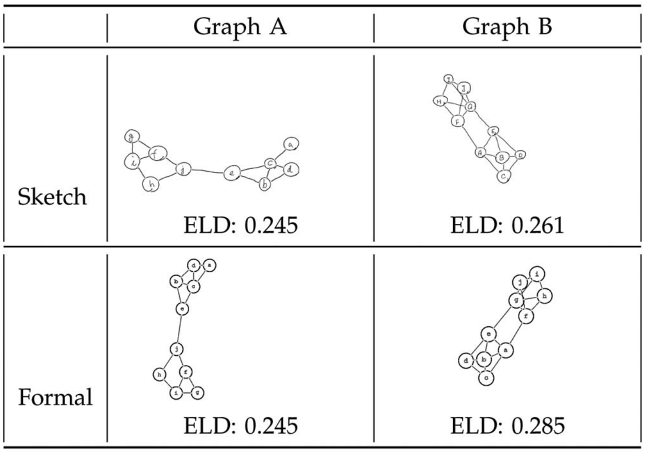


Fig. 5. Graphs A and B laid out using the SketchNode spring algorithm, and their edge length distributions.

presentation of edges along grid lines, may have meant that participants were biased away from this feature.

Participants were reluctant to produce final drawings with curved or bent lines: although only 45 of the 68 drawings comprised only straight lines, the overall percentage of straight lines is high (92.3 percent). Curved and bent lines are typically used at the edge of the drawing to avoid edges crossing nodes or other edges (e.g., 4Bs, 17As, 9Bf, and 14Bf). There was only one bent edge in all 17 formal drawings of graph A. SketchNode straightens associated curved edges when nodes are moved in sketch mode, but this affected the nature of the final drawings for only two participants (4 and 14). In both these cases, the participants drew the whole graph first before moving any nodes, indicating that they were unconcerned with the initial shape of the graph (including the curved edges) as they knew that they were going to subsequently adapt the whole drawing to make the layout more acceptable.

The edge length distributions are small (0.224-0.243), indicating that there was tendency for the participants to draw edges of equal length—but these are not as small as those reported by van Ham and Rogowitz [3] (0.1 for the graph with one connecting edge). Our ELD results are comparable, however, to those of the graph drawings produced by the SketchNode spring algorithm (0.207-0.285), an algorithm which has producing similar edge lengths as one of its aims (Fig. 5).

4.1.2 Comparison between Interface Modes

In comparing performance between the two interface modes, data from the two graphs were aggregated. In considering the extent of bent edges and horizontal and vertical edges, percentage values were computed (using the total number of edges in each graph as the maximum) and the mean taken. It is more difficult to measure the extent of crossed edges in a graph drawing, as the upper limit of the number of possible edge crossings depends on the graph structure; the formula suggested by Purchase [17] was used to derive a value to represent the percentage of crossing edges in a graph drawing—these were used to aggregate the crossings data for the two graphs using a mean.

The final column in Table 2 shows the p-values derived from the independent-measures t-tests: these values show the probability that the different mode of interaction (sketch or formal) affected the measured feature of the drawing.

Using the typical probability level of $p = 0.05$, two results are significant, and one approaches significance:

- Mean cluster distance ($p = 0.036$). The mean distance between the clusters in the formal drawings was significantly less than in the sketch drawings.
- Number of drawings with all edges straight ($p = 0.021$). There were significantly more formal drawings with all edges straight than sketch drawings.
- Mean percentage of edges straight ($p = 0.068$, approaches significance). There were a greater percentage of straight edges in formal drawings than in sketch drawings. Only 21 of a total of 541 of the edges drawn in formal mode are not straight.

There are therefore two main conclusions we can make about the differences between the graphs as drawn in the two modes: the clusters were closer together in the formal drawings, and there were more straight edges in the formal drawings. Neither of these results is particularly surprising; they suggest that the formal mode encouraged the creation of a more constrained, regular, and careful drawing than the sketch mode, which permitted more flexibility and expressive freedom. The surprising result is that there was no significant difference between the extent to which a grid formation was used, where we might have expected there would be more use of horizontal and vertical edges in the formal drawing. This was not the case: even the sketch participants made extensive use of an underlying grid.

4.2 The Process: How Were the Graphs Drawn?

Analyzing the screen cast videos proved very revealing: it showed that the final product seldom represented the layout strategy used by the participant when drawing the graph, and that aesthetic criteria emphasized by participants early in the creation process were often compromised as the graph grew in size.

The screen casts were analyzed for the following features (Table 3):

- The order and timing of drawing the nodes.
- If and when nodes or sub-graphs were moved to different locations on the panel.
- Use of straight or curved lines.
- The variation in the length of edges.
- Approximate alignment to a horizontal/vertical grid.
- Evidence of participants analyzing the adjacency list and planning ahead before drawing, suggested by substantial pauses.

In all cases, the participant was deemed to have used the drawing process if it had been used at least once in the drawing of either graph.

4.2.1 The Nature of the Drawing Process

As expected, Graph B took significantly longer to draw than Graph A, as it had more edges ($p < 0.001$), and the time taken to draw the graphs varied considerably between participants (Fig. 6).

Analysis of the screen casts revealed that there was seldom a clear break between the process of creating the drawing (i.e., representing all the information in the adjacency list) and the process of arranging the drawing so as to make it "easy to understand." In most cases, node

TABLE 3
Features of the Drawing Process, Showing How Many Participants Adopted the Different Strategies

	Strategy	Sketch Number of participants (out of 17)		Formal Number of participants (out of 17)		p-value when comparing sketch and formal modes
Drawing nodes	Draw all nodes first	6	One participant used a mix of the strategies	6	One participant changed her strategy between graphs	1.000
	Draw nodes where convenient for the placement of the next edge	12		12		1.000
Moving nodes	Move nodes during creation	10	One participant changed strategy between graphs. One participant moved nodes during creation, but not sub-graphs	16	Ten participants moved nodes during creation, but not sub-graphs	0.014*
	Move sub-graphs during creation	9		9		1.000
	Move nodes after creation	8		11		0.315
Edges	Mainly straight lines	14	This process feature is derived independently of the number of straight lines in the final product, as SketchNode straightens curved edges when nodes are moved in sketch mode.	16	The default in formal mode is for lines to be drawn straight, although bends can be inserted; one participant, 13, made excessive use of bends.	0.301
	Only use curved edges to avoid crossings	10	These participants favoured straight lines, and only introduced curved lines when necessary	9	These participants did not use bends unless necessary; of these, three subsequently removed these bends	0.739
	Use similar length edges during creation	11		2		0.001*
Alignment	Favour horizontal and vertical edges during creation	10		7		0.318
Analyse	Plan ahead	5	This was evidenced by excessive pauses, or self-reporting	4	This was evidenced by excessive pauses, or self-reporting	0.708

positioning decisions were made during the creation process as participants placed new nodes close to their connecting node. This means that no timing data could explicitly be associated with the process of graph layout (as opposed to graph creation). This is unlike the research of van Ham and Rogowitz [3] who, because graph creation was not part of the task that they set their participants, have clear data on the time taken for the graph layout process and are therefore able to make layout time comparisons between different graphs and different initial layouts.

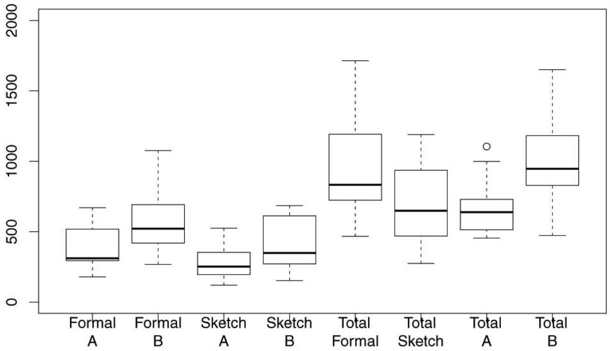


Fig. 6. The median and range (in seconds) of the time taken for the participants to draw the graphs.

Fig. 7 shows snapshots of the creation of five drawings, each demonstrating a different strategy. 12As has all nodes drawn first, and then moved around as necessary, to allow

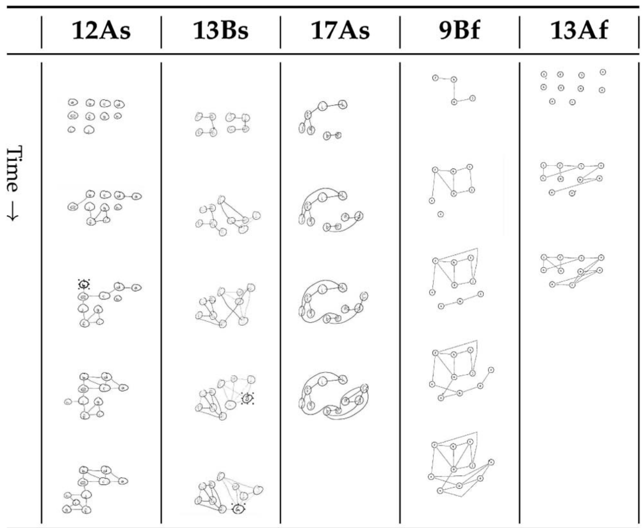


Fig. 7. Snapshots of the graph creation process, showing five different strategies. The snapshots were taken from the screen cast videos whenever there was a substantial change to the drawing. Further examples are in the Appendix.

for easy insertion of edges; 13Bs has nodes drawn where convenient for creation of the next edge, avoiding initial crossings, introducing crossings as the graph becomes complex, and finally relocating nodes to remove crossings; 17As and 9Bf have nodes placed where convenient, and never relocated, using curved or bent edges to avoid edge crossings. 13Af has all nodes drawn before edges; nodes are never relocated and there is no attempt to remove crossings.

There were clearly two different strategies in the creation of the graph drawing: drawing all the nodes first and then moving them as necessary when creating edges, or drawing nodes where convenient. The latter was favored in both sketch and formal modes. Participants tended to place the next node they came across in the adjacency list near to the node it needed to be connected to so as to avoid crossings.

There were also two different strategies when it came to producing the final layout of the graph: movement of nodes and sub-graphs during creation of the graph, and movement after the whole graph was complete: the former was typically favored. There were even some participants who did not move the nodes at all once they had been placed conveniently next to their first adjoining node (participants 7s, 8s, 10s, and 13f), even if this meant introducing edge crossings or curved or bent edges.

The use of these different strategies (and the fact that only eight sketch and 11 formal participants moved nodes after the graph was created), suggests reluctance on the part of some participants to move nodes once placed, and a possible misunderstanding of the task (which was to arrange the final graph so as to make it easy to understand). It appeared that some participants may have focused on ease of process (i.e., ease of creating the graph) rather than on the form of the final product. While our experiment removed the layout bias of van Ham and Rogowitz [3] and Dwyer et al. [4] work by asking participants to draw the graph from scratch, asking them to effectively do two tasks (creation and layout) makes it more difficult (and in some cases impossible) to focus our analysis just on the layout process.

Analyzing the participants' drawing and layout processes allowed us to see where layout aesthetics had been favored during the creation process, but abandoned later when the graph became more complex. These aesthetics are therefore not evident in the final product, even though they were considered important by the participants.

Most noticeable of these was the tendency for participants to place their nodes as if on the intersections of an underlying unit grid, and to draw edges vertically and horizontally. So while only 10 sketch and 13 formal of the resultant graph drawings had evidence of a grid-like formation (representing 16 out of 34 participants), further four sketch and two formal participants attempted to use horizontal and vertical lines during the creation process, but abandoned this feature as more edges were added. There are therefore seven sketch and 11 formal drawings that do not show evidence of an underlying grid, but in the drawing of which the participants used an underlying grid formation that was later abandoned.

Similarly, most participants attempted to use only straight edges while drawing the graphs, and the curved or bent edges tended to be the very last edges to be drawn. Many participants also attempted to use edges of similar length during the creation process: a feature that is evident in very few of the resultant drawings (and is not obvious in the edge length distribution figure reported above).

Nine participants performed some pre-drawing analysis and planned ahead. Planning ahead is, of course, unnecessary, as any nodes could be moved to a more appropriate position at any time. This again suggests a reluctance of the participants to move nodes and change the overall layout of the drawing.

The Appendix shows snapshots of the creation of two pairs of graph A (3As/4As, 2Af/17Af) and two pairs of graph B (15Bs/16Bs, 16Bf/18Bf) each of which has similar final products, but different creation strategies.

4.2.2 Comparison between Interface Modes

The drawing methods used were classified per participant, rather than per graph, and so each drawing method is considered to have been followed by a participant if it had been used at least once in the drawing of either graph. The final column in Table 3 shows the p-values derived from the independent-measures t-tests: these values show the probability that the different mode of interaction (sketch or formal) affected the measurement.

Using the typical probability level of $p = 0.05$, two results are significant:

- Moving nodes during creation ($p = 0.014$). There were significantly more participants using the formal mode moving nodes during the creation process than those using the sketch mode.
- Use similar length edges during creation ($p = 0.001$). There were significantly more participants using the sketch mode using edges of similar length during the drawing of the graph than those using the formal mode.

The formal drawings took significantly longer than the sketch drawings ($p < 0.001$): this is not unexpected, as the procedure for labeling nodes in formal mode takes longer than in sketch mode, requiring the use of an on-screen keyboard. The time taken to draw the graphs varied particularly in formal mode (Fig. 6).

There are therefore two main conclusions we can make about the differences between the graph drawing process in the two modes. First, the formal participants moved nodes around more during the creation process than the sketch participants. We suggest that this is due to a familiarity with formal interfaces that allow for regular objects to be selected and dragged and dropped, for example, in creating diagrams in Powerpoint or Word. The sketched nodes may not have exhibited the same drag-and-drop affordance, due to their unfamiliar nature and the pen-and-paper metaphor used.

Second, the sketch participants tended to use lines of a similar length while drawing the graph more than the formal participants. We suggest that this result is related to the first: the formal participants were more willing to move nodes during the creation process, and therefore, the initial length of the edges they drew was immaterial, as the edge length would naturally change when nodes were relocated. The sketch participants, on the other hand, were more reluctant to move nodes, and therefore, the length of their drawn edge was important—the pen-and-paper metaphor meant that they perceived their sketch marks as more permanent than the participants in the formal mode.

4.3 The Preferences: What Did the Participants Think?

In the post-experiment interview, the participants were encouraged to articulate their strategy in drawing the graphs, both in terms of the product and the process. No

TABLE 4
The Effect of Demographics

	Gender			CS/graph experience		
	F	M	p-value when comparing gender	Yes	No	p-value when comparing experience
Participants	13	21		16	18	
Graphs	26	42		32	36	
Product (number of graphs)						
Mean percentage of crosses	1.6%	1.5%	0.872	2.5%	0.53%	0.009*
Mean percentage of straight edges	85.8%	96.4%	0.023*	87.6%	97.7%	0.027*
Mean percentage of edges horizontal/vertical	27.2%	38.4%	0.056*	26.4%	42.8%	0.003*
Graphs clearly drawn with a grid arrangement in mind	6	17	0.145	6	17	0.001*
Drawings with two clusters	17	26	0.776	20	23	0.168
Process (number of participants)						
Draw all nodes first	4	8	0.675	3	9	0.015*
Move nodes after creation	7	12	0.856	8	11	0.164
Favour horizontal and vertical edges during creation	2	15	0.001*	7	10	0.180
Planning in advance	4	5	0.667	5	4	0.860
Mean time, for both graphs (secs)	374	309	0.213	318	352	0.500
Preferences (number of participants stating preferences)						
No crossings	9	19	0.121	12	16	0.010*
Grid configuration	1	4	0.379	2	3	0.544
Straight edges	4	8	0.675	7	5	0.654
Special use of node with highest degree	4	4	0.449	2	6	0.074

specific features of graphs were suggested to the participants in this interview, and they were encouraged to describe the features in their own words.

All but four of the sketch participants emphasized the need to avoid edge crossings; three of these said that avoiding crossings was not important, one of whom said that edge crossings are acceptable if the cross is at right angles. Two of those who said that edge crossing was not important specifically said that avoiding edges crossing nodes was more important. Of the participants who used the formal interface, only two did not explicitly mention the avoidance of crossings. One of the formal participants said that it was acceptable to include crossings if doing so would avoid making edges bent.

Despite the fact that approximately one third of the drawings were clearly drawn with horizontal and vertical edges, and with an orthogonal grid in mind, only five participants explicitly mentioned a preference for horizontal or vertical alignment of edges, with one also favoring 45 and 60 degree angles. Five (sketch) and seven (formal) participants said that straight lines were preferred, while one (sketch) participant specifically said that they were not important. Other features mentioned were maintaining a similar distance between nodes (two sketch, four formal), spreading the nodes out (two sketch, three formal), putting the nodes in groups (clusters) (two sketch, three formal), and symmetry (one sketch and one formal). No sketch participants mentioned keeping the edges short, but four formal participants did.

With respect to the graph creation process, four sketch and four formal participants specifically mentioned their use of the node with highest degree: four participants placed it centrally, one participant placed it at the middle, and another participant just made sure that the node with the highest degree was drawn first. Two participants placed the node with the highest degree at the top, with one (sketch mode) mentioning a desire to draw the graphs in a tree structure.

Two sketch participants said that they preferred to place the nodes in alphabetical order at the beginning of the drawing process.

When asked why they drew the graphs the way that they did, most sketch participants used phrases like "most logical" (3), "easier" (4), and "neater" (4). Two participants specifically said that they looked ahead, making room for future edges that would be added. Of the formal participants, the most common phrases used were "simple" (3), "vertical/horizontal lines" (5), and "easy to understand" (3).

4.4 Demographic Effects

The important demographic properties of Computer Science/graph experience, and gender both have appropriate splits to allow for separate analysis (Table 4). No distinction is made between formal or sketched graph drawings.

The final column in Table 4 shows the p-values derived from the independent-measures t-tests: these values show the probability that the different gender or extent of CS/graph experience affected the product, process, and preferences.

Using the typical probability level of $p = 0.05$, two gender results are significant; one approaches significance:

- Mean percentage straight edges ($p = 0.023$). More men than women produced drawings with straight edges.
- Mean percentage horizontal and vertical edges ($p = 0.056$, approaches significance). More men than women produced drawings with a high proportion of horizontal and vertical edges.
- Favor horizontal and vertical edges during graph creation ($p = 0.001$). More men than women used horizontal and vertical edges during graph creation.

These data suggest that the male participants favored straight edges and the horizontal and vertical edges required for a grid-based graph layout (both during the creation process and in the final product) more than the female participants, showing a desire for a more ordered, regular, and structured graph drawing.

Using the typical probability level of $p = 0.05$, six experience results are significant:

- Mean percentage of edge crossings ($p = 0.009$). Participants with CS experience produced graphs with fewer edge crossings than those without CS.
- Preferences for no crossings ($p = 0.010$). More participants with CS experience expressed a preference for removing edges crossings than those without.
- Mean percentage of edges straight edges ($p = 0.027$). Participants with CS experience produced graphs with more straight edges than those without.
- Mean percentage of horizontal and vertical edges ($p = 0.003$). Participants with CS experience produced graphs with more horizontal and vertical edges than those without.
- Graphs drawn with a grid in mind ($p = 0.001$). More participants with CS experience produced graphs based on an underlying unit grid than those without.
- Drawing nodes first ($p = 0.015$). More participants with CS experience drew all the nodes first than those without.

Computer science and graph experience therefore made a difference in several factors: edge crossings (both product and preference), straight edges, horizontal and vertical edges, and grid-like graph drawings. More interesting is their tendency to draw all the nodes first, indicating that these participants recognized the clear distinction between the graph (represented by the relational structure) and the graph drawing (represented by the visual nodes and edges), and were aware that the graph layout process can be considered independently from the graph creation process.

5 Discussion

In addressing our primary research question *Which graph drawing aesthetics do people favor when creating their own drawings of graphs?*, it is not surprising that the minimization of edge crossings has again been revealed as most important aesthetic. The data presented here give more weight to the principle of fixing edges and nodes to an underlying unit grid than previously been shown to be the case in empirical studies of graph drawing comprehension [1]. This latter point is surprising, suggesting that users would like to see their graphs fixed to a grid, even if doing so does not necessarily assist in improving their performance in graph reading tasks.

Dwyer et al.'s [4] results show that the two most preferred graph drawings were the spring version and a user-generated layout not obviously conforming to a grid formation. Few of our user-generated drawings could be said to be comparable to a spring layout (possibly 3As, 15As, 19Bs, 4Af, 5Af, 8Bf, 12Af). Few of the Dwyer et al. [4] user-generated layouts used horizontal or vertical edges, although there is no information on whether such edges were used during the layout process. Their highly orthogonal drawing was one of the least preferred: we suggest that this was because the Dwyer et al. [4] graphs were substantially larger than ours, and their orthogonal drawing included several long bent edges on the outer face.

Only one of the published graph drawings from van Ham and Rogowitz [3] shows any indication that the participants were interested in horizontal or vertical edges: we suggest that this is due to the layout bias (circular or spring) in the initial drawing presented to the participants, and the acknowledged fact that the horizontal labels made the presentation of horizontal edges awkward.

When considering creation strategies, two methodological issues stood out. First, a final graph drawing often does not reveal the aesthetic principles that the creator has tried to adhere to during the drawing process, and much can be missed by concentrating on the product rather than the process. The tendency for users to use a unit grid during the graph creation process, even if it was abandoned later, suggests that users would welcome visual grid facilities in graph editing tools. While not all graphs can be drawn on a grid (e.g., those with any node of degree greater than 4), the fact that many participants tended to use a grid during the process of creation (independent of the final form of the product), suggests that such a facility would be useful.

Second, Graph Drawing researchers tend to make a clear distinction between the process of creating a graph and the algorithm for laying it out: the abstract graph structure exists independently from its visual depiction. This does not appear to be the case for our participants: the action of presenting relational information in as best a way as possible using a graph drawing encompasses both the initial visual representation of the nodes as well as their relocation. Even those participants who drew all the nodes first typically moved nodes while they were adding the edges. Some graph layout tools (e.g., GraphViz [18]) take a graph structure as input, and produce an automatically generated graph diagram that is not editable; others (e.g., Dunnart [19]) allow for automatic layout support to be integrated with the graph creation process—our results suggest that the latter model matches users' graph drawing processes better than the former. Complete automatic layout during graph creation may be inappropriate for the interactive creative process; allowing participants to "pin" nodes to locations and laying out the rest of the drawing around these pinned nodes may be more useful.

In addressing our secondary research question: "*Do the aesthetics favored differ between two different interface styles?*," we find fewer differences than we might have expected, given that the modes are very different in style, and one (the formal mode) is more familiar than the other. The only aesthetics difference between the two styles in the product (the graph drawing itself) is a tendency for the participants using the formal interface to avoid bent lines, and in the process, for sketch participants to use similar edge lengths. There were no obvious differences in the preference opinions between the participants for each mode. We found, however, that the participants in the formal mode were more willing to move nodes, a possible consequence of the similarity of this mode to other familiar and commonly used systems.

Analysis of our demographic data suggests that male participants with a computing science or graph background tend to favor straight lines, few crossings, and a grid-like formation, and appear to recognize the distinction between the creation of a relational graph structure and the process of laying it out.

6 Conclusion

While this experiment has successfully addressed some of the limitations of the work of van Ham and Rogowitz [3] and Dwyer et al. [4], it has itself been subject to some methodological constraints. The strategy employed by most of our participants was to interleave the creation and layout processes, making it difficult to analyze the process of layout separately (as both van Ham and Rogowitz [3] and Dwyer et al. [4] did), and both our graphs were small. Seven of the total of 27 participants did the experiment in both modes: despite the six month delay between experiments, their prior experience may have affected their performance. Future work would, of course, entail the drawing of larger graphs (possibly using a digital whiteboard or similar technology), and could use abstract graph reading tasks so as to give participants a specific idea as to what "easy to understand" could mean.

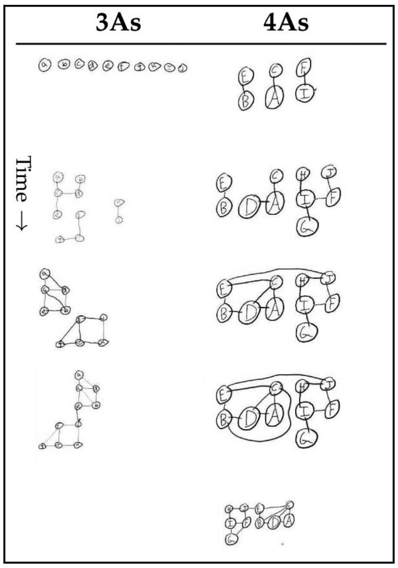
The van Ham and Rogowitz [3] experiment considered how users would manipulate graph drawings so as to improve their layout; Dwyer et al. [4] investigated how users would lay graphs out so as to be appropriate for specific tasks. The research reported here extends this work significantly with the use of two drawing interface modes, by removing any initial layout bias, by asking participants to create the drawings as well as lay them out, and by including consideration of key participant demographics. While some results (e.g., the importance of minimizing crossings) clearly follow several previous findings, we have also

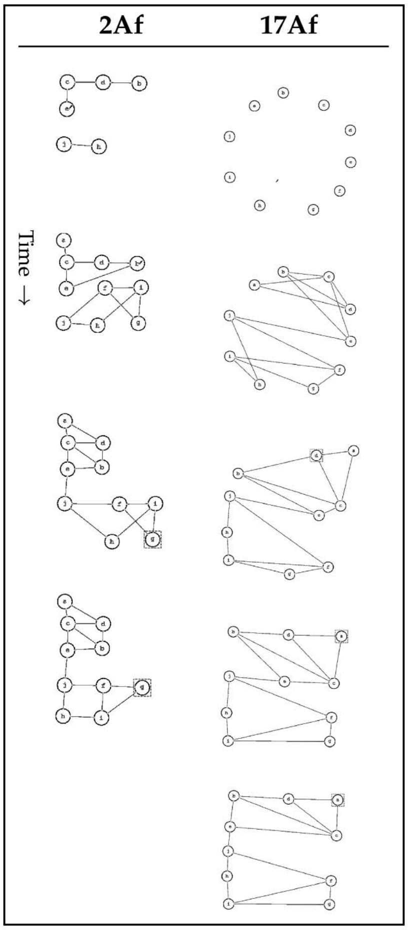
- established the importance of a grid-based layout, and
- shown that the layout of a graph drawing should not simply be judged by the product, but should also be considered in the light of the creation process.

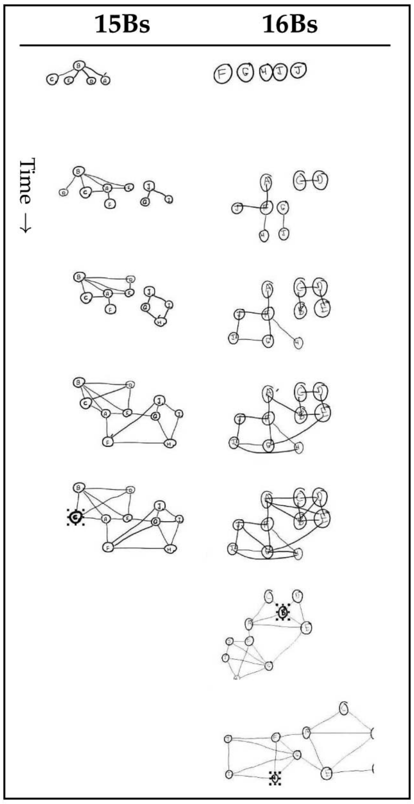
Our results demonstrate the importance of providing grid facilities in graph editing tools, and integrating the automatic layout process with that of graph creation.

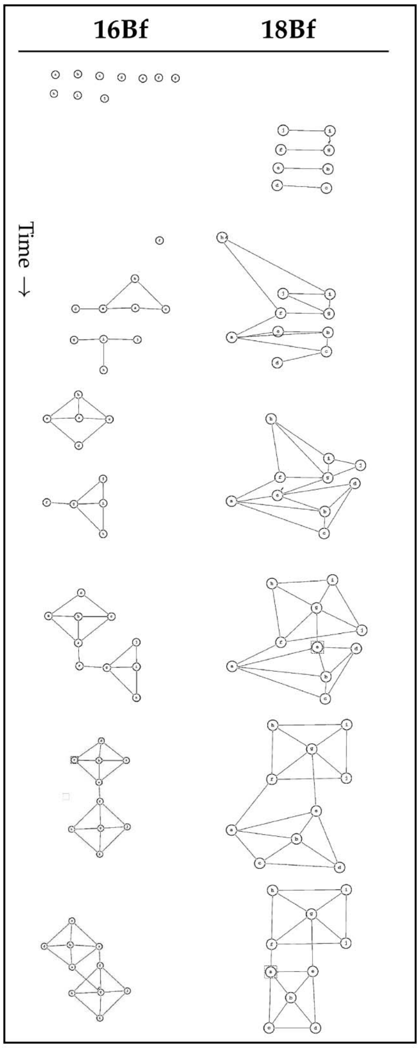
Appendix

The pairs of graph drawings 3As/4As, 15Bs/16Bs, 2Af/17Af, and 16Bf/18Bf are similar in style, although the process of drawing in each case is different.









ACKNOWLEDGMENTS

The authors are grateful to the participants, Hong Yul Yang, who implemented SketchNode; to Rose Baker, who assisted with the sketch equipment; and to John Hamer, who helped with the editorial process. Ethical clearance for this study was granted by the University of Auckland in 2009.

REFERENCES

[1] H. Purchase, "Which Aaesthetic Has the Greatest Effect on Human Understanding?," *Proc. Fifth Int'l Symp. Graph Drawing,* pp. 248-261, 1997.
[2] W. Huang, "Using Eye Tracking to Investigate Graph Layout Effects," *Proc. Int'l Asia-Pacific Symp. Visualization (APVIS '07),* pp. 97-100, 2007.
[3] F. van Ham and B. Rogowitz, "Perceptual Organization in User-Generated Graph Layouts," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1333-1339, Nov./Dec. 2008.
[4] T. Dwyer, B. Lee, D. Fisher, K. Quinn, P. Isenberg, G. Robertson, and S. North, "A Comparison of User-Generated and Automatic Graph Layouts," *IEEE Trans. Visualization and Computer Graphics,* vol. 15, no. 6, pp. 961-968, Nov./Dec. 2009.
[5] H. Purchase, B. Plimmer, R. Baker, and C. Pilcher, "Graph Drawing Aaesthetics in User-Sketched Graph Layouts," *Proc. 11th Australasian User Interface Conf.,* pp. 80-88, 2010.
[6] G. Battista, P. Eades, R. Tamassia, and I. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs.* Prentice Hall, 1998.
[7] M. Coleman and D.S. Parker, "Aaesthetics-Based Graph Layout for Human Consumption," *Software—Practice and Experience,* vol. 26, no. 12, pp. 1415-1438, 1996.
[8] H. Purchase, R. Cohen, and M. James, "Validating Graph Drawing Aaesthetics," *Proc. Third Int'l Symp. Graph Drawing,* pp. 239-250, 1995.
[9] W. Huang, S. Hong, and P. Eades, "Layout Effects on Sociogram Perception," *Proc. 12th Int'l Symp. Graph Drawing,* pp. 262-273, 2006.
[10] C. Ware, H. Purchase, L. Colpoys, and M. McGill, "Cognitive Measurements of Graph Aaesthetics," *Information Visualization,* vol. 1, no. 2, pp. 103-110, 2002.
[11] W. Huang, S. Hong, and P. Eades, "Effects of Sociogram Drawing Conventions and Edge Crossings in Social Network Visualization," *J. Graph Algorithms and Applications,* vol. 11, no. 2, pp. 397-429, 2007.
[12] E. Gansner, Y. Koren, and S. North, "Graph Drawing by Stress Majorization," *Proc. 12th Int'l Symp. Graph Drawing,* pp. 239-250, 2005.
[13] F.B. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon, "ManyEyes: A Site for Visualization at Internet Scale," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1121-1128, Nov./Dec. 2007.
[14] B. Plimmer, H. Purchase, and H. Yang, "SketchNode: Intelligent Sketching Support and Formal Diagramming," *Proc. Conf. Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OzCHI '10),* pp. 136-143, 2010.
[15] yWorks GmbH, "yEd Graph Editor," www.yworks.com, Sept. 2010.
[16] "Morae Usability Software," www.TechSmith.com/Morae, Aug. 2009.
[17] H. Purchase, "Metrics for Graph Drawing Aaesthetics," *J. Visual Languages and Computing,* vol. 13, no. 5, pp. 501-516, 2002.
[18] J. Ellson, E. Gansner, L. Koutsofios, S. North, and G. Woodhull, "GraphViz — Open Source Graph Drawing Tools," *Proc. Ninth Int'l Symp. Graph Drawing,* pp. 594-597, 2002.
[19] T. Dwyer, K. Marriott, and M. Wybrow, "Dunnart: A Constraint-Based Network Diagram Authoring Tool," *Proc. 16th Int'l Symp. Graph Drawing,* pp. 420-431, 2009.

Helen C. Purchase is a senior lecturer in the School of Computing Science at the University of Glasgow. While her main interest is the evaluation of the visual presentation of graphs, she also takes part in several empirical research projects investigating a variety of visual stimuli.

Christopher Pilcher received the BSc (Hons) degree from the Department of Computer Science at The University of Auckland. He works as a software developer. His interests include the visual presentation of graphs, and web application programming.

Beryl Plimmer is a senior lecturer in the Department of Computer Science at The University of Auckland. The focus of her research is Stylus-based interaction—this ranges from developing better digital recognition engines to the cognitive effects of different visualizations.