

# User Preference of Graph Layout Aesthetics: A UML Study

Helen C. Purchase, Jo-Anne Allder, and David Carrington

School of Computer Science and Electrical Engineering  
The University of Queensland  
Brisbane 4072, Australia  
{hcp, joanne, davec}@csee.uq.edu.au

**Abstract.** The merit of automatic graph layout algorithms is typically judged on their computational efficiency and the extent to which they conform to aesthetic criteria (for example, minimising the number of crossings, maximising symmetry). Experiments investigating the worth of such algorithms from the point of view of human usability can take a number of different forms, depending on whether the graph has meaning in the real world, the nature of the usability measurement, and the effect being investigated (algorithms or aesthetics). Previous studies have investigated performance on abstract graphs with respect to both aesthetics and algorithms, finding support for reducing the number of crossings and bends, and increasing the display of symmetry.

This paper reports on preference experiments assessing the effect of individual aesthetics in the application domain of UML diagrams, resulting in a priority listing of aesthetics for this domain. The results reveal a difference in aesthetic priority from those of previous domain-independent experiments.

## 1 Introduction

The success of automatic graph layout algorithms which display relational data in a graphical form is typically measured by their computational efficiency and the extent to which they conform to aesthetic criteria (for example, minimising the number of crossings, maximising symmetry). In addition, designers of these algorithms often claim that by conforming to these aesthetic criteria, the resultant graph drawing helps the human reader to understand the information embodied in the graph. However, little research has been performed on the usability aspects of such algorithms: do they produce graph drawings that make the embodied information easy to use and understand? Is the computational effort expended on conforming to conventional aesthetic criteria justifiable with respect to better usability?

Usability studies investigating the merit of graph drawing algorithms from a human perspective can take several different forms, depending on the nature of the graph (syntactic or semantic), the nature of the usability measurement (preference or performance), and the nature of the effect being investigated (algorithms or aesthetics).

- *Syntactic* graph drawing experiments use a graph structure that has no meaning in the real world: it is merely an abstract collection of nodes with relationships between them. *Semantic* graph drawing experiments use a graph within a particular application domain: in this case, the graph has meaning in the real world (for example, a transport network, or a data-flow diagram).
- *Preference* experiments ask the subjects to state their preference of one drawing over another. *Performance* experiments require subjects to perform a particular task (or tasks) using a given graph, the data collected is the extent of the subjects’ success in performing the task.
- Two possible effects on usability may be investigated in graph drawing experiments: the effect of individual *aesthetics* (e.g. reducing crossings, maximising symmetry) and the effect of the use of different *algorithms* (producing drawings conforming to different aesthetics to varying degrees).

Two previous studies have investigated syntactic performance. The first experiments considered the effect of individual aesthetics, and found support for reducing the number of crossings and bends, and increasing the display of symmetry. However, no support was found for maximising the minimum angle or increasing orthogonality (Purchase, 1997). The second experiment considered the effect of eight different algorithms, and revealed that it is difficult to say that one algorithm is ‘better’ than another in the context of syntactic understanding of the abstract graph structure (Purchase, 1998).

This paper reports on two semantic, preference experiments that investigated the effect of individual aesthetics. The application domains are the presentation of two types of UML diagrams: class diagrams and collaboration diagrams.

All these experiments are part of a larger project, the aim of which is to perform a thorough empirical investigation of the aesthetics underlying graph layout algorithms, and the algorithms themselves, in an attempt to influence the future design of graph layout algorithms through empirical human (rather than computational) experimentation.

## 2 Experimental Scope and Definition

### 2.1 The Application Domain: UML Diagrams

The Unified Modeling Language (UML) (Booch, Rumbaugh and Jacobson, 1998) was chosen as the semantic domain for these preference, aesthetics experiments. Many different methods and models have been proposed to capture a complete specification of requirements and a comprehensive design representation in a formal software engineering process. UML provides a mainly graphical notation to represent the artifacts of software systems. The notation is relatively new but it is rapidly being adopted as the accepted notation for object-oriented analysis and design. UML incorporates notations to describe systems at various levels of abstraction. UML diagrams can be used to model requirements, designs, implementations and tests. Since these diagrams are a means of communication between customers, developers and others involved in the software engineering

process, it is critical that the diagrams present information clearly. Appropriate layout of these diagrams can assist in achieving this goal.

UML uses several different types of graph drawings that aim to describe a system to meet the users' needs at reasonable cost. Two UML diagram types, class and collaboration, were selected for the experiments.

Class diagrams describe the types of objects in the system, and the relationships between them. These relationships are either subtypes (representing inheritance) or associations (representing other types of relationship).

Collaboration diagrams show the interaction of objects and the sequence of events by numbering the events in the order in which they occur (along the arcs), referring to objects as nodes in the graph.

## 2.2 Aims

The aim of the experiment is to identify an ordered list of the aesthetic features preferred by subjects when embodied in UML class and collaboration diagrams. Such a list can indicate to interface designers of CASE tools the most suitable way to lay out their diagrams for the best response from users. It will also form the basis for more extensive experiments concentrating on performance of users in software engineering tasks.

The main focus of the study aimed to identify the 'subjectively pleasing' aesthetics in graph drawings. Evaluation of the graph drawings was done solely according to human, individual preference. No consideration was given to performance with respect to a task or correctness of interpretation.

## 2.3 Aesthetics Investigated

Using graphs from a semantic domain instead of an abstract graph structure introduces additional secondary notations that are particular to the formal semantic notation. Secondary notations are layout or graphical cues that tend not to be part of the formal notation (e.g. adjacency, clustering, white space) (Petre, 1995). Thus, while this experiment included some graph drawing aesthetics as advocated by designers of generic layout algorithms, it also included investigation of other layout features specifically related to the standard UML notation.

For each experiment (class and collaboration), a suitable subset of aesthetic features was identified. These were selected based on emphases in the literature, and as ones that could feasibly be applied to UML diagrams. Many of them were also considered in the prior experiments (Purchase 1997, 1998) and most could be used for both types of diagram.

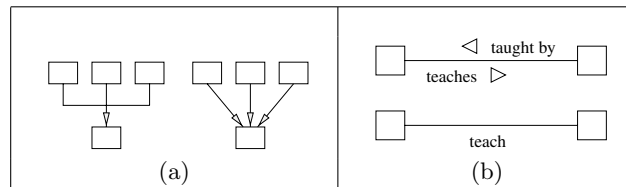
Six aesthetics were evaluated for both class and collaboration diagrams:

- *minimize bends* (the total number of bends in polyline edges should be minimized (Tamassia, 1987))
- *minimize edge crossings* (the number of edge crossings in the drawing should be minimized (Reingold and Tilford, 1981))

- *orthogonality* (fix nodes and edges to an orthogonal grid (Tamassia, 1987; Papakostas and Tollis, 2000))
- *width of layout* (the physical width of the drawing should be minimised (Coleman and Stott Parker, 1996))
- *text direction* (all text labels should be horizontal, rather than a mixture of horizontal and vertical) (based on Petre, 1995)
- *font type* (all text fonts should be the same, rather than using different fonts for different types of labels) (based on Petre, 1995)

For UML class diagrams, two additional secondary notation features were investigated. Both versions of each notation have been found in published examples of UML notation (See Figure 1).

- *inheritance* (inheritance lines should be joined prior to reaching the super-class, rather than being represented as separate arcs)
- *directional indicators* (arcs should be labelled with two relationship labels and directional indicators, rather than one)

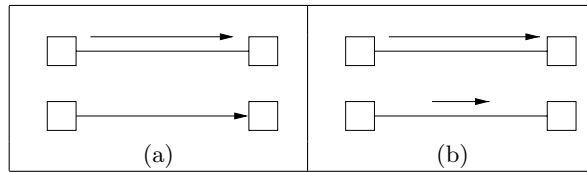


**Fig. 1.** The UML class diagram secondary notation features investigated, showing both alternatives: (a) inheritance, (b) directional indicators.

For UML collaboration diagrams, two additional secondary notation features were investigated. In both cases, the one option (long arrows adjacent to the arcs) is standard UML notation (See Figure 2).

- *adjacent arrows* (all arcs are undirected with an adjacent arrow indicating the direction of the message sent, rather than all arcs being directed)
- *arrow lengths* (the arrows adjacent to the arcs should be the same length as the arcs, rather than shorter than the arcs)

**Usability measuring method.** As these were our first experiments performed on layout aesthetics and secondary notations for UML diagrams, preference was chosen as the method of usability measurement. While increased preference does not necessarily correspond with improved performance, beginning with a preference study enables the most important layout features to be identified before a more substantial performance study is performed.



**Fig. 2.** The UML collaboration diagram secondary notation features investigated, showing both alternatives: (a) adjacent arrows, (b) arrow lengths.

By asking whether they prefer one UML drawing to another, subjects are likely to be anticipating the use of these drawings for a software engineering task: their responses are therefore likely to be related to their perceived usefulness of the diagrams.

Three experiments were performed:

Experiment 1: determined preferences for aesthetics embodied in UML class diagrams (70 subjects)

Experiment 2: determined preferences for aesthetics embodied in UML collaboration diagrams (90 subjects)

Experiment 3: a more focussed study which refined the results of the first two experiments, focussing on particular aesthetics that gave surprising results (6 subjects)

## 2.4 Methodology

A basic UML class diagram (depicting the relationships between students, lecturers, tutors and administrative staff, see Figure 3) and a basic UML collaboration diagram (depicting the procedure followed for organising honours students' seminars, see Figure 4) were created.

In both cases, the graph structures were complex enough to enable an appropriate and varied manipulation of the nodes and arcs within the diagram, but not so complex that the diagram would take a long time to understand.

The class diagram had 14 classes and 18 relationships, and the collaboration diagram had 12 objects and 17 messages.

Each diagram was drawn 16 times (twice for each aesthetic), with all the information within the drawing remaining constant: only the layout of the diagrams was altered. Each representation of the graph was drawn with attention to a specific graph-drawing aesthetic or secondary notation choice.

Graph drawings were grouped in pairs emphasising the contrast between the diagrams: one graph drawing in the pairwise comparison contained a higher presence of the aesthetic while the other graph drawing contained a lower presence. For example, one diagram was highly orthogonal while the paired diagram had minimal orthogonality. In the absence of computational metrics for measuring the presence of the UML and secondary notation features in the drawings, and due to the large number of aesthetics being investigated concurrently, the other



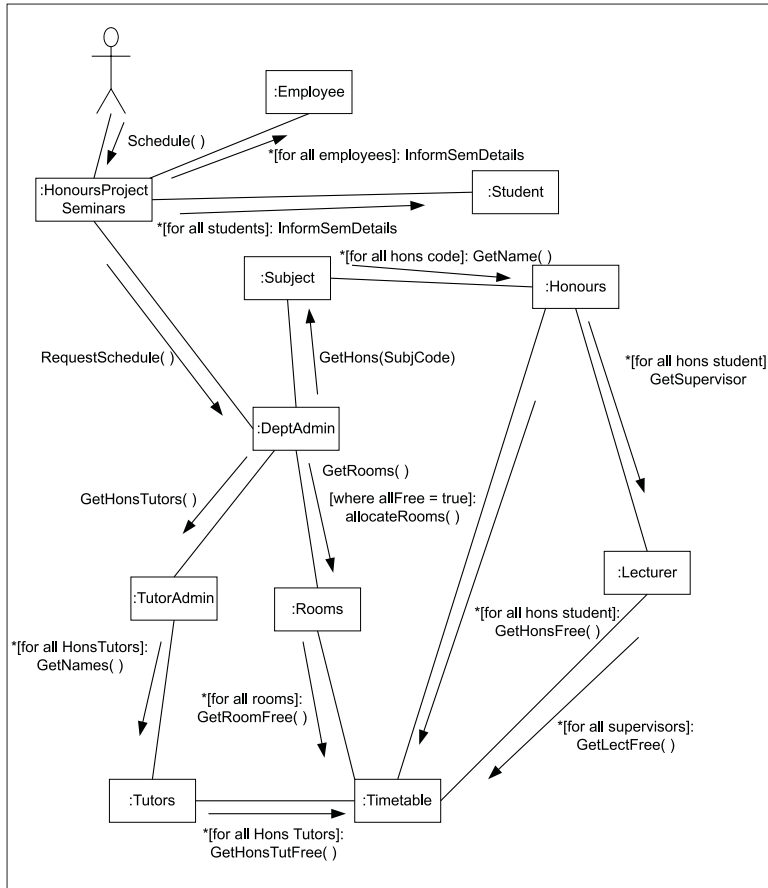


Fig. 4. A UML collaboration diagram used in experiment 2.

founding factors which may have biased the simple preference quantitative data, we collected additional qualitative and ranking data.

**Subjects.** Seventy student volunteers from the University of Queensland participated in the class diagram evaluation; ninety students participated in the collaboration diagram evaluation. The same experimental methodology and materials were used for both experiments.

All participants were third or fourth year Information Technology students who, although not generally proficient with UML, have experience with similar notations (e.g., Fusion, dataflow diagrams, entity relationship diagrams and Booch diagrams).

Prior to both experiments, pilot experiments with a small number of subjects were performed to check for problems in experimental materials and procedures.

**Materials.** Each subject was presented with an individual evaluation booklet designed to be completed without evaluator assistance and without a time limit. The booklet had the following structure:

- Instructions, the aim of the project, the task required, and an example.
- A questionnaire requesting information about the subject (prior knowledge and use of graph drawings, year of study etc.).
- A UML tutorial sheet, identifying the key points of the diagram under consideration (class or collaboration). This sheet was detached from the booklet so that subjects did not have to turn back to refer to it.
- Eight (facing) pairs of graphs drawings. For each pair, there was a space where the subjects had to indicate their preferred drawing, as well as write a brief explanation for their choice. The explanation was intended to identify any issues influencing the choice that had not been considered by the evaluator, and to provide qualitative data to support the quantitative data. The diagrams were presented in a random order in the booklet, in an attempt to counter any familiarity effect: after seeing several different representations of the same UML diagram, the subject may develop a deeper understanding of the information, and may therefore make different preference choices.
- A ranking sheet where subjects needed to rank the three diagrams that they most preferred (1-3), and the three that they least preferred (14-16). A separate large sheet, showing all 16 (reduced in size) diagrams on one page, was provided to assist in this ranking.

There was no time limit, and the subjects could go back and change any previous answers if they wished. Most subjects completed the task in about 20 minutes.

**Data Analysis.** The preference questions were analysed by calculating a percentage preference for each aesthetic, with the significance of the result computed using a standard binomial distribution. A result was considered significant (ie. not attributable to chance or random selections) if its probability was less than 0.05.

The written explanations for preferences were analysed by determining the percentage of subjects who stated that the targeted aesthetic comparison influenced their choice: this allowed us to identify whether there were any possible confounds (ie. other aesthetics unintentionally affecting the result) in the quantitative data obtained by the subjects' preferences.

The final ranking question was analysed to identify preferred diagrams by computing an overall weighted preference value for each diagram. A weight of 3 was given for a first choice, a weight of 2 for a second choice, a weight of 1 for a 3rd choice, a weight of -1 for a 14th choice, a weight of -2 for a 15th choice and a weight of -3 for a 16th choice.



aesthetic choice	% preference
fewer crosses	93%
fewer bends	91%
horizontal labels only	86%
joined inheritance arcs	76%
narrower	73%
more orthogonal	61%
no font variation	61%
directional indicators	60%

**Fig. 5.** Aesthetic preference results for UML class diagrams (all results significant).

### 3 Results

#### 3.1 Results: UML Class Diagrams

The results for UML class diagrams are shown in Figure 5. By analysing the subjects' stated reasons for their preferences, the only class diagram aesthetic result that appeared to be affected by confounding factors was horizontal labels, where many subjects who preferred the horizontal labels drawing referred to direction of information flow. Both the crosses and bends aesthetic results were unaffected by other factors, while those subjects who did not like the orthogonal diagram did so because of the increased number of bends. Most of those subjects who preferred independent inheritance arcs, a wider diagram, variation in font or no directional indicators did so merely because of personal preference.

The results of the ranking question are shown in Figure 6, which presents the weighted overall ranking value for each of the 16 drawings.

#### 3.2 Results: UML Collaboration Diagrams

The results for UML collaboration diagrams are shown in Figure 7. By analysing the subjects' stated reasons for their preferences, the only collaboration diagram result that appeared to be affected by confounding factors was longer arrows, where many subjects who preferred the longer arrows referred to diagram structure. Both the crosses and adjacent arrows aesthetic results were unaffected by other factors, while those subjects who did not like the orthogonal diagram did so because of increased bends. Most of those subjects who preferred a wider diagram or variation in font did so merely because of personal preference. The results for the use of horizontal labels only and bends were not significant, so no conclusions can be drawn for these two aesthetics as these results could be attributable to chance.

The results of the ranking question are shown in Figure 8, which presents the weighted overall ranking value for each of the 16 drawings.

class diagram drawing	weighted ranking value
joined inheritance arcs	54
narrower	53
directional indicators	42
fewer bends	41
horizontal text only	33
highly orthogonal	29
no directional indicators	21
wider	5
not orthogonal	3
no font variation	1
font variation	0
separate inheritance lines	-5
fewer crosses	-19
horizontal and vertical text	-39
many bends	-73
many crosses	-145

**Fig. 6.** Weighted ranking values for all sixteen class diagrams.

aesthetic choice	% preference
fewer crosses	91%
no adjacent arrows	90%
longer arrows	82%
no font variation	70%
more orthogonal	63%
narrower	57%
horizontal labels only	54% †
fewer bends	53% †

**Fig. 7.** Aesthetic preference results for UML collaboration diagrams († indicates a non-significant result).

## 4 Follow-Up Experiment

Experiment 3 was a follow-up experiment that focussed on the aesthetics common to both class and collaboration diagrams. Its aim was to investigate some of the unexpected results from experiments 1 and 2 that may have been due to confounding factors in the diagrams, and to create a final priority list of aesthetics. The subjects' comments had indicated that the direction of flow of information had sometimes influenced their preference. Direction of flow was an aesthetic feature that had not originally been targeted and which was introduced into experiment 3, with the following definition:

- *flow* (directed arcs should point in a consistent direction, preferably top-to-bottom and left-to-right) (Eades and Sugiyama, 1990)

class diagram drawing	weighted ranking value
highly orthogonal	101
longer arrows	77
no adjacent arrows	76
horizontal and vertical labels	30
horizontal labels only	27
shorter arrows	12
no font variation	3
fewer crosses	-2
font variation	-6
narrower	-7
not orthogonal	-10
wider	-11
fewer bends	-30
many bends	-31
adjacent arrows	-36
many crosses	-196

**Fig. 8.** Weighted ranking values for all sixteen collaboration diagrams.

The original UML diagrams were modified to correct any obvious confounding factors. For example, the original collaboration diagram with few bends had some crossing arcs that had been identified as influencing some subjects' preference for the diagram with more bends and no crossing arcs.

Six separate smaller experiments comprised this investigation. They were performed in intensive, focussed interviews where each subject was questioned about which aspects of a set of diagrams influenced their preferences. Six subjects took part, each being questioned about all six diagram sets, and providing extensive qualitative data with which to form a prioritised list of aesthetics.

The format of each of the six experiments differed according to the specific investigation of the experiment 1 and 2 results that were being considered. The general procedure was that subjects were shown two or three different diagrams, and were asked questions about their layout. Most of the diagrams used in experiment 3 were ones that had been used in experiment 1 or 2, although some were altered to target specific aesthetics.

**The components of experiment 3.** Two categories of unexpected quantitative results were investigated in these six smaller, focussed interview experiments. First, we investigated some surprising overall weighted ranking values. Second, we investigated some aesthetics for which the percentage preferences differed between the class and collaboration diagrams. In each case, we reviewed the diagrams that had been used for experiments 1 and 2 and tried to determine whether there were any confounding factors that may have led to these surprising results.

**Overall weighed ranking values:** In both the class and collaboration diagram experiments, the weighted ranking for the diagram with least crosses was negative. It was unlikely that this diagram was ranked according to its lack of crosses: careful inspection of both “least crosses” diagrams revealed that both appeared to be less orthogonal than the other fifteen diagrams in the corresponding set. The first two experiments of experiment 3 focussed on these diagrams, and supported our view that subjects’ preference for the many crosses diagram over the few crosses diagram in experiments 1 and 2 may have been influenced by the lack of orthogonality and inconsistent flow direction in the “least crosses” diagram.

**Difference in percentage preference between class and collaboration diagrams:** Four differences in the initial preferences for the aesthetics were addressed: bends, font variations, text direction and width.

- *Bends:* The results for experiment 2 did not indicate a significant preference for the collaboration diagram with the least bends. Inspection of the pair of collaboration diagrams related to the bends aesthetic revealed that the diagram with more bends had a more consistent direction of flow, and the one with no bends had a single crossed arc. The third experiment of experiment 3 focussed on these diagrams, and supported our view that the direction of flow and the single cross influenced subjects’ preference decisions.
- *Font variation:* Using consistent font was preferred more in the collaboration diagram than in the class diagram, and it was felt that this could have been because the collaboration diagram that used a variety of fonts included a more unusual font (cursive) than the corresponding class diagram (italic). This issue was addressed in the fourth experiment by producing identical diagrams with no font variation, and with more subtle font variation (bold). The results supported our view that the subjects’ dislike of font variation is dependent on the type of fonts used.
- *Text direction:* The diagram that only used horizontal text was preferred to a much greater extent in class diagrams than in collaboration diagrams. On inspection of the collaboration diagram with both vertical and horizontal text, it appeared to have a more orthogonal shape. The fifth experiment confirmed that this orthogonality could have affected the subjects’ preferences in the first experiment.
- *Width:* There was a stronger preference for narrow width in class diagrams than in collaboration diagrams. Although the qualitative data from experiments 1 and 2 suggested that preference for width may be an inexplicable subjective opinion, we also thought the amount of information associated with arcs in the collaboration diagram may have affected this preference decision. The sixth experiment confirmed this: while some subjects could not explain why they preferred the narrower or wider diagram, those that could explain their preference did so by referring to the amount of information in the collaboration diagram.

## 5 Conclusions

The follow-up experiments provided rich qualitative data which shed insight on the initial quantitative data. By analysing the extensive interview comments provided by the subjects, we concluded that the priority order of the aesthetics common to both diagram types is: arc crossings, orthogonality, information flow, arc bends, text direction, width of layout and font type. Of the UML-specific aesthetics, we concluded that joined inheritance arcs and directional indicators are preferred for class diagrams. For collaboration diagrams, no adjacent arrows are preferred (although this preference is incompatible with UML notation). This list provides a useful starting point for further studies on UML diagram layout aesthetics with respect to performance in a related task.

## 6 Discussion

The previous syntactic performance experiment found support for reducing the number of crossings and bends, and for increasing the orthogonality. Information flow and width were not considered.

The results of this semantic preference experiment confirm that the evidence is overwhelmingly in favour of reducing the number of arc crossings as the most important aesthetic to consider. While the results of the syntactic experiments did not highlight orthogonality as being important, in the domain of UML diagrams, this aesthetic moves up the priority list to second place: this is a clear signal that algorithms that are designed for abstract graph structures, with no consideration of their ultimate use, will not necessarily produce useful visualisations of semantic information.

This study is only the first step in assessing the usability of graph drawings produced by layout algorithms when used in application domains. Future work includes investigating aesthetics and algorithms with respect to measures of performance in UML related tasks, and extensions of this methodology to other domains. There is an increasing demand for software environments that can assist users in graph-based application tasks, for example, software engineering, social and transport network analysis, database design. Empirical research into the effectiveness of graph layout algorithms from a usability point of view can ensure that such software environments serve their users effectively.

**Acknowledgements.** We are grateful to the students in the School of Computer Science and Electrical Engineering at The University of Queensland who participated in these experiments. This research was partly funded by an Australian Research Council grant.

## References

1. G. Booch, J. Rumbaugh, and I. Jacobson. *The Unified Modeling Language User Guide*. Addison-Wesley, 1998.

2. M.K. Coleman and D. Stott Parker. Aesthetics-based graph layout for human consumption. *Software – Practice and Experience*, 26(12):1415–1438, 1996.
3. P. Eades and K. Sugiyama. How to draw a directed graph. *Journal of Information Processing*, 13(4):424–437, 1990.
4. A. Papakostas and I.G. Tollis. Efficient orthogonal drawings of high degree graphs. *Algorithmica*, 26(1):100–125, 2000.
5. M. Petre. Why looking isn’t always seeing: Readership skills and graphical programming. *CACM*, 38(6):33–44, June 1995.
6. H. C. Purchase. Which aesthetic has the greatest effect on human understanding? In G. Di Battista, editor, *Proceedings of Graph Drawing Symposium 1997*, pages 248–259. Springer-Verlag, Rome, Italy, 1997. LNCS, 1353.
7. H. C. Purchase. Performance of layout algorithms: Comprehension, not computation. *Journal of Visual Languages and Computing*, 9:647–657, 1998.
8. E. Reingold and J. Tilford. Tidier drawing of trees. *IEEE Transactions on Software Engineering*, SE-7(2):223–228, 1981.
9. R. Tamassia. On embedding a graph in the grid with the minimum number of bends. *SIAM J. Computing*, 16(3):421–444, 1987.