Journal of Visual Languages and Computing 24 (2013) 262–272

Contents lists available at SciVerse ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

Improving multiple aesthetics produces better graph drawings ☆

Weidong Huang [a,*], Peter Eades [b], Seok-Hee Hong [b], Chun-Cheng Lin [c]

[a] CSIRO ICT Centre, Marsfield, NSW 2122, Australia
[b] School of Information Technologies, University of Sydney, NSW 2007, Australia
[c] Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan

ARTICLE INFO

Available online 10 January 2012

Keywords:
Graph drawing
Force-directed algorithms
Aesthetic criteria
Crossing angles
Angular resolution
Evaluation
Effectiveness

ABSTRACT

Many automatic graph drawing algorithms implement only one or two aesthetic criteria since most aesthetics conflict with each other. Empirical research has shown that although those algorithms are based on different aesthetics, drawings produced by them have comparable effectiveness.

The comparable effectiveness raises a question about the necessity of choosing one algorithm against another for drawing graphs when human performance is a main concern. In this paper, we argue that effectiveness can be improved when algorithms are designed by making compromises between aesthetics, rather than trying to satisfy one or two of them to the fullest. We therefore introduce a new algorithm: BIGANGLE. This algorithm produces drawings with multiple aesthetics being improved at the same time, compared to a classical spring algorithm. A user study comparing these two algorithms indicates that BIGANGLE induces a significantly better task performance and a lower cognitive load, therefore resulting in better graph drawings in terms of human cognitive efficiency.

Our study indicates that aesthetics should not be considered separately. Improving multiple aesthetics at the same time, even to small extents, will have a better chance to make resultant drawings more effective. Although this finding is based on a study of algorithms, it also applies in general graph visualization and evaluation.

1. Introduction

Graphs, defined as a set of vertices and a set of edges that connect the vertices, have been widely used to model network data for various purposes. Research in graph drawing concerns the problem of constructing geometric representations of graphs. That is to design an algorithm that takes a graph as an input and calculates the positions of vertices to optimize a set of pre-defined layout requirements. The final representations of graphs are usually in the form of so-called node-link diagrams.

According to Di Battista et al. [7], layout requirements used in algorithm design can be classified into three fundamental parameters of graph drawing: drawing conventions, aesthetics and constraints. Drawing conventions are normally common practices or requirements of real-life applications. For example, draw each edge as a straight line, or draw each edge as a chain of horizontal and vertical line segments. Constraints are rules that only apply to subsets of a graph or parts of a drawing, rather than the entire graph or drawing. For example, place a given vertex close to the center of a drawing, or place a subset of vertices close to each other. Aesthetics are a set of visual properties that algorithms are required to achieve in the final drawings, as much as possible, in order to improve readability. Examples of aesthetic criteria include the following:

- Minimum number of edge crossings.
- Maximum size of crossing angles.
- Uniform edge lengths.

☆ This paper has been recommended for acceptance by Shi Kho Chang.

* Corresponding author. Tel.: +61 2 9372 4614.
E-mail address: tony.huang@csiro.au (W. Huang).

1045-926X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jvlc.2011.12.002

- Maximum angular resolution of vertices.
- Even distribution of vertices.

1.1. Two experiments of purchase

The past two decades have seen a fast growing body of research dedicated to designing algorithms to construct aesthetically pleasing drawings of graphs [7]. For excellent reviews on graph drawing algorithms, see [7,29]. While judgement of the quality of a drawing is subjective, it is generally believed that drawings that conform to the aesthetic criteria should be more effective in conveying the embedded information to the viewer. This belief is also supported by empirical work mostly done by Purchase. In particular, in her seminal work, Purchase and her colleagues [34] examined the effects of three aesthetics: crossings, bends and symmetry. For each aesthetic, the same graph was drawn three times with the value of the aesthetic in consideration being varied (see Fig. 1). Then the users were asked to perform the same graph reading tasks with the three drawings. The task performance was measured as the number of errors and they found that, for example, more errors were made when there were more crossings. In other words, increasing the number of crossings decreased the readability of graph drawings.

It is often tempting to optimize aesthetic criteria as many as possible in the same drawing in order to achieve the best possible readability. However, this can be practically difficult to achieve. Firstly, optimization of even a single aesthetic can sometimes be computationally difficult. For example, minimization of edge crossings is NP-complete [16]. As a result, a number of algorithms take a heuristic approach by which the aesthetic in consideration is not necessarily optimized in the resulting drawings. Secondly, most of the aesthetics are mutually exclusive; it is difficult, if not impossible, to implement all of the aesthetics *to the fullest* at the same time. For example, look at the two drawings shown in Fig. 2. If we want to draw the graph with maximum symmetries, then edge crossings are necessary (left). On the other hand, minimizing the number of crossings can only be achieved at the cost of symmetry (right). As a result, many automatic graph drawing algorithms aim to draw graphs satisfying one or two aesthetics [9].

Despite the fact that algorithms may be based on different aesthetic criteria, Purchase [32] has shown in another user study that these algorithms produce visualizations with similar levels of effectiveness. In this study, eight different algorithms were compared based on human performance. These algorithms were of a great variety in terms of aesthetic criteria that each of them aimed to satisfy. A single graph that had 17 vertices and 29 edges was drawn by the algorithms resulting in eight stimuli (two examples of the stimuli were shown in Fig. 3). The stimuli included drawings produced by force-directed algorithms with few edge crossings, planar grid drawings with many sloped edges, orthogonal grid drawings with minimum edge bends, drawings with even distribution of vertices and drawings with maximum symmetries. Fifty-five computer science students participated in

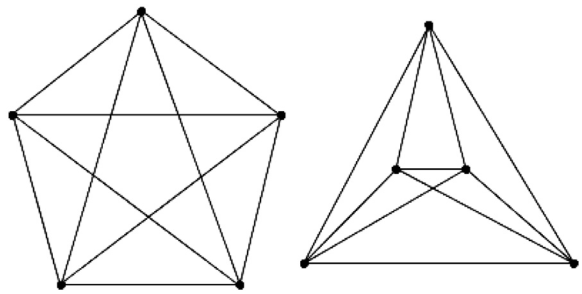

Fig. 2. Two drawings of the same graph. Left: maximum degree of symmetry; right: minimum number of crossings.

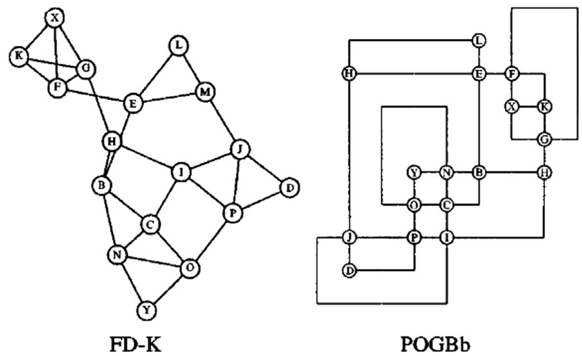

Fig. 3. Two examples of the drawings used in the study of Purchase [32].

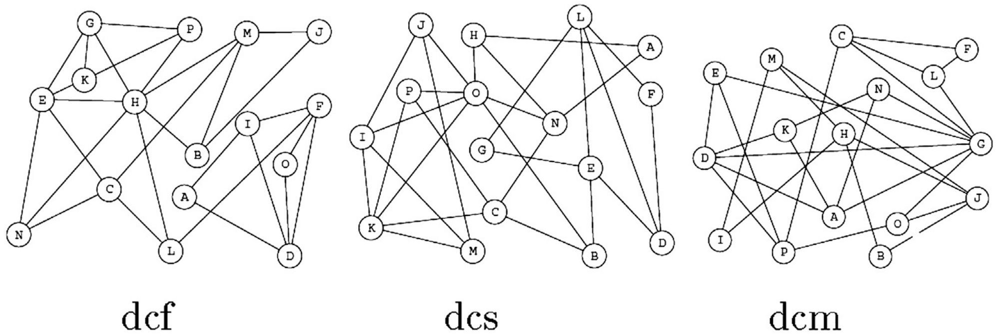

Fig. 1. Three crossing drawings of a graph with varied numbers of crossings [34].

the study. The participants were asked to view the drawings and answer the following three questions:

1. How long is the shortest path between two given nodes?
2. What is the minimum number of nodes that must be removed to disconnect two given nodes?
3. What is the minimum number of edges that must be removed to disconnect two given nodes?

A custom-built system was used to display the drawings so that the participants could perform tasks online. Times and responses were recorded. The results showed that with the exception of one algorithm, there were no statistical differences found between the algorithms in regard to either response time or accuracy. The study demonstrated that despite that each algorithm aimed for different drawing criteria, these algorithms produced drawings with comparable effectiveness.

1.2. Motivations

According to Purchase [32], the comparable effectiveness is likely due to the interactions between aesthetics. Indeed, since graph components (vertices and edges) are interlinked with each other, moving the components to improve one aesthetic will inevitably change the status of other aesthetics. Further, if that aesthetic is to be implemented to the fullest, as shown in Fig. 2, other aesthetics will likely be made worse as a result. The study of Purchase et al. [34] showed that when one specific aesthetic is considered, satisfying it could lead to improved human performance. However, when different aesthetics are considered across graph drawings, or algorithms, interactions between aesthetics could come into play, resulting in drawings being equally effective, as demonstrated in the study above.

The comparable effectiveness raises a question about the necessity of choosing one algorithm against another for drawing graphs when human performance is a main concern. Is there any way that helps us to design more effective algorithms? In this paper, we argue that algorithms can be more effective by making compromises between multiple aesthetics. That is to improve each of the aesthetics to a certain extent at the same time, instead of trying to satisfy only one or two of them to the fullest. The work presented in this paper is intended to validate this argument.

1.3. Our results

To validate this argument, we need an algorithm that improves multiple aesthetics. We therefore introduce a new force-directed algorithm: *BIGANGLE*. This algorithm is developed based on a classical spring algorithm, called *Classical*, with two additional forces: cosine force and sine force. We first demonstrate that inclusion of these two forces in Classical leads to multiple aesthetics being improved at the same time. Then we validate our argument by presenting a user study, which compares drawings produced by these two algorithms in respect to human graph reading performance, cognitive load and visualization efficiency. The contributions of this work can be summarized as follows:

- We have introduced cosine force, which is to increase the size of crossing angles.
- We have introduced sine force, which is to increase the angular resolution of vertices.
- We have introduced and implemented a force-directed algorithm, which improves multiple aesthetics at the same time.
- We have conducted a human study, which was the first to examine the collective effect of multiple aesthetics on human graph comprehension.
- We have presented empirical evidence showing that improving multiple aesthetics produces better graph drawings.
- We have evaluated two graph drawing algorithms in ways that have not been used in prior studies of the same kind: (1) we evaluated algorithms both on visual aesthetics of drawings and on human graph comprehension; (2) apart from performance measures, cognitive load and visualization efficiency were used; (3) a substantial amount of graphs were used.

1.4. Organization

This paper is an extension of its conference version [23]. The remainder is organized as follows. In Section 2, we describe the two algorithms in detail. The statistics of the resulting drawings are also presented and compared. We then present our user study in Section 3 with details of design, data analysis and results, followed by a discussion on the experimental findings and their implications for future research in Section 4. Finally in Section 5, we conclude the paper with a brief summary.

2. The two algorithms

We consider the force-directed methods for the purpose of validation. Force-directed algorithms treat graphs as physical systems in which vertices are replaced with equally charged rings repelling each other while edges are replaced with springs (see Fig. 4 for the spring embedder model). Springs pull connected rings together when stretched, while they push rings apart when compressed. Starting with an arbitrary placement of vertices, the algorithm calculates the

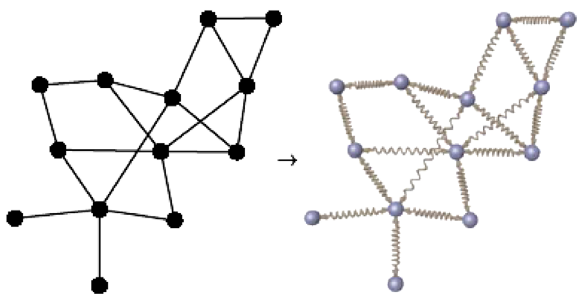

Fig. 4. The spring embedder model (modified from Brandes [3]).

combined force on each of the vertices and moves them accordingly. The final positions are determined by repeating this process for a fixed number of times. The resulting pictures can be aesthetically pleasing [3,12].

2.1. Classical

Classical is a spring algorithm introduced in the book of Di Battista et al. [7]. In this algorithm, the force exerted by springs follows Hooke's law:

$$f_s = k_s(d-l) \tag{1}$$

and the repulsive force between all rings follows an inverse square law:

$$f_r = k_r/d^2 \tag{2}$$

where k_s and k_r are constants, d is the Euclidean distance between two rings and l is the natural length of the spring.

Therefore, given a graph $G=(V,E)$, the combined force applied on vertex v is

$$F(v) = \sum_{(u,v)\in E} f_{s,uv} + \sum_{(u,v)\in V\times V} f_{r,uv} \tag{3}$$

where $f_{s,uv}$ denotes the spring force and $f_{r,uv}$ denotes the repulsive ring force.

The spring force is to ensure that edges are as equal to the natural spring length as possible, while the ring force is to make sure that vertices are not too close to each other.

2.2. BIGANGLE

The beauty of the spring embedder model is that we can simply add up force functions, each of which aims to maximize a specific aesthetic criterion [3,40]. We expected that those forces could work together resulting in each aesthetic being improved to a certain extent.

In addition to the above-mentioned spring force and ring force, BIGANGLE includes two extra forces: one is called *cosine force* that is to increase the size of crossing angles, and the other is called *sine force* that is to maximize the angular resolution of vertices.

How cosine force works is shown in Fig. 5(1). Suppose that two edges (a,b) and (c,d) cross at an angle of θ. If $\theta \neq 90^\circ$, then the cosine force is applied on each of the endpoints so that θ approaches 90°. For example, consider vertex a: the force is applied along the line parallel to edge (c,d) with a magnitude determined by the size of θ. If θ is acute, vertex a is pulled toward vertex d. If θ is obtuse, vertex a is pulled toward vertex c. If $\theta = 90^\circ$, no force is applied. The magnitude of the force is computed as follows:

$$f_{cos} = k_{cos} \cos\theta \tag{4}$$

where k_{cos} is a constant.

How sine force works is shown in Fig. 5(2). Suppose that vertex a has at least two incident edges. Let ϕ be the optimal angle ($=360^\circ/deg(a)$), and θ be the angle formed by a pair of two neighboring edges (a,b) and (a,c). If $\theta \neq \phi$, then the sine force is applied on vertices b and c so that θ approaches ϕ. The magnitude of the force is computed as follows:

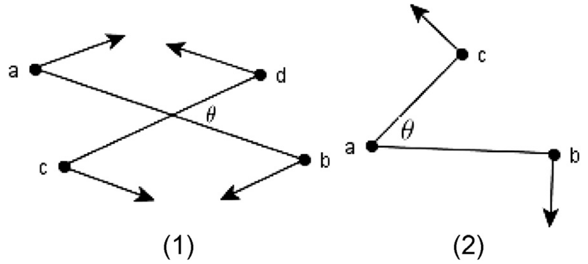

Fig. 5. Illustration of cosine force (1) and sine force (2). Lines with arrows are forces applied on the vertices and the arrows indicate force directions.

$$f_{sin} = k_{sin} \sin((\phi-\theta)/2) \tag{5}$$

where k_{sin} is a constant.

Therefore, given a graph $G=(V,E)$, let $V' \subseteq V$ be the set of vertices in which each vertex has at least one incident edge being crossed, and let $V'' \subseteq V$ be the set of vertices in which each vertex is a neighbor of another vertex that has at least two incident edges. The combined force applied on vertex v is

$$F(v) = \sum_{(u,v)\in E} f_{s,uv} + \sum_{(u,v)\in V\times V} f_{r,uv} + \sum_{(c,v)\in C\times V'} f_{cos,cv} + \sum_{(n,v)\in N\times V''} f_{sin,nv} \tag{6}$$

where $f_{s,uv}$ denotes the spring force; $f_{r,uv}$ denotes the repulsive force; $f_{cos,cv}$ denotes the cosine force; $f_{sin,nv}$ denotes the sine force; $C \subseteq E \times E$ denotes the set of pairs of edges that cross, in which each pair includes an incident edge of vertex v; $N \subseteq E \times E$ denotes the set of pairs of neighboring edges that are incident on neighbors of vertex v, in which each pair includes an incident edge of vertex v.

It should be noted that to increase crossing angle and angular resolution, the corresponding forces can also be implemented using linear functions. For example, an alternative to the cosine force can be a linear force calculated by subtracting θ from 90°. The reason why the cosine force is used in BIGANGLE comes from the original experiment of Huang et al. [27]. The experiment shows that readability increases with crossing angles. The correlation was nonlinear; we felt that it may be close to a cosine function. While the experiment does not formally prove this, we felt that the intuition was worth being embedded in the algorithm.

2.3. Comparison of the resulting drawings

In this sub-section, we measure and compare the quality of drawings produced by BIGANGLE and Classical. We demonstrate that the introduction of the two extra forces leads to seven aesthetics being improved.

Force-directed algorithms have been widely used in various application domains to explore real world problems (e.g., [11,20,30]). Real world graphs often have little structure and tend to be random [43]. Therefore we chose testing graphs from "Rome Graphs" to produce drawings. The Rome graphs [17] are a set of random benchmark graphs commonly used for testing the efficiency of graph drawing algorithms.

We chose graphs that were connected and had size between 10 and 50, which gave us 6138 graphs in total with an average size of 32. These graphs were chosen simply because (1) it is commonly acknowledged that force-directed algorithms were best suitable for connected graphs with a size up to 50 [3], and (2) drawing graphs with size larger than 50 is unlikely to be useful for graph perception or reading, without the help of interaction techniques [42].

We have implemented both BIGANGLE and Classical in Java for the comparison purpose. Experiments were performed on a 2.4 GHz laptop with 2.99 GB RAM. Starting with the same initial random placement, each graph was drawn with BIGANGLE and with Classical respectively. To compare the quality of drawings produced by the two algorithms, the following aesthetic properties were measured for each drawing:

1. Number of crossings (# of crossings).
2. Average size of crossing angles (angle size).
3. Standard deviation of crossing angles (angle dev.).
4. Average of edge lengths (edge length).
5. Standard deviation of edge lengths (edge dev.).
6. Angular resolution (angular res.).
7. Average of standard deviations of angular resolution (angular dev.).

Among the above-mentioned aesthetics, angular resolution is a measure of how incident edges of each vertex are distributed. Usually an even distribution of incident edges is desirable. In this paper, angular resolution is measured as an average of differences between the smallest angle and the optimal angle for each vertex [14]. A smaller difference implies a better angular resolution. Average of standard deviations of angular resolution is an average of standard deviations of all angles formed by any two neighboring incident edges for each vertex. A smaller average value indicates a better angular resolution. Standard deviation of edge lengths is a measure of uniformity of edges. A smaller value means more uniform edges. Standard deviation of crossing angles is a measure of uniformity of crossing angles. A smaller standard deviation implies more uniform crossing angle sizes.

For each aesthetic measure, we computed the average across the drawings for BIGANGLE and for Classical respectively. The experimental results including running time are summarized in Table 1. As can be seen from the table, on average, BIGANGLE produced improved drawings in terms of the seven measured aesthetic criteria, compared to Classical. Given that both the cosine force and sine force were designed specifically for increasing crossing angles and for improving angular resolution, it is surprising that they also improved other aesthetics. In particular, BIGANGLE reduced the number of crossings by 26%, increased the size of crossing angles by 5%, shortened the total edge length by 7% and improved the angular resolution by 15%. Also the drawings produced by BIGANGLE were more uniform in size of crossing angles (14%), edge length (13%) and angular resolution (14%).

Table 1
Averages of the aesthetic and time measures for the Rome drawings.

Measurement	BIGANGLE	Classical
# of crossings	23	31
Angle size (deg.)	69.09	65.83
Angle dev. (deg.)	12.78	14.82
Edge length	2.04	2.20
Edge dev.	0.52	0.60
Angular res. (deg.)	54.72	64.47
Angular dev. (deg.)	52.22	60.96
Running time (s)	6.56	1.13

Given the improvements produced by BIGANGLE, the next question is whether these improvements together are significant enough to make a difference in human performance. To demonstrate the effectiveness of BIGANGLE, we conducted a controlled experiment with real users which we present in the next section.

3. The user study

3.1. Design

To compare the effectiveness of the two algorithms, we obtain one set of BIGANGLE drawings and one set of Classical drawings. The two sets of drawings are of the same graphs. We ask users to perform a typical graph reading task with these two sets of drawings. Their task performance (measured as response time and accuracy) and mental effort devoted for the task are recorded during the experiment. In this study, the measure of *visualization efficiency* of Huang et al. [24] is also considered. This measure (E) is calculated based on the recorded performance and mental effort data, using the equation below:

$$\mathrm{E} = \frac{A - T - R}{\sqrt{3}} \tag{7}$$

In the above equation, A, T and R are standardized z-scores of the accuracy, time and effort data, respectively. Visualization efficiency gives us the insight of cognitive gain (response accuracy) relative to cognitive cost devoted (response time and mental effort) in performing a cognitive task. As shown in Eq. (7), high efficiency is achieved when high performance accuracy is attained in association with low mental effort and a short response time. In contrast, low efficiency occurs when low accuracy is associated with high mental effort and a long response time.

Therefore, the experiment employs a within-subject design. The independent variable is algorithm type, which has two conditions: BIGANGLE and Classical. The dependent variables include response time, accuracy, mental effort and visualization efficiency. Based on the testing data obtained, we compute mean values of each dependent measure in the two conditions and test whether there is a statistical difference at the significance level of 0.05.

3.2. Stimuli

We randomly chose one hundred graphs from the testing pool of the Rome graphs used in Section 2.3. For each of the graphs, the corresponding BIGANGLE and Classical drawings were used as experimental stimuli. Therefore there were 100 pairs of BIGANGLE and Classical drawings (200 drawings in total). It is worth mentioning that since these drawings were generated by force-directed algorithms, it was possible that in some of them, a vertex or an edge was too close to another edge for the viewer to discern. However, since we were comparing effectiveness of two algorithms, the drawings were used without any additional layout fine-tuning.

3.3. The task

The task was to determine the length of the shortest path between two vertices. In the field of graph theory, a path is an alternating sequence of unique vertices and edges, starting at a vertex and ending at a vertex. The length of a path is the number of edges in it. It is possible that there are two or more paths between the same two vertices. The shortest path is the path that has the least number of edges. In general, there could be two or more shortest paths for the same two vertices. Looking for shortest paths is a component of many graph reading tasks. Therefore, the shortest-path search task is considered as being representative and has been widely used in studies of graph visualizations (e.g., [43]). For each subject, we first randomly chose the two vertices for each graph with the following constraints:

1. There was only one shortest path between the two vertices. This was to ensure that the subjects searched for the same path if the same two vertices were specified.
2. The path length was between 3 and 6 inclusive. This was to ensure that the task was not too simple or too difficult.

The same vertices were used for the two drawings of the same graph. Therefore, for the same pair of BIGANGLE and Classical drawings, the target path could be different from subject to subject, but was the same within the subject. This helped to make fair and thorough comparisons between the drawings of the two algorithms.

3.4. Subjects

Forty-three subjects participated in the experiment. These subjects had normal or correct-to-normal vision. They were undergraduate students from the School of Information technologies at the University of Sydney studying Computer Science or Information Systems. At the time of their participation, most of them indicated that they had no prior experience related to graphs while some mentioned that they had basic ideas about graphs. They were paid $20 each for their participation, with the best performer being awarded an extra $30.

3.5. Online system

The drawings were displayed by an online system. The system was designed to highlight the two specified vertices as red and display the drawing in the center of the screen with a resolution of 1240×768. When drawings were mapped to the screen, the same scale was used for each pair of BIGANGLE and Classical drawings so that the relative difference between the two drawings was maintained for each aesthetic measure.

When it was started, the system displayed one of the highlighted vertices first. The subject was asked to look at the vertex and press the space bar on the keyboard to show the whole drawing. This was to make sure that every subject had already identified one of the highlighted vertices when the drawing was completely visible. However, the subject was free to start path searching from either of the highlighted vertices. Once the answer was determined, the space bar was pressed; the picture disappeared and the answer screen appeared. On the right side of the answer screen, there were six boxes with each representing a possible answer. The subject was asked to click on one of the boxes. There was also another set of nine boxes for the subject to click on to indicate the mental effort devoted for the drawing just viewed. The mental effort was rated based on a 9-point scale ranging from 1: “very very low effort” to 9: “very very high effort”. After the two answers were given, the subject pressed the space bar to proceed. Then the new vertex of the next drawing was displayed. This process is illustrated in Fig. 6 and was repeated for each drawing.

For each trial, the subject had to view the 200 drawings one by one. The order of these drawings was random with one constraint. The constraint was that each pair of BIGANGLE and Classical drawings should be separated with at least one drawing from another pair. For each subject and each drawing, responses and the time taken for path search were recorded in real time by the system. The time started when the drawing was completely shown and ended when the space bar was pressed.

3.6. Procedure

The experiment was conducted in a quiet computer laboratory room. All subjects performed their tasks independently. There were four laptops which allowed at most four subjects to participate in the study at one time. All the laptops had the same hardware and software specifications. First, subjects were given a set of experimental documents to get familiar with graph concepts and node-link diagrams, the online system, the task and the procedure. After that, the subjects were given a chance to practice the system, ask questions and sign the consent form.

During the introduction session, the subjects were told to perform tasks as accurately and as quickly as they could. They were also informed that the best performer would be chosen based on the following procedure: (1) find the top five subjects who were most accurate; (2) choose the quickest from the top five.

When ready, the subjects indicated to the experimenter and started to perform tasks online. The pace of the experimental session was controlled by the subjects. For example, a rest could be taken when the answer screen appeared if their eyes were tired. During the experiment, there were also two compulsory 2-min breaks that took place at the time when a half and three quarters of the

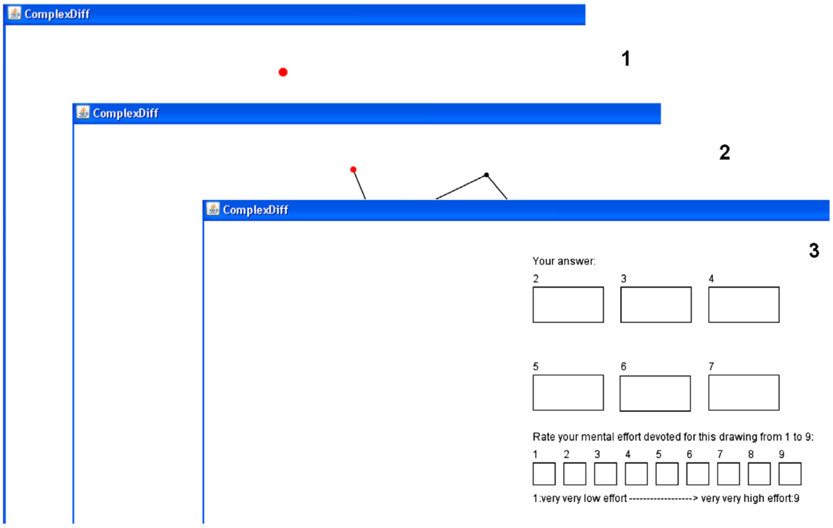

Fig. 6. Illustration of the task process for a drawing.

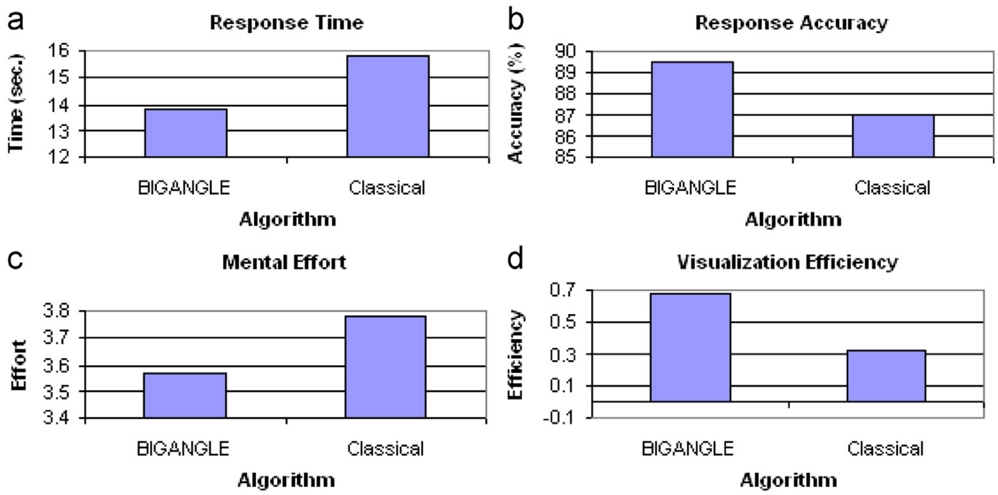

Fig. 7. Averages of response time, accuracy, effort and efficiency.

drawings had been viewed respectively. The system displayed a message indicating the break, and could only proceed after 2 min had passed. After the online task, the subjects were informed of the purpose of the experiment, paid and thanked. The whole experiment took about 80 min on average, including the time taken for introduction, breaks and debriefing.

3.7. Hypotheses

Based on the argument made in Section 1, we made hypotheses as follows:

1. Drawings of BIGANGLE take less time to read than those of Classical.
2. Drawings of BIGANGLE induce fewer errors than those of Classical.
3. Drawings of BIGANGLE take less mental effort to read than those of Classical.
4. Drawings of BIGANGLE are more efficient to read than those of Classical.

3.8. Results

The results are illustrated in Fig. 7. It can be seen that all the four dependent measures were consistently in favor of BIGANGLE. That is, the subjects were faster and more accurate to complete their task with less effort with BIGANGLE drawings than with Classical drawings. Visualization efficiency data indicated that BIGANGLE induced higher cognitive efficiency.

In particular, the subjects spent 13.86 s with BIGANGLE drawings on average. That is 12.33% faster compared to the average time of 15.81 s spent with Classical drawings.

A paired t test indicated that this difference was statistically significant, $t = -7.418$, $p < 0.01$.

The average accuracy with BIGANGLE was 89.54%, which is 2.86% higher than that with Classical (=87.05%). A paired t test indicated that this difference was statistically significant, $t = -3.665$, $p = 0.01$.

The average mental effort with BIGANGLE was 3.57, which is 5.56% more than that with Classical (=3.78). A paired t test indicated that this difference was statistically significant, $t = -7.823$, $p < 0.01$.

The average efficiency for BIGANGLE drawings was 0.68 while it was 0.33 for Classical drawings. That is an improvement of 106%. A paired t test indicated that this difference was statistically significant, $t = 9.125$, $p < 0.01$.

3.9. Discussion

The user study confirmed our hypotheses that BIGANGLE was more effective than Classical in terms of response time, accuracy and mental effort. The visualization efficiency measure also indicated that overall, drawings produced by BIGANGLE were about twice as efficient as those by Classical.

There are two surprising highlights in our work: (1) although cosine force is designed for increasing crossing angles and sine force is designed for improving angular resolution, inclusion of them in BIGANGLE results in other aesthetics being improved as well. (2) Although improvements made in BIGANGLE drawings are not notably significant with 26% in the number of crossings being the biggest, the four human performance measures consistently indicate that BIGANGLE is more effective.

Our finding that the two algorithms are different in effectiveness and in visualization efficiency is particularly interesting given the general perception that when unstructured random graphs are used, it is unlikely that there will be any difference between algorithms in human performance [43]. However, this finding should not come as a surprise if we consider that in perceiving or reading graphs, human performance is actually a reflection of individual effects of and interaction effects between all known and unknown aesthetics. When individual aesthetics are universally improved, the positive benefits can add up and the combined effects are likely to be positive. In contrast, if the aim is to bring only one or two aesthetics to extremes, the benefits of the extremes are likely to be offset by the consequence of other aesthetics being worsened, due to the fact that some aesthetics can be satisfied only at the expense of others.

The two drawings in Fig. 8 give us a further example. The drawing in Fig. 8(1) was produced by an algorithm implemented in OGDF [5], which draws a planar graph on a grid without crossings, while the other drawing was produced by BIGANGLE. It can be seen that there are crossings in the BIGANGLE drawing while there are no crossings at all in the OGDF drawing. However, it is fairly safe to say that tracing paths with the BIGANGLE drawing should be more straightforward than with the OGDF one.

It is important to note that we only measured seven aesthetic criteria in this paper. It is not clear whether other aesthetics, both known and unknown, had been improved by BIGANGLE. However, our user study demonstrated that the improvements in these measured aesthetics were significant enough to make BIGANGLE a better algorithm for producing effective drawings.

It is also worth noting that there have been many force-directed algorithms existing that aim to improve multiple aesthetics. For example, an algorithm by Fruchterman and Reingold [15] draws graphs following generally accepted aesthetics including evenly distributed vertices, minimized edge crossings and uniform edge lengths. Although this algorithm was included in Purchase's study for algorithm comparison, its performance was not notably better compared with other algorithms [32]. Perhaps improvements made in it were not big enough to make a difference.

Further, the simulated annealing work of Davidson and Harel [6] makes compromises between aesthetic criteria. It does a direct calculation of several aesthetic criteria and attempts to reduce a weighted sum of them. Brandenburg et al. [2] performed an extensive empirical analysis of some widely cited force-directed methods including the Davidson–Harel method and it was found that those

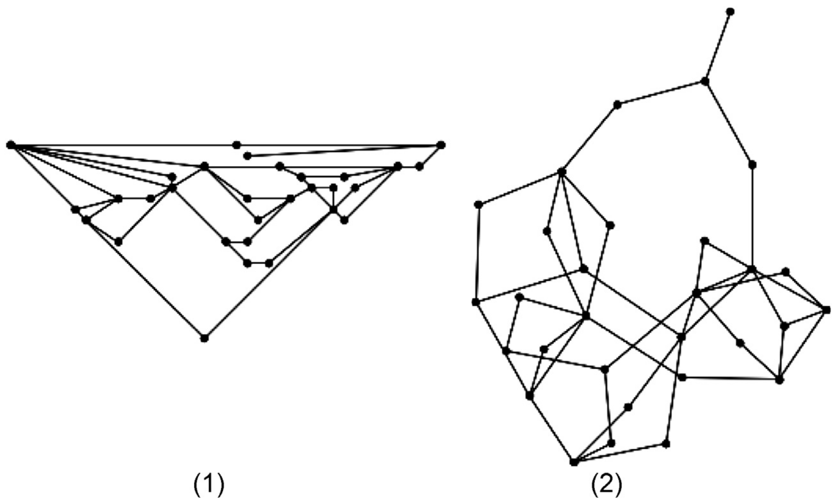

Fig. 8. Two drawings of the same graph (1) produced by OGDF and (2) produced by BIGANGLE.

methods produce drawings with remarkably similar layouts. In this paper we conducted comparisons between BIGANGLE and Classical because the improvements made in BIGANGLE were based on Classical. Although both BIGANGLE and the Davidson–Harel method take multiple aesthetics into consideration, they are essentially different: BIGANGLE is based on the spring embedder model while the Davidson–Harel uses simulated annealing. Aesthetics considered in these two methods are also different. It would be interesting to compare these two methods computationally (in terms of multiple aesthetic criteria in the resultant drawings) and empirically (in terms of human graph reading performance). However, this is beyond the scope of this paper.

4. General discussion

The main focus of human studies in graph drawing has been on validation and reinforcement of existing aesthetics (e.g., [1,25,34]). Recent studies have shown a growing interest in proposing new aesthetics based on graph reading behavior and drawing preference of end users. For example, van Ham and Rogowitz [41] asked users to create best possible drawings based on initial layout and found that users often arranged vertices belonging to the same cluster structure in a convex hull represented by the cluster's edges. Dwyer et al. [10] asked users to optimize existing layouts and user generated layouts were compared with layouts generated by graph drawing algorithms. One of their findings was that user preferred layout could be different from the layout that is best for task performance. Huang [21,22] asked users to perform a shortest-path search task while observing their eye movements and found that increasing crossing angles improved readability of graph drawings. Purchase et al. [38] asked users to draw graphs based on adjacency lists and found that aligning vertices and edges to an underlying grid was important. Yu et al. [45] asked user to draw their personal networks and found that users made use of both vertex position and line length to differentiate roles and express intimacy.

Indeed, individual aesthetics have proved to be important in their own right in designing human-centered automatic graph drawing algorithms and visualization systems. However, the current study suggests alternative research directions, which we formulate as below.

4.1. An aesthetic of overall visual quality

The current study investigates the collective effect of commonly applied aesthetics and indicates that aesthetics should not be considered separately. It is necessary to consider the effect of an aesthetic criterion, together with other visual factors. Take a look at two drawings in Fig. 3. According to Purchase [32], the FD-K drawing was intended to maximize symmetry with no bends, while the POGb drawing was drawn with maximum orthogonality, no crossings and minimum bends. Given the largest effect of crossings found in prior research, the POGb drawing would be expected to have a better performance, which was in fact not the case in that study. Perhaps this was because the effect of crossings in grid-based layout was not as strong as that in force-directed layout. In other words, effect of an aesthetic may change when layout requirements change.

A number of experiments have been conducted in the graph drawing literature to evaluate performance of specific algorithms in terms of drawing quality. In those experiments, drawing quality was mainly evaluated using either the values of pre-specified aesthetic criteria (e.g., [2,8,19]), expert judgement of the authors (e.g., [18]), or criteria specific to the algorithms in consideration (e.g., [4,46]), with few exceptions using performance of real users (e.g., [31,32]). Measuring individual aesthetics gives us a sense of the extent to which the drawing conforms to them. However, drawings produced by different algorithms satisfy different sets of layout requirements including drawing conventions, constraints and aesthetics. This makes it hard to compare drawings in respect to readability, or overall visual quality, based on individual aesthetics. For example, it is difficult to judge whether a drawing with no crossings, but having long edges with many bends is better or more readable than a drawing with a few crossings, but having short uniform straight lines.

When overall visual quality is a main concern, an aesthetic that measures it may come handy and be more preferable [36]. Based on the discussion above, we therefore propose to measure overall visual quality as a function of aesthetics, with the function itself changing with drawing conventions and constraints used for drawing construction. That is, given a specific set of drawing conventions and constraints, overall visual quality (y) can be theoretically expressed using aesthetics (x) as below:

$$y=\sum_{i=1}^{n} f(x_i)+\sum_{j,k=1}^{n} f(x_j y_k)+e \tag{8}$$

In Eq. (8), the first term is the sum of aesthetics, the second is the sum of interaction factors and the third is the error term that includes measuring errors and aesthetics that are not considered. To make this equation practically useful, some assumptions must be made first. For example, assume that there is a linear function between overall visual quality and individual aesthetics. Further, to make the comparison of overall visual quality between drawings meaningful, it is necessary to measure aesthetics based on the same scale. One way of doing that is to measure aesthetics as usual and standardize the raw measurement data into z-scores. Another way is to use the metrics proposed by Purchase [33] which measure all aesthetics as a real number between 0 and 1 inclusive. It should be noted that Ware et al. [44] derived a regression equation based on user performance data and aesthetics, which predicts the time needed for a task. Eq. (8) is different in that it measures overall visual quality. The usability and reliability of our equation needs empirical validation, and can be validated based on correlations between user perceived and equation based values of visual quality. It is hoped that using this single quality indicator, visualization designers can quickly evaluate options available and determine whether further improvement is needed.

4.2. Beyond ranking: quantitative priority list of aesthetics

While often limited by being able to satisfy only one or two aesthetic criteria to the fullest, many algorithms also take into account other aesthetics, when possible, in their attempt to produce visually pleasing and easy-to-read drawings. Eq. (8) indicates that a similar level of visual quality can be achieved by considering different sets of aesthetics. Since different aesthetics affect graph readability to different extents, it is necessary to understand the relative importance between aesthetics in order to determine what aesthetics to consider. Such understanding is important in two ways. First, we may want to implement the least number of aesthetics and achieve greatest readability. Based on the relative importance, we are able to choose the most important ones from the candidate aesthetics. Second, when aesthetics conflict, relative importance helps us to decide which one should be given priority.

Attempts have been made to build priority lists of aesthetics in the literature based on their relative importance. These lists are made for different domains and using different methods. Himsolt [19] compared 12 algorithms, in which the aesthetic ranking was obtained by observing the quality of drawings. Purchase et al. [37] conducted a questionnaire study for UML diagrams, in which the priority order was made based on user preference. Purchase [35] conducted a controlled experiment comparing five aesthetics, in which performance data were used to determine the perceived importance of aesthetics for abstract graphs. Other versions of priority listing were also proposed based on theoretic discussions and personal judgements of the authors (e.g., [13,26,39]).

One limitation of the existing lists is that they only give us ordering. Quantitative information on the distance between aesthetics is missing. As an initial step to fill this gap, Huang and Huang [28] conducted an experimental study exploring relative importance between crossing number and crossing angle, in which the quantitative information was obtained based on multiple regression analysis. Such quantitative information is important since we can use it to estimate the cost of choosing one aesthetic against another, such as "to achieve the same benefit of removing one crossing, we need to increase the crossing angle by at least 10 degrees". Further, with this information Eq. (8) can be refined by assigning weights (w) to aesthetics to determine the overall visual quality more accurately:

$$y = \sum_{i=1}^{n} w_i f(x_i) + \sum_{j,k=1}^{n} w_j w_k f(x_j y_k) + e \tag{9}$$

5. Conclusion

Graph visualizations and drawing algorithms are often designed to optimize one or two important aesthetics. Empirical research has showed that although algorithms are based on different aesthetics, drawings produced by them have comparable effectiveness. In this paper, we propose and demonstrate that graph drawings can be more effective if compromises are made between aesthetics.

We have presented a force-directed algorithm: BIGANGLE. This algorithm adds two new forces to a classical spring algorithm: Classical. Statistical data indicate that the introduction of these two forces results in multiple aesthetics being improved at the same time. A user study is conducted comparing the readability of drawings produced by these two algorithms. The results show that users perform better with BIGANGLE drawings. It is also found that BIGANGLE drawings induce lower cognitive load and are significantly more efficient with respect to the measure of visualization efficiency.

The finding of our user study indicates that aesthetics should not be considered separately in algorithm design and visualization evaluation. To be more specific, improving aesthetics as many as possible at the same time, even to small extents, will have a better chance to make resultant drawings more readable. Our study suggests two research directions: measuring overall visual quality and building priority lists of aesthetic criteria with quantitative information.

Clearly, further research is needed to explore the feasibility and benefits of making compromises between aesthetics. It is hoped that more research will be conducted in line with human-centered algorithm and visualization design.

References

[1] D. Archambault, H.C. Purchase, B. Pinaud, Animation, small multiples, and the effect of mental map preservation in dynamic graphs, IEEE Transactions on Visualization and Computer Graphics 17 (2011) 539–552.
[2] F. Brandenburg, M. Himsholt, C. Rohrer, An experimental comparison of force-directed and randomized graph drawing algorithms, in: Proceedings of Symposium on Graph Drawing, GD '95, Lecture Notes in Computer Science, vol. 1027, Springer-Verlag, 1996, pp. 76–87.
[3] U. Brandes, Drawing on physical analogies, in: M. Kaufmann, D. Wagner (Eds.), Drawing Graphs: Methods and Models, Lecture Notes in Computer Science, vol. 2025, Springer-Verlag, 2001, pp. 71–86.
[4] U. Brandes, C. Pich, An experimental study on distance-based graph drawing, in: Proceedings of 16th International Symposium on Graph Drawing, GD 2008, Lecture Notes in Computer Science, vol. 5417, Springer-Verlag, 2009, pp. 218–229.
[5] M. Chimani, C. Gutwenger, M. Jünger, K. Klein, P. Mutzel, M. Schulz, The open graph drawing framework, in: 15th International Symposium on Graph Drawing, GD' 07, Sydney, 2007.
[6] R. Davidson, D. Harel, Drawing graphs nicely using simulated annealing, ACM Transactions on Graphics 15 (4) (1996) 301–331.
[7] G. Di Battista, P. Eades, R. Tamassia, I. Tollis, Graph Drawing: Algorithms for the Visualization of Graphs, Prentice-Hall, Upper Saddle River, NJ, 1998.
[8] G. Di Battista, A. Garg, G. Liotta, R. Tamassia, E. Tassinari, F. Vargiu, An experimental comparison of four graph drawing algorithms, Computational Geometry: Theory and Applications 7 (1997) 303–325.
[9] H.A.D. do Nascimento, P. Eades, C.F.X. de Mendonça Neto, A multi-agent approach using a-teams for graph drawing, in: ISCA Conference on Intelligent Systems, 2000, pp. 39–42.
[10] T. Dwyer, B. Lee, D. Fisher, K.I. Quinn, P. Isenberg, G. Robertson, C. North, A comparison of user-generated and automatic graph layouts, IEEE Transactions on Visualization and Computer Graphics 15 (2009) 961–968.
[11] T. Dwyer, K. Marriott, M. Wybrow, Integrating edge routing into force-directed layout, in: Proceedings of 14th International Symposium on Graph Drawing, GD 2006, Lecture Notes in Computer Science, vol. 4372, Springer-Verlag, 2007, pp. 8–19.

[12] P. Eades, A heuristic for graph drawing, Congressus Numerantium 42 (1984) 149–160.
[13] H. Eichelberger, Nice class diagrams admit good design, in: Proceedings of the 2003 ACM Symposium on Software visualization, SoftVis '03, ACM, New York, NY, USA, 2003, pp. 159–167.
[14] B. Finkel, R. Tamassia, Curvilinear graph drawing using the force-directed method, in: Proceedings of 12th International Symposium on Graph Drawing, GD 2004, Lecture Notes in Computer Science, vol. 3383, Springer-Verlag, 2004, pp. 448–453.
[15] T. Fruchterman, E. Reingold, Graph drawing by force-directed placement, Software: Practice and Experience 21 (1991) 1129–1164.
[16] M.R. Garey, D.S. Johnson, Crossing number is np-complete, SIAM Journal on Algebraic and Discrete Methods 4 (1983) 312–316.
[17] GDToolkit, 2009. GDToolkit: Graph Drawing Toolkit, 〈http://www.dia.uniroma3.it/~gdt/〉.
[18] S. Hachul, M. Jünger, An experimental comparison of fast algorithms for drawing general large graphs, in: Proceedings of the Graph Drawing, Springer-Verlag, 2005, pp. 235–250.
[19] M. Himsolt, Comparing and evaluating layout algorithms within graphed, Journal of Visual Languages and Computing 6 (1995) 255–273.
[20] S.-H. Hong, D. Merrick, H. Nascimento, The metro map layout problem, Journal of Visual Language and Computing 17 (3) (2006) 203–224.
[21] W. Huang, An Eye Tracking Study into the Effects of Graph Layout, Technical Report 639, University of Sydney, 2009.
[22] W. Huang, P. Eades, How people read graphs, in: Proceedings of the 2005 Asia-Pacific Symposium on Information Visualisation, APVis '05, vol. 45, Australian Computer Society, Inc., Darlinghurst, Australia, 2005, pp. 51–58.
[23] W. Huang, P. Eades, S. Hong, C. Lin, Improving force-directed graph drawings by making compromises between aesthetics, in: IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC '10, IEEE, 2010, pp. 176–183.
[24] W. Huang, P. Eades, S.-H. Hong, Measuring effectiveness of graph visualizations: a cognitive load perspective, Information Visualization 8 (3) (2009) 139–152.
[25] W. Huang, S.-H. Hong, P. Eades, Layout effects on sociogram perception, in: Proceedings of 13th International Symposium on Graph Drawing, GD '05, Springer, 2006, pp. 262–273.
[26] W. Huang, S.-H. Hong, P. Eades, Effects of sociogram drawing conventions and edge crossings in social network visualization, Journal of Graph Algorithms and Applications (2007) 397–429.
[27] W. Huang, S.-H. Hong, P. Eades, Effects of crossing angles, in: Proceedings of IEEE Pacific Visualization Symposium 2008, 2008, pp. 41–46.
[28] W. Huang, M. Huang, Exploring the relative importance of crossing number and crossing angle, in: Proceedings of the Second International Symposium on Visual Information Communication, VINCI '10, ACM, New York, NY, USA, 2010, pp. 10:1–10:8.
[29] M. Kaufmann, D. Wagner (Eds.), Drawing Graphs, Lecture Notes in Computer Science, vol. 2025, Springer, Berlin, 2001.
[30] P. Kumar, K. Zhang, Node overlap removal in clustered directed acyclic graphs, Journal of Visual Languages and Computing 20 (2009) 403–419.
[31] M. Pohl, M. Schmitt, S. Diehl, Comparing the readability of graph layouts using eyetracking and task-oriented analysis, in: Computational Aesthetics in Graphics, Visualization, and Imaging, 2009.
[32] H. Purchase, Performance of layout algorithms: comprehension, not computation, Journal of Visual Languages and Computing 9 (1998) 647–657.
[33] H. Purchase, D. Carrington, J. Allder, Empirical evaluation of aesthetics-based graph layout, Empirical Software Engineering 7 (3) (2002) 233–255.
[34] H. Purchase, R. Cohen, M. James, Validating graph drawing aesthetics, in: Proceedings of Symposium on Graph Drawing, GD '95, Lecture Notes in Computer Science, vol. 1027, Springer-Verlag, 1996, pp. 435–446.
[35] H.C. Purchase, Which aesthetic has the greatest effect on human understanding? in: Proceedings of the Fifth International Symposium on Graph Drawing, GD '97, Springer-Verlag, London, UK, 1997, pp. 248–261.
[36] H.C. Purchase, Metrics for graph drawing aesthetics, Journal of Visual Languages and Computing 13 (2002) 501–506.
[37] H.C. Purchase, J.-A. Allder, D.A. Carrington, User preference of graph layout aesthetics: a UML study, in: Proceedings of the Eighth International Symposium on Graph Drawing, GD '00, Springer-Verlag, London, UK, 2001, pp. 5–18.
[38] H.C. Purchase, C. Pilcher, B. Plimmer, Graph drawing aesthetics—created by users, not algorithms, IEEE Transactions on Visualization and Computer Graphics 18 (1) (2012) 81–92.
[39] K. Sugiyama, Graph Drawing and Applications for Software Engineering and Knowledge, World Scientific, 2002.
[40] K. Sugiyama, K. Misue, A simple and unified method for drawing graphs: magnetic-spring algorithm, in: Proceedings of Graph Drawing, 1994, pp. 364–375.
[41] F. van Ham, B. Rogowitz, Perceptual organization in user-generated graph layouts, IEEE Transactions on Visualization and Computer Graphics 14 (2008) 1333–1339.
[42] C. Ware, R. Bobrow, Motion to support rapid interactive queries on node-link diagrams, ACM Transactions on Applied Perception 1 (1) (2004) 3–18.
[43] C. Ware, P. Mitchell, Visualizing graphs in three dimensions, ACM Transactions on Applied Perception 5 (1) (2008) 1–15.
[44] C. Ware, H. Purchase, L. Colpoys, M. McGill, Cognitive measurements of graph aesthetics, Information Visualization 1 (2002) 103–110.
[45] L. Yu, J.V. Nickerson, B. Tversky, Discovering perceptions of personal social networks through diagrams, in: Proceedings of the Sixth International Conference on Diagrammatic Representation and Inference, Diagrams '10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 352–354.
[46] F. Zaidi, D. Archambault, G. Melançon, Evaluating the quality of clustering algorithms using cluster path lengths, in: Proceedings of the 10th Industrial Conference on Advances in Data Mining: Applications and Theoretical Aspects, ICDM '10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 42–56.