# A User Study in Similarity Measures for Graph Drawing \*

Stina Bridgeman and Roberto Tamassia

Center for Geometric Computing
Department of Computer Science
Brown University
Providence, Rhode Island 02912-1910
{ssb,rt}@cs.brown.edu

**Abstract.** The need for a similarity measure for comparing two drawings of graphs arises in problems such as interactive graph drawing and the indexing or browsing of large sets of graphs. This paper builds on our previous work [3] by defining some additional similarity measures, refining some existing ones, and presenting the results of a user study designed to evaluate the suitability of the measures.

#### 1 Introduction

The question of how similar two drawings of graphs are arises in many situations. In interactive graph drawing, the graph being visualized changes over time and it is important to preserve the user's "mental map" [12] so she does not spend a lot of time relearning the drawing after each update. Animation can provide a smooth transition between drawings, but is of limited use if the drawings are too different. In layout adjustment applications, an existing drawing is modified to improve an aesthetic quality without destroying the user's mental map.

Another application is in indexing or browsing large sets of graphs. The SMILE graph multidrawing system (Biedl et. al. [1]) produces many drawings of a graph, and its graph browser arranges the drawings so that similar ones are near each other to help the user navigate the system's responses. Similarities between drawings can also be used as a basis for indexing and retrieval. In character and handwriting recognition, a written character may be transformed into a graph and compared to a database of characters to find the closest match.

Let M be a similarity measure defined so that measure's value is 0 when the drawings are identical. In order to be useful, M should satisfy three properties:

**Rotation:** Given drawings D and D',  $M(D, D'_{\theta})$  should have the minimum value for the angle a user would report as giving the best match, where  $D'_{\theta}$  is D' rotated by an angle of  $\theta$  with respect to its original orientation.

<sup>\*</sup> Research supported in part by the National Science Foundation under grants CCR-9732327 and CDA-9703080, and by the U.S. Army Research Office under grant DAAH04-96-1-0013.

J. Marks (Ed.): GD 2000, LNCS 1984, pp. 19–30, 2001.

<sup>©</sup> Springer-Verlag Berlin Heidelberg 2001

**Ordering:** Given drawings D, D', and D'', M(D, D') < M(D, D'') if and only if a user would say that D' is more like D than D'' is like D.

**Magnitude:** Given drawings D, D', and D'',  $M(D, D') = \frac{1}{c}M(D, D'')$  if and only if a user would say that D' is c times more like D than D'' is like D.

This paper describes a user study intended to address rotation and ordering, and to explore a method for addressing magnitude. Data cannot be collected directly for the magnitude part because it is very difficult to assign numerical similarity values to pairs of drawings, much more so than judging ordering (Wickelgren [18]). As a result, other data (e.g., response times) presumed to be related to similarity must be collected instead. The assumption of the suitability of the data can be partially tested by determining if using the data to make ordering decisions between drawings is consistent with the user responses.

This study improves on our previous work [3] in several ways:

- More Experimental Data: A larger pool of users (103 in total) was used for determining the "correct" behavior for the measure.
- Refined Ordering Part: Users made only pairwise judgements between drawings rather than being asked to order a larger set.
- Addressing of Magnitude Criterion: Magnitude was not addressed in [3].
- More Realistic Drawing Alignment: The previous drawing alignment method allowed one drawing to be scaled arbitrarily small with respect to the other; the new method keeps the same scale factor for both drawings.
- Refinement of Measures: For those measures computed with pairs of points, pairs involving points from the same vertex are skipped.
- New Measures: Several new measures have been included.

We describe the experimental setup in Section 2, the measures evaluated in Section 3, the results in Section 4, and conclusions and directions for future work in Section 5.

## 2 Experimental Setup

This study focuses on similarity measures for orthogonal drawings of nearly the same graph. "Nearly the same graph" means that the graphs differ by only a small number of vertex and edge insertions and deletions. The focus on orthogonal drawings is motivated by the availability of an orthogonal drawing algorithm capable of producing many drawings of the same graph, and by the amount of work done on interactive orthogonal drawing (e.g., Biedl and Kaufmann [2], Fößmeier [8], Papakostas, Six, and Tollis [14], and Papakostas and Tollis [15]).

The graphs used were generated from a base set of 20 graphs with 30 vertices each, taken from an 11,582-graph test suite. [7] Each of 20 base graphs was drawn using Giotto [17]. Forty modified drawings were created by adding a degree 2 and a degree 4 vertex to separate copies of each base drawing. Each modified drawing is identical to its base drawing except for the new vertex and its adjacent edges, placed as a user might draw them in an editor. Finally, four new drawings were

produced for each modified drawing using InteractiveGiotto [4]. The new drawings range from very similar to the base drawing to significantly different.

The experiment consisted of three parts, to address the three evaluation criteria. In all cases, the user was asked to respond as quickly as possible without sacrificing accuracy, and the user's response and time to answer were recorded. Each trial timed out after 30 seconds if the user did not respond.

Rotation Part. The rotation part directly addresses the rotation criterion. The user is presented with a screen as shown in Figure 1. The eight drawings on the right are different orientations of the same new (InteractiveGiotto) drawing; the one drawing on the left is the corresponding base drawing. The eight orientations are rotations by the four multiples of  $\pi/2$ , with and without an initial flip around the x-axis. For orthogonal drawings, only multiples of  $\pi/2$  are meaningful since clearly rotation by any other angle is not the correct choice. The vertices are not labelled to emphasize the layout of the graph over the specifics of vertex names.

The user chooses the orientation that looks most like the base drawing. She may click the "can't decide" button if she cannot make a decision.

**Ordering Part.** The ordering part directly addresses the ordering criterion. In this part, the user is presented with a screen as shown in Figure 2. The two rightmost drawings are two different new drawings of the same modified drawing; the leftmost is the corresponding base drawing. The vertices are not labelled.

The user chooses which of the two rightmost drawings looks most like the base drawing. She may click on "can't decide" if she cannot make a decision.

**Difference Part.** The difference part addresses the magnitude criterion by gathering response times on a task, with the assumption that the user will complete the task more quickly if the drawings are more similar. Figure 3 shows the screen presented to the user. The drawing on the right is one of the InteractiveGiotto-produced drawings; the drawing on the left is the corresponding base drawing.

The user identifies the vertex in the right drawing missing from the left drawing. Vertices have random two-letter names, because the task is too difficult with unlabelled vertices; labels are often important in real-world situations. Corresponding vertices in drawings in a trial have the same name, but the names are different for separate trials to prevent the user from learning the answer.

A total of 103 students completed the three parts as part of a homework assignment in a second-semester CS course at Brown University. They had some familiarity with graphs through lectures and a programming project.

The students used an online system. A writeup was presented explaining how to use the system, and the directions were summarized each time it was run. Each of the three parts was split into four runs, so the students would not have to stay focused for too long without a break. The first run of each part was a practice run. In later runs, graphs were assigned to the student randomly so that 1/3 of the students worked with each graph. Within each run, the sequence of the individual trials and the ordering of the right-hand drawings in the rotation and

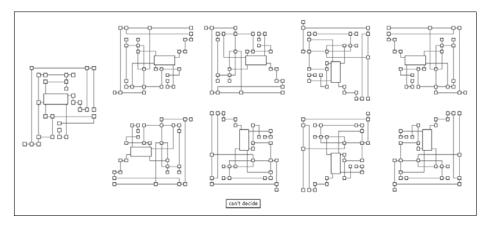
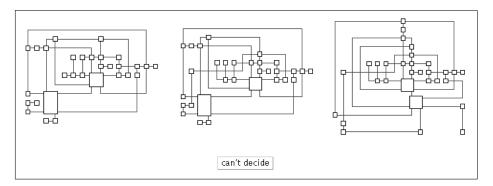
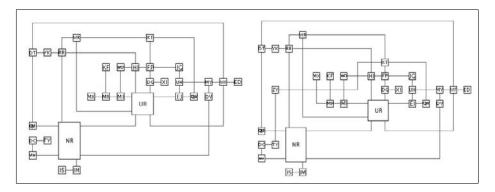


Fig. 1. The rotation part.



 $\textbf{Fig. 2.} \ \ \textbf{The ordering part}.$ 



 ${f Fig.\,3.}$  The difference part.

ordering parts was chosen randomly. After the students completed all of parts, they answered a short questionnaire about their experiences. The questions were:

- 1. (Ordering and Rotation) What do you think makes two drawings of nearly the same graph look similar? Are there factors that influenced your decisions? Did you find yourself looking for certain elements of the drawing in order to make your choice?
- 2. (**Difference**) What factors helped you locate the extra vertex more quickly? Did you compare the overall look of the two drawings in order to aid your search, or did you just scan the second drawing?
- 3. (All Parts) As you consider your answers, think about what this means for a graph drawing algorithm that seeks to preserve the look of the drawing. What types of things would it have to take into account?

#### 3 Measures Evaluated

All of the measures evaluated in this study are listed below. Most are the same as or similar to those described in [3]; the primary difference is that all have been scaled to have a value between 0 and 1, with 0 indicating identical drawings. The upper bound is often based on the worst-case scenario and may not be achievable by an actual drawing algorithm. Formal definitions and more extensive motivations for many of the measures can be found in [3]; only those measures which are new or have changed significantly are given more full treatment below.

First, we consider some preliminaries.

Corresponding Objects. The graphs in the drawings being compared are assumed to be the same; if not, only the common subgraphs are used. Thus, each vertex and edge of G has a representation in each of the drawings, and it is meaningful to talk about the *corresponding vertex or edge* in one drawing given a vertex or edge in the other drawing.

**Point Set Selection.** Most of the measures are defined in terms of point sets derived from the edges and vertices of the graph. Points can be selected in many ways; inspired by North [13], one point set contains the four corners of each vertex. A second point set, suggested by feedback from the study (section 4), contains only corner points near the edge of the drawing. Like vertices and edges, each point in one drawing has a corresponding point in the other drawing.

A change from the previous experiment [3] is the exclusion of pairs of points derived from the same vertex. This can have a great effect on measures which involve nearest neighbors, for example, because a point's nearest neighbor will often be another corner of the same vertex, which does not convey much information about how that vertex relates to other vertices in the drawing. This is not explicitly written in the definitions below for clarity of notation, but it should be assumed unless stated otherwise.

Drawing Alignment. For measures comparing coordinates between drawings, the value of the measure is very dependent on how well the drawings are aligned. Previously, drawings were aligned by simultaneously adjusting the scale and translation of one with respect to the other, which could reduce one drawing to a very small area if the drawings did not match well. The new alignment method treats the scale and translation factors separately. With orthogonal drawings, there is a natural underlying grid which can be used to equalize the scale. The translation factor is then chosen to minimize the distance squared between corresponding points. This alignment method is intended to better match how a person might align the drawings, since it does not seem likely that someone would mentally shrink or enlarge one drawing greatly with respect to the other.

Suitability for Ordering vs. Rotation and Ordering. Some measures do not depend on the relative rotation of one drawing with respect to the other. They are included even though they fail the rotation test because not all applications require determining the proper rotation for drawings. Also, a successful ordering-only measure could be combined with one which performs poorly on the ordering task but rotates well to obtain a measure which is good at both. Measures suitable for ordering only are marked [order only] below.

**Notation.** In the following, P and P' refer to point sets for drawings D and D', respectively, and  $p' \in P'$  is the corresponding point for  $p \in P$  (and vice versa). Let d(p,q) be the Euclidean distance between points p and q.

The first group of measures measure the degree of matching between the point set as the maximum mismatch between points in one set and points in another.

Undirected Hausdorff Distance. The undirected Hausdorff distance is a standard metric for determining the quality of the match between two point sets. It does not take into account the fact that the point sets may be labelled.

**Paired Hausdorff Distance.** The *paired Hausdorff distance* is an adaptation of the undirected Hausdorff distance for labelled point sets, and is defined as the maximum distance between two corresponding points:

$$\operatorname{phaus}(P,P') = \max_{p \in P} \, d(p,p')$$

The position measures are motivated by the idea that the location of the points on the page is important, and points should not move too move far between drawings.

**Average Distance.** Average distance is the average distance a point moves between drawings.

**Nearest Neighbor between.** Nearest neighbor between is based on the idea that a point's original location should be closer to its new position than any other point's new position. The weighted version considers the number of points closer to the point's original location than the point's new location rather than simply whether or not the point is closest.

Relative position measures are based on the idea that the relative position of points should not change. "Relative position" includes both the distance between the points and the angles, though each measure is concerned with only one property.

Orthogonal Ordering. Orthogonal ordering measures the change in angle between pairs of points. In the constant-weighted version, all changes of angles are weighted equally; in the linear-weighted version changes in the north, south, east, west relationships are weighted more heavily than changes in angle which do not affect this relationship.

**Ranking.** The *ranking* measure considers the relative horizontal and vertical position of the point. (Ranking is a component of the similarity measure used in SMILE [1].) Let  $\operatorname{right}(p)$  and  $\operatorname{above}(p)$  be the number of points to the right of and above p, respectively.

$$\operatorname{rank}(P, P') = \frac{1}{\operatorname{UB}} \sum_{p \in P} \min\{ | \operatorname{right}(p) - \operatorname{right}(p') | + | \operatorname{above}(p) - \operatorname{above}(p') |, \operatorname{UB} \}$$

$$UB = 1.5(|P| - 1)$$

Of note here is that the upper bound is taken as 1.5(|P|-1) instead of the actual worst-case value 2(|P|-1) because it scales the measure more satisfactorily.

Average Relative Distance [order only]. The average relative distance is the average change in distance between pairs of points.

 $\lambda$ -Matrix [order only]. The  $\lambda$ -matrix model is used by Lyons, Meijer, and Rappaport [10] to evaluate cluster-busting algorithms. It is based on the concept of order type used by Goodman and Pollack [9], where two sets of points P and P' have the same order type if, for every triple of points (p,q,r), they are oriented counterclockwise if and only if (p',q',r') are also oriented counterclockwise.

The next group of measures are guided by the philosophy that each point's neighborhood should be the same in both drawings. They do not explicitly take into account either the point's absolute position or its position relative to other points.

Nearest Neighbor Within [order only]. For nearest neighbor within, a point's neighborhood is its nearest neighbor. The weighted version includes the number of points closer to the point than its nearest neighbor, whereas the unweighted version considers only whether or not the nearest neighbor remains the same.

 $\epsilon$ -Clustering [order only].  $\epsilon$ -clustering defines the neighborhood for each point to be its  $\epsilon$ -cluster, the set of points within a distance  $\epsilon$ , defined as the maximum distance between a point and its nearest neighbor.

Separation-Based Clustering [order only]. In the separation-based clustering measure, points are grouped so that each point in a cluster is within some distance  $\delta$  of another point in the cluster and at least distance  $\delta$  from any point not in the cluster. The intuition is that the eye naturally groups things based on the surrounding whitespace.

Formally, for every point p in cluster C such that |C| > 1, there is a point  $q \neq p \in C$  such that  $d(p,q) < \delta$  and there is no point  $r \notin C$  such that  $d(p,r) < \delta$ . If C is a single point, only the second condition holds.

Let clus(p) be the cluster to which point p belongs.

$$sclus = 1 - \frac{|S_I|}{|S_U|}$$

where

$$S_I = \{ (p,q) \mid p,q \in P, \operatorname{clus}(p) = \operatorname{clus}(q) \text{ and } \operatorname{clus}(p') = \operatorname{clus}(q') \}$$
  
 $S_U = \{ (p,q) \mid p,q \in P, \operatorname{clus}(p) = \operatorname{clus}(q) \text{ or } \operatorname{clus}(p') = \operatorname{clus}(q') \}$ 

Edge measures are based on the graph's edges.

**Shape.** The *shape* measure treats the edges of the graph as sequences of north, south, east, and west segments. The sequences are compared using the edit distance. The normalized version of the measure uses the algorithm of Marzal and Vidal [11] to adjust for the length of the sequence.

#### 4 Results

For rotation and ordering, each measure was evaluated according to what fraction of the time its choice (determined by its value for the drawings the user is choosing between) agreed with the user's response. If the measure had the same minimum value for two or more drawings, the trial was marked as a tie and was only considered correct if the user clicked the "can't decide" button or if the task timed out, even if one of the possibilities agreed with the user's choice.

Figure 4 shows the fraction of trials for which the rotation criterion was satisfied. Only unweighted nearest neighbor between had a significant number of cases where the user picked one of the drawings but the measure reported a tie. The column labelled "mode" shows the percentage of the trials for which the user's choice agreed with the most common choice (the mode) for that set of drawings. Since a single measure will always select the same rotation for the same pair of drawings, this value is the best a measure could expect to do.

It is disappointing that the correctness for the mode is below 50% since it means even the best measure will tend to rotate drawings incorrectly much of the

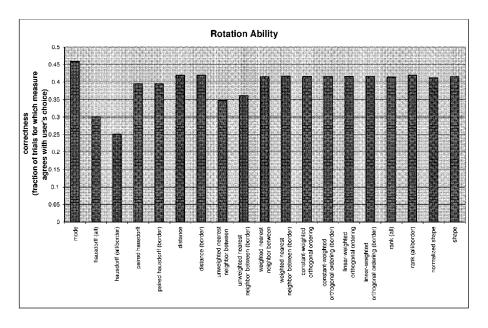


Fig. 4. Results for the rotation part. "border" indicates that the border point set was used; "all" indicates that pairs of points derived from the same vertex were included.

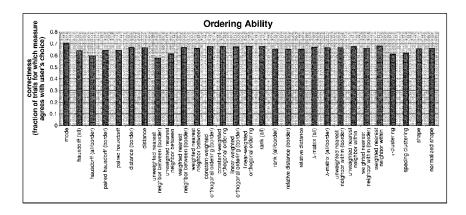


Fig. 5. Results for the ordering part. "border" indicates that the border point set was used; "all" indicates that pairs of points derived from the same vertex were included.

time. The results are better when the new drawing is more similar to the base drawing — the mode correctness averages 64% for the most similar drawings.

Figure 5 shows the fraction of trials for which the order criterion was satisfied. Tie results had only a slight effect on the unweighted nearest neighbor between, unweighted nearest neighbor within, and shape measures. The column labelled "mode" again shows how often the users agree with the most common opinion.

Most of the measures perform quite well when compared to the mode, and the most notable exceptions include those measures that performed most poorly on rotation. Also, as expected, the measures generally performed better when one of the drawings was clearly more like the base drawing than the other.

The goal in the difference part was to use the user's response times as an indicator of similarity, with the idea that a user can locate the new vertex faster if the drawings are more similar. To test the validity of this, the times on the difference part were used to order the pairs of drawings used in the ordering task. The results were very unsatisfactory, achieving only 45% correctness (compare to Figure 5, where even the worst measure reached nearly 58% correctness). As a result, the times on the difference task are not a good indicator of similarity and are not suitable for evaluating measures with respect to the magnitude criterion.

The responses to the final questionnaire yielded several interesting notes. As might be expected, the answers about what makes two drawings look similar included a sizable percentage (35%) who said preserving the position, size, number of large vertices was important and another large percentage (44%) who said they looked for distinctive clusters and patterns of vertices, such as chains, zigzags, and degree 1 vertices. More surprising was that 44% of the students said that borders and corners of the drawing are more important than the interior when looking for similarity. This is supported by research in cognitive science indicating that people often treat filled and outline shapes as equivalent, focusing primarily on the external contour (Wickelgren [18]). Many of these students mentioned the importance of "twiddly bits around the edges" — distinctive clusters and arrangements of vertices, made more obvious by being on the border. Related comments included that the orientation and aspect ratio of the bounding box should remain the same, and that the outline of the drawing should not change. Another sizable group (34%) commented that the "general shape" of the drawing is important.

On the difference part, several users said that the task was difficult and the system timed out frequently. The usefulness of the "big picture" view — looking at the overall shape of the drawing — was contested, with nearly equal numbers reporting that the overall look was useful in the task, and that it was confusing and misleading. About 16% of the users made limited use of the overall look, using it on a region-by-region basis to quickly eliminate blocks that remained the same and falling back on simply scanning the drawing or matching corresponding vertices and tracing edges when the regions were too different. Another 24% used vertex-by-vertex matching from the beginning. Similar-sized groups discovered and exploited shortcuts based on how the modified drawings was constructed (20%), and reported searching for vertices with extra edges rather than searching for new vertex directly (28%).

The most common answers about what a graph drawing algorithm should take into account to preserve the look of the drawing echoed those from the rotation/ordering question: maintaining vertex size and shape, the relative positions of vertices, the outline of the drawing, and clusters.

#### 5 Conclusions and Future Work

The results from the rotation and ordering parts show that, for the most part, there is not a large difference in the performance of the tested measures. However, it is interesting to note that the worst-performing measures (undirected Hausdorff distance, unweighted nearest neighbor between,  $\epsilon$ -clustering, and spacing clustering) give the least weight to absolute and relative point positions, suggesting that absolute and relative point positions are indeed important to similarity. It also suggests that point positions are less significant for ordering because the worst measures did not perform as badly with respect to the mode.

In the ordering task, the lack of difference between the full point set and the borders-only point set for the better measures seems to mesh well with the comments about the border being very important in the look of the drawing. More study is needed to determine if the results are because the borders really are more important, or if the degree of change in the borders is representative of the change in the whole drawing. This could be tested by comparing drawings where the border is largely unchanged but the interior is very different and those where the border is greatly changed but the interior is not. For the rotation task, only the lack of difference for the orthogonal ordering measures supports the students' comments, since most of the measures (except Hausdorff distance and orthogonal ordering) are already more sensitive to the borders of the drawing — rotation causes border points to move farther, giving them more weight.

The difficulty of the difference part suggests that the amount of difference between the drawings that is considered reasonable varies greatly with the task. When the user only needs to recognize the graph as familiar, the perimeter of the drawing and the position and shape of few key features are the most important. When trying to find a specific change, however, the drawings need to look very much alike or some other cues (change in color, more distinctive vertex names, etc.) are needed to highlight the change. The failure of using the times from the difference task to evaluate the magnitude criterion means that more study is needed to evaluate the measures in this way.

The responses on the questionnaire suggest several possible directions for future investigation. Large vertices are identified as being especially important, which could lead to measures which weight changes in the position and size of large vertices more heavily than other vertices.

Another major focus was clusters of vertices — both the presence of clusters in general, and the presence of specific shapes such as chains and zigzags. The relatively poor showing of the cluster-based measures indicates that they are not making use of clusters in the right way. The importance of specific shapes suggests an approach related to the drawing algorithms of Dengler, Friedell, and Marks [6] and Ryall, Marks, and Shieber [16]. These algorithms produce drawings employing effective perceptual organization by identifying Visual Organization Features (VOFs) used by human graphic designers. VOFs include horizontal and vertical alignment of vertices, particular shapes such as "T" shapes, and symmetrically placed groups of vertices. VOFs can also be used to identify features in an existing drawing that may be important because they adhere to a parti-

cular design principle. This is related to the work of Dengler and Cowan [5] on semantic attributes that humans attach to drawings based on the layout, such as that symmetrically placed nodes have common properties. A similarity measure might measure how well those structures are preserved, and an interactive graph drawing algorithm could focus on preserving the structures.

### References

- T. Biedl, J. Marks, K. Ryall, and S. Whitesides. Graph multidrawing: Finding nice drawings without defining nice. In GD '98, volume 1547 of LNCS, pages 347–355. Springer-Verlag, 1998.
- 2. T. C. Biedl and M. Kaufmann. Area-efficient static and incremental graph darwings. In ESA '97, volume 1284 of LNCS, pages 37–52. Springer-Verlag, 1997.
- S. Bridgeman and R. Tamassia. Difference metrics for interactive orthogonal drawing. J. Graph Alg. Appl. to appear.
- S. S. Bridgeman, J. Fanto, A. Garg, R. Tamassia, and L. Vismara. Interactive-Giotto: An algorithm for interactive orthogonal graph drawing. In GD '97, volume 1353 of LNCS, pages 303–308. Springer-Verlag, 1997.
- E. Dengler and W. Cowan. Human perception of laid-out graphs. In GD '98, volume 1547 of LNCS, pages 441–443. Springer-Verlag, 1998.
- E. Dengler, M. Friedell, and J. Marks. Constraint-driven diagram layout. In Proc. IEEE Sympos. on Visual Languages, pages 330–335, 1993.
- G. Di Battista, A. Garg, G. Liotta, R. Tamassia, E. Tassinari, and F. Vargiu. An experimental comparison of four graph drawing algorithms. *Comput. Geom. Theory Appl.*, 7:303–325, 1997.
- 8. U. Fößmeier. Interactive orthogonal graph drawing: Algorithms and bounds. In GD '97, volume 1353 of LNCS, pages 111–123. Springer-Verlag, 1998.
- J. E. Goodman and R. Pollack. Multidimensional sorting. SIAM J. Comput., 12(3):484–507, Aug. 1983.
- K. A. Lyons, H. Meijer, and D. Rappaport. Algorithms for cluster busting in anchored graph drawing. J. Graph Algorithms Appl., 2(1):1–24, 1998.
- A. Marzal and E. Vidal. Computation of normalized edit distance and applications. IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(9):926–932, Sept. 1993.
- K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. J. Visual Lang. Comput., 6(2):183–210, 1995.
- S. North. Incremental layout in DynaDAG. In GD '95, volume 1027 of LNCS, pages 409–418. Springer-Verlag, 1996.
- A. Papakostas, J. M. Six, and I. G. Tollis. Experimental and theoretical results in interactive graph drawing. In GD '96, volume 1190 of LNCS, pages 371–386. Springer-Verlag, 1997.
- A. Papakostas and I. G. Tollis. Interactive orthogonal graph drawing. *IEEE Trans. Comput.*, C-47(11):1297–1309, 1998.
- K. Ryall, J. Marks, and S. Shieber. An interactive system for drawing graphs. In GD '96, volume 1190 of LNCS, pages 387–393. Springer-Verlag, 1997.
- R. Tamassia, G. Di Battista, and C. Batini. Automatic graph drawing and readability of diagrams. *IEEE Trans. Syst. Man Cybern.*, SMC-18(1):61-79, 1988.
- W. A. Wickelgren. Cognitive Psychology. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1979.