

Shape-Based Quality Metrics for Large Graph Visualization

*Peter Eades*¹ *Seok-Hee Hong*¹ *An Nguyen*¹ *Karsten Klein*²

¹University of Sydney, Australia

²University of Konstanz, Germany

Abstract

The scalability of graph layout algorithms has gradually improved for many years. However, only recently a discussion has started to investigate the usefulness of established quality metrics, such as the number of edge crossings, in the context of increasingly larger graphs stemming from a variety of application areas such as social network analysis or biology. Initial evidence suggests that the traditional metrics are not well suited to capture the quality of corresponding graph layouts. We propose a new family of quality metrics for graph drawing; in particular, we concentrate on larger graphs. We illustrate these metrics with examples and apply the metrics to data from previous experiments, leading to the suggestion that the new metrics are effective.

| | | | | |
|----------------------------|--------------------------------|-------------------------|--|---------------------------|
| Submitted: January 2016 | Reviewed: May 2016 | Revised: August 2016 | Reviewed: November 2016 | Revised: December 2016 |
| | Accepted: December 2016 | Final: December 2016 | Published: January 2016 | |
| | Article type: Regular paper | | Communicated by: E. Di Giacomo and A. Lubiw | |

Research supported by Australian Research Council grant DP160104148, LP110100519, Tom Sawyer Software, and NewtonGreen Technologies.

E-mail addresses: peter.eades@sydney.edu.au (Peter Eades) seokhee.hong@sydney.edu.au (Seok-Hee Hong) angu5603@uni.sydney.edu.au (An Nguyen) karsten.klein@uni-konstanz.de (Karsten Klein)

1 Introduction

Several of the earliest papers on Graph Drawing (for example, [24, 25, 26]) discussed requirements for a “good” visualization of a graph. For example, Tamassia et al.[25] state:

Aesthetics: We use the term aesthetics to denote the criteria that concern certain aspects of readability. A well-admitted aesthetics, valid independently from the graphic standard, is the minimisation of crossings between edges. Also, in order to avoid unnecessary waste of space, it is usual to keep the area occupied by the drawing reasonably small. When the grid standard is adopted, it is meaningful to minimize the number of bends (turns) along the edges, as well as their total length.

We prefer the term *quality metrics* rather than “aesthetics”.

These early quality metrics were stated in terms of geometric properties of the layout. The underlying and often unstated assumption that these geometric properties of layout measure the “goodness” of a graph drawing was unchallenged until the experiments of Purchase et al. [22]. These experiments showed that human task performance is correlated with some of the previously defined quality metrics. A conclusive result was that task times and error rates were both correlated with the number of *edge crossings*. Subsequent experiments have confirmed and refined these initial results [13, 19, 20, 21, 29]. All these early experiments used relatively small graphs as stimuli; and the validity of the results for larger graphs was not tested.

Human experiments with larger graphs began recently [15, 16]. In particular it has been pointed out that edges and vertices become “blobs” in large graph drawings such as the biological network in Fig. 1; almost all the edge crossings are hidden in the blobs. Any causal relationship between readability and edge crossings seems unlikely. Further, a graph drawing can display “structure” despite having a large number of edge crossings; see, for example, Fig. 2. In this paper we propose a quality metric for large drawings such as Figs. 1 and 2.

Although it is seldom explicitly stated as a quality metric for graph drawing, *stress* is often used as such. There are various measures of stress (for example, see [7, 8, 10, 12]); the most commonly used is to define the stress in a drawing D of a connected graph $G = (V, E)$ as

$$\sigma = \sum_{u,v \in V} w_{uv} (d_G(u, v) - d_{\mathbb{R}^2}(D(u), D(v)))^2 \quad (1)$$

where $d_G(u, v)$ is the graph theoretic distance between u and v , $d_{\mathbb{R}^2}(D(u), D(v))$ is the Euclidean distance between the locations $D(u)$ and $D(v)$ of u and v , and w_{uv} is a constant.

Stress appears to measure the “faithfulness” of a graph drawing [17, 23], in the following sense. Informally, a drawing D of a graph G is *faithful* if G can be determined from D . For a large graph G , a low stress drawing D such as in Fig. 1

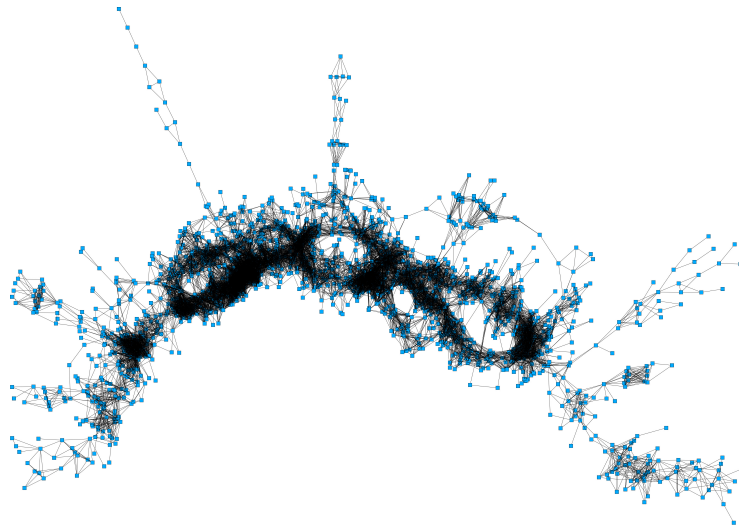


Figure 1: *Crossings can be hidden in a drawing of a large graph. This drawing of a RNA sequence graph has very dense local structures, but clearly visible global structure.*

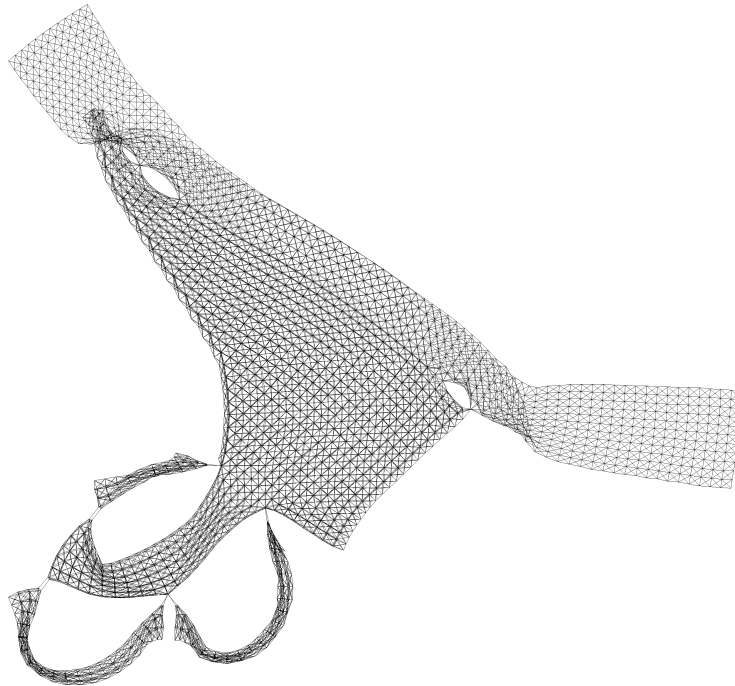


Figure 2: *This drawing of the `data` graph from the Walshaw graph library [28] clearly shows the graph’s “structure”, despite a large number of crossings.*

may not completely determine G . However, the low value of stress indicates that the Euclidean distances between vertices are (approximately) proportional to the graph-theoretic distances in the graph (see equation (1)); we say that such a drawing is (approximately) *distance faithful*.

Quality metrics are significant firstly because they measure success or failure of a graph drawing method. Most importantly, they can be used as optimisation goals in graph visualisation methods. For example, algorithms that aim to draw graphs with a small number of crossings, a small number of edge bends, and low energy/stress are well established in the academic literature [3] and in commercial graph visualization tools. New quality metrics, such as proposed in this paper, potentially can be used with optimisation algorithms to give new visualisation methods.

This paper proposes a new family of quality metrics for graph visualization, especially for large graph drawings. Here, by “large”, we mean that the graphs are large enough to make “blobs” such as in Fig. 1 inevitable. This includes dense graphs with a few hundred vertices as well as sparse graphs with a few thousand vertices.

The proposed metrics are based on the notion of the “shape” of a set of points in \mathbb{R}^2 . Our proposal, simply stated, is that *a drawing is good if the shape of the set of vertex positions is similar to the original graph*.

In Section 2 we describe this notion more precisely and illustrate with examples. In Section 3 we give some empirical indication that the metrics are valid, based on data sets from previous experiments [2, 16]. Section 4 concludes with a discussion and some open problems.

2 Shape-based Metrics

Fig. 3 summarises our proposal. The quality of a drawing D of a graph G is the similarity between G and the “shape” of the set of vertex locations of D . The “shape” is expressed as a graph, called a “shape graph”.

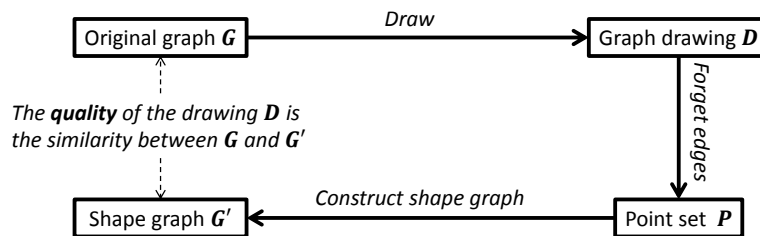


Figure 3: *Shape-based quality metrics.*

To make these notions more precise, we need to examine the notion of the shape of a set of points, and the notion of similarity between two graphs.

Note that we have an underlying assumption that the shape of the graph

drawing is the same as the shape of the set of vertex locations. For large graphs, this is a reasonable assumption, as vertices tend to be so close together that edges are hardly visible. In Fig 4, for example, the graph drawing is almost indistinguishable from its set of vertex locations.

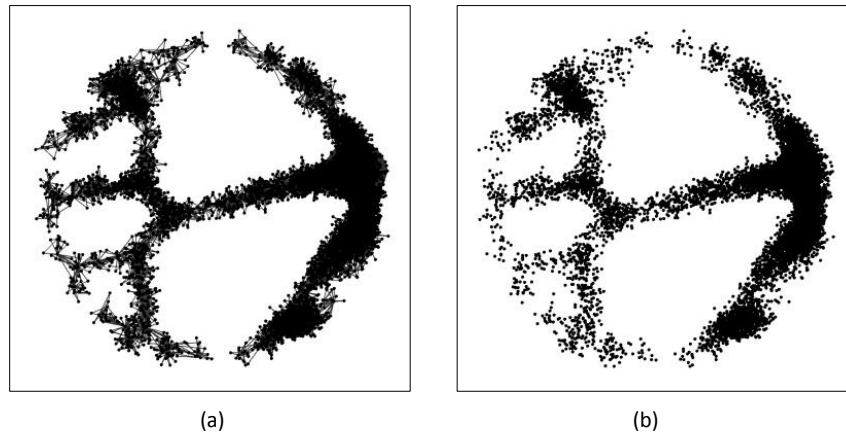


Figure 4: (a) a graph drawing D ; (b) the set of vertex locations of D .

2.1 Shape Graphs

Informally, a *shape graph* for a set of points P is a geometric graph with vertex set P that captures the “shape” of P in some sense.

The classical example of a shape graph is the α -*shape* [5]. When $\alpha = 0$ the α -shape is the convex hull; in general, alpha shapes generalize the concept of the convex hull. For $\alpha > 0$, the α -shape graph for a set of points P contains a straight-line edge between a pair of points if and only if the two points can be touched by an open disc of radius α^{-1} that contains no points of P ; for details see [5]. Note that α -shapes capture the shape of the *boundary* of P , and not the *internal structure* of P . For this paper we need a concept of shape that captures the internal structure of a set of points.

A more suitable kind of shape graph is a “proximity graph”: an edge is placed between two points $p, q \in P$ if p is “close to” q in some sense. There are many kinds of proximity graphs (see [27]); some examples are below:

- The *k-nearest neighbours graph* has a (directed) edge from point $p \in P$ to point $q \in P$ if the number of points $r \in P$ with $d(p, r) < d(p, q)$ is at most $k - 1$.

- Various *triangulations* and *quadrilateralizations*: for example, the Delaunay triangulation, greedy triangulations, and minimum-weight quadrilateralizations.
- The *Gabriel graph* (GG) has an edge between distinct points $p, q \in P$ if the closed disc which has the line segment pq as a diameter contains no other elements of P .
- The *relative neighbourhood graph* (RNG) has an edge between distinct points $p, q \in P$ if there is no point $r \in P$ such that $d(p, r) \leq d(p, q)$ and $d(q, r) \leq d(p, q)$.
- A *Euclidean minimum spanning tree* (EMST) is a minimum spanning tree of P where the weight of the edge between each pair of points is the Euclidean distance.

Note that many of these shape graphs are *local* in that the existence of an edge between two points is determined by a local neighbourhood of those points. Other shape graphs, such as the Euclidean minimum spanning tree and minimum weight quadrilateralizations, are *global*.

In Section 3 below, we examine quality metrics based on the Euclidean minimum spanning tree, the Gabriel graph, and the relative neighborhood graph respectively; each of these shape graphs can be computed in $O(n \log n)$ time using standard algorithms [18]. However, our remarks apply in principle to any shape graph.

2.2 Graph Similarity

Suppose that $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are two graphs with the same vertex set. A simple measure for the similarity of G_1 and G_2 is the *mean Jaccard similarity*:

$$MJS(G_1, G_2) = \frac{1}{|V|} \sum_{u \in V} \frac{|N_1(u) \cap N_2(u)|}{|N_1(u) \cup N_2(u)|}, \quad (2)$$

where $N_i(u)$ is the set of neighbours of u in G_i for $i = 1, 2$. It is straight-forward to compute the mean Jaccard similarity in linear time.

Note that $0 \leq MJS(G_1, G_2) \leq 1$. Also, if G_1 and G_2 share many edges, then $MJS(G_1, G_2)$ is close to 1; if they share very few edges then $MJS(G_1, G_2)$ is close to 0.

More complex measures for graph similarity include graph edit distance [9], and measures based on the notion that the similarity of two vertices u and u' depends on the similarity of their neighbours (see, for example, [14]). However, these metrics are computationally expensive and do not scale beyond a few thousand vertices; mean Jaccard similarity can be computed in linear time and performs well in the experiments described below in Section 3.

2.3 The Shape-based Metrics

We can now explicitly define our proposed metrics. Suppose that D is a drawing of a graph G ; we want to measure the quality of D . Let P denote the set of vertex locations of D , and suppose that μ is a shape graph function (that is, μ takes a set of points and produces a shape graph on this set of points). Further, let η be a graph similarity function, that is, η takes two graphs as input and returns a positive real number that indicates the similarity between these two graphs. Then we define the quality metric $Q_{\mu,\eta}$ by

$$Q_{\mu,\eta}(D) = \eta(G, \mu(P)).$$

Throughout this paper we use the mean Jaccard similarity for graph similarity, and so we abbreviate $Q_{\mu,\eta}$ to Q_μ .

The time to compute Q_μ depends on the choice of μ ; for all such choices μ explicitly described in this paper, Q_μ can be computed in time $O(n \log n)$.

2.4 Related Metrics

Our proposed metrics are, in spirit, related to the “graph theoretic scagnostics” approach to scatterplots (see [30]). Scagnostics measure global shape characteristics of scatter plots based on proximity graphs. Using EMST, α -hull, and the convex hull to characterize the global shape, these measures enable quantitative statements regarding shape, trend, density, outlier, and coherence characteristics of a scatterplot.

In the case that the shape graph μ is a k -nearest neighbor graph, the “neighborhood inconsistency” [8] and “neighborhood preservation precision” [7, 8] metrics used by Gansner et al. are also related. These two metrics have a different motivation to ours: rather than measure the general notion of shape, they attempt to measure whether neighbours in the layout coincide with neighbours in the graph. Nevertheless, we can regard the “neighborhood inconsistency” as an example of a local shape-based metric when the shape graph μ is a k -nearest neighbor graph, and the similarity function η is based on the “stochastic neighbor embedding” of Hinton and Roweis [11].

2.5 Bounds

It is clear that if D is a drawing of a graph G , then for every choice of μ ,

$$0 \leq Q_\mu(D) \leq 1.$$

More precise bounds may be obtained for more specific cases. As an example, we can compute an upper bound for the Euclidean minimum spanning tree based metric Q_{EMST} as follows. Consider the graph $G = (V, E)$ with n vertices and m edges; we assume that $m > n$. Suppose that $V = \{1, 2, \dots, n\}$ and d_i is the degree of vertex i in G ; we assume without loss of generality that $1 \leq d_1 \leq d_2 \leq \dots \leq d_n$.

Since $m > n$,

$$\sum_{i=1}^n d_i > 2(n-1);$$

let ℓ^* be the smallest integer such that

$$\sum_{i=1}^{\ell^*} d_i > 2(n-1).$$

Suppose that $T = (V, E')$ is a Euclidean minimum spanning tree of the locations of vertices in the drawing D of G , and d'_i is the degree of i in T . Equation (2) implies that

$$Q_{EMST}(D) < \frac{1}{n} \sum_{i=1}^n \frac{d'_i}{d_i},$$

and since $\sum_{i=1}^n d'_i = 2(n-1)$, it is straightforward to deduce that

$$Q_{EMST}(D) < \frac{\ell^* + 1}{n}. \quad (3)$$

We can refine this bound. For $k = 1, 2, \dots, n$ we define

$$f(k) = n - k + \sum_{i=1}^k d_i.$$

Now $f(k)$ is nondecreasing in k and $f(n) = 2|E| > 2(n-1)$. Let k^* denote the smallest integer such that $f(k^*) > 2(n-1)$. Further let

$$r = \sum_{i=k^*+1}^n \frac{1}{d_i},$$

and

$$s = \frac{n + k^* - 2 - \sum_{i=1}^{k^*-1} d_i}{d_{k^*}}.$$

Then we can deduce the following upper bound:

$$Q_{EMST}(D) \leq \frac{k^* - 1 + s + r}{n}. \quad (4)$$

For specific families of graphs, more specific bounds can be obtained. For example, if G is regular of degree $d > 1$, then from (3) we can deduce:

$$Q_{EMST}(D) < \frac{1}{n} \left\lceil \frac{2(n-1)}{d} \right\rceil \approx \frac{2}{d}.$$

Similar bounds can be derived for the metrics Q_{GG} and Q_{RNG} based on the Gabriel graph and the relative neighbourhood graph.

2.6 Some Examples

Although our proposal is principally aimed at large graphs, we first describe an easy example using a smaller graph, for illustrative purposes. Consider the graph drawing D_0 in Fig. 5(a). The set P_0 of vertex locations of D_0 is shown in Fig. 5(b). A Euclidean minimum spanning tree T_0 on P_0 is shown in Fig. 5(c).

Our proposal is that the quality $Q_{EMST}(D_0)$ of the graph drawing D_0 can be measured as the similarity between the (combinatorial) graphs in Figs. 5(a) and (c). Using the mean Jaccard similarity in Equation (2), we can calculate the value $Q_{EMST}(D_0) = 0.60$. The upper bound on Q_{EMST} for this graph, given by the inequality (3), is 0.68. The comparatively high value of $Q_{EMST}(D_0)$ expresses the fact that the “shape” of the drawing D_0 is similar to the graph G . Intuitively, the graph drawing D_0 is a reasonably faithful representation of the graph G .

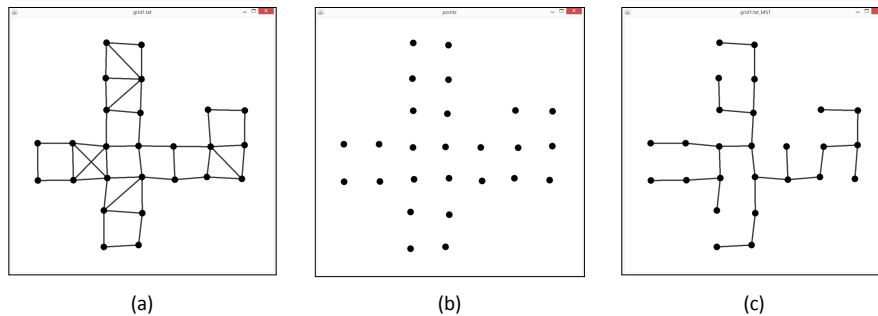


Figure 5: (a) A graph drawing D_0 . (b) The set P_0 of vertex locations of D_0 . (c) A Euclidean minimum spanning tree T_0 on P_0 .

A larger example is illustrated with two drawings D_a and D_b in Fig. 6 of a graph `blobs1001`, which has 1001 vertices and 7537 edges. Both drawings are computed with the `organic` layout tool of `yFiles` [1], but with different settings. The drawing D_a in Fig. 6(a) is computed using `yFiles` “quality/time ratio” set to minimise time (at the cost of quality); the drawing D_b in Fig. 6(b) has this ratio set to maximise quality (at the cost of time).

One can compute $Q_{EMST}(D_a) = 0.088$, and $Q_{EMST}(D_b) = 0.100$. This confirms the intuition that the quality of D_b is a little higher than that of D_a . Further, the upper bound from (3) for `blobs1001` is 0.165, so both drawings are reasonable but perhaps not optimal.

3 Three Experiments

In this section, we describe three tests of the shape-based quality metrics. In the first test we progressively deform a good drawing and compute the shape-based quality metrics. The second and third tests investigate how the shape-based

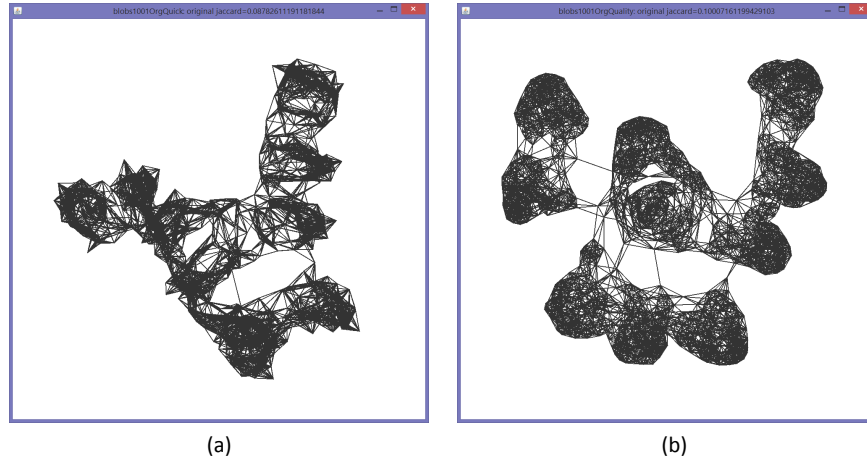


Figure 6: Two drawings of the graph `blobs1001`: (a) drawing D_a computed using `yFiles` “quality/time ratio” set to minimise time (at the cost of quality); (b) drawing D_b with “quality/time ratio” set to maximise quality (at the cost of time).

quality metrics perform on two specific data sets from previous experiments [2, 16].

3.1 Progressive Deformation

Consider drawing D_0 in Fig. 5(a) of the small graph G , as described in Section 2.6. We examine what happens when we progressively deform the drawing to make it worse. Suppose that D_δ is formed from D_0 by moving each vertex in a random direction by a random distance in the range $[0, \delta w]$, where w is the width of the screen. Drawings D_δ for $\delta = 0.1, 0.2$, and 0.5 are shown in Fig. 7.

For $\delta = 0.1$, the shape of the drawing is fairly close to G ; that is, the minimum spanning tree T_δ shares quite a few edges with D_δ . The value $Q_{EMST}(D_{0.1}) = 0.42$ is reasonably high. As δ increases, the shape graph T_δ is less similar to G , and the values of $Q_{EMST}(D_\delta)$ fall. For $\delta = 0.5$ the shape of the drawing shows no resemblance to G , and $Q_{EMST}(D_\delta)$ is low. Intuitively, as the drawing becomes worse, the shape of the set of points differs more and more from the graph.

For a larger example, we consider a graph `stringyBlobs` with 2736 vertices and 15103 edges; a drawing `stringyBlobsOrganic` of `stringyBlobs` using the `yFiles` *organic* layout [1] is in Fig. 8. The graph `stringyBlobs` is globally sparse and tree-like, but it has some dense “blobs”; this structure is faithfully shown by the organic layout in Fig. 8.

In a progressive deformation of `stringyBlobsOrganic`, we moved vertices

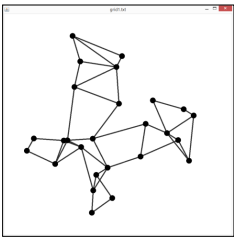
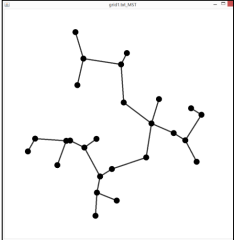
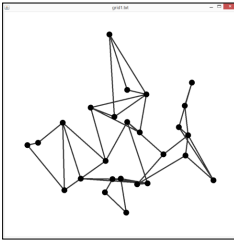
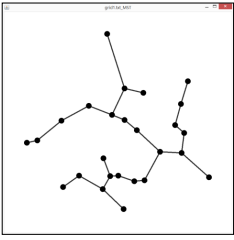
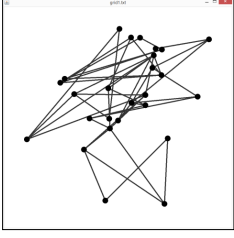
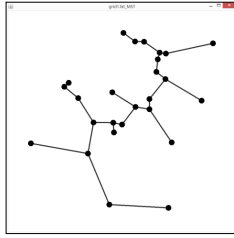
| δ | D_δ | T_δ | $Q_{EMST}(D_\delta)$ |
|----------|--|---|----------------------|
| 0.1 |  |  | 0.42 |
| 0.2 |  |  | 0.34 |
| 0.5 |  |  | 0.07 |

Figure 7: The drawing D_δ in the second column is formed from the drawing D_0 in Fig. 5 by moving each vertex in a random direction by a random distance in the range $[0, \delta]$. The graph T_δ in the third column is a Euclidean minimum spanning tree of the vertex locations of D_δ .

randomly by a distance of $0.005 * screenSize$ over 30 steps; steps 5, 10, and 15 are shown in Figs. 9, 10, and 11. As the drawing is deformed, the “blobs” merge and split, and the “stringy” parts become tangled; the drawing displays less structure. With further deformation, the drawing becomes more or less a single blob.

The values of the metric Q_{EMST} are charted in Fig. 12. As expected, the metric decreases as the drawing is deformed and displays less structure.

In fact, progressive deformation of other large graphs produces similar results. As another example, progressive deformation of the drawing D_b in Fig. 6(b) yields the chart in Fig. 13.

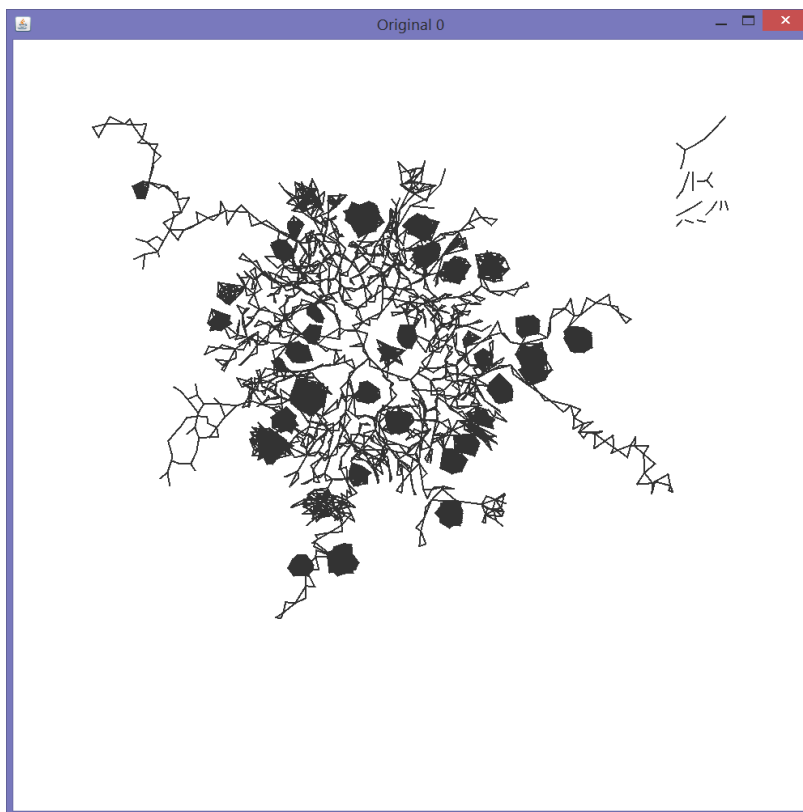


Figure 8: A drawing *stringyBlobsOrganic* of the graph *stringyBlobs*, drawn with the *yFiles* organic layout tool [1].

3.2 The “Untangling” Data Set

3.2.1 The GION experiment

Marner et al. [16] introduced a new method called *GION* for supporting interaction with graph drawings on large displays. The user study of [16] focussed on the task of *untangling* a graph drawing: subjects were presented with a graph drawing (a Fruchterman-Reingold layout [6]), and were simply asked to untangle the layout. Eight RNA sequence graphs were used, ranging from 1159 to 7885 vertices. These graphs are locally dense, but globally very sparse. Their global structure is often tree-like, perhaps path-like. There were 16 subjects.

The experimental system captured, for each subject and each graph, a snapshot drawing every 5 seconds; the snapshot at time t is denoted by D_t . Two

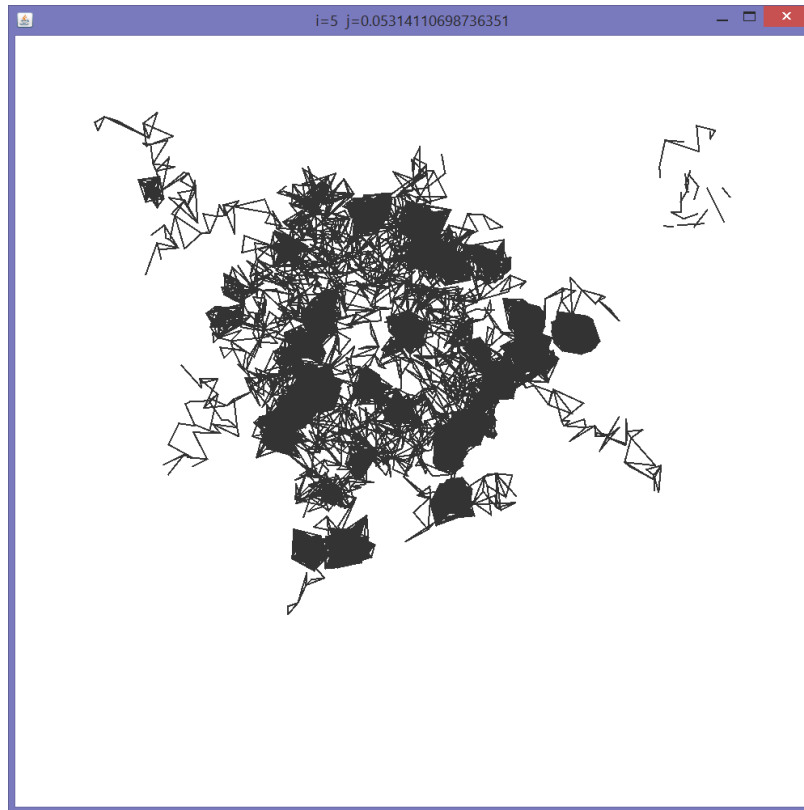


Figure 9: *Progressive deformation of the drawing `stringyBlobsOrganic` in Fig. 8: after 5 steps of deformation.*

such snapshot graph drawings are shown in Fig. 14(a) and (b).

The main result of the experiment, reported in detail in [16], was that the GION method is better in several ways than more traditional interaction methods.

3.2.2 Shape-based metrics and the GION data set

The GION experiment provides a large data set recording how users tried to untangle graph drawings (8 graphs 16 users 24 snapshot drawings). We can re-use this data to check our shape-based quality metrics. For each snapshot D_t , we computed the number $\chi(D_t)$ of edge crossings, the (scaled) stress $\sigma(D_t)$, and the metrics $Q_{EMST}(D_t)$, $Q_{GG}(D_t)$, and $Q_{RNG}(D_t)$, respectively based on

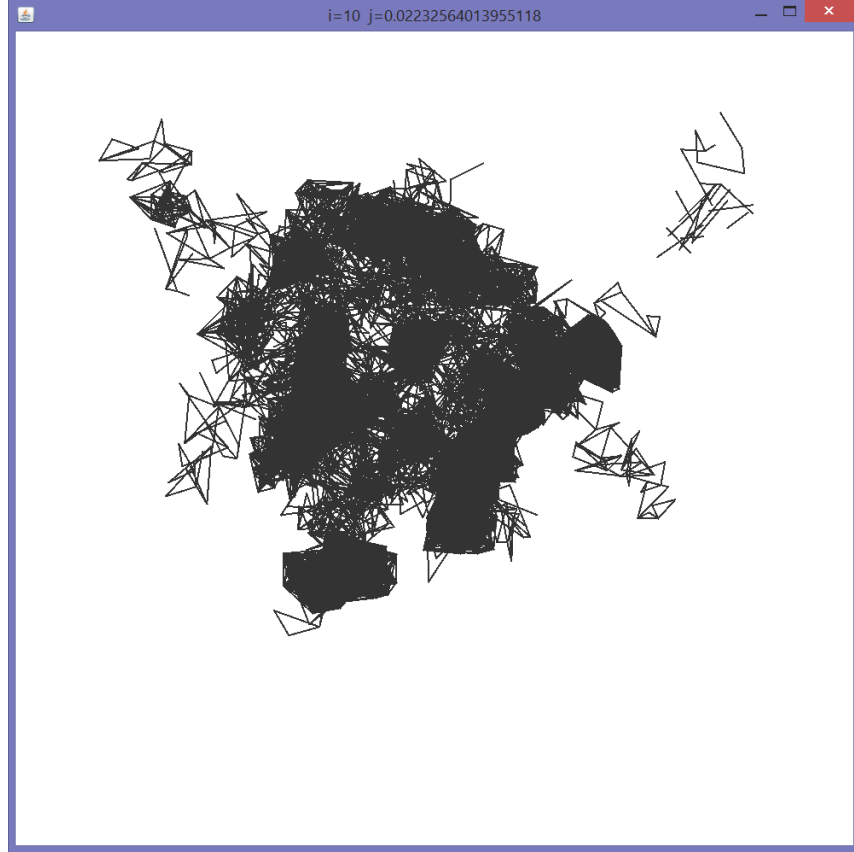


Figure 10: *Progressive deformation of the drawing `stringyBlobsOrganic` in Fig. 8: after 10 steps of deformation.*

Euclidean minimum spanning tree, Gabriel graphs, and relative neighborhood graphs.

Commonly-held graph drawing wisdom is that $\chi(D_t)$ and $\sigma(D_t)$ decrease with the quality of the graph drawing. We expect that quality increases as the graph is untangled, and so we expect that $\chi(D_t)$ and $\sigma(D_t)$ decrease with t .

Note that the shape-based quality metrics $Q_{EMST}(D_t)$, $Q_{GG}(D_t)$, and $Q_{RNG}(D_t)$ are expected to increase with t . To make the comparison between these metrics easier, we place them on a comparable scale by inverting and normalising crossings and stress, as follows.

We define

$$\bar{Q}_\chi(D_t) = \frac{M_\chi - \chi(D_t)}{M_\chi}, \quad \bar{Q}_\sigma(D_t) = \frac{M_\sigma - \sigma(D_t)}{M_\sigma},$$

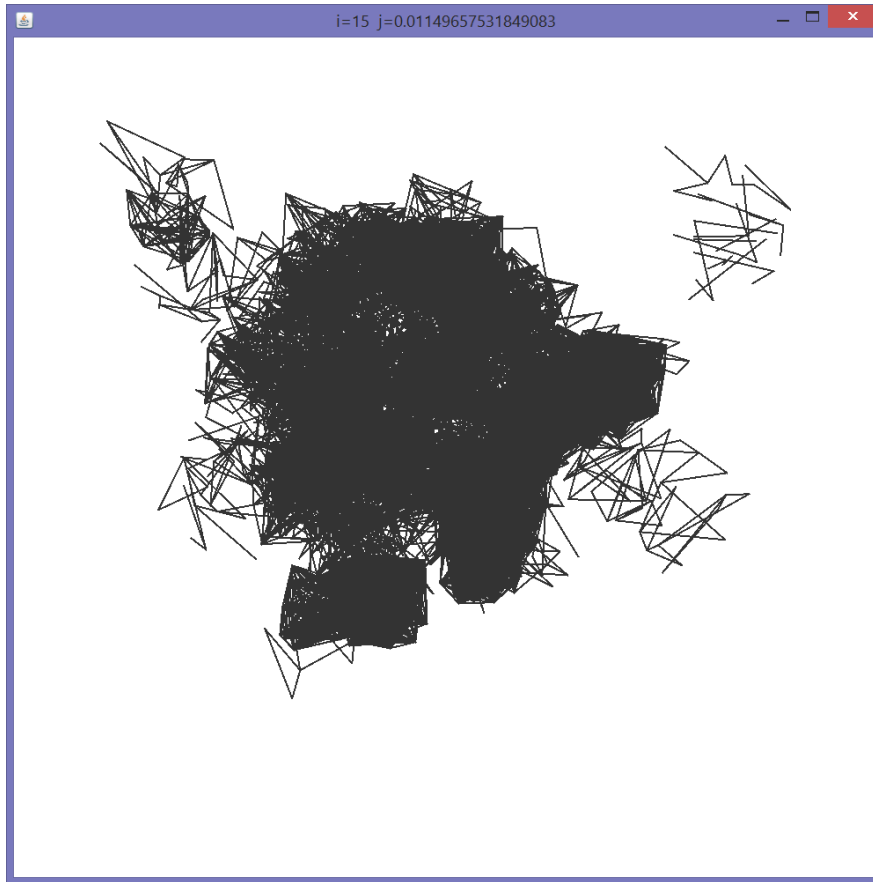
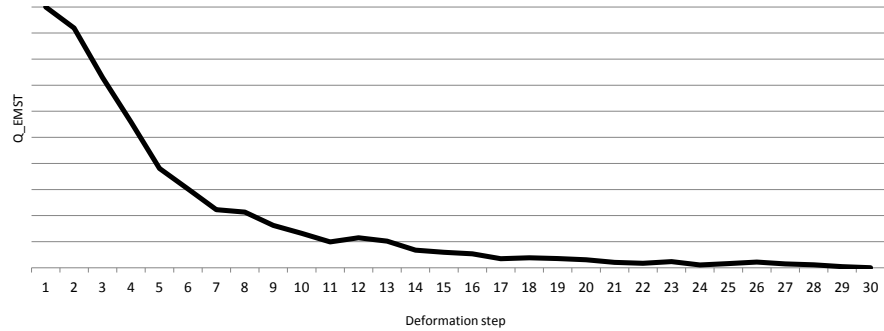
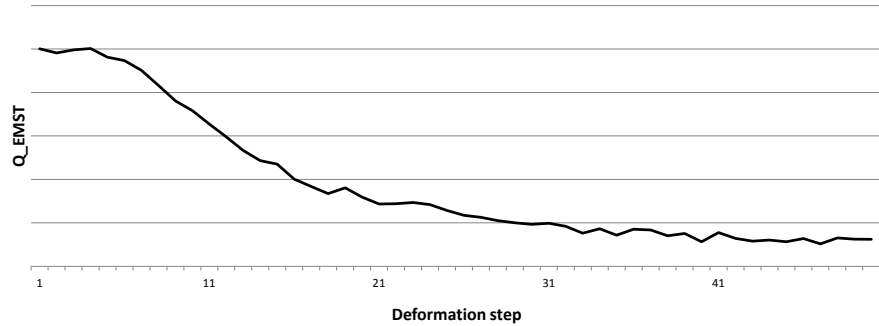


Figure 11: *Progressive deformation of the drawing `stringyBlobsOrganic` in Fig. 8: after 15 steps of deformation.*

where $M_\chi = \max_t \chi(D_t)$ and $M_\sigma = \max_t \sigma(D_t)$.

Note that $\bar{Q}_\chi(D_t)$ (respectively $\bar{Q}_\sigma(D_t)$) decreases from 1 to 0 as the number of crossings (respectively stress) increases from 0 to the maximum M_χ (respectively M_σ). For the shape-based metrics, we simply linearly normalise Q_{EMST} (respectively Q_{GG} and Q_{RNG}) to give \bar{Q}_{EMST} (respectively \bar{Q}_{GG} and \bar{Q}_{RNG}) so that it increases from 0 to 1 as the (shape-based) quality of the drawing increases.

Intuitively, one may expect that the drawing improves in quality as the untangling proceeds. However, the results reported in [16] were counterintuitive in terms of crossings and stress: as the subjects untangled the graph drawings, there was a tendency to *increase* both crossings and stress (that is, both \bar{Q}_χ and \bar{Q}_σ decreased).

Figure 12: Q_{EMST} values as the drawing *stringyBlobsOrganic* is deformed.Figure 13: Q_{EMST} values as the drawing D_b in Fig. 6(b) of *blobs1001* is deformed.

In contrast, we found that \bar{Q}_{EMST} , \bar{Q}_{GG} , and \bar{Q}_{RNG} all *increased* as the subjects untangled the drawings. The charts in Fig. 15 show \bar{Q}_χ , \bar{Q}_σ , \bar{Q}_{EMST} , \bar{Q}_{GG} , and \bar{Q}_{RNG} , averaged over all subjects, for the first 3 of the 8 graphs. The horizontal axis is time t ; the vertical axis shows the values of the metrics. For graphs #1 and #2, both crossings and stress increase with t (that is, $\bar{Q}_\chi(D_t)$ and $\bar{Q}_\sigma(D_t)$ decrease). In contrast, \bar{Q}_{EMST} , \bar{Q}_{GG} , and \bar{Q}_{RNG} increase. Graphs #4, #5, #6, #7, and #8 showed very similar patterns to graphs #1 and #2. Graph #3 was a little different in that crossings decrease (and thus \bar{Q}_χ increases), albeit chaotically.

Overall, the data from the untangling experiment shows that both crossings and stress metrics became worse as the subjects untangled the graphs, but the shape-based metrics became better. With some provisos (see Section 4 below), this suggests that the shape-based metrics are better than crossings and stress for measuring untangling.

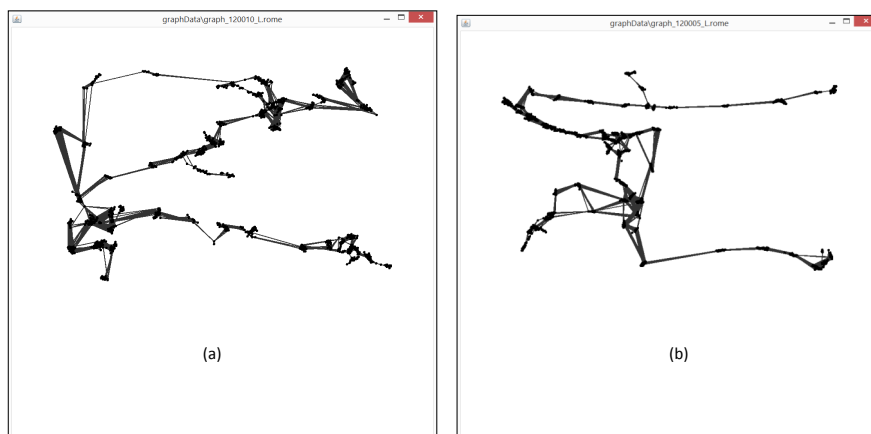


Figure 14: Two snapshots from the GION experiment.

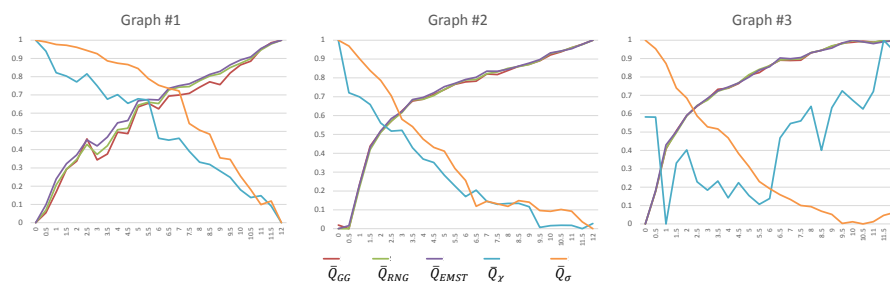


Figure 15: Metrics against untangling.

3.3 The “Preference” Data Set

3.3.1 The “preference” experiment

Chimani et al. [2] report an experiment at the University of Osnabrück aimed at determining whether human *preferences* in graph drawing correlates with crossings and stress. There were two follow-up experiments, at the Graph Drawing conference in 2014, and at the University of Sydney. The design and results of all three experiments were similar; see [2]. Here we investigate the data from the University of Sydney experiment, aiming to determine whether shape-based metrics are correlated with preference.

This experiment had 40 subjects. Each subject was presented with 20 “instances”. Each instance displayed a pair of drawings of the same graph, as in the screenshot in Fig. 16.

There is a slider bar at the bottom of the screen, and the subject indicates which of the pair of drawings he/she prefers by sliding to the left or right. The slider bar has a scale on the left from 5 to 1 and on the right from 1 to 5, with

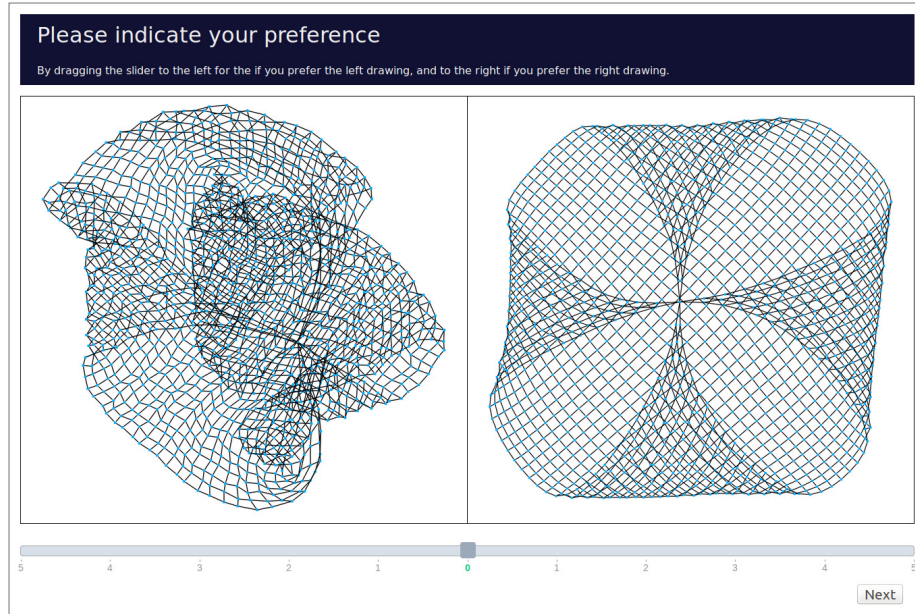


Figure 16: *Example of a typical “instance” (a graph pair shown to participants) for the preference experiment.*

zero in the middle. The slider bar is used to give a score to the drawing that the subject prefers, indicating how much the subject prefers it. A score of 5 on the left indicates a strong preference for the drawing on the left, a score of 5 on the right indicates a strong preference for the drawing on the right, and a score of zero indicates no preference.

A total of 118 graphs, ranging in size from small (25 vertices and 29 edges) to moderately large (8000 vertices and 15580 edges), were used. Five drawings for each graph were generated, and the instances were chosen randomly. For details, see [2].

The results for a particular quality metric Q_μ are expressed in terms of the “ Q_μ -ratio”, defined as follows. Consider an instance consisting of two drawings D_{left} and D_{right} of a graph G , such as in Fig. 16. Let $Q_\mu(D_{left})$ (respectively $Q_\mu(D_{right})$) be the value of the Q_μ metric for D_{left} (respectively for D_{right}). We define the Q_μ -ratio for this instance as

$$\frac{\max(Q_\mu(D_{left}), Q_\mu(D_{right}))}{\min(Q_\mu(D_{left}), Q_\mu(D_{right}))}.$$

If the Q_μ -ratio is approximately 1, then (according to the quality metric Q_μ) D_{left} has approximately the same quality as D_{right} . We expect that the subject prefers D_{left} in about 50% of such instances; our experiments showed that this was true for all the metrics under investigation.

If the Q_μ -ratio is significantly larger than 1, then we expect that most subjects prefer the drawing with the higher quality (according to the quality metric Q_μ). Further, as the Q_μ -ratio increases, we expect that more and more subjects prefer the drawing with higher quality. To make this precise, we need to define some further terms.

For each quality metric Q_μ and each instance I we compute a score $S_\mu(I)$ as follows. Suppose that for this instance, the subject gives a score of x ($0 \leq x \leq 5$). If the subject chose the drawing with a higher value of the quality metric Q_μ , then $S_\mu(I) = x$; otherwise $S_\mu(I) = -x$. The expectation that most subjects prefer the drawing with the higher quality becomes an expectation that in most instances, $S_\mu(I)$ is positive.

For each metric Q_μ , we chart the median of $S_\mu(I)$ over all instances I against the Q_μ -ratio in Fig. 17. The charts for crossings and stress are shown in Fig. 17(a), and for EMST, GG, and RNG in Fig. 17(b).

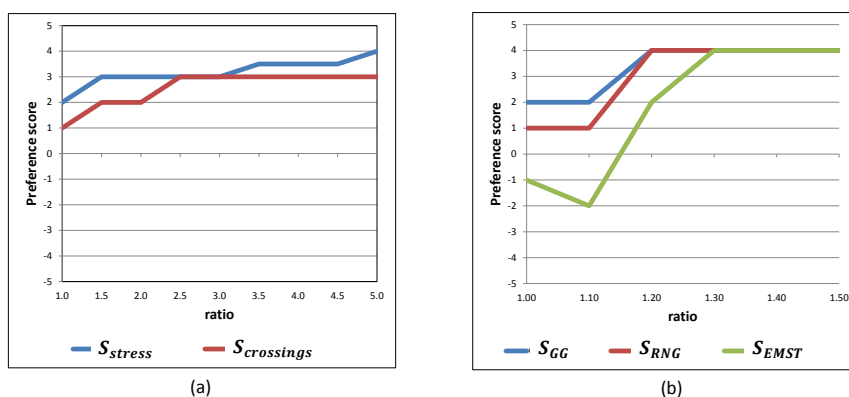


Figure 17: *Stress and crossing ratios, shape graph ratios, and preferences.*

For both crossing and stress, there is adequate data for ratios from 1 to 5; however, the data for ratios larger than 4.5 is small (less than 20 instances) and the results at this end of the spectrum must be treated with caution.

Crossings. Overall, there is a slight preference for fewer crossings (median over all instances is +1). As the crossing ratio increases, the median preference for the drawing with fewer crossings increases. When the crossing ratio is above 2.5 the median preference for the drawing with fewer crossings is +3, and stays steady at +3 as the crossing ratio increases beyond 2.5.

Stress. Overall, there is a preference for lower stress (median over all instances is +2). As the stress ratio increases, the median preference for lower stress rises; it hovers between +3 and +4 when the stress ratio is above 4.

3.3.2 Shape-based metrics and the “preference” data set

For EMST, GG, and RNG, there is adequate data for ratios from 1 to 1.5; but the data for ratios larger than 1.45 is small (less than 20 instances) and the results at this end of the spectrum must be treated with caution.

EMST. The median preference for the drawing with higher value of \bar{Q}_{EMST} is chaotic when the EMST-ratio is less than 1.2. The preference rises to +4 when the EMST-ratio rises from 1.2 to 1.3, and remains at +4 as the EMST-ratio increases beyond 1.3.

GG. Overall, there is a preference for drawings with a higher value of \bar{Q}_{GG} (median over all instances is +2). The preference for the drawing with higher value of \bar{Q}_{GG} rises smoothly with GG-ratio. When the GG-ratio is above 1.2 the median preference for the drawing with higher value of \bar{Q}_{GG} is +4, and remains at +4 as the GG-ratio increases beyond 1.2.

RNG. Overall, there is a preference for drawings with a higher value of \bar{Q}_{RNG} (median over all instances is +1). The preference for the drawing with higher value of \bar{Q}_{RNG} rises smoothly with RNG-ratio. When the RNG-ratio is above 1.2 the median preference for the drawing with higher value of \bar{Q}_{RNG} is +4, and remains at +4 as the RNG-ratio increases beyond 1.2.

One can conclude that people prefer drawings with fewer crossings, lower stress, and higher values for the shape-based metrics Q_{EMST} , Q_{GG} , and Q_{RNG} . More significantly, the preference for better GG and RNG based metrics is stronger than the preference for fewer crossings and lower stress.

Further, note that the overall preference for EMST-based metrics seems unreliable when the EMST-ratio is small; this suggests that EMST is not as good a model as GG and RNG.

4 Conclusion and Open Problems

This paper proposes a new family of metrics, aimed at measuring the quality of large graph drawings in terms of their shape.

4.1 Empirical validation

The proposal that the shape-based metrics are good measures of the quality of a graph drawing is supported by the “progressive deformation” experiment as described in Section 3.1.

The data from both the “untangling” experiment and the “preference” experiment also support the proposal; there is some indication that the shape-based metrics are better than crossings and stress. However, the support from these human experiments has some significant limitations:

- Neither experiment was designed to test the shape-based metrics. To safely validate the new metrics, further study is needed.
- The “untangling” experiment used a very specific kind of graph: RNA sequence graphs, which are locally dense with a global “tree-like” structure. For more general classes of graphs, further experimentation would be useful.
- The experiments use the notions of “untangledness” and “preference” as proxies for ground truth quality. It would be useful to test the metrics in a *task-oriented* experiment.

Designing experiments to fully validate shape-based metrics remains an open problem. In particular, we hypothesise that time and error of *tasks* on large graphs (see [15]) is related to shape-based metric values. The design of experiments to test this hypothesis is difficult. A significant problem is to determine which tasks are appropriate for large graph visualization; further difficulties arise because the results of such an experiment could be highly dependent on the specific tasks used.

4.2 Stress and shape-based metrics

It is tempting to suggest that stress and shape-based metrics are related. However, the relationship may not be strong. As a simple example, see the two drawings in Fig. 18. Here we would argue that Fig. 18(a) and Fig. 18(b) convey the structure of graph equally, and that the higher stress of Fig. 18(b) does not mean that it is a worse drawing.

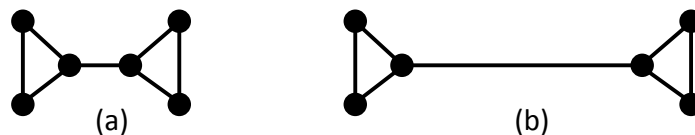


Figure 18: *Two drawings with different stress values, but the same shape-based metric values.*

In a slightly larger example, consider the two drawings in Fig. 19; this graph has 295 vertices and 931 edges. The stress in Fig. 19(a) is much larger than that in Fig. 19(b), yet the shape-based metrics return almost the same value. Again we believe that Fig. 19(a) and Fig. 19(b) convey the structure of graph equally (perhaps (a) is better), and that the lower stress of Fig. 19(b) does not mean that it is a better drawing.

Our belief here (that in both Fig. 18 and Fig. 19, drawings (a) and (b) convey the graph structure equally) is just based on intuition; testing this belief empirically remains an open problem.

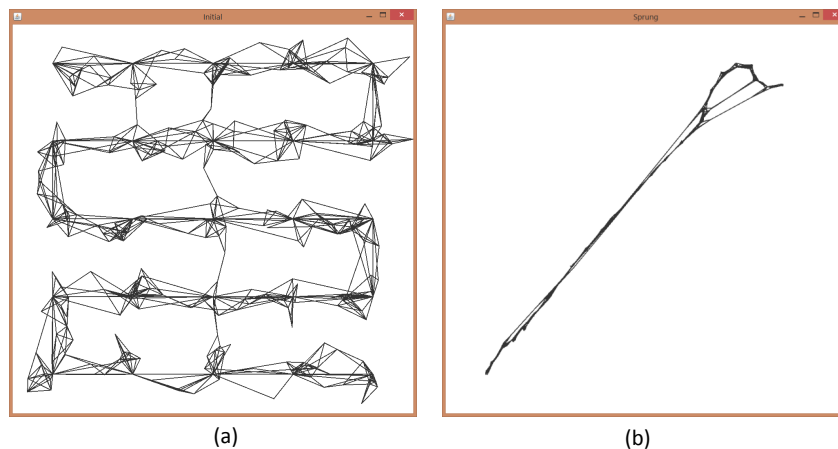


Figure 19: *Two more drawings with different stress values, but the same shape-based metric values.*

A further interesting open problem is to investigate whether there is any *mathematical* relationship between stress and shape-based metrics.

4.3 Algorithm evaluation

We believe that shape-based metrics can be used to compare graph drawing algorithms, especially for large graphs. We conjecture that the one reason that energy and force directed methods are universally used for large graphs is because they show shape better than other methods (for example, circular layout, orthogonal layout). This conjecture is currently untested.

4.4 Optimisation

Algorithms that draw graphs to optimize shape-based metrics are unknown. Note that (as with most graph layout problems) optimisation problems of this kind are typically NP-hard. For example, it is clearly NP-hard to find a drawing which optimises Q_{EMST} (see [4]). Thus approximation approaches are in order.

References

- [1] yworks. <https://www.yworks.com/>.
- [2] M. Chimani, P. Eades, P. Eades, S.-H. Hong, W. Huang, K. Klein, M. Marner, R. T. Smith, and B. H. Thomas. *People Prefer Less Stress and Fewer Crossings*, volume 8871 of *LNCS*, pages 523–524. Springer Berlin Heidelberg, 2014.
- [3] G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice-Hall, 1999.
- [4] P. Eades and S. Whitesides. The realization problem for euclidean minimum spanning trees in NP-hard. *Algorithmica*, 16(1):60–82, 1996. doi:10.1007/BF02086608.
- [5] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–558, 1983. doi:10.1109/TIT.1983.1056714.
- [6] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Softw., Pract. Exper.*, 21(11):1129–1164, 1991. doi:10.1002/spe.4380211102.
- [7] E. R. Gansner, Y. Hu, and S. Krishnan. COAST: A convex optimization approach to stress-based embedding. In S. K. Wismath and A. Wolff, editors, *21st International Symposium on Graph Drawing, GD 2013*, volume 8242 of *LNCS*, pages 268–279. Springer, 2013. doi:10.1007/978-3-319-03841-4_24.
- [8] E. R. Gansner, Y. Hu, and S. C. North. A maxent-stress model for graph layout. *IEEE Trans. Vis. Comput. Graph.*, 19(6):927–940, 2013. doi:10.1109/TVCG.2012.299.
- [9] X. Gao, B. Xiao, D. Tao, and X. Li. A survey of graph edit distance. *Pattern Analysis and Applications*, 13(1):113–129, 2010. doi:10.1007/s10044-008-0141-y.
- [10] S. Hachul and M. Jünger. Drawing large graphs with a potential-field-based multilevel algorithm. In J. Pach, editor, *12th International Symposium on Graph Drawing, GD 2004*, volume 3383 of *LNCS*, pages 285–295. Springer, 2004. doi:10.1007/978-3-540-31843-9_29.
- [11] G. E. Hinton and S. T. Roweis. Stochastic neighbor embedding. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 857–864. MIT Press, 2003. URL: <http://papers.nips.cc/paper/2276-stochastic-neighbor-embedding.pdf>.

- [12] Y. Hu and Y. Koren. Extending the spring-electrical model to overcome warping effects. In P. Eades, T. Ertl, and H. Shen, editors, *IEEE Pacific Visualization Symposium, PacificVis 2009*, pages 129–136. IEEE Computer Society, 2009. doi:10.1109/PACIFICVIS.2009.4906847.
- [13] W. Huang, S. Hong, and P. Eades. Effects of crossing angles. In *IEEE Pacific Visualization Symposium 2008, PacificVis 2008*, pages 41–46. IEEE Computer Society, 2008. doi:10.1109/PACIFICVIS.2008.4475457.
- [14] G. Jeh and J. Widom. SimRank: a measure of structural-context similarity. In *Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 538–543. ACM, 2002. doi:10.1145/775047.775126.
- [15] S. G. Kobourov, S. Pupyrev, and B. Saket. Are crossings important for drawing large graphs? In C. A. Duncan and A. Symvonis, editors, *22nd International Symposium on Graph Drawing, GD 2014*, volume 8871 of *LNCS*, pages 234–245. Springer, 2014. doi:10.1007/978-3-662-45803-7_20.
- [16] M. R. Marner, R. T. Smith, B. H. Thomas, K. Klein, P. Eades, and S. Hong. GION: interactively untangling large graphs on wall-sized displays. In C. A. Duncan and A. Symvonis, editors, *22nd International Symposium on Graph Drawing, GD 2014*, volume 8871 of *LNCS*, pages 113–124. Springer, 2014. doi:10.1007/978-3-662-45803-7_10.
- [17] Q. H. Nguyen, P. Eades, and S. Hong. On the faithfulness of graph visualizations. In S. Carpendale, W. Chen, and S. Hong, editors, *IEEE Pacific Visualization Symposium, PacificVis 2013*, pages 209–216. IEEE, 2013. doi:10.1109/PacificVis.2013.6596147.
- [18] F. Preparata and M. Shamos. *Computational Geometry - an Introduction*. Springer-Verlag, 1985.
- [19] H. C. Purchase. Which aesthetic has the greatest effect on human understanding? In G. D. Battista, editor, *5th International Symposium on Graph Drawing, GD '97*, volume 1353 of *LNCS*, pages 248–261. Springer, 1997. doi:10.1007/3-540-63938-1_67.
- [20] H. C. Purchase, J. Allder, and D. Carrington. Graph layout aesthetics in uml diagrams: User preferences. *Journal of Graph Algorithms and Applications*, 6(3):255–279, 2002. doi:10.7155/jgaa.00054.
- [21] H. C. Purchase, D. A. Carrington, and J. Allder. Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering*, 7(3):233–255, 2002.
- [22] H. C. Purchase, R. F. Cohen, and M. I. James. Validating graph drawing aesthetics. In F. Brandenburg, editor, *Symposium on Graph Drawing, GD*

- '95, volume 1027 of *LNCS*, pages 435–446. Springer, 1995. doi:10.1007/BFb0021827.
- [23] T. Sugibuchi, N. Spyrtatos, and E. Simonenko. A framework to analyze information visualization based on the functional data model. In *13th International Conference on Information Visualisation, IV 2009*, pages 18–24. IEEE Computer Society, 2009. doi:10.1109/IV.2009.56.
- [24] K. Sugiyama, S. Tagawa, and M. Toda. Methods for visual understanding of hierarchical system structures. *IEEE Trans. Systems, Man, and Cybernetics*, 11(2):109–125, 1981. doi:10.1109/TSMC.1981.4308636.
- [25] R. Tamassia, C. Batini, and G. Di Battista. Automatic graph drawing and readability of diagrams. *Technical Report, Università di Roma La Sapienza*, 01.87, 1987.
- [26] R. Tamassia, C. Batini, and M. Talamo. An algorithm for automatic layout of entity-relationship diagrams. In C. G. Davis, S. Jajodia, P. A. Ng, and R. T. Yeh, editors, *3rd Int. Conf. on Entity-Relationship Approach (ER'83)*, pages 421–439. North-Holland, 1983.
- [27] G. Toussaint. *Computational Morphology*. North Holland, 1988.
- [28] C. Walshaw. A multilevel algorithm for force-directed graph drawing. In J. Marks, editor, *8th International Symposium on Graph Drawing, GD 2000*, volume 1984 of *LNCS*, pages 171–182. Springer, 2000. doi:10.1007/3-540-44541-2_17.
- [29] C. Ware, H. C. Purchase, L. Colpoys, and M. McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2):103–110, 2002. doi:10.1057/palgrave.ivs.9500013.
- [30] L. Wilkinson, A. Anand, and R. L. Grossman. Graph-theoretic scagnostics. In J. T. Stasko and M. O. Ward, editors, *IEEE Symposium on Information Visualization (InfoVis 2005)*, page 21. IEEE Computer Society, 2005. doi:10.1109/INFOVIS.2005.14.