Eurographics Conference on Visualization (EuroVis) 2017
J. Heer, T. Ropinski and J. van Wijk
(Guest Editors)

Volume 36 (2017), Number 3

Measuring Symmetry in Drawings of Graphs

E. Welch and S. Kobourov

University of Arizona, Tucson, Arizona

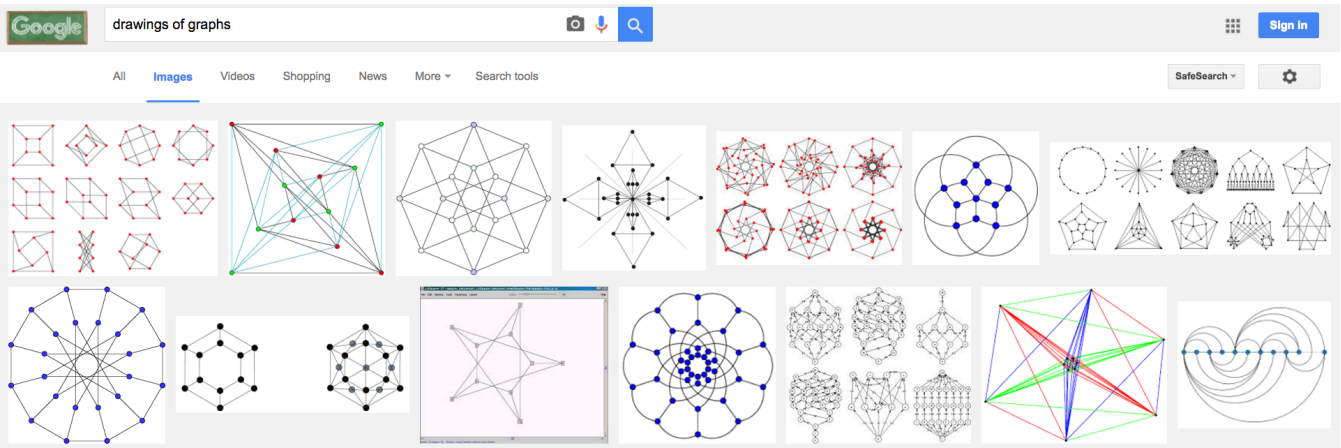

Figure 1: *The top images returned by Google-Image-Search for "drawings of graphs." Note that the overwhelming majority are graphs that are "symmetric" and also drawn in such a way as to highlight the underlying symmetries.*

Abstract

Layout symmetry is an important and desired features in graph drawing. While there is a substantial body of work in computer vision around the detection and measurement of symmetry in images, there has been little effort to define and validate meaningful measures of the symmetry of graph drawings. In this paper, we evaluate two algorithms that have been proposed for measuring graph drawing symmetry, comparing their judgments to those of human subjects, and investigating the use of stress as an alternative measure of symmetry. We discuss advantages and disadvantages of these measures, possible ways to improve them, and implications for the design of algorithms that optimize the symmetry in the layout.

1. Introduction

Symmetry in nature and in art attracts our attention and influences our perception of beauty. Multiple studies in cognitive psychology consider the role that symmetry plays in our perception of the world [GIA13, Tre10]. One explanation of the importance of symmetry comes from the Gestalt theory of human perception [WR44], which generally refers to the effect that when we see a group of objects, we perceive their entirety before we see each object individually. Even though the objects are separate entities, our cognitive system tries to make sense of them by grouping them into a whole, with a characteristic shape that is more salient than its constituent parts [Wes99]. As symmetrical objects possess many similar components, they are easily grouped together and perceived as a unit by our cognitive system [Tre10, War12]. Symmetry perception itself is an integral part of object recognition, the process by which specific objects are distinguished from the background [Tre10].

We study how the Gestalt principle of symmetry relates to visual aspects of graph drawings (e.g., node proximity, edge lengths), and how symmetries are embodied in the most popular graph layout methods (force-directed and multi-dimensional scaling). Specifically, we would like to identify effective measures for evaluating the symmetry of a given graph layout and consider how to use such measures to develop layout algorithms that optimize the symmetries in a given graph. The rationale for this work is that aesthetically appealing and symmetric layouts lead to more effective and more memorable visualizations of relational data [Pur97, MPWG12].

The detection of symmetry in 2D and 3D objects is a challenging computational problem that is undertaken by the scientific direction of *computational symmetry*. Early attempts to measure symmetry of 2D objects date back to 1932 [Bir32]. In spite of decades of effort by the fields of computer vision and computer graphics, there are still no widely applicable "symmetry detectors" for real images [LHKG10]. A survey of existing algorithms and approaches for symmetry detection can be found in [LHKG10, MPWC13].

The human desire for order and symmetry extends naturally to graph drawing and has been explicitly used to lay out large graphs by allowing human agents to select and arrange subgraphs, then algorithmically combining these components [YCHZ12]. Indeed, symmetry constitutes one of the most favorable features for graph drawing. In an empirical study by Kieffer et al. [KDMW16], participants were asked to adjust drawings of graphs manually until they felt that the layouts "looked good" and clearly conveyed the connections between nodes. An analysis of the results revealed that reflectional symmetry was among the features that were emphasized in the drawings and preferred by the other participants. It has also been confirmed that a higher level of symmetry increases the understandability of a graph [PCJ95, PCJ97, PCA02], and makes a layout more memorable [MPWG12]. The force-directed method, a very popular algorithm for graph visualization, is known to produce "symmetric" visualizations [Tam07]. It is also worth noting that many of the top results returned by Google Images when searching for "graph drawing" depict perfectly symmetric graph drawings; see Fig. 1.

In this paper, we evaluate two algorithms designed to measure the symmetry of a graph drawing by inspecting their performance on a collection of graphs from the AT&T undirected graph set [ATT]. We make some general observations about their performance, and compare their judgments of graph drawings to the judgments made by a sample of human subjects. We also consider the measure of stress (which is minimized by multi-dimensional scaling methods) as a candidate for symmetry detection, and compare its performance to those of the other measures with respect to human judgment. After an analysis of the performance of each measure and a discussion of their advantages and disadvantages , we consider ways in which they may be improved, as well as implications for the design of layout methods that maximize symmetry.

2. Background

Symmetry Perception. In a review of the research on the visual perception of symmetry in humans and animals, Giannouli [GIA13] reports the following findings, among others:

- symmetry supports the separation of objects from the background and the perception of an object's orientation in space
- symmetrical objects are more easily encoded, recognized, and recalled
- symmetry perception appears at 4 months of age, and preference for mirror symmetry appears across cultures at 12 months
- symmetry, followed by complexity and familiarity, is the "major determining factor in aesthetic judgments"
- mirror symmetry (reflective symmetry across a vertical axis) is detected faster than rotational symmetry or reflections across other axes
- symmetry detection is automatic, accurate, and does not require attention except in the case of complex (non-composite) visual stimuli, for which a non-automated, point-to-point comparison is made

Thus, humans have extensive experience making use of symmetry in their everyday perception of objects in the world, and can be expected to make reasonable judgments about the presence of symmetry in drawings of graphs. Another review of the research on symmetry echoes the importance of symmetry detection in object formation, and characterizes human symmetry detection as automatic, on-going, and "noise-resistant" [Tre10]. Human vision has long been known to involve a two-step process of preattentive primitive feature perception followed by a focused effort to join these primitives into coherent objects [Tre85]. The fact that symmetry perception is largely preattentive suggests that symmetric objects are either identified as primitives by the visual system in and of themselves, or can rapidly be identified as objects based on numerous direct mappings amongst their primitive features. In contrast, the detection of symmetry in more complicated stimuli may require, in the language of the Gestalt literature, effortful "mechanical and intellectual operations" [Wes99].

Graph Symmetry. Note the important distinction between detecting symmetries in a graph versus detecting symmetries in a layout of a graph. In graph-theoretic terms, finding symmetries in a given graph is known as computing automorphisms and the automorphism group of the graph. Layout methods that attempt to display computed automorphisms as geometric symmetries go as far back as the 1985 Lipton, North, and Sandberg paper [LNS85]. Since the underlying problem is at least as computationally hard as the graph automorphism problem, heuristic methods have also been considered [DF99]. Fox et al. [FLP07] describe a method for finding "nearly symmetric" subgraphs using a two-phase algorithm that finds "flaws" in subgraphs and repairs them. Buchheim and Jünger [BJ03] detect "fuzzy" symmetry using integer programming, allowing edge deletions and insertions and assigning penalties to these operations in the construction of the ILP. In a nice survey of the field, Hong and Eades [HE05] discuss the problem of drawing a graph with a layout isometry that displays an underlying graph automorphism, and provide a sketch of the proof that there is an $O(n)$ algorithm for drawing planar graphs with no edge crossings and displaying "as much symmetry as possible."

Graph Layout Symmetry. There has been little work on drawing a graph in a way that maximizes the symmetry in the drawing, independent of the presence or absence of total axial or rotational automorphisms. Chuang and Yen [CY02] describe an approach for drawing (asymmetric) graphs by contracting edges until a symmetric subgraph is obtained. A force-directed algorithm is used to lay out the resulting subgraph, then the contracted edges are reinserted, followed by a second round of force-directed adjustment. Eades and Lin [EL00] provide general theoretical argumentation and evidence that force-directed (spring) layout methods can display graph symmetry.

User Evaluations. Purchase [PCJ97, Pur97] has studied the impact of graph drawing aesthetics on human performance, and found that a greater degree of symmetry led to improvements in response time for graph-theoretic tasks. While other aesthetics considered

in these evaluations (e.g., minimizing edge crossings) have a clear and objective measure, there is no such obvious way to measure the symmetry of graph layouts. To the best of our knowledge, no effort has been made to validate measures of graph layout symmetry by comparison with the human perception of symmetry.

3. Measures of Graph Drawing Symmetry

There are two earlier methods for measuring the symmetries in a given graph layout [Pur02, Kla14]. Both return a numeric value in the range $[0,1]$ to indicate the extent to which the drawing is considered symmetric, with 1 corresponding to a perfectly symmetric drawing. Before describing these in more detail, we mention that *stress*, the optimization function for classical multi-dimensional scaling, could also be considered as a proxy for symmetry, given earlier results that show theoretical and practical correlations between the two [EL00].

3.1. Purchase Measure

The measure defined by Purchase [Pur02] considers only reflective symmetry, ignoring rotational symmetry. To compute the measure, an axis of potential symmetry is generated between every pair of graph vertices (where edge crossings are also considered special "vertices"). For each such axis, a symmetric subgraph, consisting of all the edges that are indicent on vertices mirrored across the axis within a predefined *tolerance*, is computed. The value for axis is determined (mostly; see [Pur02]) by the convex hull area of the subgraph. The final symmetry score is a ratio involving the sums of the values for all nontrivial axes. Note that this measure depends upon the precise locations of graph vertices, and is designed to accommodate multiple axes of reflective symmetry, in an effort to capture both "local" and "global" symmetries. The computational complexity is $O(n^7)$, where n is the number of vertices in a graph.

A notable feature of Purchase's algorithm appears when we examine the axes of symmetry generated during the analysis: at any tolerance, a layout will often generate more axes than a human is likely to perceive; see Fig. 2.

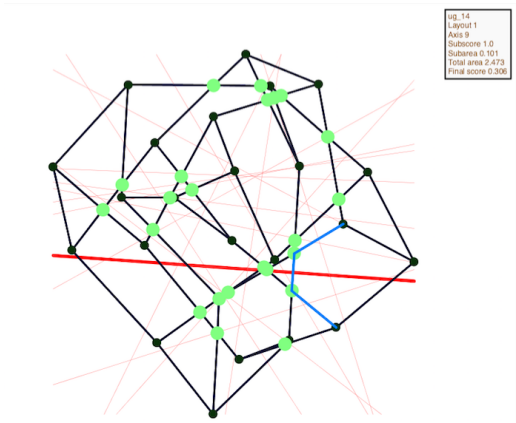

Figure 2: *All axes of symmetry identified by the Purchase metric (thin red lines) along with a particular axis (thick red line) and its corresponding subgraph (blue edges). It seems unlikely that a human considering the symmetry of a layout would perceive reflections across so many axes.*

3.2. Klapaukh Measure

Citing perceived limitations in Purchase's symmetry measure, Klapaukh [Kla14] suggests an *edge-based* measure that explicitly calculates reflection, rotation and translation symmetries. One layout ranked as more symmetric than another if it has higher scores in two of these measures. (Note that this ranking system is not transitive; for instance, consider three layouts with corresponding scores (0.9, 0.8, 0.7), (0.8, 0.7, 0.9), and (0.7, 0.9, 0.8).) We focus here on the detection of reflective symmetry, both because there is no natural way to combine these three measures, and because the graphs we considered, when drawn using a force-directed layout did not seem to exhibit significant visible or measurable translational or rotational symmetry. Moreover, in the layouts with higher values for these measurements, there did not seem to be a corresponding sense of greater symmetry.

Klapaukh [Kla14] suggested that Purchase's symmetry detection measure was limited in that it focused on vertices, ignoring edges that may be perceived as reflected across an axis even though their endpoints do not line up due to small differences in their orientation and length. Borrowing ideas from a computer vision algorithm for symmetry detection [LE06], the Klapaukh measure encodes edges as SIFT features, and uses each edge and each pair of edges to generate potential symmetry axes (a potential axis maps an edge onto itself, or the midpoint of one edge onto another). Each symmetric axis is assigned a *quality score*, based on how well the edges map onto one another with respect to their lengths and orientation. The axes are quantized, and those that are sufficiently similar are taken as a single axis – the quantization, like Purchase's tolerance, is motivated by the stability of human symmetry perception against small amounts of noise. Each of these combined axes votes for itself, using the sum of the quality scores computed for the axes that were combined into it, to determine the best N axes. The final symmetry score is a normalized sum, over these best axes, of the number of edges that vote for each axis. (Note that all examples in [Kla14] use $N = 1$; we do the same in our study.) This approach is also computationally expensive.

3.3. Stress

Several traditional methods for drawing graphs are based on minimizing a suitably-defined energy function of the graph layout. Layouts with minimal energy tend to be aesthetically pleasing and to exhibit symmetries. One of the most popular and commonly used energy functions, called "stress," is defined as the variance of edge lengths in the drawing [GKN05,KÇ09]. Assume a graph $G = (V,E)$ is drawn such that p_i is the position of vertex $i \in V$. Denote the distance between two vertices $i, j \in V$ by $||p_i - p_j||$. The energy of the graph layout is measured by

$$\sum_{i,j \in V} w_{ij}(||p_i - p_j|| - d_{ij})^2, \tag{1}$$

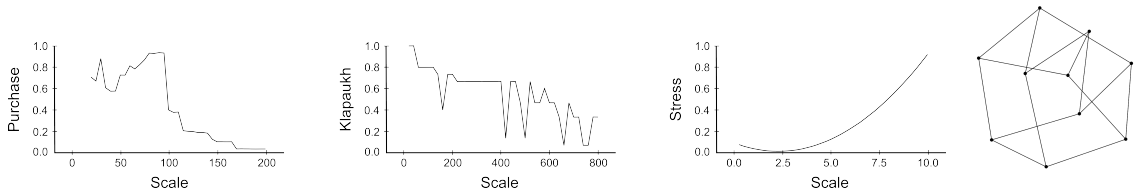

Figure 3: *The Purchase, Klapaukh, and stress measures of the graph layout on the right vary in response to the scale of the layout. Note that the Purchase and Klapaukh measures respond erratically to scale, while stress behaves predictably.*

where d_{ij} is the ideal distance between vertices i and j, and w_{ij} is a weight factor. Lower stress values correspond to a better layout. While there are several variants of the stress expression, we use the most conventional one in which an ideal distance d_{ij} is defined as the length of the shortest path in G between i and j, and utilize the conventional weighting factor of $w_{ij} = \frac{1}{d_{ij}^2}$.

Following previous work [GHN13,KPS14], we also scale the drawing before computing its stress. To this end, we find a scalar s that minimizes $\sum_{i,j \in V} w_{ij}(s||p_i - p_j|| - d_{ij})^2$. This is done to be fair to methods that do not try to fit layout distance to graph distance for all pairs of vertices.

Several earlier works argue that minimizing stress is related to optimizing symmetry in the drawing. For example, Eades and Lin [EL00] proved that the solution of a "general spring model" can uncover symmetries. Dwyer et al. [DLF*09] showed that users prefer graph layouts with lower stress. Finally, Kobourov et al. [KPS14] found a correlation between stress and the number of crossings in a drawing, and thus, stress minimization is a desirable criterion.

3.4. Robustness vs fragility to scale

Graph layout algorithms compute x, y coordinates for each vertex, but the distances between these vertices in pixels when they are rendered on a screen may be scaled depending on the resolution at which the layout is viewed. We define the *scale* of a drawing to be the maximum of the width or height of its bounding box. Some measures of a layout may be expected to vary as the scale changes, e.g., the convex hull area of a graph is quadratic in the scale of the graph. However, the symmetry of a layout should be *robust* to scaling, i.e., remain the same as the positions of all nodes vary proportionally.

We can directly address the issue of scaling in the Purchase metric by selecting a tolerance that is a ratio of the entire width of the drawing, e.g., 0.5% or 5% of the scale of the drawing. However, there is no corresponding parameter for the Klapaukh measure of reflective symmetry that makes it robust to scale; we therefore say that it is *fragile* with respect to its scale. (It is theoretically possible to adjust the size of the bins used in the quantization step in response to the graph size, but unlike the tolerance parameter, it is not obvious how this choice would be made.)

One might expect the Klapaukh measure to be monotonic in the scale of the graph, so that as the features that encode the axes of symmetry grow farther apart, and fewer are binned together in the quantization step, the best final axis would receive fewer votes. In fact, the measure is *not* monotonic in the scale size; see Fig. 3. This is likely a consequence of the quantization step, which places axes into arbitrarily determined bins in the manner of a histogram.

As a consequence of this fragility, the Klapaukh algorithm, which seems to be motivated by reasonable notions of human perception of symmetry (e.g., our perceptual system will collapse several axes of symmetry into a single axis if they do not vary much) gives measures that would appear to disagree with human judgment; see Fig. 4.

Stress also varies with scale, but as a consequence of its mathematical definition, the measure achieves a global minimum; see Fig. 3. We can therefore define a robust measure of stress, in which we use a binary search to find the scale for the layout that produces the the least stress value. For the rest of the paper, whenever we speak of the stress of a layout, we refer to this minimized stress.

We emphasize that it is not possible to similarly precompute an optimal scale before using the Klapaukh or Purchase measures on a layout, because both would converge to a maximum value of 1 for a suitably small layout, in which every node or edge is mapped onto another.

3.5. Stability vs instability to parameter values

As mentioned, the Purchase metric judges one edge to be mapped onto another across a reflective axis if their corresponding endpoints are reflected onto one another within a specified tolerance.

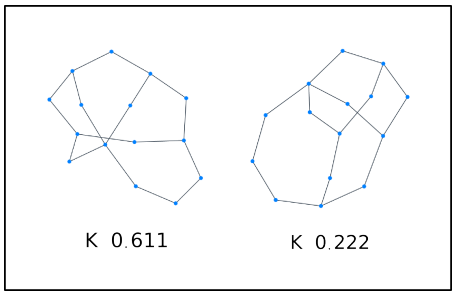

Figure 4: *Two layouts of a graph and their reflective symmetry scores according to the Klapaukh metric. Although the layout on the left received a higher score, we would expect human judgment to select the layout on the right as more symmetric, as every edge on its convex hull is rougly reflected onto another across a common diagonal axis.*

An inappropriate tolerance value can cause two layouts that appear to exhibit similar amounts of symmetry to receive sharply different scores; see Fig. 5. In the examples in [Pur02], a value of 3 pixels is given for the tolerance parameter that is used to determine whether one node is mapped onto another by an axis of symmetry. It is not clear, however, what an appropriate value for this parameter should be, or how it can be determined *a priori*. Moreover, our preliminary investigations suggest that a tolerance value that seems reasonable for one layout can result in judgments that would likely differ from those of most humans for other layouts; see Fig. 6. As we vary this parameter over a continuous range, at one end of the range there are values for which nodes might be inappropriately mapped onto one another, and at the other end, there are values for which nodes might not be mapped onto one another, but which humans may judge as roughly corresponding; see Fig. 7.

As there are no tunable parameters for the Klapaukh measure or stress, we do not need to worry about issues of stability.

3.6. Lack of agreement among metrics

To confirm that the three metrics (Purchase, Klapaukh reflective, and minimized stress, henceforth P, K, and S) do not generally agree on the level of symmetry exhibited in layouts of graphs, we generated 100 different layouts for 25 different graphs. The graphs are a subset of of the AT&T undirected graph data set [ATT] that are small ($n \leq 20$) and sparse ($m < 20$), and we created the different layouts using the Fruchterman-Reingold force-directed algorithm [FR91]. We reasoned that if two of these measurements accurately determined the amount of symmetry in a graph, there would be a strong correlation between the values they reported over this modest sample. However, when we computed the correlations between each pair of metrics (P vs K, K vs S, P vs S) over these layouts, we found no meaningful correlations for the majority of the graphs.

4. Experimental Design

Recall that we are interested in determining whether the existing measures for symmetry detection in drawings of graphs agree with a human notion of symmetry, and if so, how to use these measures to create more symmetric drawings. To this end, we designed an experiment to measure agreement between human judgment and the symmetry scores assigned by different algorithmic measures.

4.1. Graph and Layout Selection

For our user studies, we used graphs from the AT&T undirected graph data set [ATT]. For simplicity, we collapsed multi-edges to a single edge, removed disconnected graphs, and eliminated duplicate graphs. As in several other human subject studies involving graphs [PPP12, vHR08], and to keep issues related to scale and tolerance as simple as possible, we used only small ($n \leq 20$) and sparse ($m < 20$) graphs; see Table 1 for details.

We wanted to generate layouts that exhibited as much symmetry as possible, as well as layouts that exhibited less symmetry. Force-directed algorithms are capable of generating layouts of maximal symmetry [EL00], but the final layout is dependent upon the initial

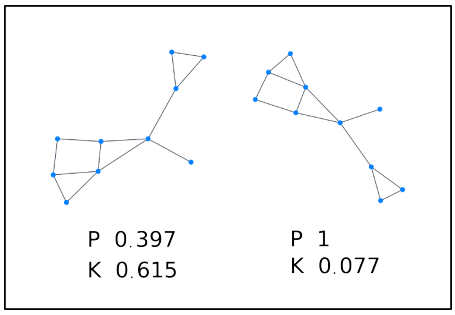

Figure 5: *Two layouts of a graph and their reflective symmetry scores according to the Purchase and Klapaukh metrics. Although the layouts seem quite similar, they received very different scores by each metric.*

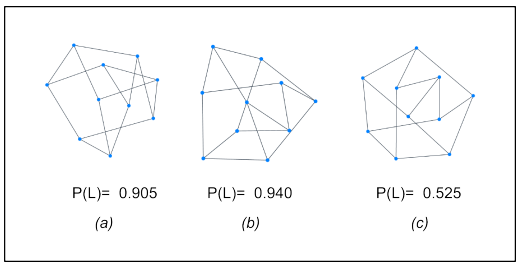

Figure 6: *(a, b) Two layouts of a graph that received much higher scores using Purchase's metric for reflective symmetry than (c), a layout that a human would likely judge as more symmetric.*

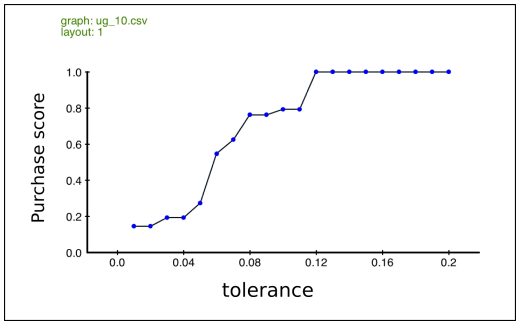

Figure 7: *The Purchase score is heavily dependent upon the selection of a tolerance parameter. It is not clear a priori how to set this parameter.*

random placement of the vertices. With this in mind, we generated 100 layouts of each graph using the `spring_layout` function in the GraphPlot library for Julia [HBHG], which implements the Fruchterman-Reingold force-directed layout algorithm [FR91]. We reasoned that with 100 randomly-generated layouts of small graphs, there would be a good chance of producing a desirable range of layouts: some more symmetric and some less symmetric.

To select the specific graphs and pairs of layouts for the pilot study,

Graph ID	nodes	edges	Purchase	Klapaukh	Stress
10	12	12	0.10-1.00	0.08-0.92	2.18-7.59
11	10	15	0.02-1.00	0.07-1.00	5.21-5.71
58	5	9	0.00-1.00	0.11-1.00	0.41-0.89
59	10	13	0.04-0.54	0.08-0.69	1.33-1.67
74	20	19	0.09-1.00	0.11-0.89	11.00-22.84
178	11	11	0.07-1.00	0.09-0.91	3.53-4.40
208	10	13	0.04-0.59	0.15-0.77	1.14-2.69
209	9	10	0.09-0.67	0.11-0.78	1.46-3.04
215	4	4	1.00-1.00	0.75-1.00	0.03-0.05
216	5	8	0.42-1.00	0.25-1.00	0.21-1.20
225	6	7	0.15-0.16	0.14-0.71	0.22-0.24
228	5	8	0.00-1.00	0.12-1.00	0.29-1.09
229	7	10	0.00-0.50	0.10-0.80	0.68-2.20
265	7	6	0.08-1.00	0.17-1.00	0.06-0.69
297	9	10	0.00-0.82	0.10-0.60	0.99-3.03
308	15	18	0.05-0.99	0.06-0.61	7.93-13.07
332	7	8	0.11-1.00	0.12-1.00	1.09-2.20
334	12	18	0.00-0.99	0.06-0.67	5.25-7.27
335	12	18	0.01-0.97	0.06-0.72	5.26-7.24
358	13	14	0.26-0.88	0.07-0.71	4.13-8.34
363	8	10	0.00-1.00	0.10-1.00	1.79-2.78
364	9	10	0.04-0.35	0.11-1.00	0.75-4.31
366	8	10	0.00-1.00	0.10-1.00	1.79-3.24
367	9	10	0.05-0.35	0.11-0.89	0.77-3.37
375	8	7	1.00-1.00	0.29-1.00	1.12-1.22

Table 1: *Features of the AT&T graphs used in the human subjects experiment, and the range of values produced by each metric for the 100 layouts generated for the graph.*

we created a custom tool that provides a visual comparison of the symmetry scores of the 100 layouts [Wel]. We used this tool to identify pairs of layouts for graphs that disagreed sharply according to Klapaukh's reflective symmetry metric and the Purchase metric; that is, P gave a much higher score to layout i than to layout j, while K gave a much higher score to layout j than to layout i. This seemed to be the simplest way to determine whether one metric outperformed the other. We found 25 such pairs.

For the formal study, we again wanted to select layouts that maximized differences in judgment, but by this time, we had discovered the strong effects of scaling on Klapaukh's measure and the tolerance parameter on Purchase's measure. In an effort to identify the most reasonable tolerance and scale for our collection of graphs, we further evaluated the symmetry scores of all the layouts of all 25 graphs over a range of scaling and tolerance parameters, each time taking their average and standard deviation. Although we had seen that the measures for individual layouts varied erratically over a range of parameters, we found that the mean and standard deviation over all layouts for each measure was monotonic in the corresponding parameter. Through a series of refinements, we came to a set of parameters so that both measures obtained an average value of 0.5 over all layouts of all graphs (0.385 tolerance for Purchase, where all graphs were scaled to a maximum width or height of 2, and 680×680 scaling for Klapaukh).

Using these values, for each graph and for each experiment type (P vs K, P vs S, K vs S), we compared all pairs of layouts to identify those for which the measures disagreed most sharply. To avoid repetitive stimuli in the experiment, we assigned each graph to only one of the three experiment types. We ultimately settled on seven pairs of layouts for each comparison of metrics, distributing the graphs to maximize the total disagreement between the metrics over all experiment types.

4.2. Pilot Study

In preparation for our main study, we recruited 23 graduate and undergraduate participants involved in data visualization (coursework or research) to participate in a pilot study. The participants were directed to a custom web survey application that presented one pair of layouts at a time, and asked them to click on the one that looked more symmetric. After seeing all 25 pairs of layouts, the participants were automatically directed to an exit survey consisting of questions about how important various factors seemed in judging symmetry, including nodes that matched, edges that matched, similar edge lengths, matching shapes (polygons), or a "gut feeling" about which layout was more symmetric. (Some of our participants later told us that verbal descriptions such as "nodes that match" were confusing, so we decided to ignore this data in the pilot study. We developed small illustrations to accompany these questions in the final study.)

Hoping to find that one of K or P would strongly outperform the other, we took the majority vote of all participants as a measure of human judgment for each layout pair. There was no clear winner, , but the subsequent inspection of the layouts and their votes led us to conclude that the measures suffer from a lack of stability and a lack of robustness, causing us to consider these issues carefully in our formal study.

One unexpected observation was that for two of the graphs, participants were unanimous in their preferences for one layout over the other, and also responded in an average time of less than one half the mean time over all layouts (2.27 seconds and 2.45 seconds, compared to an average of 5.37 seconds). This led us to revisit the idea of automatic symmetry detection in our formal study.

4.3. Experimental Design

We designed a custom web application [Wel] to guide participants through the experiment, providing task instructions and collecting responses and response times. At the beginning of the experiment, we briefly introduce graphs, the notion of a graph layout, and the experiment itself. Next, the 21 layout pairs are presented in a random order, with the layouts randomly assigned to be placed on the left or the right. For each question, the participants see the same request: "Click on the graph that looks more symmetric." The amount of time each participant takes to answer each question is logged, starting as soon as the images have finished loading on the page and ending when the subject clicks on one of the images. As in the pilot study, at the end of the survey we ask the participants to self-evaluate on how important various factors (e.g., rotation, translation, reflection, node matching, edge matching) were when they made their decisions.

We also collected basic demographic and background information,

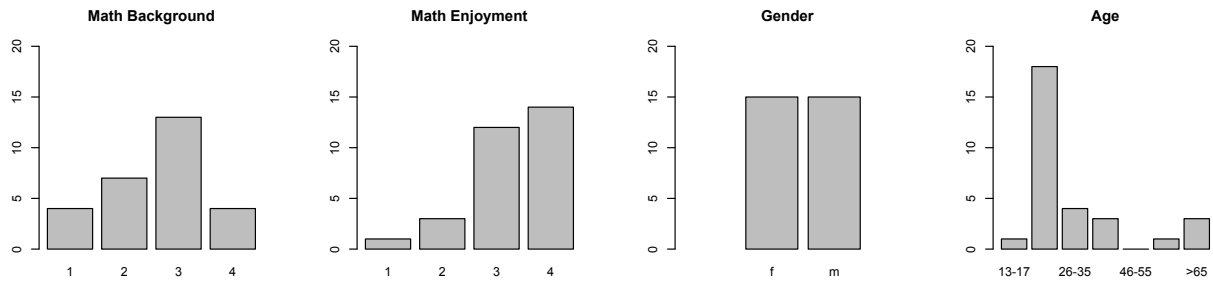

Figure 8: *Demographic characteristics of study participants. Math Background is coded as (1) K-12, (2) 100-200 level college, (3) 300-400 level college, (4) graduate. Math Enjoyment ranges from (1) "Math is the absolute worst" to (4) "Math is really great."*

such as age, gender, and mathematical background. The exit survey is provided in the supplementary material, and is also available online [Wel].

4.4. Hypotheses

On the basis of findings about human symmetry perception, certain features of the symmetry measurement algorithms and our observations of the behavior of these measures, as well as results and feedback from our pilot study, we propose the following hypotheses:

- H1: When presented with layouts that differ in their judgments by the Klapaukh and Purchase measures, human subjects will agree with the judgments of one measure over the other.
- H2: Human subjects will agree with the judgments of stress more than with the judgments of either the Purchase or the Klapaukh metric.
- H3: Human judgment of symmetry will favor layouts that demonstrate mirror symmetry (reflection across a vertical axis) of the entire drawing.
- H4: The speed of the average response for a winning layout will be correlated with the level of consensus for this layout.

H1 follows from the different ways in which P and K measure symmetry and the significant differences in scores that P and K assign to many pairs of layouts. As neither of the measures seemed to agree more with human judgment than the other in our pilot study, we can only hypothesize that one should be better.

H2 is motivated by prior work associating stress with symmetry [EL00, DLF*09]. It also seems plausible, given the discussion in Section 3; in particular, the observation that stress is more "well-behaved" than the other measures, i.e., it is not subject to issues of tolerance or quantization that seem to make the other measures sometimes give rather different results for perceptually similar layouts.

H3 is grounded in evidence that mirror symmetry about a vertical axis is perceived more readily than symmetry about other axes [GIA13]. In feedback following our pilot study, participants also cited mirror symmetry as more significant than other axial symmetries.

H4 is based on results from symmetry perception research that symmetry detection is typically preattentive [GIA13, Tre10]. Our pilot studies also found that unanimously preferred layouts had average decision times that were less than half the average time of those for all layouts. We expected to find that if two layouts had a similar amount of symmetry, costly "mechanical and intellectual operations" would be required to determine which was more symmetric. This would lead to slower results, and to "noise" in the final decisions resulting from the interference of cognitive systems required for point-to-point comparison that are not typically involved in symmetry detection.

4.5. Participants

We recruited participants via social media and personal emails. Of the 30 subjects who participated, 15 were male and 15 were female. The majority of the participants were in the 18-25 age range. Most of them reported enjoying math, and most had taken college coursework in mathematics; see Fig. 8.

5. Analysis

In order to evaluate the first two hypotheses, for each pair of measures we counted the number of votes that one measure received compared to the other and applied a binomial test. We found support for H1, as our participants agreed with the judgments of the Purchase metric more often than the Klapaukh metric (125 vs 85). Specifically, a binomial test found a significant preference for Purchase over Klapaukh ($p < 0.01$), confirming H1.

We did not find support for H2, as both P and K agreed with human judgment more than S (114 vs 96 and 134 vs 76, respectively); see Fig. 9. In fact, a binomial test would suggest a significant preference for K over S ($p < 0.01$), although there was no significant preference for P over S. This lack of transitivity initially struck us as puzzling, but as our hypothesis was that stress would outperform both P and K, we cannot affirm that K matches human judgment any more than stress.

To evaluate the third hypothesis, we examined the pairs of layouts

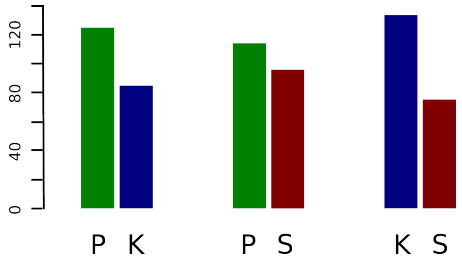

Figure 9: *On survey items in which layouts with higher P scores were compared to layouts with higher K scores, human judgment significantly preferred P over K (125 vs 85), supporting H1. There was no support for H2.*

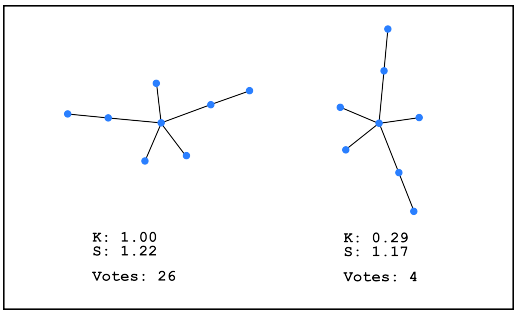

Figure 10: *An example of a pair of layouts for which the amount of symmetry does not seem to differ, but for which humans demonstrated a preference for a vertical axis of symmetry.*

with statistically significant differences in human preference, hoping to observe that differences in orientation may in some cases explain the outcome. We found some evidence in support of H3, but could not validate it statistically. Specifically, we identified two layouts in which there does not seem to be much difference between the amount of symmetry in the layout, but for which humans demonstrated preference for one over the other; see Fig. 10. However, there did not seem to be any other layouts for which humans sacrificed a compelling difference in symmetry in favor of a vertical axis.

For the fourth hypothesis, we computed the correlation between the number of votes a layout received and the average speed of the responses for the layout. We did not find support for H4, although as in the pilot study, there were some cases in which almost all of the participants made the same decision, and did so rapidly. Specifically, layouts that received more than 90% of the votes against their paired layout were decided on average in less than one half of the average time over all responses.

To determine whether there may be any bias in our results due to the demographic make-up of our participants, we inspected the voting patterns of all demographic subgroups. We also reviewed participants' responses to the questions about which features of a layout were most important in their judgments, and examined the layouts used in the study to get a sense of why the Klapaukh and Purchase measures may have made judgments that differed from those of the human subjects.

5.1. Subgroup analysis

To account for the possibility that different demographic subgroups might respond to the tasks in different ways, we examined the voting results for age, gender, levels of mathematical background, and level of mathematical enjoyment. Most of these subgroups mirrored the general agreement of *P* over *K* and *K* over *S*, with the notable exceptions of the 36-45 age group (P=6 < K=15) and the K-12 math background group (P=13 < K=15). Most groups were roughly ambivalent on *P* vs *S*, although it is interesting that the graduate level math background group strongly preferred *P* over *S* (22 vs 6). The entire subgroup analysis is available in the supplemental material [Wel].

5.2. Survey

Our exit survey confirmed that participants found reflection symmetry to be a more salient feature than rotational or translational symmetry. It also suggested that a "gut feeling" was more important than other factors in deciding which layout is more symmetric, followed by matching edges and matching polygons; see Table 2.

5.3. Examination of Layouts

We ordered the layout pairs according to the number of votes given to the "winning" drawing (i.e., the drawing with more votes) to look for reasons why the "losing" metric may have misjudged the level of symmetry in the layout [Wel]. In several cases, we suspect that the Klapaukh measure gave a score that was too low due to a problem involving the quantization step (Fig. 11). We also found three cases in which human judgment was essentially random (14 vs 16), while Purchase ascribed a difference of values greater than 0.2 to one over the other. These seem to be cases in which, although the layouts appear identical (up to rotation), there are small changes in the position of the vertices, and the Purchase metric is demonstrating its oversensitivity to tolerance.

6. Discussion and Limitations

Due to the limitations of our study, we hesitate to affirm that *P* is definitively superior to *K* in terms of agreement with human judgment. Nevertheless, we find that our study provides insights into deeper issues of layout symmetry.

6.1. Generalizability

While our main result is significant support for human preference of the rankings of the Purchase metric over the Klapaukh metric, we are wary of overgeneralizing this finding. In particular, we must distinguish between the theoretical validity of the metrics and the quality of their results for a set of parameters in practice.

We selected the layouts for our final study so as to maximize the disagreement between each pair of metrics, and to avoid researcher bias, we did not examine the layouts after selecting them. This was

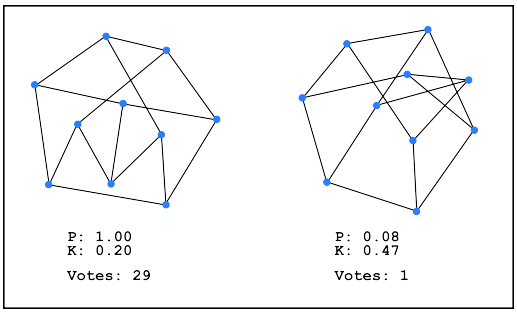

Figure 11: *An example where the Klapaukh measure likely failed to identify the highly symmetric layout (left) due to a quantization error.*

Consideration	Importance
reflection	3.87
gut feeling	3.43
matching edges	3.20
matching polygons	3.14
matching nodes	2.90
larger matching regions	2.96
rotation	2.63
similar edge lengths	2.52
translation	2.30

Table 2: *Mean values of participants' responses (1-4) to the question "How important were each of the following when you made decisions about symmetry?"*

not a random sample of layouts, and this method may have instead introduced a bias towards the set of layouts in which one or both metrics produced aberrant results due to quantization or threshold issues. For instance, a post hoc inspection of the layouts used in the P vs K experiment revealed that several layouts chosen by this method seemed to "thwart" the Klapaukh measure, which did not successfully identify matched pairs of edges that it should have according to the algorithm design. While this may mean that K tends to suffer from this fate more frequently than P, we lack sufficient evidence to support such a claim.

The possibility that the rankings of P were preferred to those of K because of quantization issues rather than the validity of the P algorithm over the K algorithm is bolstered when we examine the responses to the exit survey questions. Users ranked edge matches, the theoretical basis for the K metric, as more important than node matches, the basis for the P metric.

6.2. Multi-facted nature of symmetry

Our study suggests that, in terms of agreement with human judgment, P significantly outperforms K and K significantly outperforms stress, but P does not significantly outperform stress. One plausible explanation for this lack of transitivity is that symmetry is not a single feature that is detected more or less well by the three metrics, but rather has multi-faceted nature that is only partially captured by these measures. It is also possible that symmetry in graph layouts would be better explained by some meta-feature that could hopefully be derived from the positions and connections between nodes. In this case, an accurate measurement of symmetry would depend upon identifying this feature, rather than focusing on primitives (e.g., matching nodes, matching edges, or similarity of edge lengths).

6.3. Resilience to noise

Although we attempted to find a tolerance for the Purchase metric and a scale for the Klapaukh metric that would give the most reasonable values over our dataset, both metrics still produced some "bad" outputs, resulting from a failure to detect edges that should be matched. This raises the possibility that there may be no "correct" value for these parameters. Indeed, an "all-or-nothing" approach to discriminate matching nodes or equivalent axes may be a fundamentally flawed approximation to noise resilience.

6.4. Preattentiveness

Although we did not explicitly test the hypothesis that the fastest responses would give judgments that agreed more with any one metric over another, a post hoc binomial test finds significance for P over K, P over S, and K over S responses made in less than half the average time. This is an encouraging extension of Purchase's finding that layout symmetry improves response time for graph-theoretic tasks [PCJ97, Pur97]. Given the preattentive nature of symmetry perception and the Gestalt separation between automatic and intellectual/mechanical processes, it is worth speculating that the measure of symmetry we seek would agree more with a rapid judgment by human subjects than their careful efforts to examine individual nodes and edges. This suggests a future study in which subjects are only able to see layouts for a few seconds before making a judgment, to filter out the (potentially less meaningful) judgments based on point-to-point assessments.

7. Conclusion

In this study, we found that in cases where the Klapaukh and Purchase measures strongly disagreed on the rankings of pairs of layouts by their symmetry, human judgment agreed more often with the Purchase metric. We could not find evidence to support our suspicion that stress would outperform both P and K , even though they sometimes give scores that are clearly wrong, due to their dependence on parameters that seem difficult to tune even for a very restricted class of graphs. We found some evidence that humans will rank a layout with mirror symmetry as more symmetric than a nearly identical layout with a different orientation, and some evidence that if one layout is overwhelmingly considered to be more symmetric than another, it will be chosen more rapidly. We provide access to the survey, the data we collected, and several additional visualizations online at `http://www.gdsym.xyz/supplemental`.

Our work suggests several avenues for further inquiry. With regard to symmetry detection, it would be advantageous to design a measure of graph layout symmetry that incorporates a loss or energy

function so that layouts with small differences would have similar scores, rather than differing sharply due to thresholding or quantization effects. It would also be valuable to find a way to combine the detection of matching nodes with the detection of matching edges, perhaps by deriving composite features from layouts. It is interesting to consider the possibility of a feature that captures some notion of "shape" that could be derived from the primitives of a graph layout, as a major tenet of the Gestalt school is that we perceive forms or shapes as entities that are more than the sum of their parts. Pulling back from such a layout feature to a related graph feature could open new possibilities for graph layout algorithms with forces and constraints that operate on a higher level than individual nodes and edges.

We intend to pursue this direction by seeking connections from the field of computer vision, where significant effort has been invested in deriving features from the smallest primitives. It seems likely that verification of these and other ideas related to symmetry would be aided by human subject studies in which a time limit forces participants to rely exclusively on their preattentive judgments of symmetry.

References

[ATT] ATT: graph drawing data. http://www.graphdrawing.org/data/. Accessed: 2016-10-05. 2, 5

[Bir32] BIRKHOFF G.: *Aesthetic Measure*. Cambridge, MA: Harvard University Press, 1932. 2

[BJ03] BUCHHEIM C., JÜNGER M.: An integer programming approach to fuzzy symmetry detection. In *International Symposium on Graph Drawing* (2003), Springer, pp. 166–177. 2

[CY02] CHUANG M.-C., YEN H.-C.: On nearly symmetric drawings of graphs. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on* (2002), IEEE, pp. 489–494. 2

[DF99] DE FRAYSSEIX H.: An heuristic for graph symmetry detection. In *International Symposium on Graph Drawing* (1999), Springer, pp. 276–285. 2

[DLF*09] DWYER T., LEE B., FISHER D., QUINN K. I., ISENBERG P., ROBERTSON G., NORTH C.: A comparison of user-generated and automatic graph layouts. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6 (2009), 961–968. 4, 7

[EL00] EADES P., LIN X.: Spring algorithms and symmetry. *Theoretical Computer Science 240*, 2 (2000), 379–405. 2, 3, 4, 5, 7

[FLP07] FOX M., LONG D., PORTEOUS J.: Discovering near symmetry in graphs. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE* (2007), vol. 22, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, p. 415. 2

[FR91] FRUCHTERMAN T. M., REINGOLD E. M.: Graph drawing by force-directed placement. *Software: Practice and experience 21*, 11 (1991), 1129–1164. 5

[GHN13] GANSNER E. R., HU Y., NORTH S.: A maxent-stress model for graph layout. *Visualization and Computer Graphics, IEEE Transactions on 19*, 6 (2013), 927–940. 4

[GIA13] GIANNOULI V.: Visual symmetry perception. *ENCEPHALOS 50* (2013), 31–42. 1, 2, 7

[GKN05] GANSNER E., KOREN Y., NORTH S.: Graph drawing by stress majorization. In *Graph Drawing*, Pach J., (Ed.), vol. 3383 of *LNCS*. Springer, 2005, pp. 239–250. 3

[HBHG] HAN J., BROMBERGER S., HOFFIMAN J., GREENWELL A.: Juliagraphs/graphplot.jl: Graph visualization for julia. https://github.com/JuliaGraphs/GraphPlot.jl. Accessed: 2016-12-12. 5

[HE05] HONG S.-H., EADES P.: Symmetric graph drawing. *Handbook of Graph Drawing and Visualization, Discrete Mathematics and Its Applications* (2005). 2

[KÇ09] KOREN Y., ÇIVRIL A.: The binary stress model for graph drawing. In *Graph Drawing*, Tollis I., Patrignani M., (Eds.), vol. 5417 of *LNCS*. Springer, 2009, pp. 193–205. 3

[KDMW16] KIEFFER S., DWYER T., MARRIOTT K., WYBROW M.: HOLA: human-like orthogonal network layout. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2016), 349–358. URL: http://dx.doi.org/10.1109/TVCG.2015.2467451, doi:10.1109/TVCG.2015.2467451. 2

[Kla14] KLAPAUKH R.: *An Empirical Evaluation of Force-Directed Graph Layout*. PhD thesis, Victoria University of Wellington, 2014. 3

[KPS14] KOBOUROV S. G., PUPYREV S., SAKET B.: Are crossings important for drawing large graphs? In *Graph Drawing* (2014), Springer, pp. 234–245. 4

[LE06] LOY G., EKLUNDH J.-O.: Detecting symmetry and symmetric constellations of features. In *European Conference on Computer Vision* (2006), Springer, pp. 508–521. 3

[LHKG10] LIU Y., HEL-OR H., KAPLAN C. S., GOOL L. J. V.: Computational symmetry in computer vision and computer graphics. *Foundations and Trends in Computer Graphics and Vision 5*, 1-2 (2010), 1–195. URL: http://dx.doi.org/10.1561/0600000008, doi:10.1561/0600000008. 2

[LNS85] LIPTON R. J., NORTH S. C., SANDBERG J. S.: A method for drawing graphs. In *1st ACM Symposium on Computational geometry (SoCG)* (1985), ACM, pp. 153–160. 2

[MPWC13] MITRA N. J., PAULY M., WAND M., CEYLAN D.: Symmetry in 3d geometry: Extraction and applications. *Comput. Graph. Forum 32*, 6 (2013), 1–23. URL: http://dx.doi.org/10.1111/cgf.12010, doi:10.1111/cgf.12010. 2

[MPWG12] MARRIOTT K., PURCHASE H., WYBROW M., GONCU C.: Memorability of visual features in network diagrams. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2477–2485. 1, 2

[PCA02] PURCHASE H. C., CARRINGTON D., ALLDER J.-A.: Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering 7*, 3 (2002), 233–255. URL: http://dx.doi.org/10.1023/A:1016344215610, doi:10.1023/A:1016344215610. 2

[PCJ95] PURCHASE H. C., COHEN R. F., JAMES M. I.: Validating graph drawing aesthetics. In *Graph Drawing, Symposium on Graph Drawing, GD '95, Passau, Germany, September 20-22, 1995, Proceedings* (1995), Brandenburg F., (Ed.), vol. 1027 of *Lecture Notes in Computer Science*, Springer, pp. 435–446. URL: http://dx.doi.org/10.1007/BFb0021827, doi:10.1007/BFb0021827. 2

[PCJ97] PURCHASE H. C., COHEN R. F., JAMES M. I.: An experimental study of the basis for graph drawing algorithms. *J. Exp. Algorithmics 2* (Jan. 1997). URL: http://doi.acm.org/10.1145/264216.264222, doi:10.1145/264216.264222. 2, 9

[PPP12] PURCHASE H. C., PILCHER C., PLIMMER B.: Graph drawing aesthetics created by users, not algorithms. *IEEE Transactions on Visualization and Computer Graphics 18*, 1 (Jan 2012), 81–92. doi:10.1109/TVCG.2010.269. 5

[Pur97] PURCHASE H.: Which aesthetic has the greatest effect on human understanding? In *International Symposium on Graph Drawing* (1997), Springer, pp. 248–261. 1, 2, 9

[Pur02] PURCHASE H. C.: Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing 13*, 5 (2002), 501–516. 3, 5

[Tam07] TAMASSIA R.: *Handbook of Graph Drawing and Visualization (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC, 2007. 2

[Tre85] TREISMAN A.: Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing 31*, 2 (1985), 156 – 177. URL: http://www.sciencedirect.com/science/article/pii/S0734189X85800049, doi:http://dx.doi.org/10.1016/S0734-189X(85)80004-9. 2

[Tre10] TREDER M. S.: Behind the looking-glass: A review on human symmetry perception. *Symmetry 2*, 3 (2010), 1510–1543. URL: http://www.mdpi.com/2073-8994/2/3/1510, doi:10.3390/sym2031510. 1, 2, 7

[vHR08] VAN HAM F., ROGOWITZ B.: Perceptual organization in user-generated graph layouts. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1333–1339. 5

[War12] WARE C.: Information visualization: perception for design, 2012. 1

[Wel] WELCH E.: Supplemental material for measuring symmetry in drawings of graphs. http://www.gdsym.xyz/supplemental/. Accessed: 2017-03-10. 6, 7, 8

[Wes99] WESTHEIMER G.: Gestalt theory reconfigured: Max wertheimer's anticipation of recent developments in visual neuroscience. *Perception 28*, 1 (1999), 5–15. 1, 2

[WR44] WERTHEIMER M., RIEZLER K.: Gestalt theory. *Social Research 11*, 1 (1944), 78–99. URL: http://www.jstor.org/stable/40982002. 1

[YCHZ12] YUAN X., CHE L., HU Y., ZHANG X.: Intelligent graph layout using many users' input. *IEEE transactions on visualization and computer graphics 18*, 12 (2012), 2699–2708. 2