

Measuring Symmetry in Drawings of Graphs

Submission # 314

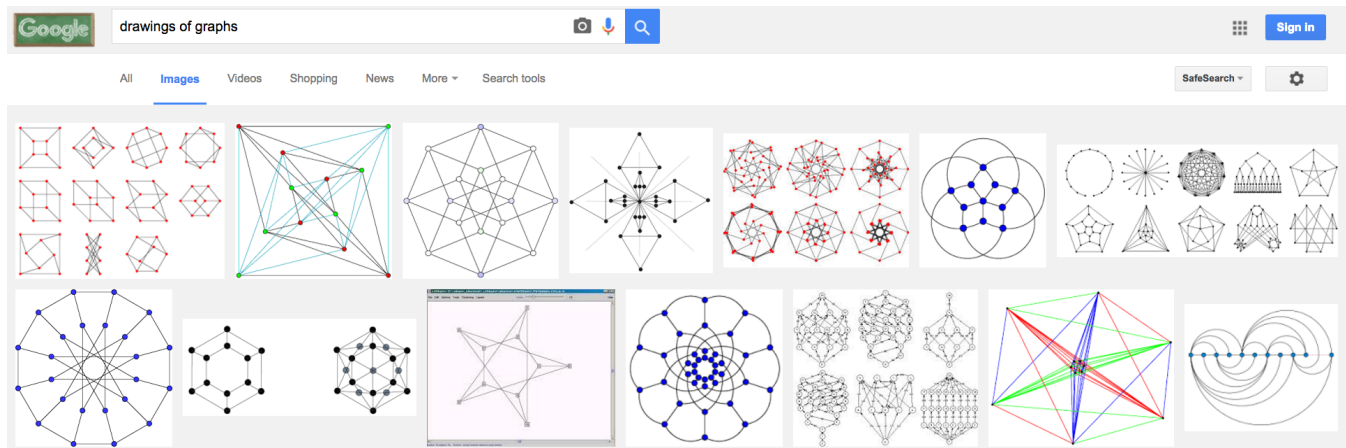


Figure 1: The top images returned by Google-Image-Search for “drawings of graphs.” Note that the overwhelming majority are graphs that are “symmetric” and also drawn in such a way as to highlight the underlying symmetries.

Abstract

Layout symmetry is one of the important and desired features in drawing graphs. 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. One explanation of the importance of symmetry comes from the Gestalt theory of human perception. Very generally, it posits that when we see a group of objects, we perceive their entirety before we see each object individually. Even when the objects are separate entities, our cognitive system tries to make sense of them by grouping them into a whole. As symmetrical objects possess many similar components, they are easily grouped together and perceived as a unit by our cognitive system. They form patterns that can be easily distinguished from the background. The recognition of patterns gives us a feeling of order and meaning.

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, more engaging, and more memorable visualizations of relational data.

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 im-

ages [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 the empirical study by Kieffer et al. [KDMW16], the 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 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 (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 and a discussion of the advantages and disadvantages of each measure, we consider ways in which they may be improved, as well as several 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, Gianoulis [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
- mirror symmetry (reflective symmetry across a vertical axis) is detected faster than rotational symmetry or reflections across other axes
- symmetry, followed by complexity and familiarity, is the “major determining factor in aesthetic judgments”
- 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
- symmetry perception appears at 4 months of age, and preference for mirror symmetry appears across cultures at 12 months

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.

Graph Symmetry: Note the important distinction between detecting symmetries in a graph and that of detecting symmetries in a layout of a graph. In graph theoretic terms, finding symmetries in a graph is known as computing automorphisms and the automorphism group of the given graph. Layout methods based on computing automorphisms go at least 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 by 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.”

Layout Graph Symmetry: There has been little work on drawing a given graph in a way that maximizes the symmetry in the drawing. 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 then used to lay out the resulting subgraph, and 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.

Measures of 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]

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. A symmetric subgraph, consisting of edges that are mirrored within a predefined *tolerance*, is computed for every axis. The value for a given symmetry axis is determined in part by the convex hull area of the subgraph, and the final symmetry score is a ratio involving the sums of the scores for all nontrivial axes. Note that this measure is dependent on the precise locations of graph vertices, and is designed to accommodate multiple axes of reflective symmetry. The computational complexity is $O(n^7)$, where n is the number of vertices in a graph.

To overcome some of the limitations of the earlier measure, Klapakouh [Kla14] suggests an edge-based measure that explicitly calculates reflection, rotation and translation symmetries. Edges are

encoded as SIFT features (a computer vision approach) and used to generate potential symmetry axes. To determine the best N axis of symmetry, the axes are quantized and those that are sufficiently similar are combined. Each axis votes for itself, using the sum of the “quality scores” of all the axes that were combined into it. The final symmetry score is a normalized sum, over all N axes, of the number of edges that vote for each axis. (Note that in all examples, [Kla14] uses $N = 1$) This approach is also computationally expensive.

3. Analysis of Symmetry Measures

Purchase Measure

An interesting feature of the Purchase measure is that it begins by “promoting” edge crossings to nodes, with the assumption that humans will make judgments based on the visual behavior of these artifacts of the layout process. As mentioned, a potential axis of symmetry is generated between every pair of graph vertices, including the promoted nodes. All edges that are reflected across an axis, as determined by the correspondence of endpoints, are combined to create a symmetric subgraph. The subgraph’s score is determined in part by the area of the subgraph, and in part by whether the vertices mapped onto one another are of the same type or not. If a true node is mapped onto a promoted node, a fractional score is given to both nodes (0.5 is used in the examples in [Pur02]), and if two nodes of the same type are mapped onto each other, they have a score of 1. The value of an edge is the product of the scores of its nodes, and the score of the subgraph is the average of the scores of its subgraphs, each weighted by its convex hull area, divided by either the convex hull area of the entire graph, or the sums of the areas of all the subgraphs, whichever is larger.

An alarming 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 human is likely to perceive, much less taken into consideration. (Fig 2)

3.1. Klapaukh Measure

Klapaukh [Kla14] suggested that Purchase’s symmetry detection measure was limited in that it focused on vertices, ignoring edges 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], Klapaukh measures reflective, rotational, and translational symmetry, and considers one layout to be more symmetric than another if it has higher scores in two of these measures. A theoretical problem with this judgment emerges immediately, in that it is not transitive: consider three layouts with 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, as there is no natural way to combine these three measures, and our preliminary investigations suggested that most layouts generated by a force-directed algorithm do not exhibit much translational or rotational symmetry, and larger values of these measures do not seem to correspond to a sense of greater symmetry.

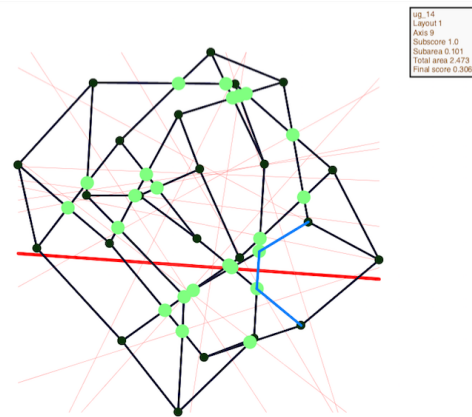


Figure 2: (a, b) All the axes (thin red lines) and a particular axis (thick red line) and its corresponding subgraph (blue edges). It is unlikely that a human considering the symmetry of a layout would consider all of the axes generated;

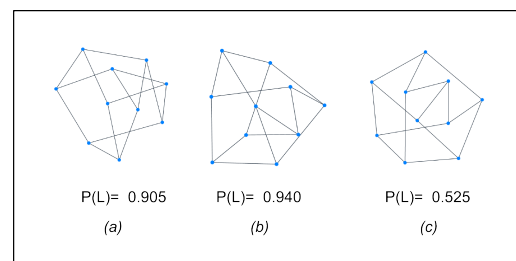


Figure 3: (a, b) Two layouts of a graph that received much higher scores using Purchase’s metric for reflective symmetry than (c), a layout that a human would likely judge as more symmetric. (d) A layout with the same structure as (c) that received a perfect score. It is likely that (c) would receive a perfect score if a larger tolerance value were chosen.

Like the Purchase algorithm, Klapaukh’s reflective symmetry detection algorithm also constructs a set of axes of symmetry; however, instead of creating an axis between every pair of nodes, it constructs two axes through every edge, and one axis between every pair of edges. Specifically, for each edge, there is an axis that passes through this edge and parallel to it, and another that is a perpendicular bisector of the edge. For each pair of edges, the constructed axis is the perpendicular bisector of the line segment connecting the centers of the two edges. Each symmetric axis is assigned a *quality score*, the product of a *scale quality score* S and an *orientation quality score* Φ . For an axis created from a single edge, the axis has a quality score of 1, since the axis maps the edge exactly onto itself. For an axis created from a pair of edges the scale quality score falls in the range $(0, 1]$, where its value is 1 if the edges are of the same length, and decreases monotonically as the difference between lengths of the two edges increases. The orientation quality score is similarly 1 if the two edges are parallel, and de-

creases monotonically as their orientations (angle they make with an arbitrary axis) diverge.

After these scores are calculated, a round of voting occurs in which each edge votes for itself with a weight determined by its quality score. The axes, which are represented as a pair of real numbers indicating their orientation and their distance to the origin, are *quantized* according to these values, and merged into composite axes that are given the votes of all their constituents. The intuition is that, as with Purchase’s tolerance, human perception is stable against small perturbations, as is clear from our ability to detect symmetry in faces in spite of their distortion due to perspective. The N axes with the most votes are used for the calculation of the final score: for each axis the number of edges that voted for an axis (with any pair of quality scores) is divided by the total number of edges, and the average score for these N axes is taken as the final score. In the examples used in [Kla14] a value of $N = 1$ is used, and this is also the default value for N in the source code that accompanies the paper; as such, we have used the same value for our preliminary investigations.

3.2. Discussion of “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 function, called “stress”, is defined as the variance of edge lengths in the drawing [GKN05, KÇ09]. Assume a graph $G = (V, E)$ is drawn with p_i being 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

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

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; in addition it is necessary because sfdp does not utilize edge length at all.

Several earlier works claim 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] show that users prefer graph layouts with lower stress. Finally, Kobourov et al. [KPS14] find a correlation between stress and the number of crossings in a drawing, and thus, stress minimization is a desirable criteria.

3.3. 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, eg, 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 account for issues of scaling with the Purchase metric by selecting a tolerance that is a ratio of the entire width of the drawing, eg, 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. One might expect that the measure would be monotonic in the scale of the graph, as features that encode the axes of symmetry grow farther apart, and fewer are binned together in the quantization step, resulting in a final axis that receives fewer votes. In fact, the Klapaukh measure is *not* monotonic in the scale size. (Fig 7). This is likely a consequence of the quantization step, which somewhat arbitrarily places axes into bins in the manner of a histogram. As a simple example, one can imagine four parallel axes that have as features their minimum distance to the origin; perhaps at a scale of 10, the distances are as 2, 2.9, 3.1, and 3.2. If the size of the bins is constant at 3, then as the scale increases continuously, the maximum number of axes binned together will vary from 2 at scale 10, (2 and 2.9 vs 3.1 and 3.2) to 4 at scale 15 (3.4, 3.5, 4.65 and 4.80), and down to 3 at scale 18.8 (3.76, 5.452, 5.828 vs 6.016).

As a consequence of this fragility, the Klapaukh algorithm, which seems to be motivated by reasonable notions of human perception of symmetry (e.g., we look at a single axis of symmetry, and our perception will collapse very similar axes of symmetry into a single axis if they do not vary much) gives measures that would appear to disagree with human judgment (Fig 5)

Stress also varies with scale, however, as we vary the scale, we find that the stress measure achieves a global minimum (Fig 7) Thus, we can speak of a robust measure of stress, which we find by using a binary search to locate the scale at which the stress is minimized, and use this measure as the stress for the layout at any scale. For the rest of the paper, whenever we speak of the stress of a layout, we refer to this minimized stress.

3.4. Stability vs instability to parameter values

As mentioned, the Purchase metric judges an edge to be mapped onto another edge across a reflective axis if their corresponding endpoints are reflected to points near one another with a specified tolerance. An inappropriate tolerance value can cause two layouts that appear to exhibit similar amounts of symmetry receive sharply different scores. (Fig 6). 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. However, no rationale is given for this choice, and it is not clear *a priori* what an appropriate value for this parameter should be. Moreover, our preliminary investigations suggest that a tolerance value that seems reasonable for one layout can for other layouts result in judgments that would likely differ from those of most humans (Fig 3). As we vary this parameter over a continuous range, at

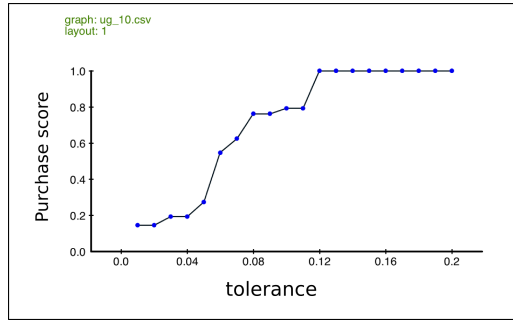


Figure 4: The Purchase score is heavily dependent upon the selection of a tolerance parameter. It is not clear a priori how to set this parameter

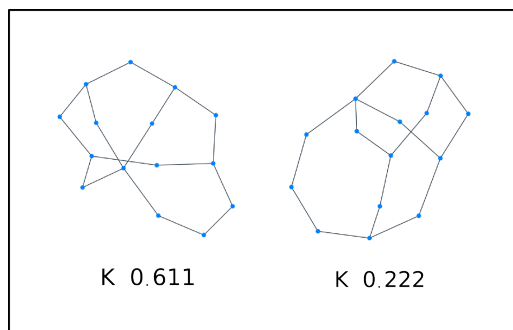


Figure 5: Two layouts of a graph and their reflective symmetry scores according to the Klapaukh metric. Although the layout on the left received a higher score, we would expect human judgment to select the layout on the right as more symmetric.

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. (Fig 4)

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

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 such measures for creating more symmetric drawings. With this in mind, we designed an experiment to measure agreement between human judgment and the symmetry scores assigned by different algorithmic measures.

4.1. Pilot Studies

To gain a sense of how human judgment might compare with the notions of symmetry underlying the Purchase and Klapaukh measures, we conducted two rounds of pilot studies. In the first, we generated 10 layouts of a single graph, which displayed as much

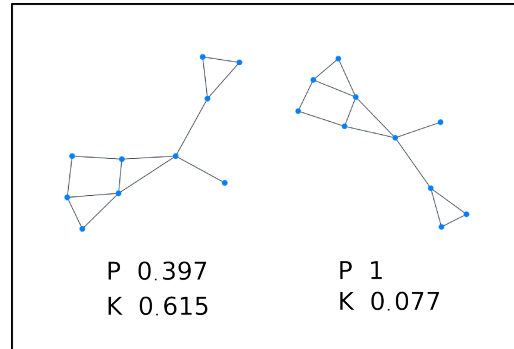


Figure 6: Two layouts of a graph and their reflective symmetry scores according to the Purchase and Klapaukh metrics. Although the layouts seem very similar, they received very different scores by each metric.

reflectional, rotational, or translational symmetries as possible. The participants were presented with a series of pairs of layouts, in two rounds. In the first round, participants were asked which layout was most aesthetically pleasing. In the second, they were asked which looked most symmetric. Based on [GIA13], we expected to find that aesthetics would agree with judgments of symmetry. We also expected that layouts with many isometries (e.g., invariant under rotations of $\pi/4$, $\pi/2$ and $3\pi/4$ radians) would be considered more symmetric than layouts with a single isometry. Informal discussions with the small number of colleagues who took part in this study suggested that reflectional symmetry was more important than the rest. Another observation was that some participants consider node placements, while others consider the edges, and still others look at larger shapes and regions. Finally, the participants suggested that a “gut feeling” might be different than that from a careful study of the layout that considers different types of symmetries and their relative importance. This seemed to agree with the distinction between automatic and point-to-point symmetry detection described in [GIA13].

In a second pilot study, we selected 25 pairs of layouts of graphs with large differences in scores by the Purchase and Klapaukh measures. The 23 participants were asked which of each pair seemed more symmetric. Hoping to find that one measure would strongly outperform the other, we took the consensus vote of all participants as a measure of human judgment for each layout pair. There was no clear winner, although analyzing the individual layouts led us to observation that the measures suffer from a lack of stability and a lack of robustness, leading us to consider these issues carefully in our formal study.

Our pilot study also featured an exit survey consisting of questions about what factors seemed most important in judging symmetry: nodes that matched, edges that matched, equal edge lengths, matching shapes (polygons), larger matching regions, reflective symmetry, rotational symmetry, translational symmetry, or a “gut feeling” about which measure was most symmetric. Our participants told us that these verbal descriptions were confusing and seemed to overlap, leading us to develop small illustrations to accompany these questions in the final study.

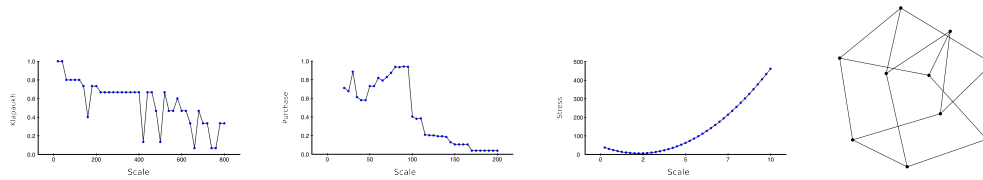


Figure 7: The two measures of symmetry proposed in the literature respond erratically to scale, while stress behaves predictably. Using a binary search strategy, we can compute a scale-invariant measure of stress for a layout.

One unexpected observation in the second pilot study was that for two of the graphs, participants were unanimous in their preferences for one of the layouts, and also responded in an average time of less than one half the mean time over all layouts (2.265 seconds and 2.453 seconds, compared to an average of 5.370 seconds). This led us to revisit the idea of automatic symmetry detection in our formal study.

4.2. Experimental Design

We designed custom application software that guides participants through the experiment, providing task instructions and collecting the 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, 21 layout pairs are presented in a random order, with the layouts randomly assigned to be placed on the right or the left. 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. As in the second 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) were when they made their decisions. We also collected basic demographic and background information (e.g. age, gender, mathematical background).

4.3. Graph Selection

Our graphs come 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. In order to keep issues related to scale and tolerance as simple as possible, we used only small and sparse graphs; see details in Table 4.3 below.

We confirmed that the three metrics do not generally agree on the level of symmetry exhibited in the layouts of graphs by evaluating the correlation between each pair of metrics on the 100 layouts of our graphs. We found no meaningful correlation between any pair of the three metrics for the majority of the graphs.

4.4. Drawing Selection

We used a force-directed layout in our study, as every initial random placement of the vertices of the same graph leads to a different layout. We thus generated 100 layouts of each graph using the `spring_layout` function in the `LightGraphs` library for Julia

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 of each metric for the 100 layouts generated for the graph.

[Bro]. To select the specific layouts for the pilot study, we created a custom tool that provided a visual comparison of the symmetry scores of the layouts. We used this tool to select pairs of layouts for graphs that disagreed sharply according to Klapaukh’s reflective symmetry metric and the Purchase metric. We found 25 such pairs.

For the formal study, we 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. We evaluated the symmetry of all the layouts of all 25 graphs for a range of parameters, each time taking their average and standard deviation. As expected, each measure was monotonic in the corresponding parameter. As previously mentioned, 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. (0.385 tolerance for Purchase, where all graphs were scaled to a maximum width or height of 2; 680 scaling for Klapaukh). Using these values, for each graph and for each experiment type, defined by a pair of measures (Purchase and Klapaukh, Purchase and stress, and Klapaukh and stress), 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.

4.5. Hypotheses

On the basis of our own experiences and the informal feedback from our pilot study, we propose the following hypotheses:

- H1: When presented with layouts that differ in their judgments by Klapaukh and Purchase measures, human subjects will agree 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.

4.6. Participants

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

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 ran similar tests for several subgroups. We also examined participants' responses to the questions about which features of a layout were most important in their judgements. We found support for H1, as human subjects preferred Purchase metric judgments over Klapaukh metric judgments at the ($p = 0.01$) level. We did not find support for H2. (Table 2)

To evaluate the third hypothesis, we examined the pairs of layouts with statistically significant differences in human preference, to observe if differences in orientation may in some cases explain the outcome. We find some evidence in support of H3, but we have not validated it statistically.

For the fourth hypothesis, we conducted a subgroup analysis with a binomial test as above on responses given faster than the average time. We also 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.

We finally examined the layouts used in the study to get a sense of why the Klapaukh and Purchase measures may have made judgments that varied sharply from those of the human subjects.

5.1. Voting

Examining the votes that each measure received for each experiment type, we found a significant preference for Purchase over Klapaukh ($p < 0.01$) and for Klapaukh over stress ($p < 0.01$). However, we did not find a significant preference for Purchase over stress. We considered subgroups based on levels of mathematical background, taking users reporting background above and below the mean (2.6), and above and below values one standard deviation away from the mean. We also examined the subgroups of subjects with greater and less than the average reported enjoyment of mathematics, and those taking more or less than the average amount of time per question. We finally considered subgroups of votes themselves based on whether the votes were made in less than or more than the average time. See Table 2).

Cross-section	n	P/K	P/S	K/S
All	30	125/85**	114/96	134/76**
MathBG > 3.5	4	16/12	22/6**	19/9
MathBG > 2.6	17	71/48*	69/60*	81/38**
MathBG < 2.6	11	44/33	39/38	43/24
MathBG < 1.7	4	13/15	12/16	13/15
MathEnj > 3.3	14	56/42	59/39*	66/32**
MathEnj < 3.3	16	69/43**	55/57	68/44**
AvgTime < 6.3	17	68/51*	62/57	79/40**
AvgTime > 6.3	13	57/34	52/39	55/36
TimePerQn < 6.3	-	73/55*	81/59*	108/49**
TimePerQn > 6.3	-	54/35*	38/39	29/31

Table 2: The number of votes in favor of the judgments of one metric over another, using seven pairs of layouts for each experiment type. A single star * indicates significance at the $p = 0.05$ level. A double star ** indicates significance at the $p = 0.01$ level. Of particular interest are the strong levels of significance for responses that were given in less the average time.

5.2. Survey

Our exit survey confirmed that reflection is a more salient feature than rotational or translational symmetry. It also showed that a "gut feeling" was more important than other factors in deciding which layout is more symmetric. Corresponding edges were more important than shapes, or regions. (Table 5.2.)

5.3. Correlation Analysis

We found no significant correlation between the number of votes a layout received and the average time that it took to decide on that layout, both when we considered the entire data set and when we subsampled the data to only include responses made in less than the

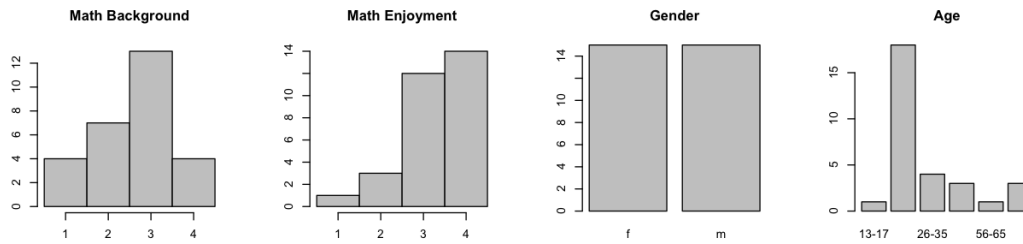


Figure 8: Demographic characteristics of study participants

Consideration	Importance
reflection	3.866667
gut feeling	3.433333
matching edges	3.2
matching polygons	3.137931
matching nodes	2.896552
larger matching regions	2.964286
rotation	2.633333
similar edge lengths	2.517241
translation	2.3

Table 3: Mean values of participants’ responses (1-4) to the question “How important were each of the following when you made decisions about symmetry?”

average time (< 6.34 seconds). However, we did find that layouts that received more than 90% of the votes against their pair layout had average times less than one half of the average time.

5.4. Examination of Layouts

We ordered the layout pairs according to the number of votes given to the drawing that received the human consensus judgment and attempt to describe here the reasons why the “losing” metric may have misjudged the level of symmetry in the layout. All the layouts, their scores by each metric, scatterplots and correlations of their scores by each pair of metrics, and number of votes given to each layout can be found online at <http://www.gdsym.xyz/supplemental/>.

Klapaukh In some cases, we suspect that the Klapaukh measure gave a score that was too low due to a problem surrounding the quantization step (Fig 9). There was also a case in which Klapaukh seems to overestimate the amount of symmetry in the drawing because the final score is given as a ratio of the number of edges that vote for an axis of symmetry, ignoring the differences in orientation between edges that have been mirrored (Fig 10).

Purchase We found one case in which Purchase missed what appears to be an obvious symmetry but still gained the majority of the votes (perhaps due to the presence of a vertical axis of symmetry, see Fig 11), and 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 are presumably

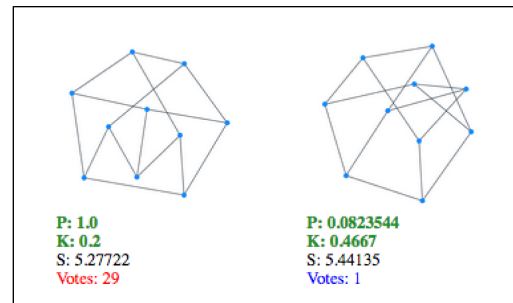


Figure 9: An example where the Klapaukh measure likely failed to identify the highly symmetric layout (left) due to a quantization error.

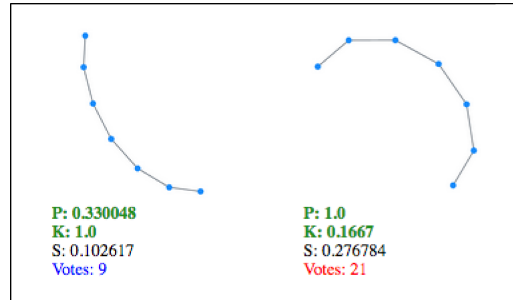


Figure 10: An example where the Klapaukh measure seems to have overestimated the amount of symmetry in the layout (left), failing to penalize a pair of edges for differences in orientation.

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.

5.5. Human Preference for Vertical Axes of Symmetry

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. (Fig 12) 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.

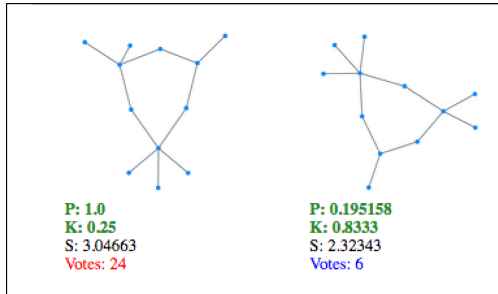


Figure 11: A survey item in which the Purchase measure provides very different judgements of two layouts, seeming to miss a strong symmetry on the right due to oversensitivity to tolerance, but won the human consensus strongly. We suspect this is the result of the vertical orientation of the axis of symmetry on the left.

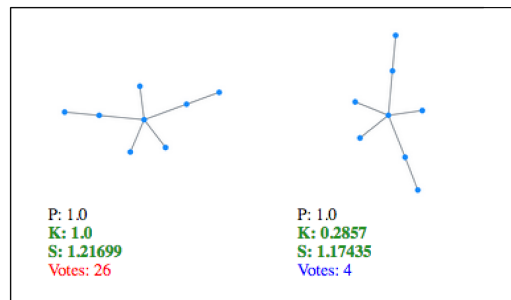


Figure 12: An example of a pair of layouts for which the amount of symmetry does not seem to differ, but for which humans demonstrated a preference for a vertical axis of symmetry.

6. Discussion

It is surprising to find that overall human preference for the judgments of Purchase metric over the Klapaukh metric, as well as the preference for the Klapaukh metric over stress were statistically significant, while there was no statistical significance for a preference for the Purchase metric over stress. When we consider the fact that a significant preference of Purchase over stress was found for the subgroup with very high enjoyment of mathematics, and the selection of responses that took place rapidly, there is plausible support for a claim that the Purchase metric captures some subtle aspect of symmetry that is easily missed by a point-by-point analysis, but may be detected automatically, or through the playful mental manipulation of someone genuinely engaged in the symmetry task. The strong preference for a “gut feeling” of whether a layout is symmetric reinforces the idea that preattentive processes are at work in our perception of symmetry. Moreover, the loss of statistically significant preference for Klapaukh over stress when we look at the subset of responses taking longer than average time suggests that the slow, point-to-point evaluation of symmetry may yield haphazard responses.

As noted in our examination of the layouts, Purchase’s and Klapaukh’s algorithms sometimes fail to find significant symmetries due to fragility to scale and instability to changes in a tolerance

parameter. While the choices for tolerance and scale used for this survey yielded results that agreed with human judgment in some cases, a more meaningful method for selecting these parameters is needed if they are to perform reliably.

Our exit survey results support the dominance of reflective symmetry over rotational and translational symmetry. This makes sense given that humans perceive reflectional symmetry more rapidly than the other forms, and force-directed layouts tend to generate layouts exhibit reflective symmetry rather than the other two. We also found some evidence that humans prefer mirror symmetry to reflective symmetry across other axes.

We provide access to the survey, the data we collected, and several additional visualizations at <http://www.gdsym.xyz/supplemental>

Limitations The layouts were chosen so as to maximize the differences in judgment between metrics. A larger study that compared a greater number of layouts, selected at random may find more significant results. The study also did not consider graphs with more than 25 edges.

7. Conclusions

The Purchase metric, as published, seems to outperform the Klapaukh measure, and succeeds at detecting significant symmetries in a greater number of cases. However, it is unclear whether this difference is due to an intrinsic feature of the algorithm in theory, or simply an artifact of the tolerance parameter being a better fit for the particular layouts in this study than the scaling parameter for the Klapaukh measure (or the particular thresholding parameters used in the quantization step). Therefore, one direction for future work is to find ways to tune these parameters to achieve the “correct” performance, in which the calculated values reflect the values that human subjects would give if directed to follow the steps, matching nodes on the basis of whether they are judged to map onto one another.

The results of our study would be of interest when designing efficient layout methods that optimize symmetries. Other than the stress measure, which can be optimized efficiently, the algorithms of Purchase and Klapaukh have high computational complexity and would be impractical even for graphs with under 100 vertices. Knowing which symmetric features are most useful, one could design simpler algorithms that optimize a suitable subset of features.

Our study suggests that humans pay more attention to reflective symmetry and consider a layout with vertical axis of symmetry “more symmetric” than an apparently identical layout with a different orientation. An obvious implication for layout algorithms, in the service of human consumption, is that as a final step of a layout algorithm, the main axis of symmetry can be inferred, perhaps borrowing techniques from Purchase’s or Klapaukh’s algorithms, and the layout rotated so that this axis is aligned vertically. This also suggests a caveat for future evaluation studies: to avoid the possibility that one layout will be chosen over another on the basis of orientation, the layouts should be presented with the main axis aligned vertically, or presented multiple times at random orientations.

References

- [ATT] ATT: graph drawing data. <http://www.graphdrawing.org/data/>. Accessed: 2016-10-05. 2, 6
- [Bir32] BIRKHOFF G.: *Aesthetic Measure*. Cambridge, MA: Harvard University Press, 1932. 1
- [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
- [Bro] BROMBERGER S.: Juliagraphs/lightgraphs.jl: An optimized graphs package for julia. <https://github.com/JuliaGraphs/LightGraphs.jl>. Accessed: 2016-12-12. 6
- [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
- [EL00] EADES P., LIN X.: Spring algorithms and symmetry. *Theoretical Computer Science* 240, 2 (2000), 379–405. 2, 4
- [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
- [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. 2, 5
- [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. 4
- [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. 4
- [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] KLAPPAUKH R.: *An Empirical Evaluation of Force-Directed Graph Layout*. PhD thesis, Victoria University of Wellington, 2014. 2, 3, 4
- [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/10.1561/06000000008>, doi:10.1561/06000000008. 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/cgfm.12010>, doi:10.1111/cgfm.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. 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
- [Pur02] PURCHASE H. C.: Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing* 13, 5 (2002), 501–516. 2, 3, 4
- [Tam07] TAMASSIA R.: *Handbook of Graph Drawing and Visualization (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC, 2007. 2
- [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