

Stress in Graph Drawings: Perception, Preference, and Performance

Gavin J. Mooney   

Monash University, Melbourne, Australia

Jacob Miller  

Technical University of Munich, Heilbronn, Germany

Michael Wybrow  

Monash University, Melbourne, Australia

Stephen Kobourov  

Technical University of Munich, Heilbronn, Germany

Helen C. Purchase  

Monash University, Melbourne, Australia

Abstract

Stress in a graph drawing has been a popular layout principle for more than two decades. Low stress drawings exhibit the property that the geometric distances between all pairs of nodes correlate with the shortest paths between them. The assumption has always been that low stress drawings are “nicer” and better support human perception and comprehension than high stress drawings. In this paper, we put these assumptions to the test. We use a normalised scale-independent and rotation-independent metric for stress; this is necessary to ensure strict controls on our experimental stimuli. We report on three experiments, exploring human perception of stress, preference for stress, and the effect of stress on a graph performance task. We conclude that people can see stress in a graph drawing, that they prefer low stress drawings, and that their performance in a shortest path task improves as stress decreases – thus empirically confirming long-standing assumptions.

2012 ACM Subject Classification Human-centered computing → Graph drawings

Keywords and phrases Graph Drawing, Graph Drawing Metrics, Stress, Visual Perception, User Study

Digital Object Identifier 10.4230/LIPIcs.GD.2025.38

1 Introduction

Since its introduction in the 1980s [12], minimising stress has remained one of the most popular layout criteria in graph drawing with implementations in GraphViz [10] and NetworkX [19]. It is among the most frequent quality metrics that researchers use to evaluate layout algorithms, being the third most used metric behind run time and edge crossings [7]. Although stress is believed among experts to reflect aesthetic, symmetric, and general high quality layouts [27], validating stress as a metric has only been considered in a handful of limited studies [5, 18]. The assumption is that humans can perceive stress, prefer low stress, and understand low stress drawings better than high stress ones. This paper reports on empirical studies that address these assumptions.

While the fundamental principle of measuring stress relates to the extent to which the geometric distances between pairs of nodes correlate with the graph theoretic distances between them, there are several different definitions of stress. Many formulae for stress depend on the scale of the drawing [24] – that is, since the physical size of a drawing determines geometric distances, it also determines the value of the drawing’s stress. These



© Gavin J. Mooney, Jacob Miller, Michael Wybrow, Stephen Kobourov, and Helen C. Purchase; licensed under Creative Commons License CC-BY 4.0

33rd International Symposium on Graph Drawing and Network Visualization (GD 2025).

Editors: Vida Dujmović and Fabrizio Montecchiani; Article No. 38; pp. 38:1–38:23

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

ambiguities in the definition of stress means that it is difficult to compare the stress of two graph drawings of different size, and not easy to create a unique graph drawing with given stress.

To investigate the human response to varying stress in graph drawings, we use a scale-independent and rotation-independent normalised metric for stress [18]. Having a robust metric like this enables rigorous formal experiments to be conducted, since it provides a means by which graph drawings of different size and scale can be quantitatively compared for the extent of stress within them.

Based on this stress metric, a large set of graph drawings with different stress values served as stimuli for our three human empirical studies. The first experiment addresses the question of whether people can see stress in a graph drawing or not, gathering qualitative interview data that validates the results of prior quantitative work [18]. We find that while participants can recognise differences in stress, they have difficulty describing what it is. Since these “perception” results may have been due to the preferences of the participants for low stress drawings, our second experiment investigated preference, asking participants to identify which of a several pairs of drawings they most preferred in a two-alternative-forced-choice experiment. We found that they did, indeed, prefer the drawing with lower stress, but that this was not the same as comparing two drawings and identifying the one with least stress. The final experiment is arguably the most important, since it focusses on actual task performance, asking participants to complete shortest path tasks. As we had hoped, the results reveal a correlation between task accuracy and stress: the lower the stress, the higher the accuracy.

We have therefore considered stress in graph drawings from three human perspectives: perception, preference and performance.

2 Background: Stress in Graph Drawings

The concept of stress in graph drawing is based on the principle that any two vertices should have a geometric distance proportional to their graph theoretic distance. While there are several variants for the definition of stress, the one proposed by Gansner et al. [10] is most popular (Equation 1).

$$\sum_{i < j} d_{i,j}^{-2} (||X_i - X_j|| - d_{i,j})^2 \quad (1)$$

where X_i is the coordinate position of vertex i and $d_{i,j}$ is the graph theoretic distance between vertices i and j . Gansner et al. [10] note that the sum of square differences over-penalises large differences and under-penalises small distances. Hence, the inclusion of the $d_{i,j}^{-2}$ “normalisation” term, weighting distant pairs proportionally to an inverse-square law; thus this definition is known as *normalised stress*. This equation has roots in statistical analysis and dimension reduction, where it is known as Multi-dimensional Scaling (MDS) or Sammon mapping [14, 22].

Stress was first introduced to graph drawing as an objective function, that is, something to be minimised to produce a desirable drawing, and several variants of Equation 1 have been proposed as objective functions in various algorithms. Gansner et al. [9] and Ortmann et al. [20] use approximation in order to avoid the expensive all-pairs-shortest-path computation that is required for normalised stress. Chen and Buja [4] and Miller et al. [16] modify the function to measure only “local” (small) distances.

When optimising stress, Ahmed et al. [1] and Devkota et al. [6] use stress-based optimisation schemes alongside other aesthetic criteria, and stress is used as an evaluation metric by Hong et al. [11] and Marmer et al. [15], as well in dynamic graph layout methods from Simonetto et al. [23] and Arleo et al. [2]. Kruiger et al. [13] and Zhu et al. [29] evaluate the stress of their layouts to determine how much “global” distances are lost.

When investigating stress, Brandes and Pich [3] evaluate layout algorithms by stress and Welch and Kobourov [27] use stress as an alternative measure of symmetry in graph layouts. Wang et al. [26] both optimise and evaluate stress. Wageningen et al. [25] extend stress as an evaluation for 3-dimensional drawings, by considering each orthographic viewpoint.

2.1 Scale-Independent Stress

Normalised stress (Equation 1) is susceptible to changes in the scale (size) of a drawing [24]: that is the same drawing will have different stress values when drawn at different sizes. While this may not be a concern when the aim of a layout algorithm is to minimise the stress in a drawing, it means that the measured stress value of a drawing is objectively meaningless outside of the specific implementation. In particular, it means that it is impossible to meaningfully compare the stress values of two graph drawings.

If we are to empirically investigate the effect of stress on human perception, preference and performance, we need a stress metric that is scale-invariant and bounded. There are three main contenders. Scale-normalised stress has been deployed in recent evaluations [25, 26], where the size of the drawing is chosen which yields the smallest normalised stress, and the Shepard goodness score measures the rank correlation between the graph-theoretic and geometric distances between vertices.

The option we have chosen is the non-metric stress of Kruskal [14], which has several desirable properties as a metric. Unlike normalised stress which is unbounded and scale-sensitive, non-metric stress lies in the range $[0, 1]$ and is scale-invariant. The two stress functions share the same overall complexity to compute (quadratic, once all-pairs shortest-path lengths are precomputed), and the difference in runtime is also small in practice. Mooney et al. [18] show that this Kruskal’s stress metric (*KSM*) is highly correlated with Gansner et al.’s [10] normalised stress.

$$KSM = 1 - \sqrt{\frac{\sum_{i < j} (||X_i - X_j|| - \hat{d}_{i,j})^2}{\sum_{i < j} ||X_i - X_j||^2}} \quad (2)$$

where X_i is the position of vertex i , and $\hat{d}_{i,j}$ is the horizontal distance to the monotonic regression line of best fit in the Shepard diagram of the drawing. A more detailed description is given in [24]. Note that, consistent with other literature on normalised evaluation metrics (e.g., [17, 18, 21]), a value of 1 represents the assumed “best” case (i.e., zero stress). Therefore, high KSM values indicate low stress.

There are few empirical studies that compare the effectiveness of stress. Computational experiments have shown that the low stress of a drawing tends to correspond to exhibiting symmetries in the graph when they exist [27]. A small-scale experiment indicates that people prefer less stress and fewer crossings [5]. However, in this experiment, it is unclear the level of difference of stress values between drawings, or even what variant of stress was deployed (stated “scaled by average edge length”). The measure of stress used relies on normalising by average edge lengths, which may invalidate comparisons between drawings where the average edge length (or scale of the overall drawing) are different. Our use of the scale-invariant and bounded KSM allows for meaningful stress comparison between graph drawings, ensuring that valid experimental stimuli can be created.

Other studies consider the length of shortest paths and Euclidean distances without referring to stress directly. Yoghoudjian et al. [28] evaluate shortest path task difficulty with respect to features of the shortest path, including a stress variant called “geodesic path deviation”. However, they find that this feature of the shortest path does not influence task difficulty as much as several other layout parameters that relate to crossings (e.g. number of edge crossings, angle of crossings, node-edge crossings).

2.2 Experimental Stimuli

Our experiments require a set of graph drawings with known stress values; a basic hill climbing algorithm was used to optimise the Kruskal Stress Metric (KSM) described in Section 2.1. Drawings were generated with distinct KSM values from 0.4 to 0.8, incrementing by 0.05. This gives nine unique stress values and nine unique differences in stress values between pairs of drawings, which we refer to as “deltas”. For example, for a pair of drawings with KSM of 0.45 and 0.55, the delta would be 0.1, thus the deltas range from 0 to 0.4. In total, 405 graph drawings were created, from 15 randomly generated graphs at three sizes (10, 25, 50 nodes). Mooney et al. [18] describe the process for creating these drawings in more detail. It is important to note that KSM assigns a value of 1 to represent the ideal case (zero stress); thus “low stress” has a high KSM value; “high stress” has a low KSM value.

This set of drawings is used in all three experiments described below. Example stimuli are shown in Figure 1.

3 Stress Perception Experiment

In this section we present results of an extended version of an experiment that investigated whether humans can perceive stress in a graph drawing [18]. The stress perception experiment validates and extends the work of Mooney et al. [18], with the inclusion of more participants and some interviews.

3.1 Methodology

Three cohorts – trained novices (25), untrained novices (25), and experts (9) – were asked to determine which of a pair of graph drawings had the lowest stress (or if they had the same stress) in an online survey offered through the “Qualtrics” platform. Three graph sizes were shown to participants, resulting in a total of 177 participant data points. Accuracy, response time, and confidence data were collected. Participants were also asked to describe the strategies that they used to determine stress in the drawings. In-person interviews with five novice participants aimed to gain a deeper understanding of participants’ strategies when comparing stress. Interview participants were also asked to explain what “stress” is; we wanted to see if participants really understood the theoretical notion of “stress”. The methodology is otherwise the same as Mooney et al. [18].

3.2 Summary of results

The results reinforced those of Mooney et al. [18], indicating that even untrained novices can identify stress, although less successfully than trained novices or experts. Surprisingly, the size of the graph (number of nodes) did not alter these results. Participants devised (visible) proxies for (invisible) stress – for example, the length of edges, the distances between nodes, extent of edge crossings, and node distribution, using words like “compactness”, “clustering”,

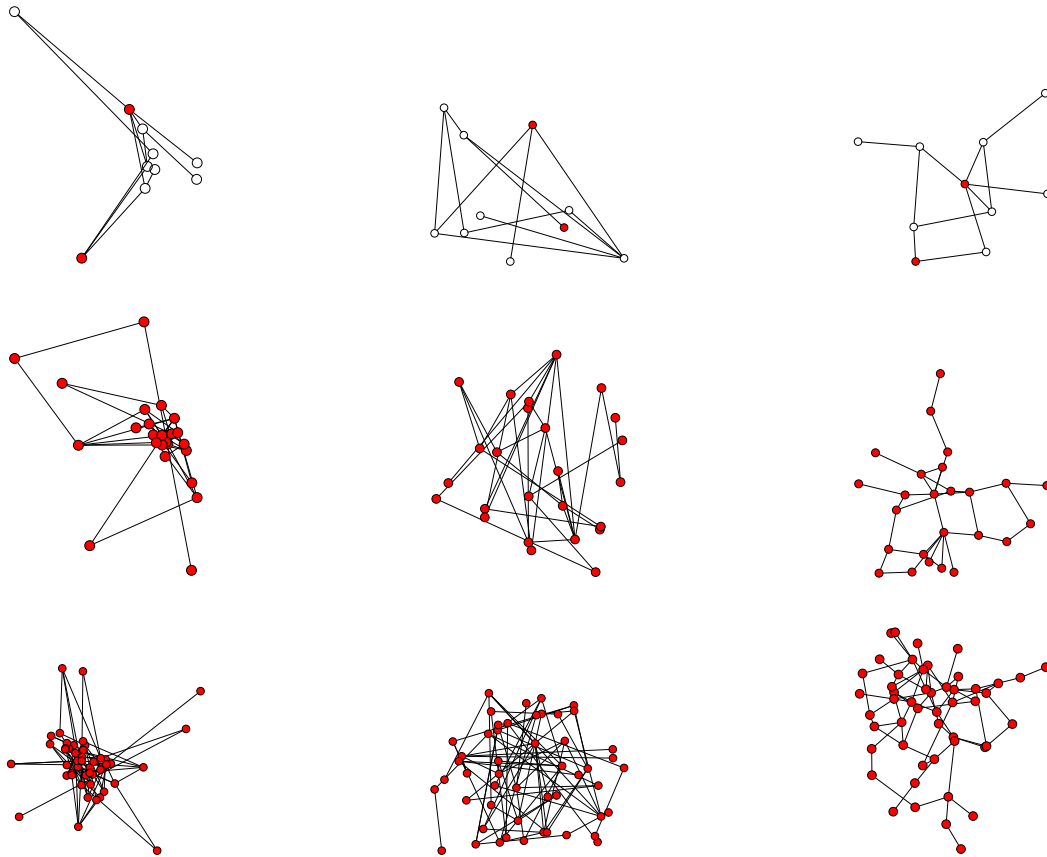


Figure 1 Example stimuli used in the experiments. The top row shows three drawings of the same 10-node graph with different highlighted end points for shortest paths at KSM values of 0.4, 0.6, and 0.8 (left to right), used in the *Shortest Path Experiment*. The middle and bottom rows show drawings of the same 25-node and 50-node graphs, respectively, at KSM values of 0.4, 0.6, and 0.8 (left to right), used in the *Perception* and *Preference* experiments.

“density” – we call these “*Stress Identification Proxies*”. Only one of the interview participants gave a reasonable definition of stress (“it’s the way the objects are positioned that matches the length of paths.”). We conclude that while it is unclear whether participants utilised the exact geometric definition of stress their performance suggests that they were able to perceive it (“I know it when I see it”).

4 Stress Preference Experiment

While the results of the Stress Perception Experiment in the previous section suggest that participants can correctly identify the difference between high and low stress drawings, it is not clear that they can “see” stress in the literal sense. Perhaps they simply chose the drawing that they preferred?

In this section, we describe a two-alternative forced choice experiment that simply asks participants to choose which of two graph drawings they prefer. We use the same stimuli from Mooney et al. [18]: 405 graph drawings with KSM stress values ranging from 0.4 to 0.8, in 0.05 increments (Section 2.2), using graphs with 10, 25 and 50 nodes and 25 participants per graph size.

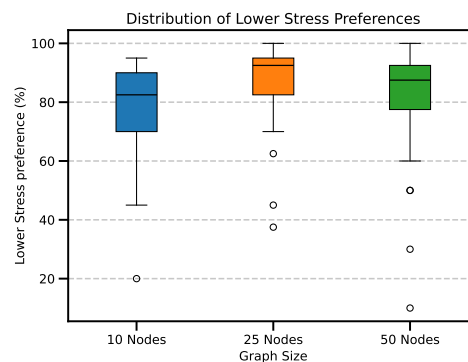
4.1 Methodology

Participants were shown 45 pairs of drawings and asked to choose their preferred drawing (left or right). We represent these choices as a binary value: 1, if the participant selected the drawing with lower stress (to match the accuracy measure for the stress perception experiment), and 0 if they chose the drawing with higher stress. To be consistent with the prior perception experiment, the trials where stress is the same in both drawings were shown to participants but the responses discarded from analysis.

Participants in these experiments were given a layperson’s description of networks, but no specific information on paths or stress. At the end of the experiment they were asked demographic questions (including their familiarity with “stress in a graph drawing”) and asked “Please describe in your own words which visual aspects of the drawings affected your choice.”

4.2 Results

75 people participated in the Stress Preference Experiment (25 per graph size). Figure 2 shows the distribution of preferences for each participant, grouped by graph size. If stress did not affect preference choice, we would expect these distributions to approximate normal distributions centred around a mean of 50%, since each data point is the percentage of times a participant chose the lower stress drawing, and 50% represents a random choice. The high median values ($>80\%$) and skewed distributions show a tendency for preference of lower stress drawings. A Wilcoxon signed-rank test for each graph size (comparing observed distributions against the median of 50% expected for a random choice) gave p-values <0.001 for all graph sizes. Thus, each distribution in Figure 2 differs significantly from random chance and we conclude stress affects preference choices.



■ **Figure 2** Box plot showing the distribution of preferences for each participant. Each data point represents the percentage of times the participant preferred the lower stress drawing. Skewed distributions show a preference for low stress drawings for all three graph sizes (10, 25, 50).

Figure 3 presents a heatmap showing the percentage of times each individual drawing was preferred, in relation to the number of times it was shown: the “preference ratio”. This shows that low stress (high KSM) drawings were chosen more frequently than high stress (low KSM) drawings. The outliers can be explained by the fact that drawings were chosen randomly and so some were only shown a few times, and with small differences in KSM values between the pairs.

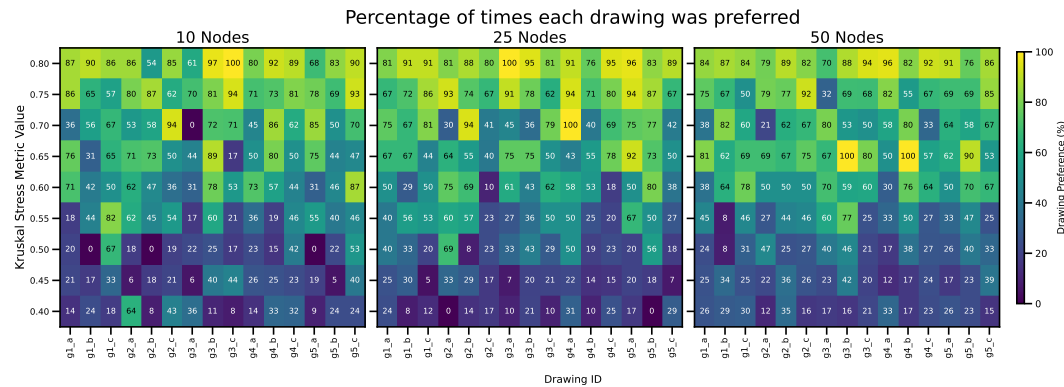


Figure 3 Heatmap showing the percentage of times each drawing was selected as the preferred choice, in relation to the number of times shown.

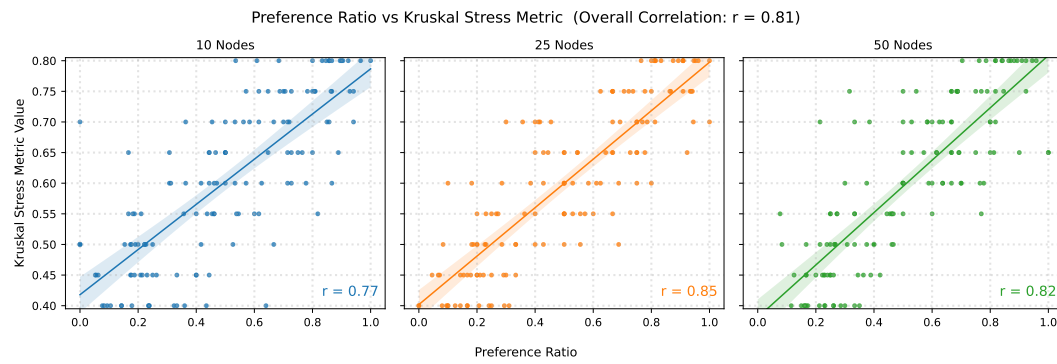


Figure 4 The Preference ratio of each drawing plotted against its KSM value. The linear regression lines have confidence intervals shown by shading. The trends show a preference for low stress (high KSM) drawings across all three graph sizes (10, 25, 50).

We plot the KSM values against the preference ratios over all graph sizes in Figure 4, revealing that the preference ratio of a drawing tends to increase as KSM increases (regardless of graph size). The Pearson correlation between preference and KSM (over all graph sizes) is 0.81.

Participants mentioned factors influencing their preference, including: more space between nodes (39), overall shape/symmetry (7), space between edges (9), fewer edge crossings (21), and less clutter/mess (10). Some responses can be interpreted as direct references to existing graph drawing metrics: “I disliked overlapping edges, and preferred evenly spaced nodes.” (Edge Crossings, Node Uniformity); while others were less specific: “How messy it appeared to my brain”. One participant mentioned multidimensional scaling directly: “I found myself preferring the pattern that looks more multidimensional. Multidimensional scaling appears more structured and provides more information about the connections between the nodes and the distances among them.”

From the participants’ responses, we identify five *Stress Preference Proxies*: Angular Resolution, Edge Lengths, Node Uniformity, Edge Crossings, and Gabriel Ratio. The correlation between these five proxies (measured with metrics defined by Mooney et al. [17]) and the KSM stress metric over all 405 stimuli are: Angular Resolution (0.61), Edge Lengths (0.78), Node Uniformity (0.69), Edge crossings (0.66) and Gabriel Ratio (0.84), suggesting

that people prefer low stress drawings. Although their decision is based on identifiable visual proxies, underlying these proxies is the unseen layout principle of “stress”. The distributions of metric values for all 405 drawings in our dataset are in Appendix B.

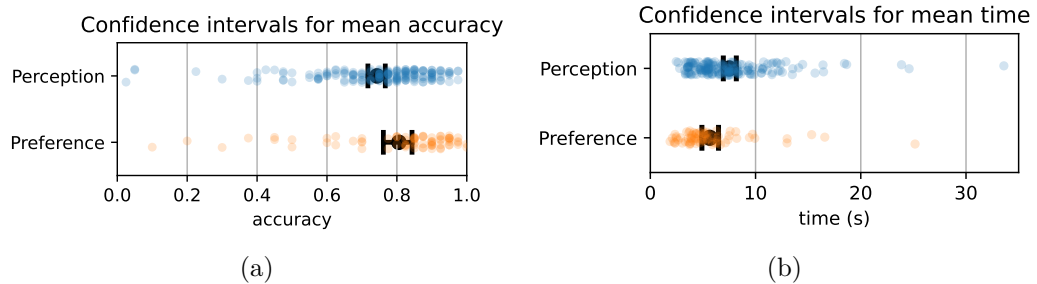
4.3 Perception vs Preference

One of our motivations for following the Stress Perception Experiment with the Stress Preference Experiment was to determine whether our results of the former were simply based on participants’ preference for lower stress drawings (rather than on perception of stress). To investigate this, we compare the 177 data points from the Perception Experiment with the 75 from the Preference Experiment.

4.3.1 Results

Figure 5 (a) shows the distribution of mean accuracy across the aggregated groups in both experiments, with bootstrapped 95% confidence intervals. We compare the distributions for similarity (excluding same-stress pairs). Note that while we use the term “accuracy” there are no objectively “correct” answers in the preference experiment. Here, accuracy refers to whether the participant selected the drawing with lower stress.

A Wilcoxon rank sum test between the Perception data and the Preference data gives a p-value of 0.0189 (less than 0.05) indicating a significant difference, albeit with a small effect size. Thus, getting the correct answer in the Perception experiment is not the same as preferring lower stress.



■ **Figure 5** Bootstrapped 95% confidence intervals for (a) mean accuracy and (b) mean response time in the Perception and Preference experiments, aggregated over all graph sizes.

Figure 5 (b) shows the distribution of mean response time across the aggregated groups in both experiments, with bootstrapped 95% confidence intervals. We compare the distributions for similarity (excluding same-stress pairs). A Wilcoxon rank sum test between the Perception data and the Preference data gives a p-value of 2.6×10^{-6} (less than 0.05) indicating a significant difference. Thus, the cognitive effort required for selecting the preferred drawing (regardless of accuracy) was less than that required for comparing stress.

4.3.2 Discussion

From the Perception Experiment, we conclude that participants can ‘see’ stress, and can identify the drawing with less stress within a pair (even if they can’t explain what ‘stress’ actually means). From the Preference Experiment, it is clear that participants prefer drawings with lower stress. Our comparative analysis suggests, however, that when asked to choose the drawing with least stress (in the earlier experiment), participants were not simply choosing

the one they most preferred – they really were considering their perception of stress. This is shown both by the differences in the choice of the least stress drawing (Figure 5 (a)) as well in the difference in the time taken to make the decision (Figure 5 (b)). The difference in response time shows that the perception task took longer than the preference task. This result suggests that the task in the Perception Experiment had higher cognitive load (with participants thinking carefully about stress in terms of graph structure and relative distances) while the Preference task was enacted more intuitively, possibly based on immediate perception rather than cognitive analysis.

We note that, while the difference in means between the two experiments is quite small, it is still statistically significant. While we conclude that perceiving stress is distinct from preferring lower stress, one is certainly a very strong predictor of the other.

5 Stress Performance

Assessing whether humans can see stress and what stress levels they prefer is all very well, but task performance is what really matters. If we are to use graphs to depict information in a meaningful way, so that it can be accurately used and understood, then stress perception and preference is insufficient: we need to know how stress affects human's ability to read the information embodied in the graph. This Stress Performance experiment uses a shortest path task to measure participant's ability to interpret the graph.

5.1 Methodology

Participants in this study, conducted via the new reVISit framework [8], were shown a series of drawings of graphs, which had two vertices highlighted in red – all other vertices were transparent to reduce uncertainty when following edges which intersect nodes. They were then asked, "What is the shortest path between the two highlighted nodes?" The possible answers available were 2, 3, 4 or 5. Participants were also given the option "I am unable to work this out".

The stimuli were sourced from the perception and preference experiments described earlier. We use a subset of drawings (243 of the original 405) to ensure an equal number of responses for each drawing. The highlighted nodes were based on shortest paths randomly selected through repeated sampling to ensure an even distribution of path lengths (2, 3, 4, or 5) across KSM values and graph sizes. Due to the structure of the 10-node graphs, their corresponding stimuli only included shortest paths up to length 4.

Participants were given a short description about graphs, graph drawings, and shortest paths. These descriptions are made available through supplemental material (Appendix A). Participants were then informed they would be shown 6 training examples, that we discard from analysis. These examples provided feedback on whether the participant answered the task correctly, and were the same for each participant.

We used a within-subjects study design, where each participant sees all conditions during the study. In this case, each participant saw one example of each stress level (9) for every size of graph (3). Each size was broken into its own block during the study, so that for a given size, all trials are shown sequentially. After seeing all trials of a size, participants could take a short break. Additionally, each size block included two additional fixed trials at the beginning which were discarded from the analysis to account for learning effects. In total, this amounts to $(9 + 2) \times 3 = 33$ trials per participant, after the training.

The exact stimuli used during the study varied per participant. We partitioned the drawings used from the earlier two studies into nine disjoint groups, and assigned participants a group number based on a Latin square, ensuring the same number of participants in each

group. The order of trials was randomised in two ways: first the size blocks were shown in random order, and second within each block the order of the stimuli was randomised. Together these aimed to reduce learning effect noise in the collected data. Finally, after all 33 trials, participants were asked to provide basic demographic information the same as in the previous two experiments.

5.2 Results & Discussion

Figure 6 (a) plots the mean accuracy against the KSM values over all graph sizes, showing a clear upward slope: accuracy increases as stress decreases (where high KSM indicates low stress). The linear fit line indicates a correlation of 0.97 with $p < 0.001$. Figure 6 (b) shows the results according to graph size: clearly the larger the graph, the harder the task. The lower correlation value for $n=50$ (0.80) compared with $n=10$ (0.94) suggests that the relative effect of stress on performance is diminished for the larger graphs. From this, we might even anticipate that for very large graphs, say $n=1000$, stress has no effect on task performance; this is an interesting avenue for future work.

Figure 7 (a) shows the mean duration of trials for each KSM value. We suggest the following explanations for the shape of this trend: Drawings with high KSM (low stress) made the task easier, and thus participants required less time to work out the answer. Drawings with low KSM (high stress) were more difficult, causing participants to realise quickly that they cannot easily find the correct answer. The middle KSM values indicate longer response times where the drawing made the task neither too easy nor too difficult. This is also supported by Figure 7 (b), which shows that the trials for $n=25$ had the longest duration ($n=10$: too easy; $n=25$: neither too easy nor too difficult; $n=50$: too difficult). Figure 8 (a) further supports this argument, showing that participants were unable to determine the answer for low KSM drawings. Figure 8 (b) shows the number of abandoned trials for each graph size, further highlighting the difficulty of the task on larger graphs.

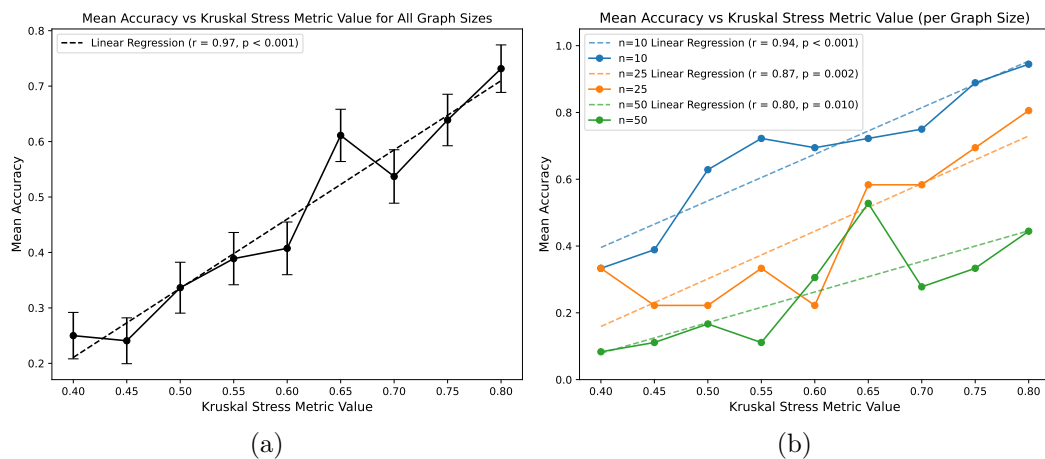


Figure 6 Relationship between mean accuracy and Kruskal stress metric values. (a) Aggregated across all graph sizes. (b) Disaggregated by individual graph sizes. Error bars represent standard error of the mean. Linear regression lines indicate a positive correlation between stress and accuracy for all graph sizes ($r = 0.97$, $p < 0.001$), with strongest correlation observed for smaller graphs ($n = 10$: $r = 0.94$, $p < 0.001$; $n = 25$: $r = 0.87$, $p = 0.002$; $n = 50$: $r = 0.80$, $p = 0.01$).

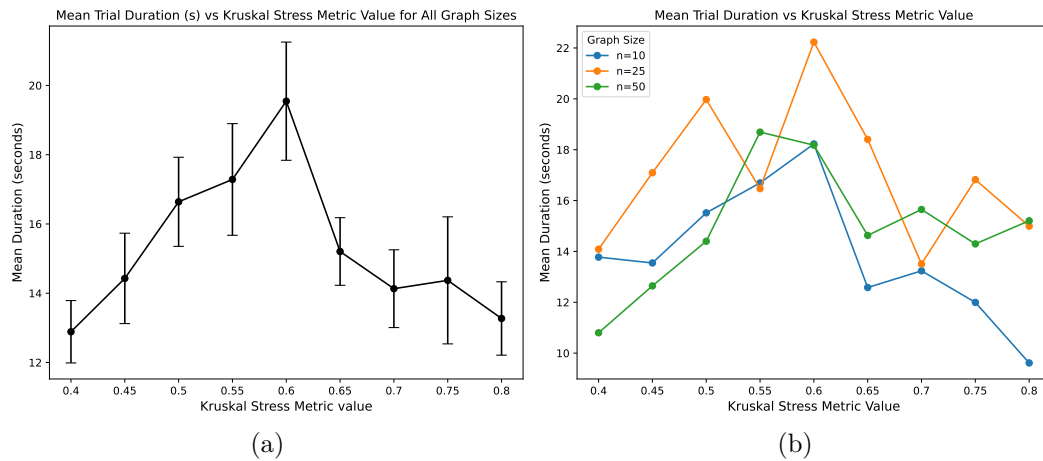


Figure 7 Relationship between mean task duration (in seconds) and Kruskal stress metric values. (a) Aggregated across all graph sizes. (b) Disaggregated by individual graph sizes. Error bars represent standard error of the mean.

5.2.1 Anomalies

There are three noticeable anomalies in Figure 6 (b):

- For $n=25$, $KSM=0.60$, the task seems to have been unusually difficult;
- For $n=50$, $KSM=0.60$, the task seems to have been unusually easy;
- For $n=50$, $KSM=0.65$, the task seems to have been unusually easy.

We investigated these anomalies by analysing the stimuli, considering the accuracy for individual drawings (Figure 9) and features of the drawings that may have made them particularly easy or difficult. For example, we looked at whether the shortest path was unique, whether either the highlighted source or target nodes were of degree 1, or if they overlapped another node, or whether an edge on the path was shorter than the radius of a node (and therefore shown within the intersection of two nodes). Most of these features were revealed to have little effect (see the heatmaps in Appendix C), but some of the data helps explain our three anomalies:

- Eight out of the nine $n=25$, $KSM=0.60$ drawings had edges on the shortest path that were directly on top of another edge (more than any other set of $n=25$ graphs when grouped by KSM). In addition, five of the nine $n=25$, $KSM=0.60$ drawings had a source or target node overlapping an unconnected edge. The combination of these two factors would have made these drawings naturally more difficult than the other $n=25$ ones.
- Fewer $n=50$, $KSM=0.60$ and $n=50$, $KSM=0.65$ drawings (three and four respectively) had source or target nodes overlapping another node (as opposed to five or more for other $n=50$ drawings grouped by KSM); there were similarly fewer drawings where a node on the shortest path overlaps another node. The combination of these two factors would have made these drawings naturally easier than the other $n=50$ ones.

While we focused on the three obvious anomalies in Figure 6 (b), our investigation revealed other sets of drawings that may have been easier than others of the same size. Having source or target nodes on the convex hull might make the task easier; this was the case for drawings of size 25 with $KSM=0.55$. Similarly, a shorter average geodesic node distance (average of the perpendicular distances between nodes on the shortest path and a straight line between the source and target) might also simplify a shortest path task; this was the case for both $n=25$, $KSM=0.55$ and $n=50$, $KSM=0.5$ drawings. These smaller variations can also be seen in the results chart in Figure 6 (a).

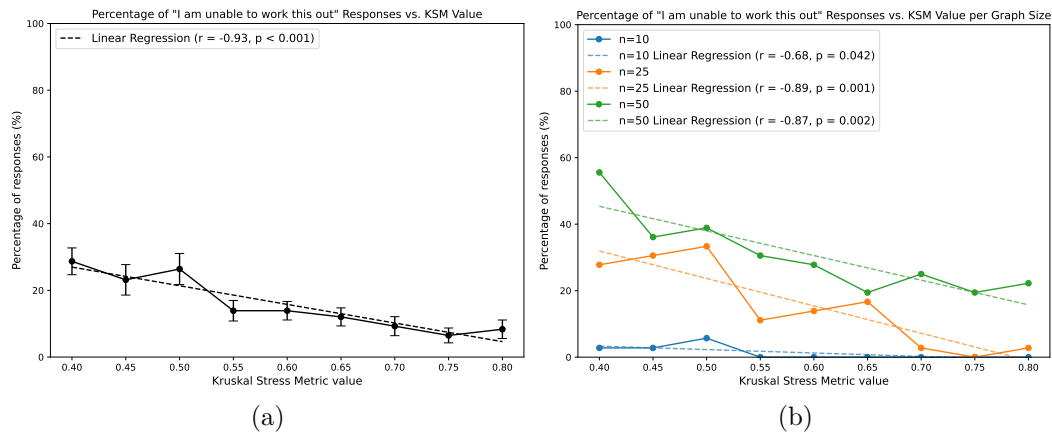


Figure 8 Relationship between the percentage of “I am unable to work this out” responses and Kruskal stress metric values. (a) Aggregated across all graph sizes. (b) Disaggregated by individual graph sizes. Error bars represent standard error of the mean. Linear regression lines indicate a negative correlation between Kruskal stress and task abandonment for all graph sizes ($r = -0.93$, $p < 0.001$).

In exploring these anomalies, one issue became increasingly apparent: *an alteration in stress in a drawing naturally affects other visual features*. If a graph drawing is generated by an algorithm that only cares about the amount of stress, then undesirable visual phenomena will appear as a matter of course: for example, overlapping nodes, overlapping edges, nodes lying on top of edges – even in low stress drawings. An investigation of the (positive) effect of (invisible) stress on performance is also therefore an implicit investigation of the (negative) effect of these other consequent (visible) features.

Our stimuli were the same as those used for the previous Perception and Preference experiments; they were not specifically created for the purposes of this shortest path Performance experiment. The source and target nodes for the shortest paths were chosen randomly, ensuring an even distribution of path lengths across stress values. We could have created specific stimuli for the Performance experiment, and carefully chosen our shortest paths to avoid visual features that may have made the task easier or harder. However this would have resulted in an inauthentic, contrived set of drawings that may have been seen as deliberately formed so that they would support our intuition about the relationship between stress and task performance.

5.2.2 Demographics & Participant Strategies

A total of 36 participants completed the demographic questionnaire. Most participants reported feeling somewhat confident (30) in their responses, with only a few expressing very confident (1) or not very confident ratings (4). The majority found the study difficult (26) or very difficult (5), suggesting a reasonable level of challenge in the task. In terms of prior exposure, most were somewhat familiar (17) or not very familiar (14) with network diagrams, with only one participant indicating strong familiarity. There were 23 males and 13 females, and participants spanned a wide age range, with the largest groups falling between 26–35 (10) and 36–45 (9) years of age.

Of the 36 participants, 29 described the strategy they used to complete the task. As this was an optional, open-ended question, most of the responses were short or essentially restated the task (“I just tried to follow the lines”, “I tried to look for the shortest paths”, “Moving my mouse and counting”).

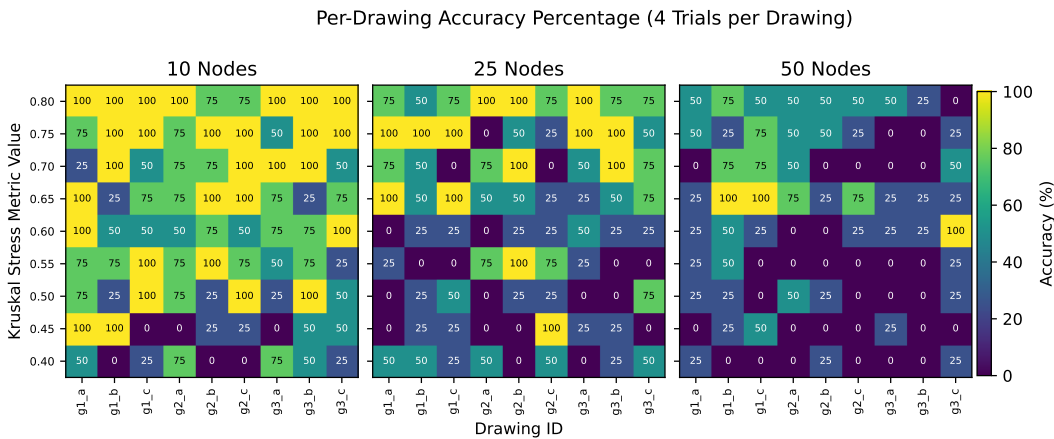


Figure 9 Per-drawing accuracy percentages (based on 4 trials per drawing) for each unique stimulus. Higher accuracy indicates better task performance. Each cell shows the mean accuracy (%) across participants for a specific drawing.

Several participants provided more detailed accounts of their strategies. Some began by scanning the whole diagram to “identify overall patterns” before following the “most intuitive connections.” Others focused on counting steps, such as one who “looked for the shortest visual route between the two red nodes and counted the steps,” or another who “tried to count the number of nodes in the shortest route” but sometimes found the diagrams too complex to work out. Some participants adapted their approach when a path became unclear: one reported they “tried from the other red dot” if their initial attempt failed, while another would abandon a path that took “over 5 hops” and try a different route. One participant highlighted the range of difficulties between trials: “Some tasks were easy to visualise. Some were impossible due to the complexity of lines making it difficult to see which node they terminated on.” These responses indicate that while most participants relied on intuitive tracing or counting, several applied more reflective or adaptive techniques when faced with visual complexity.

6 Conclusions, Limitations and Future Work

Defining a normalised stress metric [24] that permits quantitative comparison between graph drawings of different size, structure and scale has allowed us to investigate human responses to stress in three ways: the perception of stress, preference for stress, and the effect of stress on task performance.

We conclude that people (even untrained novices) can perceive the extent of stress in a graph drawing, often using visible features (for example, “node distribution”) as a proxy for the invisible “stress”, while typically being unable to accurately define what stress means. People generally prefer low stress drawings, but making a preference choice is not the same as identifying the drawing with least stress. Lower stress drawings support higher accuracy in shortest path tasks.

Our conclusions are constrained by the natural limitations of the experimental method. We used three graph sizes for comparative and generalisability purposes, but the maximum graph size is only 50. Future experiments with larger graphs may have different findings. Our intuition is that as the size of the experimental graphs increase, the effect of stress (in all of perception, preference and performance) will decrease.

The graph drawings used as stimuli were generated using an optimisation function based only on the stress metric. Other metrics could conceivably be used, resulting in drawings that avoid negative features like node or edge overlaps. Since our focus was on stress, we did not want to create artificial experimental stimuli: changing stress naturally alters other features and so any (initial) investigation of stress should simply accept the existence of these features. Nevertheless, it would be interesting for future work to explore the effect of these other visual features on perception, preference, and performance.

The Stress Performance experiment is based on only a shortest-path task, and future work should investigate whether the same results hold true for other localised tasks (e.g., number of common neighbours of two nodes) as well as global tasks (e.g., extent of clustering). Our intuition is that, like with the shortest path task, lower stress will improve performance, as long as the graphs are not much bigger than 100 nodes – at which point we expect it will become difficult for change in stress to have any effect on human performance.

Our work represents the first comprehensive empirical investigation into the effect of stress in graph drawing from a human perspective. We have investigated whether people can actually perceive stress, what their preferences are with respect to stress, and whether the amount of stress affects task performance. For many years graph drawing researchers have made assumptions about the human response to stress: this paper validates these assumptions.

References

- 1 Reyhan Ahmed, Felice De Luca, Sabin Devkota, Stephen G. Kobourov, and Mingwei Li. Multicriteria scalable graph drawing via stochastic gradient descent, (SGD)². *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2388–2399, 2022. doi:10.1109/TVCG.2022.3155564.
- 2 Alessio Arleo, Silvia Miksch, and Daniel Archambault. Event-based dynamic graph drawing without the agonizing pain. *Computer Graphics Forum*, 41(6):226–244, 2022. doi:10.1111/cgf.14615.
- 3 Ulrik Brandes and Christian Pich. An experimental study on distance-based graph drawing. In *Graph Drawing*, volume 5417 of *Lecture Notes in Computer Science*, pages 218–229. Springer, 2008. doi:10.1007/978-3-642-00219-9_21.
- 4 Lisha Chen and Andreas Buja. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association*, 104(485):209–219, 2009. doi:10.5555/2567709.2502616.
- 5 Markus Chimani, Patrick Eades, Peter Eades, Seok-Hee Hong, Weidong Huang, Karsten Klein, Michael Marnier, Ross T. Smith, and Bruce H. Thomas. People prefer less stress and fewer crossings. In *Proceedings of International Symposium on Graph Drawing (GD)*, pages 523–524. Springer, 2014.
- 6 Sabin Devkota, Abu Reyhan Ahmed, Felice De Luca, Katherine E. Isaacs, and Stephen G. Kobourov. Stress-Plus-X (SPX) graph layout. In *Graph Drawing and Network Visualization*, volume 11904 of *Lecture Notes in Computer Science*, pages 291–304. Springer, 2019. doi:10.1007/978-3-030-35802-0_23.
- 7 Sara Di Bartolomeo, Tarik Crnovrsanin, David Saffo, Eduardo Puerta, Connor Wilson, and Cody Dunne. Evaluating graph layout algorithms: A systematic review of methods and best practices. In *Computer Graphics Forum*, page e15073. Wiley Online Library, 2024.
- 8 Yiren Ding, Jack Wilburn, Hilson Shrestha, Akim Ndlovu, Kiran Gadhave, Carolina Nobre, Alexander Lex, and Lane Harrison. revisit: Supporting scalable evaluation of interactive visualizations. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pages 31–35. IEEE, 2023. doi:10.1109/VIS54172.2023.00015.

- 9 Emden R. Gansner, Yifan Hu, and Stephen North. A maxent-stress model for graph layout. *IEEE Transactions on Visualization and Computer Graphics*, 19(6):927–940, June 2013. doi:10.1109/tvcg.2012.299.
- 10 Emden R. Gansner, Yehuda Koren, and Stephen C. North. Graph drawing by stress majorization. In *Graph Drawing and Network Visualization*, volume 3383 of *Lecture Notes in Computer Science*, pages 239–250. Springer, 2004. doi:10.1007/978-3-540-31843-9_25.
- 11 Seok-Hee Hong, Peter Eades, Marnijati Torkel, Ziyang Wang, David Chae, Sungpack Hong, Daniel Langerenken, and Hassan Chafi. Multi-level graph drawing using infomap clustering. In *Graph Drawing and Network Visualization*, volume 11904 of *Lecture Notes in Computer Science*, pages 139–146. Springer, 2019. doi:10.1007/978-3-030-35802-0_11.
- 12 Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, 1989. doi:10.1016/0020-0190(89)90102-6.
- 13 Johannes F. Krüger, Paulo E. Rauber, Ricardo M. Martins, Andreas Kerren, Stephen Kobourov, and Alexandru C. Telea. Graph layouts by t-SNE. *Computer Graphics Forum*, 36(3):283–294, 2017. doi:10.1111/cgf.13187.
- 14 Joseph B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
- 15 Michael R. Marner, Ross T. Smith, Bruce H. Thomas, Karsten Klein, Peter Eades, and Seok-Hee Hong. GION: interactively untangling large graphs on wall-sized displays. In *Graph Drawing and Network Visualization*, volume 8871 of *Lecture Notes in Computer Science*, pages 113–124. Springer, 2014. doi:10.1007/978-3-662-45803-7_10.
- 16 Jacob Miller, Vahan Huroyan, and Stephen Kobourov. Balancing between the local and global structures (LGS) in graph embedding. In *International Symposium on Graph Drawing and Network Visualization*, pages 263–279. Springer, 2023. doi:10.1007/978-3-031-49272-3_18.
- 17 Gavin J. Mooney, Helen C. Purchase, Michael Wybrow, and Stephen G. Kobourov. The multi-dimensional landscape of graph drawing metrics. In *17th IEEE Pacific Visualization Conference, PacificVis 2024, Tokyo, Japan, April 23-26, 2024*, pages 122–131. IEEE, 2024. doi:10.1109/PACIFICVIS60374.2024.00022.
- 18 Gavin J. Mooney, Helen C. Purchase, Michael Wybrow, Stephen G. Kobourov, and Jacob Miller. The perception of stress in graph drawings. In *32nd International Symposium on Graph Drawing and Network Visualization (GD 2024)*, pages 21–1. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.GD.2024.21.
- 19 Networkx documentation. https://networkx.org/documentation/stable/reference/generated/networkx.drawing.layout.kamada_kawai_layout.html.
- 20 Mark Ortmann, Mirza Klimenta, and Ulrik Brandes. A sparse stress model. *Journal of Graph Algorithms and Applications*, 21(5):791–821, 2017. doi:10.7155/jgaa.00440.
- 21 Helen C. Purchase. Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing*, 13(5):501–516, 2002. doi:10.1006/jvlc.2002.0232.
- 22 John W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on computers*, 100(5):401–409, 1969. doi:10.1109/T-C.1969.222678.
- 23 Paolo Simonetto, Daniel Archambault, and Stephen G. Kobourov. Drawing dynamic graphs without timeslices. In *Graph Drawing and Network Visualization*, volume 10692 of *Lecture Notes in Computer Science*, pages 394–409. Springer, 2017. doi:10.1007/978-3-319-73915-1_31.
- 24 Kiran Smelser, Jacob Miller, and Stephen Kobourov. “Normalized stress” is not normalized: How to interpret stress correctly. In *2024 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pages 41–50. IEEE, 2024. doi:10.1109/BELIV64461.2024.00010.
- 25 Simon van Wageningen, Tamara Mchedlidze, and Alexandru Telea. An experimental evaluation of viewpoint-based 3D graph drawing. In *Computer Graphics Forum*, page e15077. Wiley Online Library, 2024. doi:10.1111/cgf.15077.

- 26 Xiaoqi Wang, Kevin Yen, Yifan Hu, and Han-Wei Shen. SmartGD: A GAN-Based graph drawing framework for diverse aesthetic goals. *IEEE Transactions on Visualization and Computer Graphics*, 2023. doi:10.1109/TVCG.2023.3306356.
- 27 Eric Welch and Stephen Kobourov. Measuring symmetry in drawings of graphs. *Computer Graphics Forum*, 36(3):341–351, June 2017. doi:10.1111/cgf.13192.
- 28 Vahan Yoghoudjian, Yalong Yang, Tim Dwyer, Lee Lawrence, Michael Wybrow, and Kim Marriott. Scalability of network visualisation from a cognitive load perspective. *IEEE transactions on visualization and computer graphics*, 27(2):1677–1687, 2020. doi:10.1109/TVCG.2020.3030459.
- 29 Minfeng Zhu, Wei Chen, Yuanzhe Hu, Yuxuan Hou, Liangjun Liu, and Kaiyuan Zhang. DRGraph: An efficient graph layout algorithm for large-scale graphs by dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1666–1676, 2021. doi:10.1109/TVCG.2020.3030447.

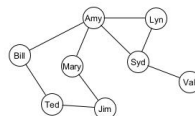
A Training Materials

Participants in the Perception experiment were shown training materials (1), (2) and (3). Participants in the Preference experiment were only shown training material (1). Participants in the Shortest Path experiment were shown training materials (1) and (2).

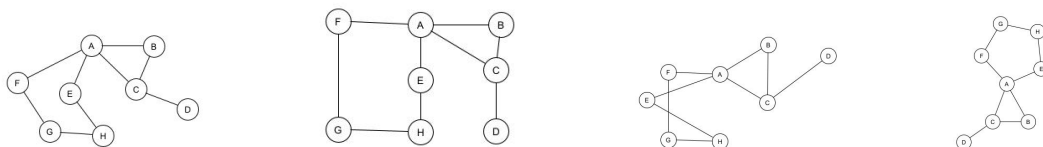
A.1 Training Material (1)

Definitions: networks and network drawings

A network is made up of objects and connections. For example, this social network depicts people (represented as circles) and friendships (represented as lines between the circles). Amy has four friends; Ted has two.

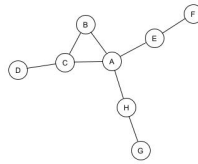


The same network can be drawn in many different ways by changing the position of the objects. For example, here are four drawings of the same network.



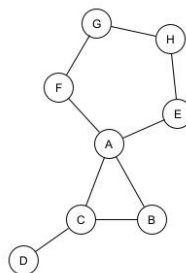
A.2 Training Material (2)

A “path” is a series of steps between objects. For example, in the network below, the length of the path between G and F is 4; the length of the path between A and D is 2 or 3 (depending on whether you go through B or not).



The “shortest path” is, as its name suggests, the shortest path when there is more than one way to get from one object to another.

For example, in the network below, the shortest path between F and B is 2 (going through A, but not C or G/H/E); the shortest path between D and H is 4 (going through C/A, but not through B or F/G).

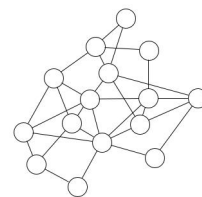
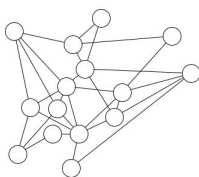


A.3 Training Material (3)

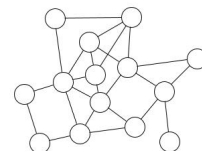
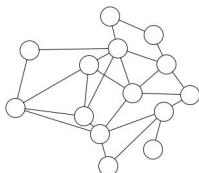
Definitions: visual properties of network drawings

Given that there is more than one way to draw a network, we can distinguish between them by their ‘visual properties’.

For example, the drawing on the left has ‘tighter angles’ than the one on the right. These are both drawings of the same network.



And the drawing on the right has “more symmetry” than the one on the left. These are both drawings of the same network.

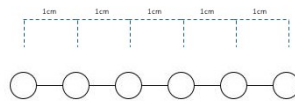


In this experiment, we are interested in the visual property of “stress”.

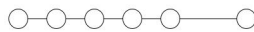
38:18 Stress in Graph Drawings: Perception, Preference, and Performance

A drawing has **low stress** if the distance between pairs of objects is proportional to the length of the shortest path between them.

In its simplest form, the following network drawing has very low stress: the distance between each pair of objects is directly proportional to the (shortest) path between them.



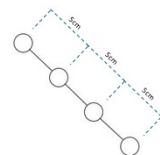
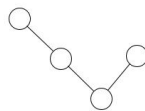
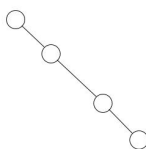
We just need to move one of the objects to **increase the stress** – the distance between the two objects at each end (6cm) is now longer proportional to the length of the path between them (5).



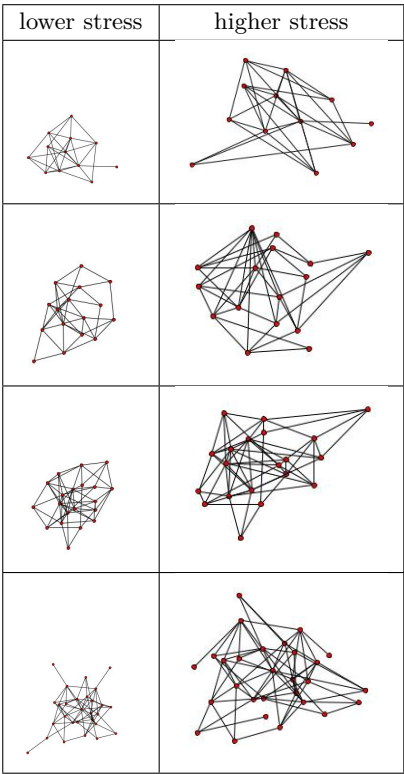
The same simple network can be drawn with even **higher stress**, where there is barely any relationship between the distance between the objects and the length of the paths between them.



Similarly, here is another network with very low stress, with two versions of higher stress.



Of course, this more-or-less-stress judgement becomes more difficult with larger networks. Here are some more examples showing the same networks drawn with different amounts of stress. In all cases, the network on the left has less stress than the network on the right. Thus, the network on the left maintains the distance/path relationship between pairs of objects better than the one on the right.



Of course, we cannot assess the differences in stress by doing all the distance and path length calculations in our head! But we can get a ‘feeling’ as to when one drawing has less stress than another.

Before you start the experiment, we will ask you to make your own judgements, and let you know whether you are correct or not.

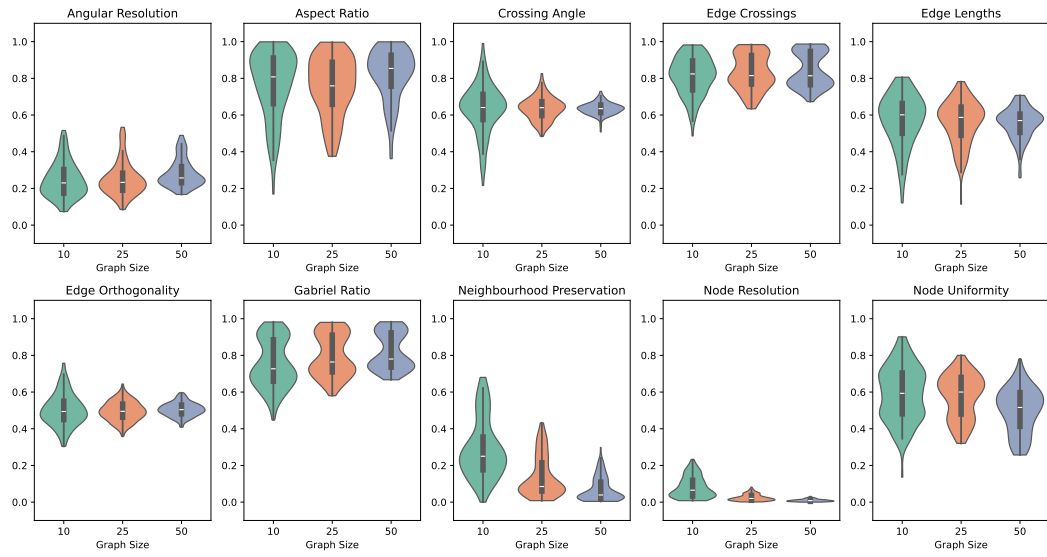
B Stimuli Metric Distributions and Correlations

We plot the distributions of 10 aesthetic metrics (defined by Mooney et al. [17]) for our 405 stimuli drawings (Figure 14). These distributions show that smaller graphs generally have a wider range of possible metric values. This can be explained by the fact that many of the metrics, such as Edge Crossings and Gabriel Ratio, compare a measurable value (e.g., the number of edge crossings) to an estimated maximum value (e.g., the total number of possible edge crossings). The estimated maximum typically grows much faster than the measured values as graph size increases.

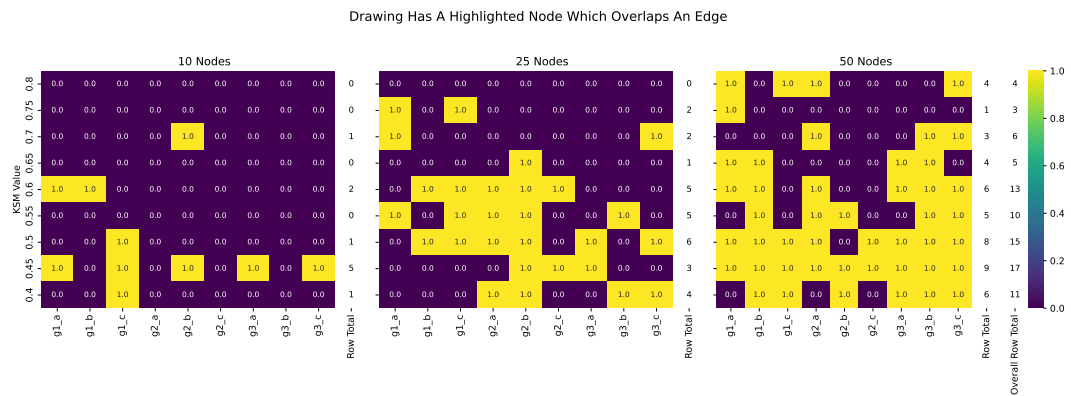
C Detailed Analysis of the Difficulty of the Task for Each Drawing

The following heatmaps (Figures 15–23) show particular properties of each drawing with respect to the difficulty of the shortest path task. Higher row totals indicate greater difficulty for each KSM.

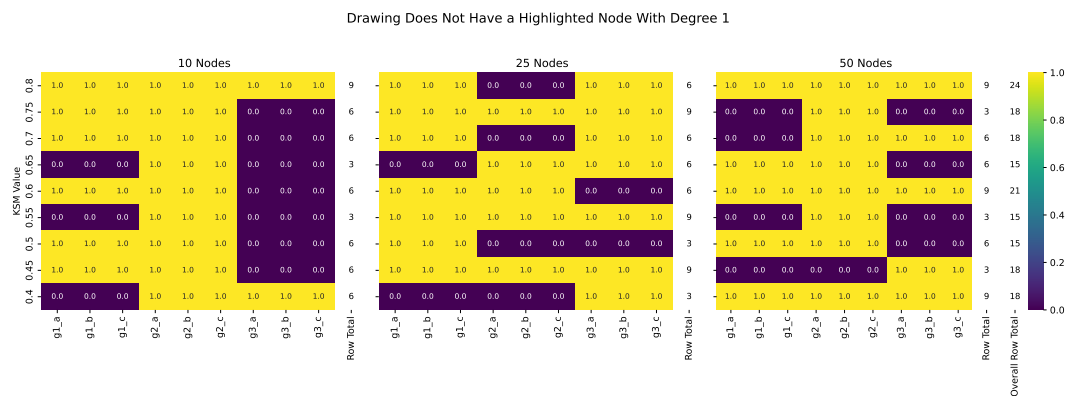
38:20 Stress in Graph Drawings: Perception, Preference, and Performance



■ **Figure 14** Distribution of metric values for the experiment stimuli, grouped by graph size.



■ **Figure 15** Drawings where a highlighted node overlaps another edge.



■ **Figure 16** Drawings where no highlighted node has degree 1.

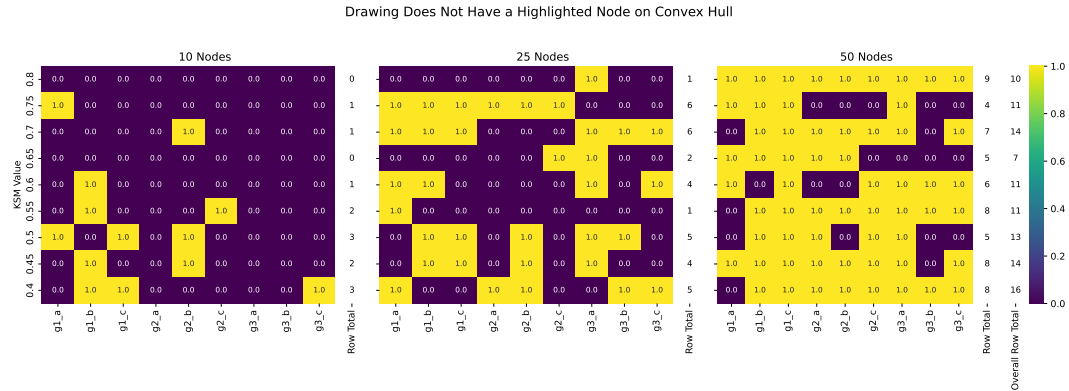


Figure 17 Drawings where no highlighted node is on the convex hull.

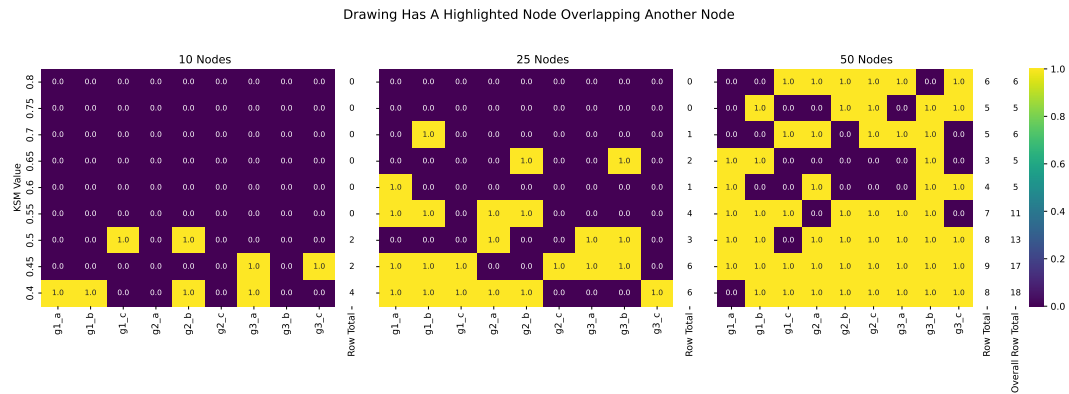


Figure 18 Drawings where a highlighted node overlaps another node.

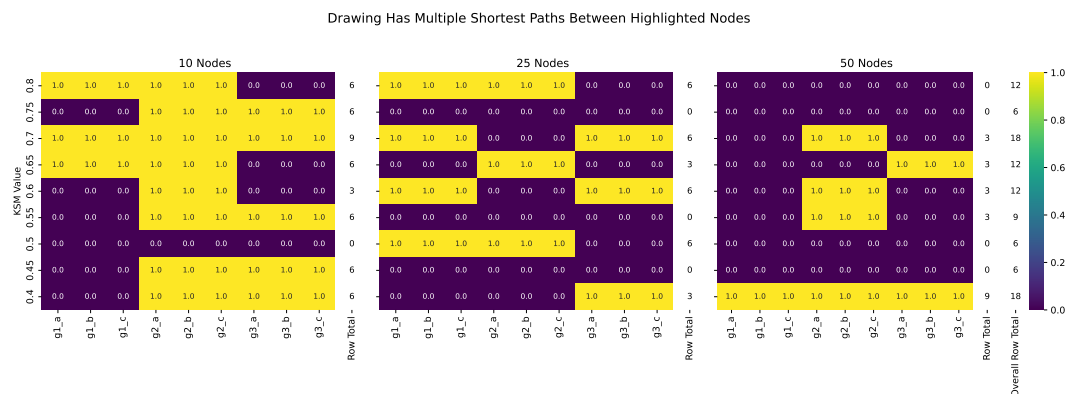


Figure 19 Drawings where there are multiple shortest paths between the highlighted nodes.

38:22 Stress in Graph Drawings: Perception, Preference, and Performance

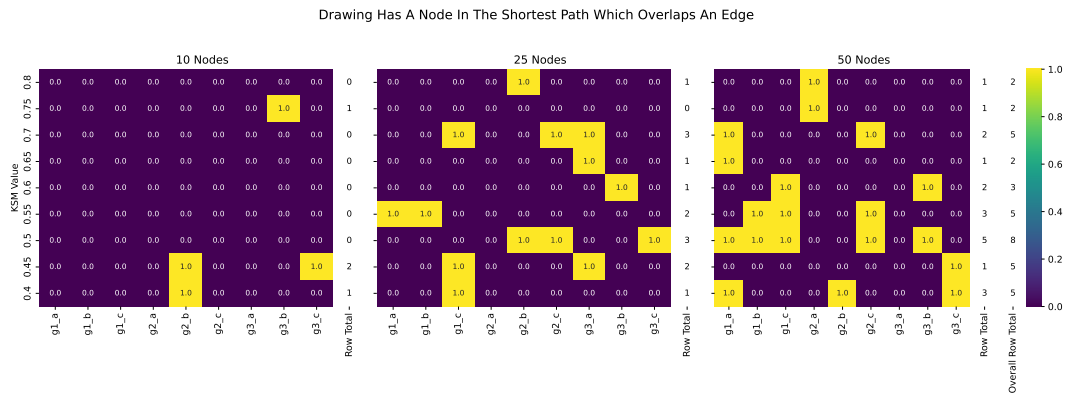


Figure 20 Drawings where a node on the shortest path overlaps an edge.

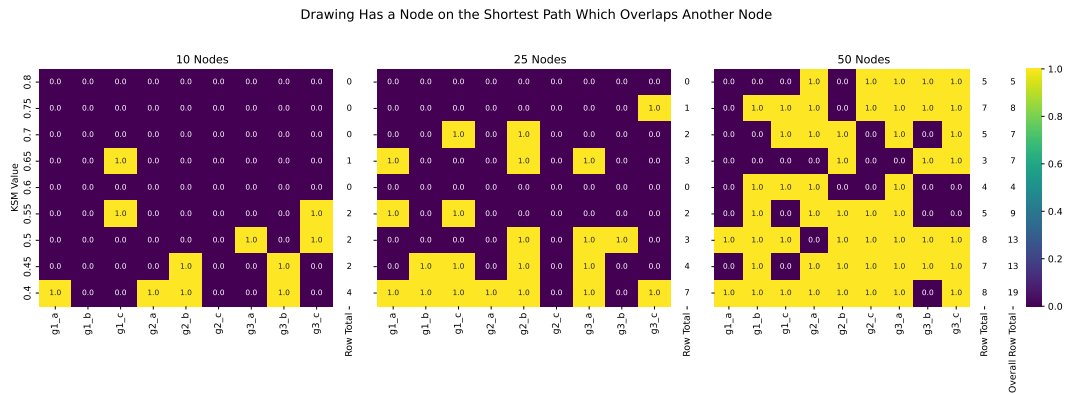


Figure 21 Drawings where a node on the shortest path overlaps another node.

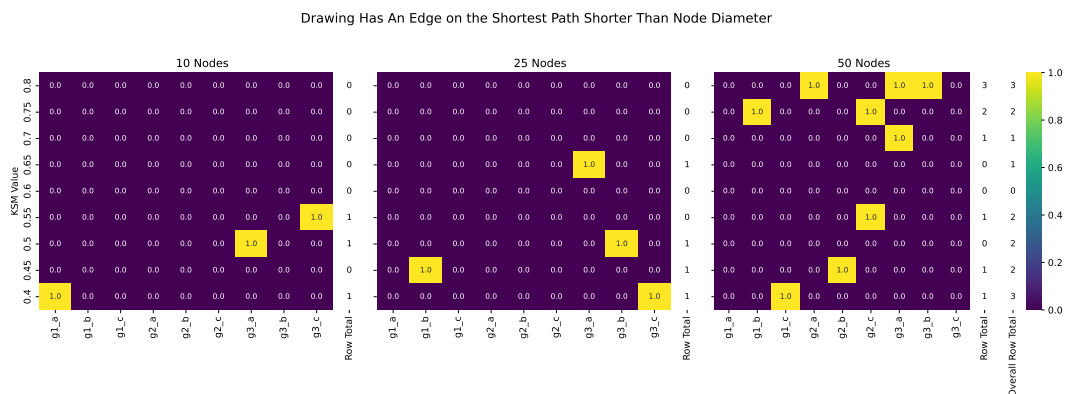


Figure 22 Drawings where an edge on the shortest path is shorter than the diameter of the nodes.

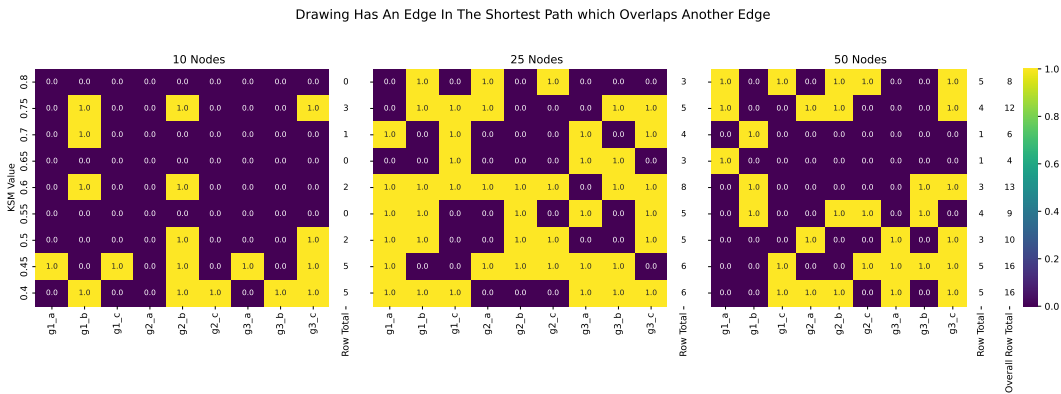


Figure 23 Drawings where an edge in the shorest path overlaps another edge.