The Perception of Stress in Graph Drawings

Gavin J. Mooney ✉ ⌂ ⓘ
Monash University, Australia

Helen C. Purchase ✉ ⌂ ⓘ
Monash University, Australia

Michael Wybrow ✉ ⌂ ⓘ
Monash University, Australia

Stephen G. Kobourov ✉ ⌂ ⓘ
University of Arizona, United States of America

Jacob Miller ✉ ⌂ ⓘ
University of Arizona, United States of America

─── **Abstract** ───────────────────────────────

Most of the common graph layout principles (a.k.a. "aesthetics") on which many graph drawing algorithms are based are easy to define and to perceive. For example, the number of pairs of edges that cross each other, how symmetric a drawing looks, the aspect ratio of the bounding box, or the angular resolution at the nodes. The extent to which a graph drawing conforms to these principles can be determined by looking at how it is drawn—that is, by looking at the marks on the page—without consideration for the underlying structure of the graph. A key layout principle is that of optimising 'stress', the basis for many algorithms such as the popular Kamada & Kawai algorithm and several force-directed algorithms. The stress of a graph drawing is, loosely speaking, the extent to which the geometric distance between each pair of nodes is proportional to the shortest path between them—over the whole graph drawing. The definition of stress therefore relies on the underlying structure of the graph (the 'paths') in a way that other layout principles do not, making stress difficult to describe to novices unfamiliar with graph drawing principles, and, we believe, difficult to perceive. We conducted an experiment to see whether people (novices as well as experts) can see stress in graph drawings, and found that it is possible to train novices to 'see' stress—even if their perception strategies are not based on the definitional concepts.

1 Introduction

Algorithms for depicting graphs are based on conformance to one or more 'layout principles' (or 'aesthetics'). These principles are chosen so that the resulting drawing is considered 'good', in terms of being easier to read or understand when compared to a random layout that does not take any useful principles into account. Some empirical work has been undertaken to determine whether these principles really do enhance human understanding, with support found for reducing the number of edge crossings [19, 20], depicting symmetry [19], and having wide edge crossing angles [13].

Many of these principles can easily be related to the process of visual perception. For example, the principle that edges adjacent at a node should subtend as wide an angle as possible relates to the limitations of visual acuity, as does the need to keep adequate space between nodes and edges; the principle of depicting symmetric subgraphs in symmetric form refers to the Gestalt law of symmetry; bent edges are contrary to the Gestalt law of continuity; an aspect ratio that is extremely different from the Golden Ratio can result in a sense of visual imbalance.

Optimising 'stress' in a graph drawing (e.g. by [11, 14, 30]) is a common layout principle that is much-cited by graph drawing researchers. It relates to the extent to which the geometric distance between each pair of nodes in the graph drawing is proportional to the shortest graph-theoretic path between them—over the whole graph drawing. The definition of stress therefore relies on calculations associated with the underlying structure of the graph (the 'paths') in a way that other layout principles do not.

So, while most layout principles can clearly relate solely to visual perception, 'stress' cannot—it requires analysis of the graph structure as well. This does not mean that there has been no empirical work that refers to stress, simply that it has not done so with explicit use of the word 'stress'. Marner et al. conducted an experiment where participants were asked to untangle large graphs on a wall-display, with the support of a novel algorithmic technique for moving several nodes at once [17]. They found that the drawings created by users did not have lower stress or fewer edge crossings when compared with the graphs drawn using the Fruchterman & Reingold algorithm [9], and they question the importance of these two metrics for creating graphs suitable for human understanding. Their instructions to their participants make no explicit reference to the property of stress.

Chimani et al. claim that 'people prefer less stress and fewer crossings' [5]. Their two-alternative-forced-choice experiment asked participants to choose their drawing of preference, with the stimuli carefully chosen with variations in both stress and edge crossings. While it was shown that participants tended to prefer the drawings with lower stress (57%) and fewer crossings (65%), the former is an implicit variable (unseen and unexplained to the participants) and the latter is explicit—that is, it is obviously seen.

Stress-preference experimental papers like these make use of the computational stress measure in their data analysis, but do not educate the participants about the concept of stress—it is treated as a 'hidden' feature, that cannot be 'seen' by participants. This paper represents the first attempt to determine whether people can 'see' stress in a graph drawing when the concept is explained to them. This is important for graph perception experiments that attempt to determine the most effective (or most preferred, or most efficient) layout algorithms for human understanding. Stress is a key feature of many common algorithms—if we cannot be sure that participants understand what it is, how can we assess the extent to which they value it in graph drawings?

Anecdotally, graph drawing researchers often claim that they can 'see stress' - i.e., that it is as immediately perceptible as other layout principles. While this may be the case for those fluent in graph drawing principles forms (who have an internally and possibly intuitive understanding of the stress principle), it is unclear that 'stress' can be explained sufficiently well to non-experts that they too can perceive stress in a graph drawing.

In this paper, we explore whether it is possible to 'see' a feature of a graph drawing that is implicitly defined by both graph structure as well as visual form. We conduct a series of human experiments, asking participants to distinguish between pairs of graph drawings, identifying which has lower stress. While our results necessarily depend on the quality of the explanation of stress given to novices, we find that even experts find the task challenging.

2   Background

2.1   Applications of Stress in Graph Drawing

Kamada and Kawai [14] first introduced the idea of using stress in graph drawing in the late 80s, with the motivation that a good drawing should accurately represent something about the graph structure. In this case, that drawn distances be proportional to graph-theoretic

distances. The Kamada-Kawai layout algorithm has seen several improvements such as stress majorization [10, 11] and stochastic gradient descent [30]. These most recent versions are among the most popular layout algorithms today, as can be seen by their implementation in libraries like NetworkX (kamada_kawai_layout), GraphViz (Neato), yEd (Organic), and others. Stress-based optimisation of multiple graph layout aesthetics is considered by Ahmed et al. [2] and Devkota et al. [6].

Kruiger et al. [15] and Zhu et al. [32] generalize the classical dimensionality reduction algorithm t-SNE [26] and include stress as one of the evaluation criteria. Other papers that optimise stress-related functions and evaluate stress include those by Zhong et al. [31] and Xue et al. [29].

Regardless of whether an algorithm explicitly optimises stress, it is very commonly used as a metric to evaluate the quality of graph drawings and layout algorithms. A recent survey of graph layout algorithms [7] shows that it is the third most common metric employed in the GD community, behind running time and number of edge crossings. Stress is used in the evaluation of graph layout algorithms in papers by Hong et al. [12] and Marmer et al. [17], as well in dynamic graph layout methods, as in Simonetto et al. [22] and Arleo et al. [3]. Brandes and Pich [4] evaluate several graph layout algorithms based on how well they optimise stress and Welch and Kobourov show that stress can be used as an alternative measure of symmetry in graph layouts [28].

2.2   Stress Definitions

The definition of stress is much older than the Kamada-Kawai algorithm, with roots in statistical analysis. Stress as it is known in the graph drawing community began with Torgerson, who proposed a technique now known as metric Multi-Dimensional Scaling (MDS) [25]. Torgerson aims to provide a low dimensional representation of a set of objects on which distances are given. Importantly, little restriction is given on where these distances come from, e.g. they may be from a traditional metric space $\mathbb{R}^d$, responses to a Likert scale, or indeed, graph-theoretic distances. However the distances are obtained, they are collected into a matrix $D \in \mathbb{R}^{n \times n}$, with the cell $d_{i,j}$ containing the distance between object i and object j. MDS aims to find a matrix $X \in \mathbb{R}^{n \times 2}$ where X_i represents a low dimensional coordinate of object i such that distances between rows in X are exactly the distances in D. This is not possible in general, so the deviation of these distances are measured as a function:

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

Note that this function is differentiable with respect to X, so gradient-based optimisation schemes can be employed to find a local minimum of the equation.

Sometime later, Kruskal defined what is now known as non-metric MDS [16], with the motivation that preservation of exact distances is often very difficult and too restrictive. In the non-metric variant, it is instead important to maintain the *rank* or ordering of distances i.e. from each object, the object that is first, second, and third closest in the original space should still be first, second, and third closest in the projection respectively. This non-metric stress is defined with a deceptively similar function:

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

However, the matrix $\hat{D}$ is not the original distance matrix from the input. Instead, a Shepard diagram [21] is formed by forming a set of coordinates for each pair of objects i, j: $c_{i,j} = (||X_i - X_j||, d_{i,j})$. A Shepard diagram created from an ideal drawing would have all points lying exactly on a straight line, since this would correspond to all input distances being exactly all output distances. To measure the deviation from this line, Kruskal performs a monotonic regression with $\binom{n}{2}$ points to best fit the diagram. The matrix $\hat{D}$ is defined such that $\hat{d}_{i,j}$ is the distance in the x-coordinate to the fitted line from $c_{i,j}$.

In terms of information, the metric stress of equation 1 is more restrictive than the non-metric stress of equation 2. However, both are measuring how well distances are maintained in the output. Non-metric stress has the additional advantage that the resulting number will be in the range 0-1 and is hence more suitable as an evaluation metric. Scale can impact normalised stress values and several scale-invariant stress measures are detailed in recent papers by Ahmed et al. [1], Smelser et al. [23], and Wang et al. [27].

3   Methodology

3.1   Experimental Methodology

Our overriding Research Question is 'Can people see stress in a graph drawing?' As a simple yes/no question, this is tricky to address directly: it does not make sense to simply ask a participant to look at a graph drawing and to state whether it 'is' or 'is not' stressed. We therefore address this question by investigating whether people can see the difference in graph drawings with different stress values; thus, the research question becomes "Can people see differences in stress between two drawings of the same graph?"

Our methodology was mostly exploratory and incremental. We first addressed the research question by explaining the concept of stress to novices, giving them some training on the task (with feedback), and collecting performance data (accuracy, time, confidence) on 45 trials - over three different graph sizes. Each trial consisted of a pair of drawings of the same graph, with the participants being asked to indicate which has lower stress.

The results of this initial experiment were very encouraging, and so our follow-up study attempted to see how the extent of training affects performance; in this case, the participants received the same explanation of stress, but did not receive feedback on the training questions. We also asked some of our graph drawing expert collaborators to do the experiment as well; they did not have any training session.

Section 3.4 explains the experimental procedure in more detail.

3.2   Measuring Stress

To test the perception of stress, we need a set of graph drawings which have different values of stress. To create these, we must define a measurement of stress which is independent of graph size, structure, and drawing space. While the general definition (i.e. the metric definition of equation 1 and its commonly used variants), can be used to calculate the stress of a given drawing, it does not meet this criteria. For example, using this definition, Welch and Kobourov show that the stress of a given drawing varies with respect to the scale of the drawing space [28]. This means graph drawings with different geometric scales cannot be compared.

To create the stimuli, we use the definition of stress provided by Kruskal [16] (equation 2). While intended for measuring the quality of multi-dimensional scaling techniques, the method is applicable to measuring stress in graph drawings and has numerous benefits. Firstly, the

measure is normalised, and hence applicable to graphs of different size and structure. This definition also has the benefit of being independent of the rotation and geometric scale of the drawing space. To integrate this measure with other graph drawing metrics, such as those described by Mooney et al. [18], we subtract the Kruskal stress value from 1. This means that, as a metric, a value of 1 represents the extreme that is intuitively assumed to be good—in this case, zero stress. For the rest of this section, we refer to this definition as the 'Kruskal Stress Metric' (KSM).

Hopefully, Section 2.2 leaves one convinced that Kruskal's definition is an appropriate measure of stress. To further justify this choice, we used the large dataset described by Mooney et al. [18], and calculated the Kruskal stress of nearly half a million graph drawings, then compared this to a normalised stress function based on the Kamada and Kawai definition [14]. We found the correlation between the two definitions to be 0.871. This strong correlation highlights the similarity in the respective definitions.

3.3 The stimuli

We ran three concurrent experiments, for graphs with node count of 10, 25 and 50. For each graph size, we created five random graphs using the Erdős–Rényi model [8], with the constraint that the graph is connected and the number of edges is less than two times the number of nodes. This constraint ensures that drawings do not end up as 'hairballs'. Preliminary testing when creating the stimuli showed that denser graphs (and hence denser drawings) tend to have higher stress than sparser graphs, and so, inclusion of them would limit the range of possible stress values.

We create graph drawings with KSM values in the range of 0.4 to 0.8, with intervals of 0.05 (nine drawings in a set). For each graph size, we generate three sets of drawings using a basic hill climbing algorithm. The evaluation function for this algorithm is the absolute difference between the KSM of the current drawing and the target KSM value. Nodes are initially positioned randomly within a unit grid. Each iteration updates the drawing by selecting a random node, and moving this node to a new random position within the initial bounding box of the drawing space. This new random position is also bounded by a circle of decreasing radius (as the number of iterations increases) centred on the chosen node's current position. If the new drawing has KSM closer to the target value, it is taken to be the current drawing, otherwise it is discarded. The algorithm terminates when the KSM of the drawing is within 0.01 units of the target value. Example stimuli are shown in Table 1.
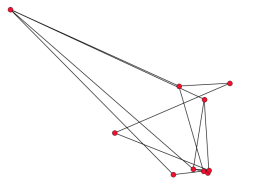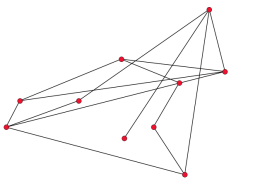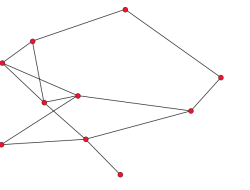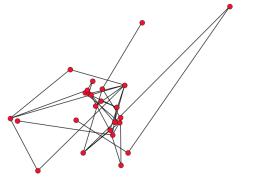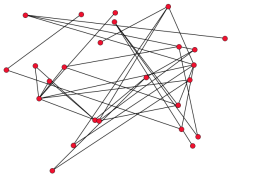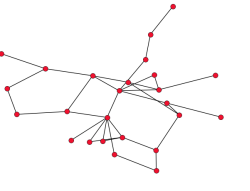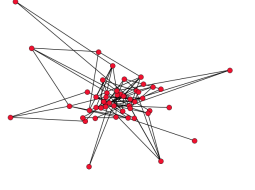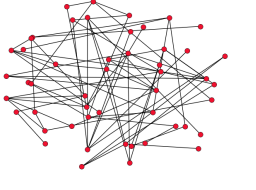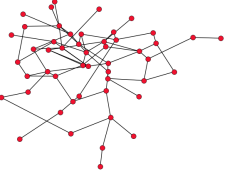
3.4 Experimental Procedure

The methodology for each novice experiment is the same, but with different stimuli (corresponding to the three different graph sizes). The participants are first shown an overview of the aims of the experiment and asked for their consent to participate. Next, participants are shown a page which outlines the key concepts of graph drawings and stress (details available in the online version of the paper). We use layman terms such as 'network' instead of 'graph', 'object' instead of 'node', and 'connection' instead of 'edge'. This page explains the following:

- What a graph is, including definitions for nodes and edges.
- What a graph drawing is, and that one graph can have many different drawings.
- That different graph drawings have different visual properties, such as symmetry.
- What is meant by a 'path' and 'shortest path' between two nodes.
- What stress means in relation to graph drawing, with very basic examples of low and high stress drawings.

■ **Table 1** Example drawings of different KSM values. Each row consists of different drawings of the same graph. Higher KSM values indicate lower stress.

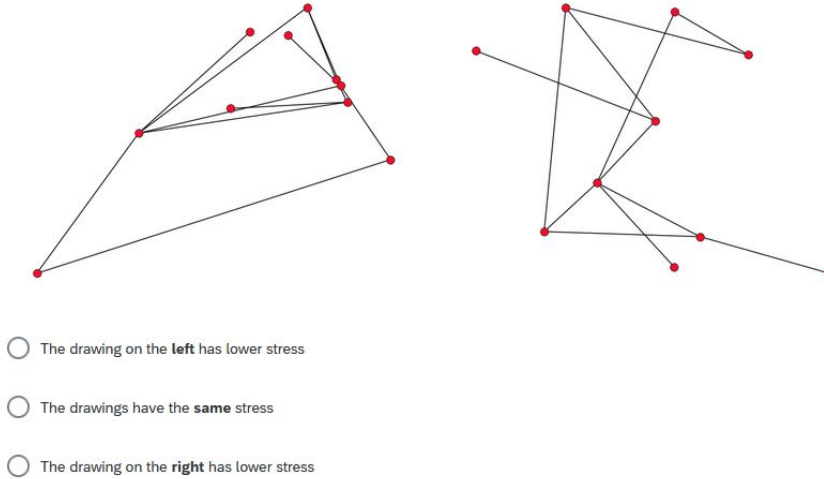|		KSM		
		0.4	0.6	0.8
Size	10			
	25			
	50			

- Four pairs of drawings with low and high stress.
- That the participants should not spend a long time trying to calculate the stress of a drawing, but should rely on their immediate perception.

Participants are then shown nine predetermined pairs of drawings and asked to choose which has lower stress (or that they have the same stress). The nine drawings are of graphs that are the same size as the given experiment with differences of KSM ranging from 0.4 to 0, decrementing by 0.05 (in that order). In the first set of experiments there are 25 participants for each graph size (75 total). These 75 participants are given feedback immediately after each pair to inform them if they were correct or, if not, what the correct response was. Where feedback is given, over 50% of the responses to these training questions must be correct to continue to the main experiment. In the second set of experiments there are 10 participants for each graph size (30 total). These 30 participants are given no feedback on the nine training questions.

In the main part of the experiment, participants are shown 45 pairs of drawings (trials), with KSM values in the range 0.4–0.8, and asked to choose the drawing with lower stress (or that they have the same stress). Each pair consists of two drawings of the same graph. There are five graphs and nine unique KSM differences (0, 0.05, 0.1...0.4). The exact drawings for each KSM difference is random—e.g., for graph one, and KSM difference of 0.3, one participant may be shown drawings with KSM of 0.4 and 0.7, whilst another may be shown drawings with KSM of 0.5 and 0.8. The order in which pairs of drawings are shown is randomised for each participant, as well as which two drawings are chosen (from a pool of three) and the order (left/right) that they appear on the page. This randomness is

incorporated to lower the chance of a learning effect. The participants are not given any feedback about their choice. After each choice, the participant is asked if they were confident about their choice. An example trial is shown in Figure 1.

Figure 1 Example trial for the n=10 experiment. Here the drawing on the left has a KSM value of 0.45, while the right drawing has a KSM value of 0.7 (but this information is, of course, not shown to the participants).

After all 45 pairs have been shown, the participant is asked some follow up questions about the experiment and asked for some demographic information. See Section 3.4.1 for more details.

The expert experiment follows the same methodology as the novice experiments, however the experts are not shown the nine training pairs and complete the 45 trials for each graph size (135 total). The experts are shown the 135 trials in blocks of 45, corresponding to each graph size (10, then 25, then 50), with an opportunity for a break between blocks.

3.4.1 Data Collection

In each experiment we collect the following data:

- The number of incorrect training responses (out of nine).
- Which pairs of drawings were shown to participants and their KSM values, as well as the order they were shown in. Within each pair we also keep track of which drawing was displayed on the left and which was displayed on the right.
- The response from the participant about which drawing had lower stress (The drawing on the left has lower stress/The drawings have the same stress/The drawing on the right has lower stress).
- How confident the user was about each choice (Confident/Not confident).
- The time taken to submit an answer for each pair.

After all pairs of drawings are shown we also collect data for some additional questions and demographic information:

- The overall strategy used to determine which drawing had lower stress.
- The participant's overall confidence in their responses (Very confident/Somewhat confident/Not very confident/Not confident at all).
- How difficult the participant found the experiment (Very difficult/Difficult/Easy/Very Easy).
- How familiar the participant is with network diagrams (Very familiar/Somewhat familiar/Not very familiar/They are new to me).
- The participants' age and gender.

3.4.2   Experimental Conduct

The online survey platform 'Qualtrics' was used to set up and run the experiments, and collect the data.

Participants were recruited from the online platform 'Prolific' and paid at a rate of £9 GBP per hour. The pool of participants was limited to users over the age of 18 and residing in either the United Kingdom or Australia. Participants who were included in one study were excluded from participating in any of the others to reduce the chance of a learning effect.

For each graph size, we collected results from 25 participants who received feedback on the training, and 10 participants who did not. For the participants who received feedback on the training, those who answered more than 50% incorrectly were excluded from the main experiment. For the experiments on graphs of size 10 and 25, only one participant did not pass the training. For the experiment on graphs of size 50, three participants did not pass the training. For each experiment, we recruited more participants until we had 25 who passed the training and completed the whole experiment. Each experiment was first piloted by two participants to ensure the experiments ran smoothly and that the data was being collected correctly. No changes were made after the successful pilots.

The expert participants were invited via email, of whom ten participated in the experiment. Two of these ten responses were incomplete, leaving us with eight complete expert responses.

This research was approved by the Monash University Human Research Ethics Committee with ID 42695.

4   Results
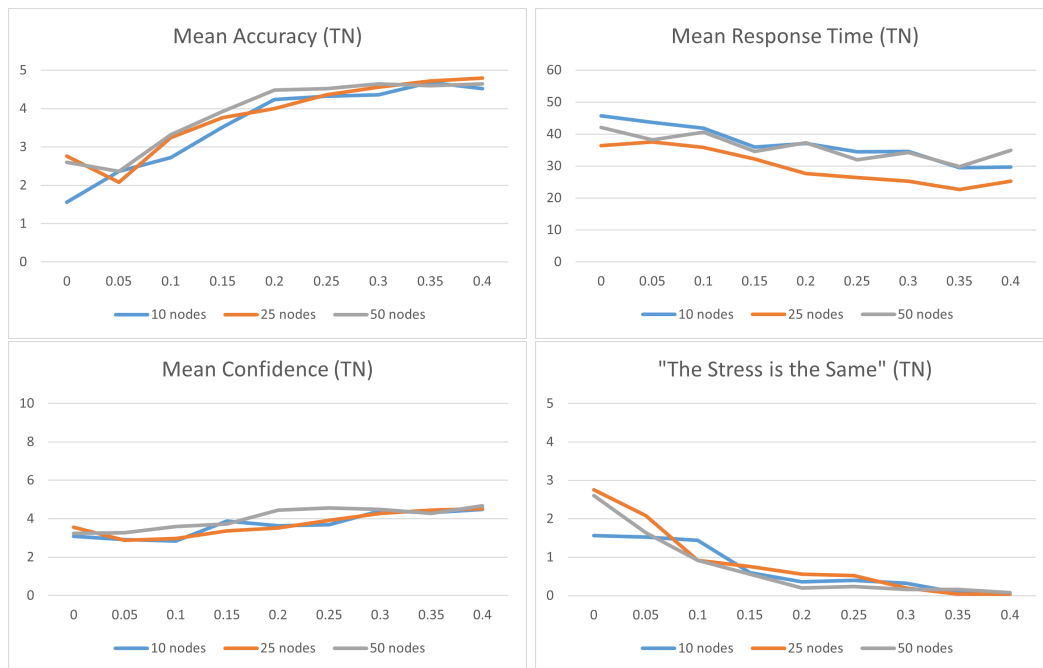
4.1   Trained Novices (TN)

We address our research question ("Can people see the differences in stress between two drawings of the same graph?") by analysing our data along two dimensions, giving two sub-questions:

- RQa: "How much does the quantifiable difference in stress (the 'delta') between two drawings of the same graph affect the perception of stress?"
- RQb: "How much does the size of the graph affect the perception of stress?"

In all cases, our dependent variables for measuring perception are accuracy, response time, and confidence. Of the 75 participants, there were 38 women, 36 men and 1 gender-diverse. The median age range was 36-45, with 16 participants between 18-25, and one over 76. Four participants said they were 'very familiar' with graph drawing; 55 claimed novice status (at most 'not very familiar').

4.1.1 Delta trends

We consider each graph size in turn, plot the average accuracy, response time, and confidence over all 25 participants, and consider trends with respect to delta (Figure 2). There were five trials per delta value, so accuracy is between 0 and 5; high confidence is recorded as 2 for each trial (so confidence per delta is between 0 and 10); time is measured in seconds. There are eight response time data points for which the response time was over 200 seconds (six for n=50, two for n=25). All other 3,367 response time data points were less than 140 seconds. These eight data points (0.24% of the total data points) were replaced with the mean of the relevant participants' other responses. Participants were given the option to indicate that the stress in the two drawings was the same. Only five of the forty-five trials were pairs with zero delta. Use of this option for other deltas indicates difficulty in distinguishing stress.

**■ Figure 2** Delta trends for Trained Novices.

4.1.2 Overall accuracy

In this case, we ignore the delta values, and focus on whether the size of the graph makes a difference to the overall mean accuracy of stress perception (Figure 3a). An independent measures t-test between the data for 10 nodes and that for 50 nodes reveals a p-value of 0.0495—a barely significant difference.

34 of the 75 participants (45%) said they found the task difficult; only three said that it was 'very easy'.

4.1.3 Discussion: Performance

With an overall accuracy rate of 75.2% (considerably higher than chance: 33% for three-way multiple choice responses), our data shows that novices who have had the notion of stress explained to them, and have had the opportunity to be trained with feedback, can indeed 'see' stress. Contrary to our expectations, the size of the graph does not make the task more
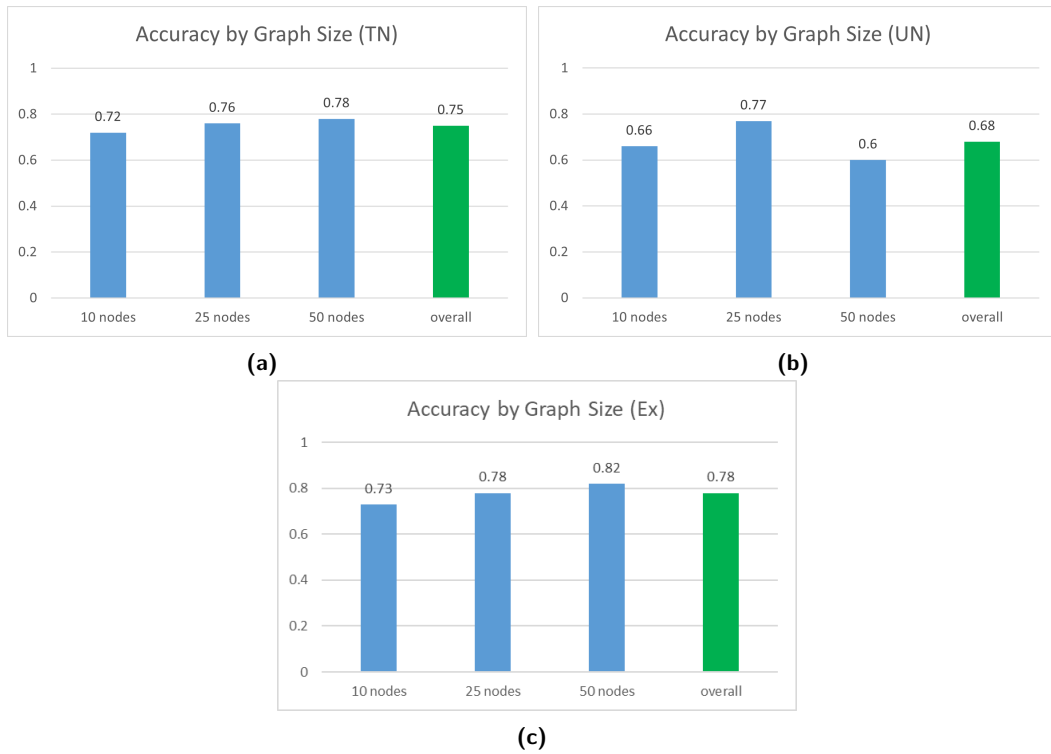
Figure 3 Overall accuracy over all trials for each participant type, with respect to graph size.

difficult, and we see expected trends in accuracy, response time and confidence with respect to difficulty (as measured by the delta). While there is some uncertainty with the lower deltas of 0, 0.05 and 0.1, by the time the delta reaches 0.15, participants are getting 80% of the tasks correct, rising as high as 96% accuracy for 25 nodes with delta=0.4. Looking at the average number of times that participants judged the stress to be the same, we can see that any ambiguity caused by deltas of 0.05 and 0.01 diminishes when the delta increases to 0.15. We were surprised by these quantitative results. We had assumed that the complexity of the notion of stress would be difficult for novices to grasp easily, and that it would be more difficult to determine in larger graphs.

4.1.4   Discussion: Strategies and Perception

We had asked the participants to explain their strategy in determining the stress differences. Many participants (24) said that they looked at the distances between the nodes, or the distribution of the nodes (5). Some were more specific, noting that they looked at the length of the edges between the nodes (11). Apart from some references to "messiness/busy-ness" (10), other features highlighted were symmetry (2), edge crossings (5) or edge closeness (4), angles (5) and "clusters" (5). Only one participant made a value judgement on the form of the drawings: "I choose the one that made me feel less stressed i.e the one that was easiest to understand". Other notable text summaries include: "The tightness of the angles, the grouping and how close it was, how "spiky" or chaotic the diagrams looked", "the ones that looked more erratic and all over the place felt more stressful" and "if the lines looked messy, it had more stress."

Our trained novices were therefore able to perceive 'stress' in a graph drawing, using the

perception of edge length, node distribution and visual edge density (i.e. the closeness and compactness of edges, including edge crossings) as proxy measures.

4.2 Untrained Novices (UN)

Our positive results for trained novices clearly depended on the nature and extent of the training offered to them: not only were they given a detailed written explanation of the stress measures (with examples), they had nine opportunities to attempt the task, with feedback given as to the correct answer for each. We conducted the experiment again, for 10 participants for each size graph, and omitting the feedback.

Of the 30 participants, there were 16 women, 13 men and 1 unknown. The median age range was 26-35, with 7 participants between 18-25, and four between 56-65. No-one claimed to be 'very familiar' with graph drawing; 24 claimed novice status (at most 'not very familiar').

Figure 4 shows the data for the untrained participants; Figure 3b shows the overall accuracy. An independent measures t-test between the data for 25 nodes and that for 50 nodes reveals a p-value of 0.0561—an insignificant difference. 23 of the 30 participants said that they found the task difficult.
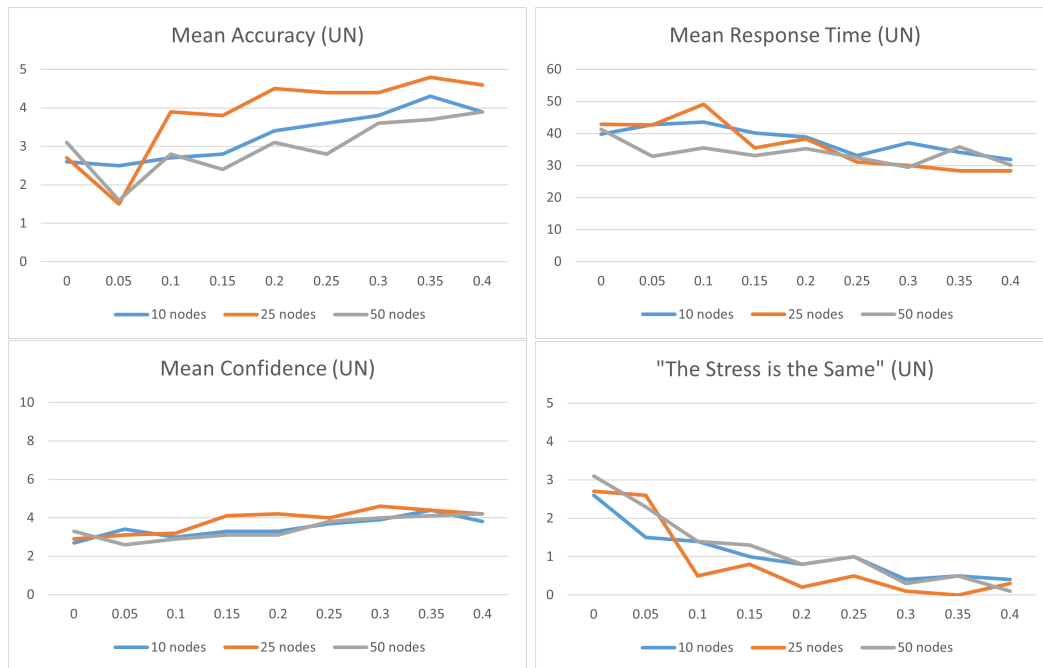
Figure 4 Delta trends for Untrained Novices.

4.2.1 Discussion

The overall trend pattern for the untrained novices is comparable to that of the trained novices: any ambiguity when delta is 0.05 or 0.1 is resolved by the time it increases to 0.15. It appears that the medium-sized graph (n=25) was easier; it is not clear why this might be the case.

The strategies used by these participants were similar to those used by the trained participants: distance between nodes and node distribution (14), 'chaotic' (5), and edge

lengths (3). There were three references to 'open space' (not mentioned by the trained participants): "Anything that looked congested or compact I found stressful to look at where as when it was more spaced with no sudden lines pointing out it looked much more relaxed." Two participants referred to analogies: "it reminded me of chemical and biological bonds, those with random arrangements where the links/bonds were more likely to break, be fragile, have weaker chain strengths. Those with consistent links/bonds where the lengths and widths are the same were stronger and less likely to separate"; "some looked like animals, which was less stressful to me. if there was a rogue line jutting out of somewhere I found this stressful. when they had more space I found it less stressful."

The overall accuracy of the untrained novices was 67.6% - still greater than chance (33%) but less than the accuracy for the trained novices (75.2%). An independent-samples t-test between the overall mean accuracy for trained and untrained novices gives a p-value of 0.009; thus, the additional training made a significant difference to the performance of the novice.

4.3   Experts

Having gathered data on how well novices can perceive stress after being presented with a rudimentary explanation of the concept, we were interested in how graph drawing experts (who already know the concept well) would fare in the same experiment. Eight graph drawing experts—known colleagues of the authors (7M, 1F, median age 36-45)—completed all the trials for all three graph sizes. The experts were shown the same written explanation of 'stress', but were not given any training.

Figure 5 shows the data for the expert participants; Figure 3c shows the overall accuracy. A repeated measures t-test between the data for 10 nodes and that for 50 nodes reveals a p-value of 0.089—an insignificant difference.


Figure 5 Delta trends for Expert participants.

4.3.1 Discussion

The trend patterns for the experts are remarkably similar to the novices. As with both trained and untrained novices, confusions between the pairs with deltas of 0.1 and less are resolved for those with values of 0.15 and above. The longer response time for the graphs with 10 nodes can be explained by the fact that these were the trials that all experts completed first, before they then looked at the larger graphs (suggesting a 'warming-up' period). The experts' confidence is remarkably low, closely matching that of the novices.

The strategies used by the experts included line length (2), distances between nodes (2), and edge crossings (2). There were, however, some less specific considerations and some clear value judgements: "drawings with Low stress just look "right"... high stress drawings look more random showing features that a human graph drawer would attempt to correct in a second iteration"; "general feeling how much the nodes are settled into place"; "how well the overall structure is highlighted"; "how well untangled they are." It appears that experts do not (like novice) need to rely so much on visual proxies, but their experience gives them an intuitive 'feel' for stress. This 'feel' is, however, not fail-safe, with the experts having only an overall accuracy of 77.5%. Figure 6 shows the relative accuracy of the three categories of participant.
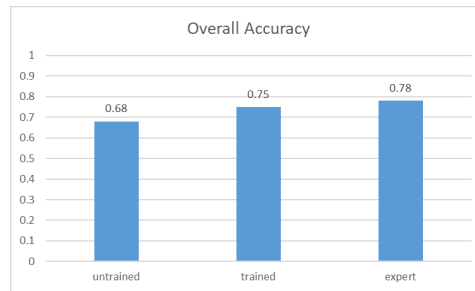
**Figure 6** Overall accuracy over all trials for each participant group.

An independent-samples t-test between the overall mean accuracy for untrained and trained novices gives a p-value of 0.009; thus, the additional training made a significant difference to the performance of the novices.

An independent-samples t-test between the overall mean accuracy for trained and expert gives a p-value of 0.361; thus, there is no significant difference in accuracy between trained novices and experts. There is also no significant difference between mean response time for trained novices (7.89s) and experts (5.86s); p=0.065.

We speculate that trained novices thought more carefully about their choices, using visual proxies they had identified during the training sessions, while the experts made rapid 'gut-feel' decisions. Despite this, all participants reported low confidence in their responses—even the experts.

4.4 Correlations Between Stress and Other Metrics

We calculated the correlations between KSM and four other visual properties present in the 'strategy' responses: edge crossings, distribution of nodes in the drawing space, average distance between nodes, and average edge length (Figure 7). For the distribution of nodes and edge crossings, we use the Node Uniformity and Edge Crossing metrics defined by Mooney et al. [18] (details available in the online version of the paper). We compute these metrics on our set of 405 stimuli drawings.
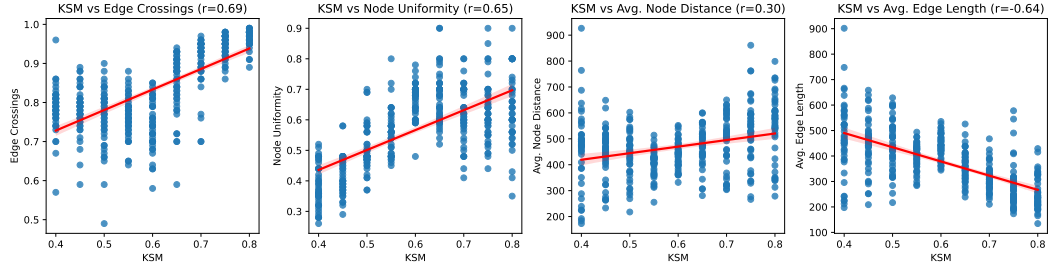
Figure 7 Correlations between KSM and other graph drawing metrics.

We find that KSM is positively correlated with the average distance between nodes (0.30), the distribution of nodes (0.65), and the number of edge crossings (0.69). KSM is negatively correlated with average edge length (-0.64). Recalling that KSM values closer to 1.0 correspond to lower stress, we can make the following observations:

- Drawings with fewer edge crossings have lower stress.
- Drawings in which the nodes are more evenly distributed around the drawing space have lower stress.
- Drawings with nodes which are farther apart on average have lower stress.
- Drawings which have shorter edge lengths on average have lower stress.

These observations suggest that some of the participants' 'strategies' are effective visual proxies for stress, though do not fully capture it.

5 Conclusion

These experiments represent the first investigation of whether people can 'see' stress, an invisible property of graph drawings that is defined by both the geometric form of the graph drawing as well the underlying structure of the graph. We find that it is indeed possible to describe the notion of stress to people unfamiliar with graphs and graph drawing in manner that allows them to perceive it. Unprompted, they are able to devise their own visible proxies for the invisible stress feature—that is, they easily identify appropriate visual features of drawings which can be used to assess stress: in particular the geometric length of edges, distances between nodes and node distribution, and 'compactness'/ 'clustering'/ 'density'. We can't be sure that the participants utilised the exact geometric definition of stress but their performance suggests that they were able to perceive its visual features. A few participants referred to the personal stress that they felt while looking at the drawings, thus transferring a mathematical concept into an emotion.

Surprisingly, experts do not perform much better than trained novices. While there is some evidence of them also using visual proxies, they also appear to rely on high-level concepts like 'untangling', 'general feeling' and 'settled nodes'—more than is evident in the strategies described by the novices.

Our findings do not differ with respect to graph size. We were surprised at this, but this could be explained by the fact that the participants focused on these overall visual proxies rather than the edges themselves. Our results therefore appear to be generalisable to graph drawings of different sizes.

This work is important for any human experimental work on the perception and use of graph drawings where stress is deemed an important feature. Not only have we shown that people can understand the concept of stress sufficiently well to be able to perceive it,

we have also identified those visual features that they consider to be most related to stress: edge length, node distribution and concentrated areas of edges and edge crossings. Future experimental work can now build on these results to investigate, for example, nuances of different implementations of stress, trade-offs between stress and other layout principles, the smallest delta for stress perception, and the form of user-generated drawings where low stress is an explicitly stated goal.

5.1 Future Work

All empirical studies are limited by necessary practical parameters. In this case, the stimuli are small and sparse graphs. A follow up experiment on larger and denser graphs would inform on how the perception of stress varies as graph sizes increase. While our results show only a small difference in terms of perception between graphs of size 10 and 50, this may not be the case in general as the size of the graph increases. However, as the size and density of the graph increases, producing drawings with low stress becomes increasingly challenging.

Our results suggest that stress deltas of 0.1 and below are confusing, but that stress differences of 0.15 and above are more discernible. These observations are also limited by the range of KSM values used in the experiment. It may be the case that the discernible difference in KSM is non-uniform; i.e., the difference between drawings with KSM of 0 and 0.15 may be more or less discernible than that of 0.85 and 1.0. A more comprehensive 'Just Noticeable Difference' methodology (such as that used by Soni et al. to measure the perception of graph density [24]) could validate this, and perhaps result in a more specific threshold. This methodology would allow us to see what the smallest perceptible difference in stress is, and see if this follows Weber's law. This requires careful creation of a broader set of stimuli as well as a larger number of participants.

Investigation into additional 'non-visual' metrics, which are not associated with the principles of visual perception, such as Neighbourhood Preservation, would provide insights into their perceptibility. This may also inform on the usefulness of such metrics as criteria for optimisation and/or evaluation of graph drawings.

References

1. Reyan Ahmed, Cesim Erten, Stephen Kobourov, Jonah Lotz, Jacob Miller, and Hamlet Taraz. Size should not matter: Scale-invariant stress metrics. *arXiv preprint arXiv:2408.04688*, 2024.

2. Reyan Ahmed, Felice De Luca, Sabin Devkota, Stephen Kobourov, and Mingwei Li. Graph drawing via gradient descent, $(GD)^2$. In *Lecture Notes in Computer Science*, pages 3–17. Springer International Publishing, 2020. `doi:10.1007/978-3-030-68766-3_1`.

3. A. Arleo, S. Miksch, and D. Archambault. Event-based dynamic graph drawing without the agonizing pain. *Computer Graphics Forum*, 41(6):226–244, 2022. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14615`, `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14615`, `doi:10.1111/cgf.14615`.

4. Ulrik Brandes and Christian Pich. An experimental study on distance-based graph drawing. In Ioannis G. Tollis and Maurizio Patrignani, editors, *Graph Drawing*, pages 218–229, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. `doi:10.1007/978-3-642-00219-9_21`.

5. Markus Chimani, Patrick Eades, Peter Eades, Seok-Hee Hong, Weidong Huang, Karsten Klein, Michael Marner, 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, Reyan Ahmed, Felice De Luca, Katherine E. Isaacs, and Stephen Kobourov. Stress-plus-x (SPX) graph layout. In *Lecture Notes in Computer Science*, pages 291–304. Springer International Publishing, 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   P Erdös and A Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

9   Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, November 1991. `doi:10.1002/spe.4380211102`.

10 E. R. Gansner, Yifan Hu, and S. 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`.

11 Emden R. Gansner, Yehuda Koren, and Stephen North. Graph drawing by stress majorization. In *Graph Drawing*, pages 239–250. Springer Berlin Heidelberg, 2005. `doi:10.1007/978-3-540-31843-9_25`.

12 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 *Lecture Notes in Computer Science*, pages 139–146. Springer International Publishing, 2019. `doi:10.1007/978-3-030-35802-0_11`.

13 Weidong Huang, Peter Eades, Seok-Hee Hong, and Chun-Cheng Lin. Improving multiple aesthetics produces better graph drawings. *Journal of Visual Languages & Computing*, 24(4):262–272, August 2013. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1045926X11000814`, `doi:10.1016/j.jvlc.2011.12.002`.

14 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`.

15 J. F. Kruiger, P. E. Rauber, R. M. Martins, A. Kerren, S. Kobourov, and A. C. Telea. Graph layouts by t-sne. *Computer Graphics Forum*, 36(3):283–294, 2017. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13187`, `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13187`, `doi:10.1111/cgf.13187`.

16 Joseph B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

17 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 *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 113–124. Springer International Publishing, 2014. `doi:10.1007/978-3-662-45803-7_10`.

18 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 Symposium (PACIFICVIS)*. IEEE, 2024.

19 Helen Purchase. Which aesthetic has the greatest effect on human understanding? In Giuseppe DiBattista, editor, *Graph Drawing*, pages 248–261. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

20 Helen C. Purchase, C. Pilcher, and B. Plimmer. Graph Drawing Aesthetics—Created by Users, Not Algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 18(1):81–92, January 2012. URL: `http://ieeexplore.ieee.org/document/5674033/`, `doi:10.1109/TVCG.2010.269`.

21 Roger N Shepard. The analysis of proximities: multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27(2):125–140, 1962.

22 Paolo Simonetto, Daniel Archambault, and Stephen Kobourov. Drawing dynamic graphs without timeslices. In *Lecture Notes in Computer Science*, pages 394–409. Springer International Publishing, 2018. `doi:10.1007/978-3-319-73915-1_31`.

23 Kiran Smelser, Jacob Miller, and Stephen Kobourov. " normalized stress" is not normalized: How to interpret stress correctly. *arXiv preprint arXiv:2408.07724*, 2024.

24 Utkarsh Soni, Yafeng Lu, Brett Hansen, Helen C. Purchase, Stephen Kobourov, and Ross Maciejewski. The perception of graph properties in graph layouts. *Computer Graphics Forum*, 37(3):169–181, 2018. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13410, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13410, doi:10.1111/cgf.13410.

25 Warren S Torgerson. Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4):401–419, 1952. doi:10.1007/BF02288916.

26 Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11), 2008.

27 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.

28 E. Welch and S. Kobourov. Measuring Symmetry in Drawings of Graphs. *Computer Graphics Forum*, 36(3):341–351, June 2017. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13192, doi:10.1111/cgf.13192.

29 Mingliang Xue, Zhi Wang, Fahai Zhong, Yong Wang, Mingliang Xu, Oliver Deussen, and Yunhai Wang. Taurus: Towards a unified force representation and universal solver for graph layout. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):886–895, 2023. doi:10.1109/TVCG.2022.3209371.

30 Jonathan X. Zheng, Samraat Pawar, and Dan F. M. Goodman. Graph drawing by stochastic gradient descent. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2738–2748, 2019. doi:10.1109/TVCG.2018.2859997.

31 Fahai Zhong, Mingliang Xue, Jian Zhang, Fan Zhang, Rui Ban, Oliver Deussen, and Yunhai Wang. Force-directed graph layouts revisited: A new force based on the t-distribution. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–14, 2023. doi:10.1109/TVCG.2023.3238821.

32 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.