

Memorability of Visual Features in Network Diagrams

Kim Marriott, Helen Purchase, Michael Wybrow, and Cagatay Goncu

Abstract—We investigate the cognitive impact of various layout features—symmetry, alignment, collinearity, axis alignment and orthogonality—on the recall of network diagrams (graphs). This provides insight into how people internalize these diagrams and what features should or shouldn't be utilised when designing static and interactive network-based visualisations. Participants were asked to study, remember, and draw a series of small network diagrams, each drawn to emphasise a particular visual feature. The visual features were based on existing theories of perception, and the task enabled visual processing at the visceral level only. Our results strongly support the importance of visual features such as symmetry, collinearity and orthogonality, while not showing any significant impact for node-alignment or parallel edges.

Index Terms—Network diagrams, graph layout, perceptual theories, visual features, diagram recall, experiment.

1 INTRODUCTION

Network diagrams (also called node-link diagrams and graphs) are widely used to visualise relational data. Visualisation relies on the human viewer building an internal cognitive structure. The design of effective visualisations relies on understanding what information is contained in this internal representation [42]. The work reported here investigates how people internalise graph drawings, and, in particular, whether visual features of network diagrams impact their internal representation.

Better understanding of the internal representation provides fundamental insight into the effective use of network diagrams for information visualisation. We wish to find which kinds of visual features are preserved in the internal representation and thus noticed at the visceral level. One reason is that these features can be effective at presenting additional information, or conversely, if a layout inadvertently contains these features, they may subconsciously detract from efficient interpretation of the network. A second reason, and our original motivation for this research, is to better understand which visual features should be preserved during interactive network visualisation applications or when showing a dynamic network. In such applications we do not want to the user to notice unnecessary changes to the layout, i.e. we wish to preserve the user's *mental map* [28] of the diagram. A third reason is that understanding which visual features are noticed is useful when designing layouts that facilitate the visual comparison of networks by emphasising the similarities and differences between them.

While designers of network layout and graph drawing algorithms focus on embodying well-known layout aesthetics in their algorithm design and empirical HCI researchers study the way in which these aesthetics influence users' interpretation of graphs, there are few studies that attempt to determine how network diagrams are represented internally. Of course, we cannot ever know the exact cognitive nature of a person's internal representation (or *mental model* [24]), but we can make reasonable judgements based on appropriate data.

Our broad experimental aim is to try to establish what cognitive structures people build when reading and internalising a graph drawing. Our starting point was to identify visual features relevant to network diagrams that we speculated would be important when these di-

agrams are internalised. Unlike most other studies, we did not use as our starting point the layout aesthetics that are commonly used by graph drawing algorithm designers (and which are assumed to lead to good performance). Rather our research is driven by existing perceptual theories. From these theories, we identified six visual features: symmetry, alignment of nodes, collinearity of edges, axis alignment of edges, parallel edges and orthogonality. Most, but not all, correspond in some way to typical graph layout aesthetics.

Our experimental task was to ask participants to reproduce a small network diagram from memory after seeing it briefly. The task was deliberately low-level and did not involve any higher level cognitive processing tasks, for example, graph interpretation or analysis. Similar recall tasks are a standard method for investigating internal mental representations, e.g. chunking [14, 18].

We performed two analyses on the data. In the first analysis we analysed how the visual features in the diagram impacted the correctness of recall. In the second analysis we investigated the extent to which the visual features were preserved in the participants' drawings.

2 BACKGROUND

2.1 Graph drawing aesthetics

The many graph layout algorithms that have been devised over several decades [4] have typically been designed in accordance with the intuitions of algorithm designers. Over the years, a set of assumed 'graph drawing aesthetics' has emerged, defining the criteria by which the 'goodness' of the drawing produced by a layout algorithm can be assessed [5, 12, 37]. Such aesthetics include, for example, a minimum number of edge crossings, as few edge bends as possible, a display of symmetric sub-structures, and large angles between edges incident at a node. Some empirical work has been done in assessing graph drawings, investigating whether the aesthetic criteria used by algorithm designers do indeed assist with comprehension. Findings include overwhelming evidence for the benefits of reducing edge crossings [22, 47, 36] and some evidence for the reduction of bends and depiction of symmetry [36], placement of important nodes at the top of the graph [23] and large angles between incident edges [22]. All these studies have been conducted by asking participants to answer graph-based questions on a variety of presented graph drawings, each carefully controlled for the aesthetic criteria.

More recently, layout of graphs has been investigated by asking participants to produce their own graph drawings from given graphs (so as to depict clusters [45] and to support given tasks [15]) and to create graph drawings from adjacency lists [38]. Like the comprehension empirical studies, these experiments all revealed support for the minimisation of edge crossings, with support for depiction of clusters [45, 38] and grid-based layout [38]. There was little support for uniform edge lengths [45].

Our work is fundamentally different from these previous studies because we are interested in understanding the internal representation of network diagrams, not the most effective way to present a network.

-
- Kim Marriott, Monash University, e-mail: kim.marriott@monash.edu.
 - Helen Purchase, University of Glasgow, e-mail: helen.purchase@glasgow.ac.uk.
 - Michael Wybrow, Monash University, e-mail: michael.wybrow@monash.edu.
 - Cagatay Goncu, Monash University, e-mail: cagatay.goncu@monash.edu.

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

As a result our starting point are the low-level visual features that perceptual theories suggest might be important, not the aesthetic criteria commonly used in graph layout.

2.2 Internalisation of diagrams

Research on internal (i.e. cognitive) representations has existed for many years: Johnson-Laird [24] proposed the existence of ‘mental models’ of logical propositions that facilitate reasoning and explanation, and much subsequent research (e.g., Stenning [43]) has been built upon this idea. Shepard and Metzler [41] demonstrated that internal models of diagrams were not, however, internally represented as propositions, but were ‘mental images’. Kosslyn and Pomerantz [25] summarise the central ideas of this imagery theory thus:

- An image is a spatial representation like that underlying the experience of seeing an object during visual perception. These images may be generated from underlying abstract representations, but the contents of these underlying representations are accessible only via generation of a surface (experienced) image.
- Only a finite processing capacity is available for constructing and representing images. This limits the amount of detail that may be activated at any one moment.

Palmer [32] addresses the problem that we cannot know the actual form of someone’s internal representation, and defines *informational equivalence* as being the extent to which two representations (internal and/or external) embody the same information. Subsequent extensive research has investigated the use of such internal diagrammatic representations in reasoning and problem solving [21, 43], and the benefit of externalizing these models to assist in learning [13, 1] and thinking [6, 8] has been demonstrated. Asking users to draw their mental models (i.e., *externalise* them) provides a means by which we can *look at* the form of these models, acknowledging, of course, the transcription process may not be exact. Palmer [32] acknowledges that *cognitive psychology can only aspire to informational equivalence between its models and the actual mental representation inside the head.*

2.3 Mental map preservation

Misue et al. [28] were among the first researchers to identify the importance of layout stability (which they called mental map preservation) in interactive network diagram visualisation and layout adjustment. They suggested a number of possible ways to do this that have since been tried: preserving orthogonal ordering (such as in [31]), minimising node movement (e.g., [19]) and preserving topology (e.g., [7, 16]).

Purchase et al. [39] studied whether limiting node movement to preserve the mental map in force-directed layouts for dynamically changing graphs had an effect on efficiency and accuracy in answering questions about the graph structure across the lifetime of the network. They found that participants performed better with low node position preservation or with a high amount, but not so well for a medium amount. This seemed to be because the medium condition often led to crowded tangled layouts whereas the low preservation condition quickly revealed changes in the network structure from relayout and the high condition preserved a reasonable spacious and stable layout.

Lam et al. [27] determined the degree to which various individual geometric transformations could be applied before impacting recognizability of small networks. While the networks in this study were not designed to contain visual features such as we discuss here, their results suggest it could be interesting to investigate how visual features impact recognizability after transformations and whether transformations that preserve visual features improve recognizability.

Our work differs from preceding research in that we wish to experimentally identify low-level visual features of a network diagram that are recalled (either sub-consciously or consciously) by the viewer and so should, where possible, be preserved in the layout.

2.4 Perception of network diagrams

There are many existing theories of perception that relate to the perception of network diagrams. Bennett et al. [5] present a comprehensive overview of all the aesthetic criteria proposed, each with support

from existing literature on perception. The three main theories that are relevant here are Norman’s levels of processing, the Gestalt laws of perceptual organisation, and pre-attentive processing (sometimes referred to as the *pop-out* phenomenon).

Norman divides the process of perception into three levels: visceral, behavioural and reflective [30]. The lowest level is visceral—this is fast, automatic, makes quick judgements, and includes the basic perceptual operations of distinguishing objects. The behavioural level is not conscious: it uses the output of the visceral level, and acts on it, but typically as a result of inherent skills built up with practise. It is this level where the *usability* or *understandability* of a stimulus is most important [5]. The highest level, reflective thought, does not have access to the low-level signals received by the visceral level: it reflects on what is happening at the behavioural level, attempts to find meaning in it, and attempts to influence it. This high level is strongly affected by context, including the culture and experience of the perceiver, and the viewing circumstances [5].

The Gestalt theorists studied the way in which we perceive the form of objects, and provided a set of perceptual rules describing fundamental perceptual phenomena and the patterns that we see in visual stimuli [10, 46]. The laws can be summarized as follows [10]:

Proximity: *Things that are close together are grouped together*

Similarity: *Things that look “similar” are grouped together*

Common Fate: *Things that appear to move together are grouped together.*

Good Continuation: *Perceptual organisation will tend to preserve spatial continuity rather than yielding abrupt changes.* For instance, four line segments in an X shape will be seen as two straight lines rather than as two V shaped lines.

Closure: *Closed shapes are preferred to open shapes.*

Relative size, surroundedness, orientation and symmetry: These effect perceptions of which objects are seen as figures on a background. *Other things being equal, the smaller of two areas will be seen as (a) figure against a larger background.* This effect is enhanced if the larger area surrounds the smaller area, the smaller object has a vertical or horizontal orientation or the smaller object is symmetric.

There is evidence that people remember figures as more symmetric and closed than they really are [44] and that symmetry may aid recall of abstract patterns [40, 26] (though see [9]).

Nesbitt and Friedrich [29] considered the Gestalt Laws in the context of the development of an automatic layout algorithm for evolving graphs, and introduce the additional rules:

Simplicity: *Every stimulus pattern is seen in such a way that the resulting structure is as simple as possible*

Familiarity: *Things are more likely to form groups if the groups appear familiar or meaningful.*

Connectedness: *Things that are physically connected are perceived as a unit.*

They do not mention symmetry, although this Gestalt rule is given some prominence by Ware [46] who regards it as a separate law: *Symmetric things appear to be grouped together.*

Pre-attentive processing occurs before conscious attention, and is thus associated with Norman’s visceral level. This processing identifies the most prominent objects in a scene: these objects are said to ‘pop-out’, and to become the focus of attention. The *pop-out* objects are typically all similar in form, and clearly different from the surrounding objects, thus conforming to the Law of Similarity. Colour is one of the most obvious pre-attentive features: if there are three yellow objects within a scene that is otherwise entirely blue, these three objects will immediately command the attention of the viewer. Ware [46] provides a comprehensive list of pre-attentive features; those that are most relevant to the research reported here are line orientation, line length, line co-linearity, spatial grouping, 2D position and shape.

3 EXPERIMENTAL METHOD

3.1 Experiment overview

We used a memory recall task to investigate the internal representation of network diagrams. We were particularly interested in investigating the effect of various visual features on their recall. This approach is supported by Palmer [33], who explains that Gestalt theorists suggested that the *goodness* property of Gestalt patterns provided an important means of investigating the way in which visual information is internally represented. Garner's extensive experimental research supported this view, including the conclusion that *pattern goodness [...] influence[s] memory processes as simple as those involved in immediate reproduction and free recall* [20].

We created eight layouts of four small graphs, each version conforming to one of six clearly defined experimental visual features. Three additional control versions conformed to none of these features. Participants were shown each one of these 44 drawings for three seconds, after which they were asked to draw what they had seen on a tablet PC, taking as long as they needed to do so.

We anticipated that network diagrams containing these visual features would be easier for participants to internalise and recall, and that we would see their presence in the drawings created by the participants.

3.2 The six experimental features

Our experimental aim was to try to establish whether the cognitive structures people build when reading graph drawings correspond to features suggested by perceptual theories. To do this, we first identified those layout features that we thought (based on our knowledge of perception) might be important in building internal cognitive structures of a graph.

We defined the following six visual features of graph drawings as our experimental conditions:

Symmetry: the nodes and edges of symmetric sub-graphs are displayed in a symmetric pattern (Gestalt law of Symmetry).

Collinearity: edges incident at the same node are drawn such that they appear as a continuous line going through the node (Gestalt law of Continuity, pre-attentive co-linearity)

Node-alignment: nodes are placed such that an imaginary straight line could be drawn through the centre of them all (Gestalt law of Proximity and Good Continuation)

Parallel: two or more edges that do not share nodes are placed parallel to each other (pre-attentive parallelism)

Horizontal-vertical: edges are, as much as possible, placed horizontally or vertically (Gestalt Law of Orientation, pre-attentive edge orientation, [2])

Orthogonal: Edges are, as much as possible, horizontal and vertical and placed at right-angles. This feature is a composite of the collinearity, parallel and horizontal-vertical features.

A seventh *control* feature was a graph representation that included none of these six features—there were therefore seven experimental conditions.

In our pilot studies, participants often mentioned familiar shapes created by closed contours (for example, a butterfly, a house, an person pushing a pram). These perceptions relate to the Laws of Familiarity and Closure [46]; however, they are impossible to control for in the production of stimuli for this experiment, as they are highly context and user dependent.

3.3 The experimental stimuli

We chose four graph structures of similar size, but with differing complexity. Graph 1 (nodes=5, edges=6) comprised two closed contours connected by one node; graph 2 (n=5, e=5) had one closed contour, and a two-node tail; graph 3 (n=5, e=6) was two closed contours connected by two nodes, and graph 4 (n=6, e=6) comprised one closed contour, and a one-node and a two-node tail attached to one of the contour nodes.

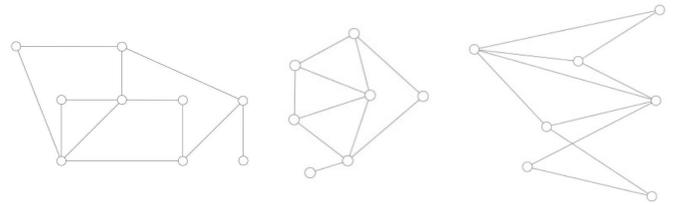


Fig. 1. Examples of large graph stimuli used in pilot tests.

Each graph was drawn once according to each of the seven experimental features. Three of the features (symmetry, horizontal-vertical, control) were also drawn with an alternative topology, and one of these (control-alternative-topology) was also drawn with one edge crossing. Table 1 shows the 44 experimental drawings.

The drawings were created carefully, ensuring that each drawing embodied only one of the six features (or none at all in the case of the control). For example, when there was more than one horizontal edge, we ensured that these were not seen as parallel; symmetry had to be avoided in all but the symmetric drawings.

3.4 Design considerations

The nodes in our stimuli are not labelled. It is easier to create a mapping between the original stimulus and a participant's drawn graph if node labelling is used. However, as the experiment was focussed on the layout of the drawings, and visual features, we did not want the process of internalising and remembering the graphs to be affected by the need to remember labels. We also felt that if labels were used (even alphabetic characters), the graphs may be interpreted as having some semantic meaning, and that any meaning assigned to the graph may impact its reproduction (either positively or negatively). We therefore used unlabelled nodes, to encourage participants to focus on the form of the graph drawing.

We originally designed the experiment with graphs of a much larger size (n=7–9, e=9–11) (see Figure 1), and gave the participants 25s to look at the drawings before drawing them. While many pilot participants were able to recreate some aspects of the stimuli, the drawings were mostly so incorrect that it was impossible to do any reasonable node mapping between the stimuli and the sketched graphs. It was clear that graphs of this size were too hard, and all pilot participants said that they thought that increasing the viewing time would not have made their task easier or improved their performance. We were very surprised by the difficulty that these participants had in comparison with our own performance in pre-pilot trials: it was clear that we had underestimated the effect that our personal expertise in graph drawing would have on task performance when compared with those without such expertise.

Larger graphs would have permitted some perception at the behavioural level. Having decided to use smaller graphs, we switched our focus to the visceral level, and fast, immediate perception. This required a substantial decrease in the time given to participants to look at the drawings, down to 3 seconds.

Participants had as much time as they liked to redraw the drawings—we did not limit this time as we wished to make sure that they had the opportunity to remember as much of the drawings as they could. However, in an attempt to capture their immediate perception of the drawings, they were advised that it would be easier to remember the diagram if they drew it quickly.

Given that the length of time for looking at the drawing and recreating it was to be so short, we needed to ensure that the memory of one drawing would not influence the next. In between each drawing, the participants were asked to count aloud the number of differences in a visual *spot the difference* puzzle. This was displayed for 10 seconds, ensuring that the visual memory buffer was cleared before presentation of the next drawing. Pilot tests revealed that it was important that these spot-the-difference puzzles were easy, as participants became frustrated if they could not find any differences in the allocated time.

Table 1. The 44 experimental drawings.

Feature	Graph 1 (g1)	Graph 2 (g2)	Graph 3 (g3)	Graph 4 (g4)
Collinear (Col)				
Control (Cont)				
Control (with an edge crossing) (ContAC)				
Orthogonal (Orth)				
Horizontal and Vertical (HV)				
Horizontal and Vertical (alternative topology) (HVA)				
Node alignment (NAlign)				
Parallel lines (Par)				
Symmetry (Sym)				
Symmetry (alternative topology) (SymA)				

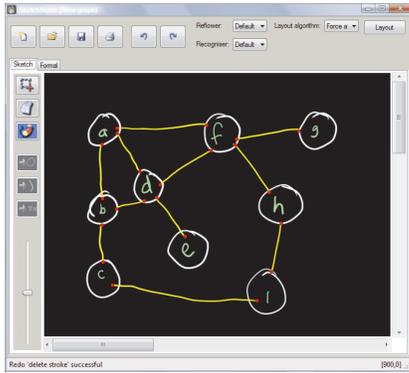


Fig. 2. SketchNode in sketch mode.

This was a within-participants experiment, with each participant’s performance using one experimental condition being compared with their own performance on the other conditions, and all participants seeing all 44 stimuli. Within participant experiments are subject to the learning effect, whereby the participants’ performance improves over time as they have more practise in the task. We addressed this by presenting 8 practise tasks at the start of the experiment—these were randomly rotated and flipped versions of a subset of the experimental stimuli, and the participants did not know that they would not be included in the data collection. We presented the stimuli in a different random order for each participant.

3.5 Experimental equipment

Our experiment focussed on the visceral level: the instantaneous perception of the drawings. The participants were shown the drawings for only 3 seconds, and then asked to recreate them from memory.

We chose 3 seconds by referring to the principle that the visual sensory buffers can store information for approximately 200 ms [11], and that information stored in visual short term memory will usually be lost within 20 seconds if there is insufficient time for processing and rehearsal of the stimulus ([34] as discussed in [3]). We wished to ensure that the participants had sufficient time to perceive the whole drawing, but insufficient time to process and rehearse it so as to better understand (and hence remember) it, ensuring that their recreation of the drawing was simply based on their initial visual perception.

It was important, therefore, that the facilities that they used to create their drawings were easy and natural to use, and that their memory was not inhibited by the use of an awkward graph drawing interface.

We used SketchNode, a graph-drawing sketch tool that allows users to draw graphs on a tablet PC using a stylus [35] (Figure 2). Nodes and edges could be drawn with a stylus on the tablet screen, laid flat, thus allowing the same hand-movements as pen-and-paper, a more natural interaction than using an editing tool. Edges could be created at any time. Unlike pen-and paper, SketchNode allows nodes (or groups of nodes) to be selected and relocated (with corresponding movement of attached edges), and nodes and edges to be erased. It thus has the advantages of pen-and-paper, as well as the benefits of graph editing software.

While we could have used a more formal graph editing system (e.g., yEd Graph Editor), and were aware that if we did so, matching the graph drawings produced against the stimuli would be easier, we chose to use a sketching system in our experiment. Our pilot studies revealed that a formal mode of interaction was too laborious, clumsy, and unintuitive for the participants to be able to draw the graphs as quickly as they needed to before their memory of the stimuli faded.

SketchNode provided several useful experimental facilities: it automatically saved the participants’ drawings as text and images, and created a log file which recorded all actions performed through the interface.

The stimuli were presented on a desktop screen, with the tablet PC placed horizontally on the desk in front. The experimental program

displayed the stimuli for 3 seconds each, displayed the filler tasks for 10 seconds each, and allowed the user to take a break after each block of eight drawings—the length of the break was determined by the participant who indicated when they were ready to proceed with the experiment again.

3.6 Experimental procedure

The experiment was conducted on a one-to-one basis between the experimenter and 25 participants recruited via the Monash University weekly news email to all staff and students. The participants varied widely in age (19 to 60, median 24, mean 28.9) and background (including Arts, Business, Biology, Education, Engineering). This was as we wished it: as our research question is based on general human perceptual theories, we did not seek participants with any particular knowledge or experience with network diagrams.

Participants were first introduced to the concept of a graph as a set of nodes connected by links, and were trained in the use of SketchNode, and given as much time as they needed to get used to it. They then did a small demonstration version of the experiment with five network diagrams created specifically for this purpose—this allowed the participant to get used to the rhythm of the experiment. Some participants did this demonstration experiment more than once, until the experimenter was sure that they were comfortable with the experimental process and the equipment.

For each task, the participants were asked to *Copy the diagram as exactly as you can*. They were also told (and later reminded) that they would find it easier if they drew the diagram quickly.

Each experiment lasted approximately 45 minutes. Of the 1100 data points, one was lost because a participant was momentarily distracted and could not remember the previous diagram, and nine were possibly affected by a minor distraction. These nine data points have been retained in the analysis.

The participants’ drawings were stored both in the xml format (.gml) used by SketchNode and as image files. A log file was generated for each participant that recorded the time and nature of all actions: creating and moving nodes and edges, and use of the *undo* facility.

4 DATA ANALYSIS AND RESULTS

4.1 Pre-analysis data preparation

It was inevitable that the hand-sketched drawings produced by the participants using SketchNode would look very different from the formal stimuli created in yEd. While SketchNode includes functionality for automatically recognising nodes and edges from the input strokes, we found that they were often misrecognised. We therefore defined principles for identifying nodes and edges in the participants’ drawings (Table 2).

4.2 Analysis 1

Analysis 1 focussed on the effect the visual features had on correct recall of the stimuli. A *correct* drawing is one in which there is a single obvious mapping between the individual nodes and edges in the drawing and those of the stimulus, and the drawing is topologically equivalent, i.e., they have the same layout if nodes are allowed to move but may not pass through edges. Note that this implies that the drawing is logically correct in that it represents the same abstract relational information.

The null hypothesis is that the visual features would not impact the correctness of recall.

Each of the 1099 participant drawings was visually scrutinised for correctness, with the decisions agreed by two members of the research team. 789 drawings (72%) were correct with respect to this criterion. The distribution of these drawings to each graph condition and feature is given in Table 4.

As the data was not normally distributed, we used non-parametric analysis methods. A Friedman test was used to determine whether there was any overall main effect of condition on performance, and,

Table 2. Principles for interpreting the sketched drawings.

Principle	Example	Outcome
We ignore the SketchNode automatic object recognition, and consider only the visual image; e.g. if nodes or edges are small, they are recognised by SketchNode as labels; long edges with many bends are recognised as nodes.		p10,g2-ContA is recognised by SketchNode as six nodes (including a large triangular one) and two labels (the smaller edges), but we considered it as five nodes and five edges. Note that we use the notation $p_j,gi-c$ to refer to the drawing by participant j of graph i with condition c .
We interpret an edge to be a sketched line that touches the boundaries of two circles (even if that line has not been recognised by SketchNode as an edge).		The internal edge that goes through a node in p16, g2-SymA is visually interpreted as two edges connecting three nodes.
We do not attempt to second-guess what a participant intended—we take the drawing as it stands.		Whilst we may guess that the participant intended to place a node where the ends of two edges touch each other in p4,g4-Cont, we still classify this drawing as incorrect.
		However, we do change the drawing when there are two edges connecting two nodes: we remove one of these edges (as in p11,g3-HVA).
If two strokes cross, but are incident at the same node, we do not consider them as crossing edges. We assume that these crossed strokes are a consequence of the free-hand drawing action.		We consider the topology of p1,g3-ContAC as if the cross did not exist and the edge connected with the node cleanly.

Table 3. Determining drawing correctness.

Error	Example	Original
No obvious single mapping between the nodes and edges (g1-hv-a)		
Not topologically equivalent (g1-contAC)		
Logically, hence topologically, incorrect (g2-cont)		

Table 4. Number of correct drawings for graph conditions and features.

Graph	Condition	Feature
Orth	80	80
Col	83	83
HV	82	150
HVA	68	-
NAlign	76	76
Par	80	80
Sym	95	184
SymA	89	-

4.3 Analysis 2

The second analysis investigated to what degree the visual features in the stimuli were preserved in the participants' drawings in order to investigate whether the visual features were recalled (either sub-consciously or consciously) by the participant. The analysis relied on mapping the nodes and edges in the stimuli to the participant's drawing so as to determine if the visual features associated with particular nodes and edges were preserved. This meant it could only be performed on those drawings that had been identified as correct in the previous analysis ($n=789$).

The null hypothesis is that the participants did not try and preserve the features but rather tried only to place the nodes as close as possible to their original position in the stimuli. This is, in part, motivated by the common practice of trying to minimise node movement in order to preserve the mental map. In the null hypothesis the placement of each node is independent, and hence if a feature is preserved it is only as a by-product of node positioning. We wished to determine if the features were preserved significantly more often than this.

We chose to use a bootstrapping test [17] because it is relatively simple and because an analytical testing of the distribution of features under the null hypothesis is difficult. The basic idea was to consider the differences (offsets) in node positions between all of a participant's drawings and the corresponding stimuli. Then use these offsets to randomly generate 1000 new sample drawings from the stimuli for each valid participants drawing. For each condition, we then used a metric designed to measure preservation of that feature and see whether the feature was preserved significantly more in the participant's drawing than would be expected from the randomly generated drawings.

We used the metrics described below in our analysis. In participant drawings the edges were not drawn as straight lines directly between nodes. In our metric calculations we used the centre of the node as its position and also as the endpoint of each edge starting or ending at the node.

if a significant effect was found, a post-hoc pairwise comparison non-parametric test (the Nemenyi test [48].) identified where the significant differences lie. In all cases, a significance value of 0.05 was used. We tested for the effect of both the 11 conditions (treating the alternate topology stimuli separately) and the 7 visual features (aggregating over those stimuli based on the same features).

The results for the 11 conditions are presented in Figure 3, and for the 7 features in Figure 4. In terms of conditions we find the following to be statistically significant:

- Cont is worse than all except HV, ContA, ContAC
- ContA is worse than all except HV, ContAC
- ContAC is worse than all except NAlign, HV, HVA
- Sym is better than HV
- SymA is better than HV

And for the feature analysis we find:

- Cont is worse than all except HV
- Sym is better than HV

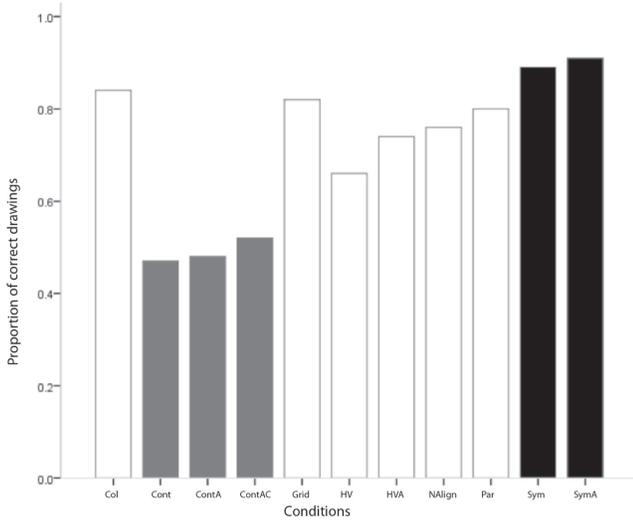


Fig. 3. Analysis 1 results for conditions. $\chi^2(df = 10, p < 0.05) = 18.31$.

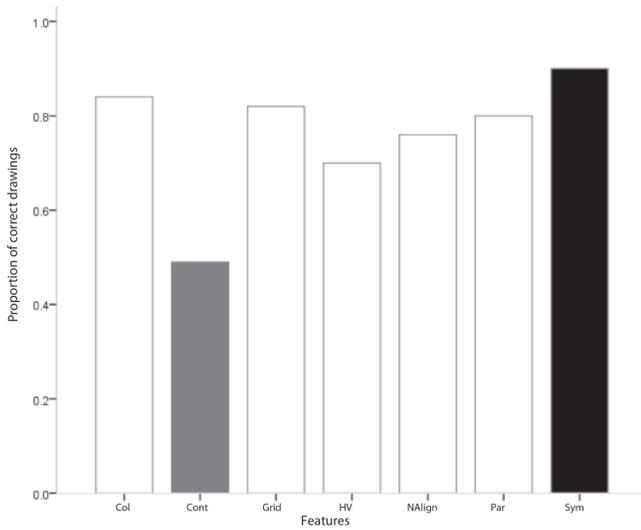


Fig. 4. Analysis 1 results for visual features. $\chi^2(df = 6, p < 0.05) = 12.59$.

Horizontal: $= abs(cos(a))$, where a is the angle between the edge(e_2) and the x-axis (Used for Orth, Horizontal and Vertical stimuli) — Figure 5(a).

Vertical: $= abs(sin(b))$, where b is the angle between the edge(e_3) and the y-axis (Used for Orth, Horizontal and Vertical stimuli) — Similarly to Figure 5(a).

Parallel: $= abs(sin(m_0 - m_4))$, where e_0 and e_4 are the parallel pairs, m_0 is the slope of the first edge (e_0) and m_4 is the slope of the second edge (e_4) (Used for Parallel) — Figure 5(b).

Collinear: $= (d_{12} + d_{20} - s)/s$, where (n_1, n_2, n_0) are the nodes which are on a continuous line s formed by the edges, d_{ij} is the length of the edge between n_i and n_j , and s is the distance between the first (n_1) and the last (n_0) node (Used for Collinear) — Figure 5(c).

Node-alignment: Similar to the collinear metric. However, the continuous line s is formed by joining the nodes rather than joining the edges (Used for Node-Align) — Figure 5(d).

Symmetry: A symmetry feature is given by a set of nodes N_S which lie on the line of symmetry and a set of node pairs N_P in which

the nodes in each pair lie on opposite sides of the line of symmetry. We compute the midpoints N_M of each of the node pairs in N_P and then find the line of total least squares fit L through the points in N_M and N_S as an estimate of the line of symmetry. The symmetry metric of this feature is the root mean square error of this fit. Figure 5(e) shows the calculation for a symmetry in a graph in which single nodes ($N_S = \{n_0, n_1, n_2, n_3, n_4\}$) should be on the symmetry line while node pairs ($N_P = \text{node}_{0,1}$) lie on the opposite sides of this line and their midpoint ($N_M = \text{mid}_{n_0-n_1}$) will be on the symmetry line. (Used for Symmetry)

Orthogonality: $= abs(1 - sin(m_i))$, where m_i is the angle between two orthogonal edges, m_1 for e_0 and e_2 , m_2 for e_3 and e_4 (Used for Orth) — Figure 5(f).

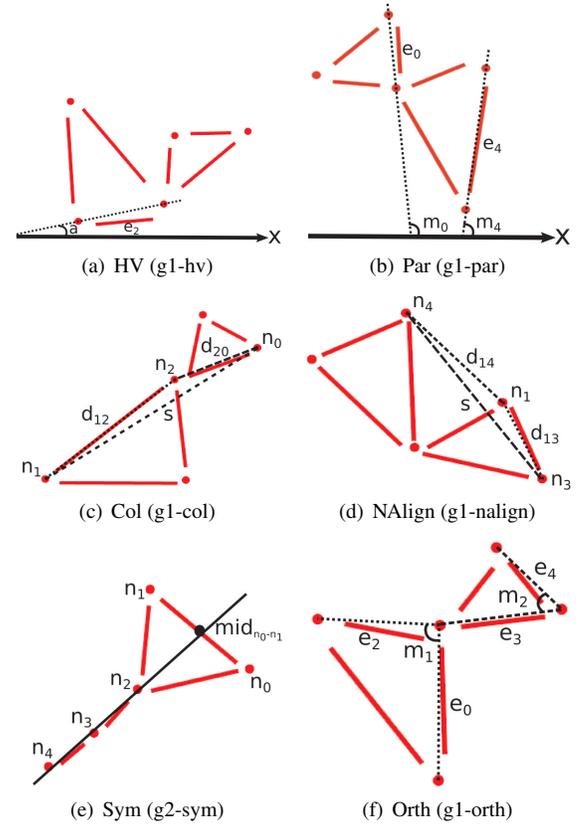


Fig. 5. Metrics used to calculate feature preservation on participant drawings: (a) horizontal and vertical, (b) parallel lines, (c) collinear, (d) node alignment, (e) symmetry, (f) orthogonal.

The analysis proceeded as follows.

1. For each correct drawing we identified the mapping between nodes in the drawing and the stimuli. Node mapping was semi-automatic with hand correction, agreed upon by two members of the research team.
2. For each correct drawing we translated and performed a uniform rescaling into the coordinate frame of the stimulus, since the participant drawings were often varying sizes, and drawn in different areas of the tablet screen. The transformation was chosen to minimize the distance between the node positions in the stimulus and the drawing. We did not perform rotation. The resulting diagrams were called the normalized valid drawings.
3. For each feature F we developed a metric $M_F(d)$ to measure the preservation of that feature in a normalized valid drawing d relative to the stimuli. This was a non-negative measure with 0 indicating full preservation and larger numbers less preservation.

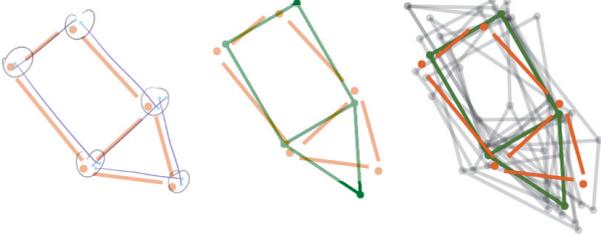


Fig. 6. Example of a participant’s drawing (left), the representation of that drawing that we use for analysis (red, overlaid), the equivalent representation of the original stimulus (green), and ten overlaid randomly generated drawings using the original stimulus combined with random (x,y) pair differences from drawings produced by that participant.

4. For each feature F we got a measure of its preservation M_F which was simply the sum of $M_F(d)$ for each normalized valid drawing d of a stimuli with that feature, across all participants.

We used bootstrapping to generate a distribution for M_F under the null hypothesis and measure the significance of feature preservation in the participants’ drawings.

5. We collected for each participant p the set S_p of (x,y) pair differences between the position of a node in the stimuli and its position in the normalized valid drawing.
6. For each valid drawing drawn by participant p for stimulus d , we generated a numbered set of 1000 random drawings $R_{p,d}$ by randomly selecting an offset pair from S_p to move each node in d , e.g., Figure 6.
7. For each feature, we calculated the set of 1000 M_F preservation measures values, where each was the sum of the metric values for each correspondingly numbered random drawing in $R_{p,d}$ with that feature.
8. For each feature, we then calculated the probability (p-value) of getting values of $M_F' < M_F$, which represents the feature not having been preserved purely by chance.

The following tables show the results of the analysis described above. These tables include the metric significance as well as average metric values for the participant and random graphs for each condition and feature, respectively. Table 5 shows that Orth, Col, Sym and SymA conditions have significance at 0.05. Table 6 shows that Orth, Col and Sym features have significance at 0.05 and HV at 0.10.

A statistically significant result here, such as for the orthogonal, collinear and symmetry features, indicates that the (correct) participant drawings exhibited significantly more presence of the given feature than was seen in the corresponding 1000 randomly generated drawings for that participant and drawing—effectively significantly more than would be expected by chance if that participant were solely reproducing the stimuli based on preserving node positions.

5 DISCUSSION

The first analysis supports our hypothesis that the presence of visual features will aid accurate recall of a network diagram. We found that all features other than horizontal and vertical (HV) performed significantly better than the control. Also, we found that symmetry performed significantly better than horizontal and vertical.

The second analysis provided support for the hypothesis that visual features in the stimulus would be recalled. Orthogonality, collinearity and symmetry are preserved with high significance ($p < 0.01$) and there is some indication ($p = 0.07$) that horizontal and vertical edges are preserved. There is no support for preservation of node-alignment or parallel edges: this suggests that these visual features were not noticed either consciously or subconsciously by many of the participants.

Table 5. Metric significance and values for graph conditions.

Condition	Significance p-value	Participant graphs Average (Std.Dev.)	Random graphs Average (Std.Dev.)
Collinear	< 0.001	0.01 (0.01)	0.03 (0.03)
HoriVert-1	0.1603	0.37 (0.07)	0.41 (0.14)
HoriVert-2	0.1275	0.51 (0.09)	0.56 (0.09)
NodeAlign	0.5036	0.01 (0.01)	0.01 (0.01)
Orthogonal	< 0.001	0.4 (0.17)	0.64 (0.22)
Parallel	0.9744	0.14 (0.05)	0.15 (0.06)
Symmetry-1	< 0.001	380.78 (173.72)	1106.62 (294.6)
Symmetry-2	< 0.001	357.94 (183.32)	1498.19 (433.21)

Table 6. Metric significance and values for graph features.

Feature	Significance p-value	Participant graphs Average (Std.Dev.)	Random graphs Average (Std.Dev.)
Collinear	< 0.001	0.01 (0.01)	0.03 (0.03)
HoriVert	0.0671	0.43 (0.09)	0.47 (0.13)
NodeAlign	0.5036	0.01 (0.01)	0.01 (0.01)
Orthogonal	< 0.001	0.4 (0.17)	0.64 (0.22)
Parallel	0.9744	0.14 (0.05)	0.15 (0.06)
Symmetry	< 0.001	369.73 (179.9)	1296.02 (396.98)

Of course, a caveat is that the node-alignment and parallel edges used in our stimuli involved only a few nodes or were visually separated by other edges: if they were more prominent it is likely that they may have been preserved.

Together, these results are useful for visualization practitioners who may wish to selectively employ or exclude certain visual features so as to emphasise and highlight particular data in their visualizations without giving implied significance to other unimportant information.

6 CONCLUSION

Our experimental conditions were carefully chosen to relate to existing theories of perception, in the expectation that using stimuli that emphasise visual features reported to be prominent in perception would assist in internalising and recalling network diagrams.

Our results confirm that presence of visual features based on well-known perceptual theories do indeed lead to better recall than for diagrams without any of these features. This was true for all the features we considered, other than horizontal and vertical edges. More interestingly, the results show that users notice—either consciously or subconsciously—certain visual features and preserve these when correctly recalling graphs they have seen. The strongest of these visual features are symmetry, collinearity and orthogonality. There is also some support for horizontal and vertical edges.

Perhaps surprisingly, two visual features based on well-established perceptual theories—node-alignment and parallel lines—are not significantly preserved by participants in their correctly recalled drawings. This could possibly mean that they did recall but participants are not consciously aware of them and thus do not reproduce these features in their drawings. Or it could be that the stimuli we used were just not complex enough to demonstrate their use effectively.

The latter point relates to experimental design: our pilot tests showed that larger graphs that showed the visual features to their advantage were too difficult for participants to remember correctly, and that displaying the stimuli for more than 3 seconds made the overall task too easy. An experiment with larger and more complex stimuli which emphasise the visual features to a greater or lesser extent would be unlikely to be able to use the same memorise-and-reproduce procedure—alternative methodologies could include asking participants if they have seen particular drawings before or asking them to rank them in order of ease of memorising.

There is still much work to be done on the analysis of this rich dataset. It will be interesting to consider common graph drawing aesthetics (e.g., angular resolution or orthogonal ordering) and see how these might be preserved between the stimuli and participant’s drawings. Analysis of the process (log and video) data will allow us to investigate the order in which the graph elements were drawn, and

whether any of the visual features were drawn during an initial period of confidence. We may also look at whether participants introduced any visual features into diagrams where these features did not exist, for example, linking objects in some way to aid memory.

We believe our work is an important first step in understanding the internal representation of network diagrams. This is interesting in its own right but also has significant implications for how to layout network diagrams in information visualisation applications. It suggests visual features that can be used to emphasise relationships between nodes in a static layout, and which visual features should be preserved when visualising dynamic network diagrams or when comparing two networks. In particular it suggests that in dynamic network diagram visualisation applications as well as preserving the approximate location of nodes it is important to preserve visual features such as symmetry, collinearity and orthogonality.

ACKNOWLEDGMENTS

We thank the participants, Hong Yul (who implemented SketchNode and prepared it for the experiment), John Hamer (who developed the experimental system), and Jane Nurton. This work was supported by the ARC Discovery Project Grants DP0987168 and DP110101390.

REFERENCES

- [1] S. Ainsworth. Deft: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(6):183–198, 2006.
- [2] S. Appelle. Perception and discrimination as a function of stimulus orientation: The oblique effect in man and animals. *Psychological Bulletin*, 78(4):266–278, 1972.
- [3] A. Baddeley and E. Warrington. Amnesia and the distinction between long- and short-term memory. *Verbal learning and verbal behavior*, 9:176–189, 1970.
- [4] G. Battista, P. Eades, R. Tamassia, and I. Tollis. *Graph Drawing: Algorithms for the Visualisation of Graphs*. Prentice Hall, 1998.
- [5] C. Bennett, J. Ryall, L. Spalteholz, and A. Gooch. The aesthetics of graph visualisation., 2007.
- [6] A. Blackwell. *Thinking with Diagrams*. Kluwer Academic, 2001.
- [7] S. S. Bridgeman, J. Fanto, A. Garg, R. Tamassia, and L. Vismara. InteractiveGiotto: An algorithm for interactive orthogonal graph drawing. In *Proceedings of the 5th International Symposium on Graph Drawing (GD'97)*, volume 1353 of *LNCS*, pages 303–308. Springer, 1998.
- [8] P. Brna, R. Cox, and J. Good. Learning to think and communicate with diagrams: 14 questions to consider. *Artificial Intelligence Review*, 15(1-2):115–134, 2001.
- [9] M. B. Brodeur, M. Chauret, G. Dion-Lessard, and M. Lepage. Symmetry brings an impression of familiarity but does not improve recognition memory. *Acta Psychologica*, 137(3):359–370, 2011.
- [10] V. Bruce, P. Green, and M. Georgeson. *Visual Perception: physiology, psychology and ecology*. Psychology Press, 2003.
- [11] S. K. Card, T. Moran, and A. Newell. *The Model Human Processor: An engineering model of human performance.*, pages 1–35. Wiley, 1986.
- [12] M. Coleman and D. Stott Parker. Aesthetics-based graph layout for human consumption. *Software—Practice and Experience*, 26(12):1415–1438, 1996.
- [13] R. Cox and P. Brna. Supporting the use of external representations in problem solving: the need for flexible learning environments. *Journal of Artificial Intelligence in Education*, 6(2-3):239–302, 1995.
- [14] A. D. de Groot. *Thought and choice in chess*. The Hague: Mouton, 1965. (Original work in Dutch published in 1946).
- [15] T. Dwyer, B. Lee, D. Fisher, K. Quinn, P. Isenberg, G. Robertson, and S. North. A comparison of user-generated and automatic graph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):961–968, 2009.
- [16] T. Dwyer, K. Marriott, and M. Wybrow. Topology preserving constrained graph layout. In *Proceedings of the 16th International Symposium on Graph Drawing (GD'08)*, volume 4517 of *LNCS*, pages 230–241. Springer, 2009.
- [17] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1994.
- [18] D. Egan and B. Schwartz. Chunking in recall of symbolic drawings. *Memory & Cognition*, 7:149–158, 1979. 10.3758/BF03197595.
- [19] Y. Frishman and A. Tal. Online dynamic graph drawing. In *Eurographics/IEEE-VGTC Symposium on Visualization*. Eurographics Association, 2007.
- [20] W. Garner. *The Processing of Information and Structure*. Lawrence Erlbaum Associates, 1974.
- [21] J. I. Glasgow, N. H. Narayanan, and B. Chandrasekaran. *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. AAAI Press, 1995.
- [22] W. Huang. Using eye-tracking to investigate graph layout effects, 2007.
- [23] W. Huang, S. Hong, and P. Eades. Effects of sociogram drawing conventions and edge crossings in social network visualisation. *Journal of Graph Algorithms and Applications*, 11(2):379–429, 2007.
- [24] P. Johnson-Laird. *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*. Cambridge University Press, 1983.
- [25] S. Kosslyn and J. Pomerantz. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 9:5276, 1977.
- [26] H.-C. Lai, S. H.-L. Chien, and W.-Y. Kuo. Visual short-term memory for abstract patterns: Effects of symmetry, element connectedness, and probe quadrant. *Journal of Vision*, 9(8), 2009.
- [27] H. Lam, R. Rensink, and T. Munzner. Effects of 2D geometric transformations on visual memory. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization (APGV'06)*, pages 119–126, 2006.
- [28] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of Visual Languages and Computing*, 6(2):183–210, 1995.
- [29] K. Nesbitt and C. Friedrich. Applying gestalt principles to animated visualisations of network data., 2002.
- [30] D. Norman. *Emotional Design: why we love (or hate) everyday things*. Basic books, 2004.
- [31] S. C. North and G. Woodhull. Online hierarchical graph drawing. volume 2265 of *LNCS*, pages 232–246. Springer, 2002.
- [32] S. Palmer. *Fundamental aspects of cognitive representations.*, pages 259–303. Erlbaum, 1978.
- [33] S. Palmer. *Goodness, Gestalt, Groups and Garner*. American Psychological Association, 1991.
- [34] L. Peterson and M. Peterson. Short-term retention of individual verbal items. *Experimental psychology*, 58:193–198, 1959. As discussed in Badelly and Warrington.
- [35] B. Plimmer, H. Purchase, and H.-Y. Yang. Sketchnode: Intelligent sketching support and formal diagramming, 2010.
- [36] H. Purchase. Which aesthetic has the greatest effect on human understanding? In *Proceedings of the 5th International Symposium on Graph Drawing (GD'97)*, volume 1353 of *LNCS*, pages 248–261. Springer, 1998.
- [37] H. Purchase, R. Cohen, and M. James. Validating graph drawing aesthetics. In *Proceedings of the 3rd International Symposium on Graph Drawing (GD'95)*, volume 1027 of *LNCS*, pages 435–446. Springer, 1996.
- [38] H. Purchase, C. Pilcher, and B. Plimmer. Graph drawing aesthetics created by users not algorithms. *IEEE Transactions on Visualization and Computer Graphics*, page (preprint), 2010.
- [39] H. Purchase and A. Samra. Extremes are better: Investigating mental map preservation in dynamic graphs. In *Diagrammatic Representation and Inference*, volume 5223 of *LNCS*, pages 60–73. Springer, 2008.
- [40] M. M. Schmore and J. T. Partington. Immediate memory for visual patterns: Symmetry and amount of information. *Psychonomic Science*, 8:421–422, 1967.
- [41] R. N. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 171:701703, 1971.
- [42] R. Spence. *Information Visualization*. ACM Press, 2001.
- [43] K. Stenning. *Seeing Reason: Image and Language in learning to think*. Oxford University Press, 2002.
- [44] B. Tversky. Visuospatial reasoning. *The Cambridge handbook of thinking and reasoning*, pages 209–249, 2005.
- [45] F. van Ham and B. Rogowitz. Perceptual organisation in user-generated gaph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1333–1339, 2008.
- [46] C. Ware. *Information Visualisation: Perception for Design*. Elsevier, 2004.
- [47] C. Ware, H. Purchase, L. Colpoys, and M. McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2):103–110, 2002.
- [48] J. Zar. *Biostatistical Analysis*. Prentice Hall, 5th edition, 2009.