On the Perception of Small Sub-graphs

Jacob Miller[1,2,3,4(✉)] , Mohammad Ghoniem[1,2,3,4] , Hsiang-Yun Wu[1,2,3,4] ,
and Helen C. Purchase[1,2,3,4]

[1] Department of Computer Science, University of Arizona, Tucson, USA
jacobmiller1@arizona.edu
[2] Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg
[3] St. Pölten University of Applied Sciences, Sankt Pölten, Austria
[4] Department of Human-Centred Computing, University of Monash, Melbourne,
Australia
helen.purchase@monash.edu

Abstract. Interpreting a node-link graph is enhanced if similar sub-graphs (or 'motifs') are depicted in a similar manner – that is, they have the same visual form. Small motifs within graphs may be perceived to be identical when they are structurally dissimilar, or may be perceived to be dissimilar when they are identical. This issue primarily relates to the Gestalt principle of similarity, but may also include an element of quick, low-level pattern-matching. We believe that if motifs are identical, they should be depicted identically; if they are nearly-identical, they should be depicted nearly-identically. This principle is particularly important in domains where motifs hold meaning and where their identification is important. We identified five small motifs: bi-cliques, cliques, cycles, double-cycles, and stars. For each, we defined visual variations on two dimensions – same or different structure, same or different shape. We conducted a crowd-sourced empirical study to test the perception of similarity of these varied motifs, and found that determining whether motifs are identical or similar is affected by both shape and structure.

Keywords: Perception · Graph Motifs · Gestalt Principles · User Study

1 Introduction

As they attempt to understand complex natural, technical or social phenomena, application domain experts use graph models to analyze intricate relationships. They often resort to graph visualization tools to navigate in and make sense of such data [22]. Much work has also looked at the automatic extraction of graph motifs [29], and at counting them [18], to characterize important graph-level properties. Graph motifs are simply local connectivity patterns, or small sub-graphs, lying within graphs; see Fig. 1(a). They can be elementary motifs, like triangles, or assembled into higher-order motifs, like chains and cycles [16]. Given their role in anticipating the behavior of the graph at hand, there is so

far little work dedicated to the visual perception of graph motifs. In particular, existing graph visualization tools are usually not designed to ensure that similar graph motifs take a similar graphical form.

Figure 1 uses a five-cycle motif to exemplify the problem of visual matching of graph motifs. Motif nodes are highlighted in black, while other nodes are grey. In Fig. 1(a), the base shape places the motif nodes on the vertices of a regular pentagon. In Fig. 1(b), the same structure is also drawn as a regular pentagon, subject to a rotation. Compared to the drawing of the base shape, one might hypothesize that users will recognize that the two motifs are identical. In Fig. 1(c), the motif is drawn as a quadrilateral, with the fifth node lying within, making it harder to match it to the base shape. In Fig. 1(d), the pentagon shape is preserved despite the addition of an edge to the motif. The rationale is that a small structural change should be commensurate to the induced change of shape. In Fig. 1(e), both the structure and the shape are different to the base motif and its shape, which users should see easily. Section 3.1 discusses the full list of motifs considered in this study (see Figs. 3, 4, 5, 6 and 7).

In this paper, we explore the perception of (identical or similar) sub-graphs by conducting an empirical study which asks participants to compare (identical or similar) depictions of sub-graphs. Our results include evidence that depicting identical sub-graphs differently makes them harder to recognise as the same, and depicting different sub-graphs similarly results in false identically judgements.

2 Background and Related Work

Similarity is essential in knowledge development as it allows us to organize principles to classify, form, and generalize concepts [40]. It also serves as a common measure for comparison purposes or tasks in data visualization [23]. Specific shapes in visualization, such as *Star Glyphs* [12], *Scatterplots* [30], and *Directed Acyclic Graphs (DAGs)* [2], have been considered as factors that influence human similarity perception. One particular usage of similarity in graph analytics is *motif analysis*, because structural motifs in graphs often act as influential building blocks in many domains [37], such as biology [33], social science [45], internet communication, and others. Therefore, visually identifying these sub-graph structures (motifs) facilitates effective comparisons among various data.

2.1 Perception of Similarity and Shape

Our research questions draw on well-established research on human perception. Ware's three levels of perceptual processing [44] start with a 'bottom-up' stage that processes low-level visual properties in parallel, identifying, e.g., colour, texture, movement etc.; it is quick, automatic and data-driven. The 'pattern recognition' second level comprises sequential processing of the scene, recognising patterns, contours, and regions. The slower third level 'top-down' phase sequentially explores the scene to identify objects, typically engaging cognition
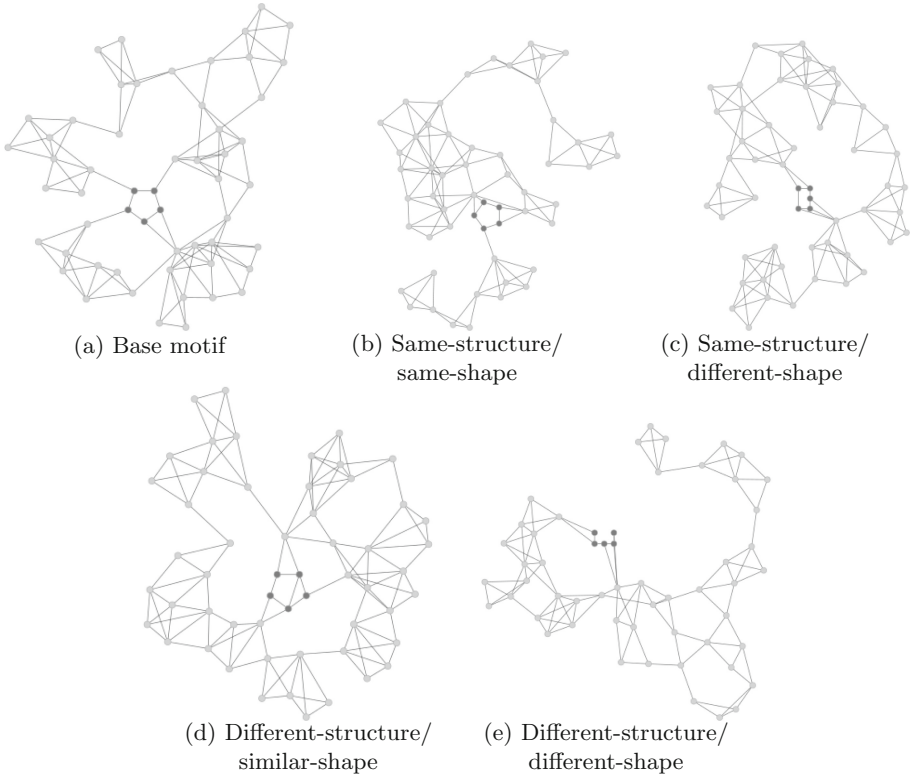
(a) Base motif

(b) Same-structure/
same-shape

(c) Same-structure/
different-shape

(d) Different-structure/
similar-shape

(e) Different-structure/
different-shape

Fig. 1. Example 'cycle' motif drawings. (a) shows the base shape and structure are regular and well-formed; nodes are placed on a pentagon. (b)–(e) show variations of shape (visual form) and/or structure (connectivity).

for the completion of a specific task. The first level includes the immediate identification of prominent objects which 'pop-out', being of obvious different visual form to those surrounding them. The visual features that result in pop-out are of varying effectiveness: colour is the most obvious one; others include texture, orientation, size, shape, curvature [39]. The Gestalt laws [8,20] describe how we see patterns and groups, and relate to the second of Ware's levels. Distinct objects may be seen to form a group by being close together (proximity), looking similar (similarity), or moving together (common fate), etc. This paper considers the quick recognition of shapes and the Gestalt law of similarity in graph drawing: if two sub-graphs are of the same structure, then depicting them in the same visual form will ensure that they can be quickly recognised as the same; if they are of similar structure, then depicting them in the same (or similar) visual form will highlight their similarity. We investigate the immediate recognition of same (or similar) sub-graphs rather than serial processing requiring the use of cognition because making similarities immediately prominent can help in gaining a better (and quicker) overall understanding of the structure of a graph at a

glance. Comparing sub-graphs with enough time to engage cognitive processes is simple; making such comparisons quickly may not be. Our research questions focus on determining whether depicting identical sub-graphs using the same (or similar) shape facilitates the recognition of their identicality. We also explore what happens when the sub-graph structure is slightly different, and when its presentation is distorted by additional forces within a force-directed algorithm.

2.2   Related Work

Prior work has investigated the perception of graph properties, such as graph density, clustering coefficient [36] and layout types [21]. Although the perception of shapes in sub-graphs *per se* has not yet been fully investigated with respect to sub-graph structure, some studies of shapes in visualization have been conducted. Gogolou et al. investigated if time series visualizations generated from automatic similarity measures are aligned with readers' similarity constraints [14]. They concluded that the selection of visualizations influences the patterns that readers consider as similar. Ballweg et al. researched the influencing factors of directed acyclic graphs (DAGs) when they are drawn as layered drawings [2]. Their study suggests that the similarity perception of DAGs is mainly affected by the number of levels in the drawings, the number of nodes on a level, and the corresponding overall shape (i.e., convex hull of the DAG), while Wallner et al. [41,42] did not find significance in the perception of overall shapes. Tarr and Pinker studied the effect of mental rotation in shape recognition, focusing on letter-like asymmetrical characters [38]. They found that trained readers can recognize the characters almost as fast at all familiar angles, while the performance varies when the character appears at novel angles.

Besides perceptual studies, several graph drawing algorithms incorporate shapes (i.e., motifs from different graph classes [1]) to improve layout readability. Yuan et al. developed an algorithm based on Laplacian constrained distance embedding to control the shapes of sub-graphs input by the users [48]. Wang et al. generalized the classical stress majorization approach [43], to ease sub-graph shape manipulation. Meidiana et al. extended the classical force-directed and stress minimization algorithms to optimize shape-based metrics measure [11] to achieve faithful drawings [28]. Interactions through aggregation [24] and simplification [10] also consider motif properties to reduce layout visual complexity.

In practice, many applications consider the visual form of motifs, including transcriptional regulation networks analysis [17], phylogenetic trees comparison [6], biological pathways diagrams [48], metro maps creation [3,46], dynamic graph analysis [7], and visual graph matching [15], for example, across distinct layers of a multilayer graph [27]. In this paper, we demonstrate the benefit of depicting motifs within graphs in a consistent, well-formed manner.

3   Methodology

3.1   Stimuli

Structural motifs often bind with semantics in applications for analysis purposes (Sect. 2.2) [16]. For example, a clique (or a bi-clique) in social networks represents strong mutual connections [45], and a cycle in biological networks indicates a circular biochemical reaction [26, 34]. We hence identified five 'motifs': small sub-graphs with well-defined graph structure, distinct visual form, easy description and clear definition. We denote these motifs (the graph structure along with node positions) as our 'base' motifs; see Fig. 2. For each base motif, we created 12 variations according to two dimensions: change of structure (Yes/No) and change of shape (Yes/No), where shape is determined by the convex hull of nodes (smallest convex set containing all nodes). There are hence four quadrants.
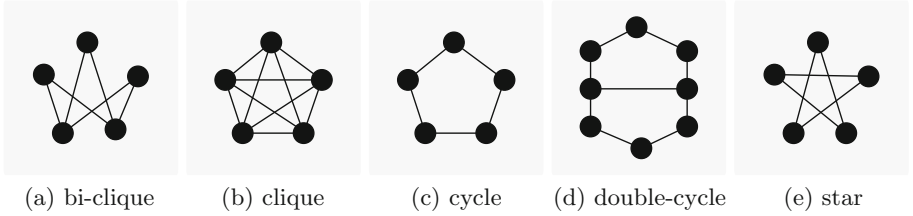

(a) bi-clique   (b) clique   (c) cycle   (d) double-cycle   (e) star

Fig. 2. The five base motifs.

The **same-structure/same-shape (SS)** variations depict the same graph structure and same visual form as the base, rotated by 90, 180, 270° (45, 90 and 135° for double-cycles which have two axes of symmetry (Fig. 5 (top left)).

The **same-structure/different-shape (SD)** variations use the same graph structure as the base motif but a different shape; we generated three different drawings by varying the relative positions of nodes (e.g. Fig. 3 (bottom left)).

The **different-structure/similar-shape (DS)** variants adopt a different graph structure to the motif; we change the number of edges (deleting one, deleting two, or adding one) while keeping node positions as in the base – giving them a similar (but not identical) shape (e.g. Fig. 3 (top right)). As edges can not be added to the clique motif, we remove three edges (Fig. 6 (top right)).

The final **different-structure/different-shape (DD)** collection adapts the original motif sub-graph as for the DS variants, but depicts them using different node positions from the base motif. The three DD variants of the base motif differed both by structure and visual form (e.g. Fig. 3 (bottom right)).

The motif variations are: *bi-clique* (Fig. 2(a)), *clique* (Fig. 2(b)), *cycle* (Fig. 2(c)), *double-cycle* (Fig. 2(d)), and *star* (Fig. 2(e)).

Each motif set therefore comprises 13 different visual motif drawings which were integrated into graphs, a total of 65 graph drawings. The graphs and graph drawings were subject to both structural and layout constraints (Sect. 3.4). Motif nodes were coloured black, all others grey.
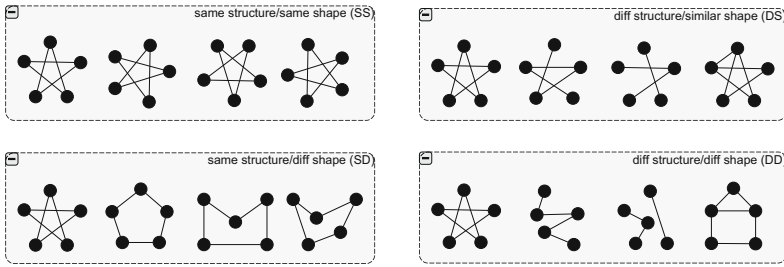
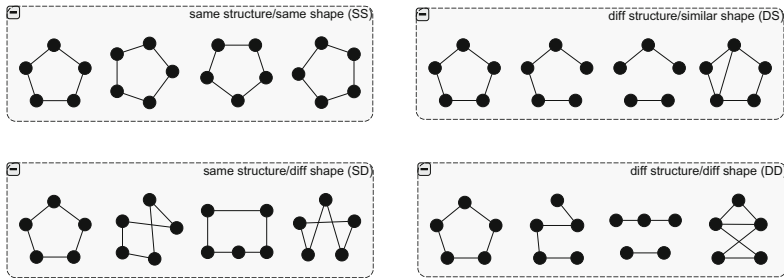**Fig. 3.** 12 variations in shape and/or structure of the *star* motif.

Fig. 4. 12 variations in shape and/or structure of the *cycle* motif.

3.2   Experimental Design

For each motif, we explored participants' ability to assess the identicality or similarity of motifs in two graph drawings, with each participant working with only one motif. In Experiment 1, we explored the effect of motif shape; in Experiment 2, we explored the effect of motif shape distortion. Although the 48 trials for Experiment 1 were interspersed with the 40 trials for Experiment 2, since their aims and stimuli were different, this first section focuses on Experiment 1.

Each trial consisted of a pair of graph drawings displayed side-by-side, with each drawing containing a highlighted motif. Participants were asked to indicate the degree of similarity between the structure of the two motifs. For each motif, the base motif graph drawing (or one of its rotations) was paired with all other 12 motif variant drawings, giving a total of 48 trials. The position of the two drawings in each trial (left or right) was randomly determined. Figure 8 shows the case of a rotated star motif on the right, with the base star motif on the left.

Since we are interested in bottom-up immediate processing, and how known shapes can be identified quickly, we limited the length of time allocated for each trial to 4 s – a duration determined by pilot testing of the complete experiment with 15 participants which considered 3, 4, and 5 s as possibilities: 4 s produced sufficient variability in accuracy to avoid floor (too hard) and ceiling (too easy) effects. Participants judged similarity on a scale by choosing among: '*identical*', '*similar*', '*not so similar*' and '*completely different*'.
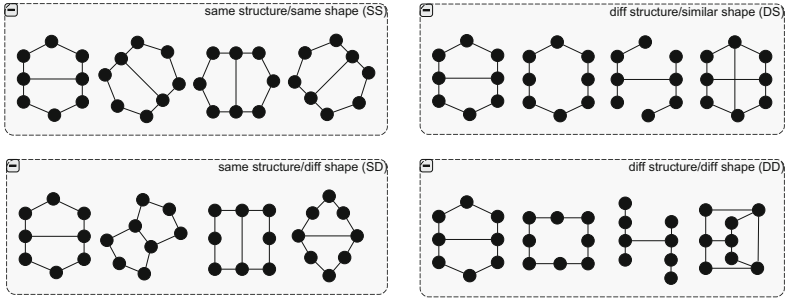
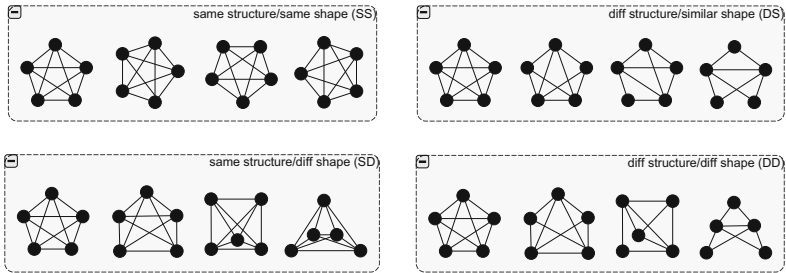**Fig. 5.** 12 variations in shape and/or structure of the *double cycle* motif.

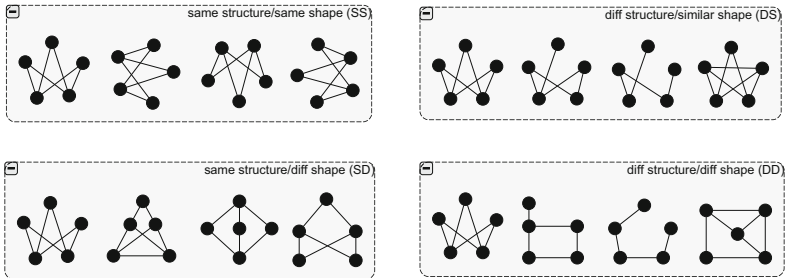**Fig. 6.** 12 variations in shape and/or structure of the *clique* motif.

Fig. 7. 12 variations in shape and/or structure of the *bi-clique* motif.

3.3   Experimental Procedure

We used Prolific [31] to recruit participants for our study and Qualtrics [32] to implement the survey. The simplicity of our experiment makes it suited for crowd-sourcing [5], enabling collection from a large number of participants per motif. Participants were paid £9 per hour, with a median completion time

of 10.4 min. They were recruited from the UK and required to use a desktop machine (no mobile devices). We ran five within-subject studies, one for each motif, and 30 participants per motif. 46% identified as women, 52% as men, 1% as non-binary/gender diverse, 1% declining to answer. 38% were 18–35, 28% were 36–45, 15% were 46–55, and 19% over the age of 55.

Participants were shown six 'practice' stimuli pairs for which data was not collected. These helped mitigate against the learning effect that can affect within-subject studies, as did the randomization of trials for each participant. The data from participants with response rate less than 85% were discarded, and missing responses from the remaining participants were not included in the calculation of mean accuracy data. Each participant saw each trial exactly once.

3.4   Implementation

Graph Generation and Motif Integration. The motifs were integrated into larger 50-node graphs, generated using a k-nearest neighbor model with k = 3 [9,25]. Each graph motif was 'stitched' into the larger graph by connecting each motif node to up to three non-motif nodes, chosen uniformly at random (though we modify the graph post-layout). Since the drawings focus on the highlighted motifs, the specific properties of the overall graphs (beyond their size) is not relevant. A different graph was created for each of the 13 stimuli, for each motif. The constraints on the graph generation were: (**GC1**) All motif nodes must have at least one edge to a non-motif node. (**GC2**) The graph must be connected.

Graph Drawing. We use multidimensional scaling (MDS) for graph layout [13,49] for its ease of use in adding constraints and reliability in representing geometric structures. The constraints on the creation of the graph drawings were: (**GD1**) Motif nodes are fixed on the plane, so they keep the desired visual form, with the rest of the graph laid out around them. This is the basis of our research. (**GD2**) No motif node may be on the periphery. (**GD3**) Edges connecting non-motif nodes must not intersect edges within the motif, since this would change the visual form of the motif and will hamper its recognition in an uncontrolled manner.

We modify the MDS algorithm with an extra parameter F, standing for the set of motif nodes whose position is fixed: $F = \{v | v$ is a node in a motif $\}$. Each member of F is 'pinned' in the plane so that their positions will not be updated, but the surrounding graph drawing will still be based on graph-theoretic distances. More formally, given a graph-theoretic distance matrix d, and a set of nodes with fixed position F, we want to find

$$\min_{X_1,\dots X_n} \sum_{i \notin F} \sum_{j \notin F} (||X_i - X_j|| - d_{i,j})^2 \tag{1}$$

If a drawing produced by Eq. 1 does not meet these requirements, we attempt to modify the drawing to fit. Since the constraints are not always simultaneously
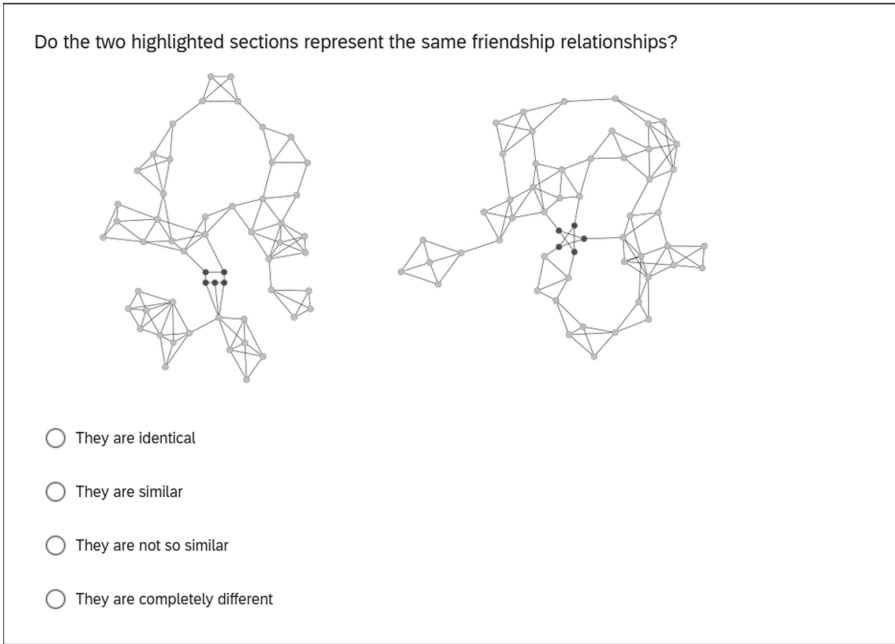
Fig. 8. Example trial – star: same-structure/different-shape (left); base (right).

satisfiable, when they conflict we discard that drawing and try again with a new random graph. To satisfy **GD2**, we compute the convex hull of the layout, and if any motif node lies on the hull, we try again.

GC1 and **GD3** often conflict. To resolve, we first look at all possible $O(|V|^2)$ line segments and categorize them as *invalid* if they violate **GD3**, i.e., if they intersect edges within the motif, or as *valid* otherwise. Then, we count and remove any *invalid* non-motif edge (an edge with an endpoint not in a motif). If some motif nodes are left without non-motif edges, we satisfy **GC1** by adding to each such motif node the edge that is *valid* and whose endpoint is nearest. Clearly, **GC1** is satisfied as we have added edges to each node which did not satisfy it, and we can verify that we did not violate **GD3** in the process.

The removal of *invalid* edges might yield a graph that is too sparse to be realistic. We add as many *valid* edges as the *invalid* edges that were removed, shortest length first. Finally, we check if **GC2** is met; if not, we iterate until all constraints are met. All code is available at https://github.com/Mickey253/graph-rep-sym.

Table 1. Participant responses (rows) mapped to graph edit distance (columns). If participant response was *similar* for GED = 2, accuracy is recorded as 0.6.

	GED = 0	GED = 1	GED = 2	GED = 3
identical	1	0.6	0.3	0
similar	0.6	1	0.6	0.3
not so similar	0.3	0.6	1	0.6
completely different	0	0.3	0.6	1

4   Results

4.1   Dependent Variable

Each trial asked the participant whether the motifs were *identical, similar, not so similar* or *completely different*. In the presented social media context, we asked: "*Do the two highlighted sections represent the same friendship relationships?*".

Measuring accuracy as a binary result (the structure is identical or not) removes subjectivity from the response, but ignores the fact that some structures are more different than others. We therefore use an accuracy measure based on graph edit distance (GED) [35]. We determine the GED between a pair of subgraphs: identical (GED = 0); one edge difference (GED = 1); two edges difference (GED = 2); three edges difference (GED = 3). For each of these GEDs, we score the participants' accuracy responses as shown in Table 1.

Experiment 1: Fixed Shape Motifs. The pairs of graph drawing stimuli presented to the participants included the base motif in one drawing, paired with each of the 12 other stimuli. For each of the five motifs, we address the questions:
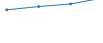
> **Q1:** Does rotation affect the perception of the same sub-graph represented using the same visual form?

We consider the *same-structure/same-shape* variations where the base motif is compared to three of its own rotations. We expect that there will be no difference in accuracy. A repeated measure ANOVA test (Table 2(a)) shows no significant differences, suggesting that the participants were able to correctly recognise the same motif, presented in the same-shape, regardless of its orientation. This result accords with that of Tarr and Pinker who studied the effect of mental rotation in shape recognition, focusing on letter-like asymmetrical characters [38].

> **Q2:** If the motif is adapted to create a sub-graph of different structure, does depicting it in a similar manner affect the ability to distinguish the difference (or would it be better to depict it using a clearly different visual form?)

We compare the results of *different-structure/similar-shape* with *different-structure/different-shape*, expecting that using a similar shape will help in identifying

Table 2. Summary of results for Experiment 1.

(a) Q1: Effect of rotation on determining whether two motifs are identical					
	bi-clique	clique	cycle	double-cycle	star
F (df=2)	0.476	0.783	0.492	0.129	0.146
p	0.623	0.459	0.613	0.879	0.865
(b) Q2: Effect of depicting different structure with similar or different shapes					
	bi-clique	clique	cycle	double-cycle	star
t (df=29)	3.289	2.168	4.230	0.923	3.780
p (1-tailed)	0.001	0.019	<0.001	0.181	<0.001
(c) Q3: Effect of using different shapes for the same sub-graphs					
	bi-clique	clique	cycle	double-cycle	star
t (df=29)	11.071	11.215	13.360	7.807	19.059
p (1-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
(d) Q4: Effect of different edit differences in different sub-graph structures. The trend line shows accuracy from no edit distance (left) to maximum edit distance (right); the accuracy scales are different for each motif					
	bi-clique	clique	cycle	double-cycle	star
F (df=3)	11.817	19.719	3.07	28.850	36.84
p	<0.001	<0.001	0.031	<0.001	<0.001
Trend line					

structural differences in the sub-graphs. A repeated measures t-test (Table 2(b)) showed that the accuracy for *different-structure/similar-shape* was significantly better than *different-structure/different-shape* for all motifs, except double-cycle.

   Q3: Does using a different layout affect perception of motifs of the same structure? If the motifs are identical, does it matter if they are depicted using different visual form – or should they be depicted in the same form?

We compare *same-structure/different-shape* and *same-structure/same-shape*, expecting that using the same shape for the same structure will produce better accuracy than using a different shape for identical structures. A repeated measures t-test (Table 2(c)) showed that accuracy was significantly better for *same-structure/same-shape* than for *same-structure/different-shape*, for all motifs.

   Q4: If different sub-graphs are depicted in similar shape, are near-similar sub-graphs incorrectly assessed as being similar?

That is, does the edit-distance between pairs of sub-graphs depicted in similar visual form affect accuracy when determining whether they are identical or not? Here, we focus on the three different variations in the *different-structure/similar-shape stimuli*, each of which has a different GED with respect to the base
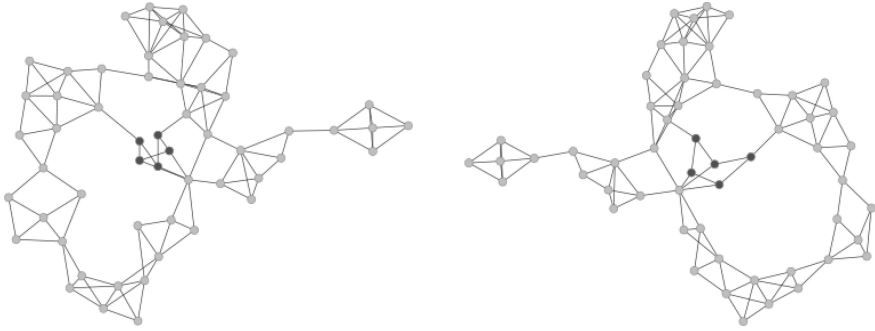
Fig. 9. (left) Cycle motif *same-structure/different-shape* variation introducing unnecessary edge crossings; (right) the same graph drawn using MDS.

motif. We also include one of the *same-structure/same-shape comparisons*, where GED = 0. We expect that where the edit distances are small, the participants were more likely to misjudge two different sub-graphs depicted similarly as identical. A repeated measures ANOVA (Table 2(d)) confirmed that there was an effect of edit distance, for all motifs, where accuracy generally increased with edit distance, with the greatest edit distance accuracy being at least that of the comparisons where GED = 0 (those where the base motif was rotated).

This result confirms our suggestion that the comparisons were made spontaneously – that is, in parallel, rather than serially. If they were serial, participants would have compared the motifs edge-by-edge, and accuracy for the smallest edit distances would have been higher. The spontaneous, parallel pattern matching means that stimuli of the most similar structure are incorrectly seen as identical.

There are a few noticeable data points in this trend analysis: for double-cycle and star, adding two edges leads to lower accuracy – both these extended motifs include at least one edge crossing. Yet, this conclusion does not extend to the bi-clique motif, although we note that the starting point of the bi-clique trend (when GED = 0) is already a low accuracy score.

4.2   Experiment 2: Unconstrained Layout

Here, we used stimuli where the motif nodes positions were 'fixed', but were distorted by MDS forces applied by surrounding nodes. For all Experiment 1 stimuli, we created a paired 'flexible' alternative, removing the three GD constraints of Sect. 3.4, using the unmodified algorithm from [49] (Fig. 9).

Our trials comprised the 'fixed' version of a motif paired with all variants of the 'flexible' version, and participants were asked to judge similarity (as in Experiment 1). In each of the four (shape × structure) quadrants (Sect. 3.1), we pair the three variants with the drawings of the same graphs rendered by an unconstrained layout, resulting in 36 pairs (4 × 3 variants (fixed) × 3 variants (flexible)). We also include the four base motif orientations: a total of 40 pairs for each motif. For each motif, our research question is:

Table 3. The effect of using well-formed shapes to match sub-graphs.

same-structure/same-shape					
	bi-clique	clique	cycle	double-cycle	star
t (df = 29)	8.692	11.611	7.380	10.485	22.254
p (one-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
same-structure/different-shape					
t (df = 29)	6.649	2.582	1.658	5.477	8.847
p (one-tailed)	<0.001	0.007	**0.054**	<0.001	<0.001

Q5: Is it easier to match identical small motifs if they have both been given a well-formed, regular shape (as opposed to one of them having their shape distorted by forces applied to nodes in the rest of the graph)?

We focus on same-structure stimuli, since the question relates to identical structures, considering the shape dimension. The *same-structure/same-shape* stimuli are rotations of the base; the *same-structure/different-shape* stimuli, while using a different shape, are still regular and well-formed. We expect that the accuracy in matching two identical motifs with well-formed shapes will be higher than for the flexible condition (when the motifs are visually distorted). A repeated measures t-test (see Table 3) showed that the accuracy when comparing two motifs with well-formed, regular shape was always significantly better than when one of the motifs had a distorted form, except in the case of the cycle. We note that the alternative well-formed, regular form of the cycle was one where two versions included unnecessary edge crossings; see Fig. 9.

5   Discussion

Our research questions addressed how visual form (shape) affects the interpretation of the abstract structure of a graph. We showed that by depicting two similar sub-graphs similarly, participants achieved high accuracy for all motifs but the double-cycle (**Q2**). The double-cycle is the largest motif of the five, and our shape variations for it are all roughly two polygons sharing a segment. One possible cause for this is that participants were able to mentally transform the base motif (a mirrored pentagon) to the other shapes (triangle, rectangle). As a result, to gain the benefit of representing similar structures as similar shapes, it may be enough that there is a simple transformation between them, e.g., rotation or translation.

The clear recommendation for visualization designers is that when supporting a task to identify sub-graphs one should ensure that identical sub-graphs are drawn similarly. We have shown that rotation does not have any effect (see **Q1**), so this is a degree of freedom in visualization design, e.g., rotation of fixed motifs can be done to reduce edge crossings while maintaining shape. However, the shape of a sub-graph *does* have an effect; depicting identical sub-graphs using

different shapes hampers the recognition of similarity (**Q3**). This means that even if two sections of a node-link diagram represent the same relationships, if they are drawn completely differently, one might never notice. When supporting this type of task, shape cannot be overlooked. Designers can also remember that evaluation of similarity is done quickly (see **Q4**), so should be careful not to depict dissimilar structures using the same shape or one might wrongly interpret the data. The results of both **Q3** and **Q4** seem to indicate that a viewer will interpret similar shapes as having similar structures, regardless of ground truth.

The results of **Q3** and **Q4** can apply beyond the simple five-node motifs we have demonstrated. For instance, graphs with large defined clusters are often depicted with each cluster being drawn as large roughly circular dense regions. If the shape of these regions is roughly the same, a viewer is likely to interpret these large structures as being equivalent. If within-cluster structure is very different between clusters, a designer would want to make this apparent by ensuring these clusters make dissimilar shapes.

The MDS layout algorithm tends to depict cycle motifs as (roughly) circles similar to our defined motif, and this is reflected in the data; see Table 3. Though the star motif is isomorphic to the cycle, comparing the star shape introduced noise that caused participants to not recognize the same structure, another point to show that it is important to represent similar structures with similar shapes. We note as well the star motif is not a planar drawing, where the cycle is. Further motif comparison can be found in the appendix.

6   Conclusion

To better understand the role of spontaneous processing on the representation of sub-graphs in node-link diagrams, we conducted a user experiment to address research questions relating to the perception of both structure (relations in the sub-graph) and shape (visual form). Amongst other findings, we conclude that presenting identical motifs using identical (or rotated) visual motif drawings makes a difference to the ease with which they can be matched.

We considered small motifs (5–8 nodes) in sparse, relatively small graphs. Further work includes determining whether the same results hold for larger motifs or graphs, or for other graph representation idioms (e.g., adjacency matrices or arc diagrams), since the same pattern-matching and bottom-up spontaneous processing issues also apply. Larger, different motifs could be studied, integration into denser graphs, or considering more (and more radical) variations of shape and structure.

Since we have shown that the shape of a motif affects how an individual recognizes it, layout algorithms that extract similar graph motifs and draw them using a similar form would highlight their similarity. Techniques exist to accommodate user constraints such as node positions [4,19,47], but combining this idea with motif extraction may present algorithmic problems and perceptual outcomes.

References

1. de Ridder et al., H.: Information system on graph classes and their inclusions (ISGCI) (2023). https://www.graphclasses.org. Accessed 19 May 2023
2. Ballweg, K., Pohl, M., Wallner, G., von Landesberger, T.: Visual similarity perception of directed acyclic graphs: a study on influencing factors. In: Frati, F., Ma, K.-L. (eds.) GD 2017. LNCS, vol. 10692, pp. 241–255. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73915-1_20
3. Batik, T., Terziadis, S., Wang, Y.S., Nöllenburg, M., Wu, H.Y.: Shape-guided mixed metro map layout. CGF **41**(7), 495–506 (2022). https://doi.org/10.1111/cgf.14695
4. Böhringer, K.F., Paulisch, F.N.: Using constraints to achieve stability in automatic graph layout algorithms. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 43–51 (1990). https://doi.org/10.1145/97243.97250
5. Borgo, R., et al.: Crowdsourcing for information visualization: promises and pitfalls. In: Archambault, D., Purchase, H., Hoßfeld, T. (eds.) Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments. LNCS, vol. 10264, pp. 96–138. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66435-4_5
6. Bremm, S., von Landesberger, T., Heß, M., Schreck, T., Weil, P., Hamacherk, K.: Interactive visual comparison of multiple trees. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 31–40 (2011). https://doi.org/10.1109/VAST.2011.6102439
7. Cakmak, E., Fuchs, J., Jackle, D., Schreck, T., Brandes, U., Keim, D.: Motif-based visual analysis of dynamic networks. In: 2022 IEEE Visualization in Data Science (VDS), pp. 17–26 (2022). https://doi.org/10.1109/VDS57266.2022.00007
8. Cherry, K.: What are the gestalt principles? An overview of the gestalt laws of perceptual organization (2023). https://www.verywellmind.com/gestalt-laws-of-perceptual-organization-2795835. Accessed 19 May 2023
9. Dong, W., Moses, C., Li, K.: Efficient k-nearest neighbor graph construction for generic similarity measures. In: Proceedings of the 20th International Conference on World Wide Web, pp. 577–586 (2011). https://doi.org/10.1145/1963405.1963487
10. Dunne, C., Shneiderman, B.: Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2013, pp. 3247–3256. Association for Computing Machinery, New York (2013). https://doi.org/10.1145/2470654.2466444
11. Eades, P., Hong, S.-H., Klein, K., Nguyen, A.: Shape-based quality metrics for large graph visualization. In: Di Giacomo, E., Lubiw, A. (eds.) GD 2015. LNCS, vol. 9411, pp. 502–514. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-27261-0_41
12. Fuchs, J., Isenberg, P., Bezerianos, A., Fischer, F., Bertini, E.: The influence of contour on similarity perception of star glyphs. IEEE TVCG **20**(12), 2251–2260 (2014). https://doi.org/10.1109/TVCG.2014.2346426

13. Gansner, E.R., Koren, Y., North, S.: Graph drawing by stress majorization. In: Pach, J. (ed.) GD 2004. LNCS, vol. 3383, pp. 239–250. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-31843-9_25

14. Gogolou, A., Tsandilas, T., Palpanas, T., Bezerianos, A.: Comparing similarity perception in time series visualizations. IEEE TVCG **25**(1), 523–533 (2019). https://doi.org/10.1109/TVCG.2018.2865077

15. Hascoët, M., Dragicevic, P.: Interactive graph matching and visual comparison of graphs and clustered graphs. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI 2012, pp. 522–529. Association for Computing Machinery, New York (2012). https://doi.org/10.1145/2254556.2254654

16. Hu, Y., Brunton, S.L., Cain, N., Mihalas, S., Kutz, J.N., Shea-Brown, E.: Feedback through graph motifs relates structure and function in complex networks. Phys. Rev. E **98**, 062312 (2018). https://doi.org/10.1103/PhysRevE.98.062312

17. Huang, W., Murray, C., Shen, X., Song, L., Wu, Y.X., Zheng, L.: Visualisation and analysis of network motifs. In: Ninth International Conference on Information Visualisation (IV 2005), pp. 697–702 (2005)

18. Jerrum, M., Meeks, K.: The parameterised complexity of counting connected subgraphs and graph motifs. J. Comput. Syst. Sci. **81**(4), 702–716 (2015). https://doi.org/10.1016/j.jcss.2014.11.015

19. Kamps, T., Kleinz, J., Read, J.: Constraint-based spring-model algorithm for graph layout. In: Brandenburg, F.J. (ed.) GD 1995. LNCS, vol. 1027, pp. 349–360. Springer, Heidelberg (1996). https://doi.org/10.1007/BFb0021818

20. Koffka, K.: Principles of Gestalt Psychology. Routeledge (1935)

21. Kypridemou, E., Zito, M., Bertamini, M.: Perception of node-link diagrams: the effect of layout on the perception of graph properties. In: Giardino, V., Linker, S., Burns, R., Bellucci, F., Boucheix, J.M., Viana, P. (eds.) Diagrams 2022. LNCS, vol. 13462, pp. 364–367. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-15146-0_32

22. von Landesberger, T.: Visual analysis of large graphs: state-of-the-art and future research challenges. CGF **30**(6), 1719–1749 (2011). https://doi.org/10.1111/j.1467-8659.2011.01898.x

23. von Landesberger, T.: Insights by visual comparison: the state and challenges. IEEE Comput. Graphics Appl. **38**(3), 140–148 (2018). https://doi.org/10.1109/MCG.2018.032421661

24. von Landesberger, T., Görner, M., Rehner, R., Schreck, T.: A system for interactive visual analysis of large graphs using motifs in graph editing and aggregation. In: 14th International Workshop on Vision, Modeling, and Visualization, pp. 331–340 (2009)

25. Lenhof, H.P., Smid, M.: Sequential and parallel algorithms for the k closest pairs problem. Int. J. Comput. Geom. Appl. **5**(03), 273–288 (1995). https://doi.org/10.1142/S0218195995000167

26. Leontis, N., Lescoute, A., Westhof, E.: The building blocks and motifs of RNA architecture. Curr. Opin. Struct. Biol. **16**, 279–87 (2006). https://doi.org/10.1016/j.sbi.2006.05.009

27. McGee, F., Ghoniem, M., Melançon, G., Otjacques, B., Pinaud, B.: The state of the art in multilayer network visualization. CGF **38**(6), 125–149 (2019). https://doi.org/10.1111/cgf.13610

28. Meidiana, A., Hong, S.H., Eades, P.: Shape-faithful graph drawings. In: Angelini, P., von Hanxleden, R. (eds.) GD 2022. LNCS, vol. 13764, pp. 93–108. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-22203-0_8

29. Micale, G., Giugno, R., Ferro, A., Mongiovì, M., Shasha, D., Pulvirenti, A.: Fast analytical methods for finding significant labeled graph motifs. Data Min. Knowl. Disc. **32**, 504–531 (2018). https://doi.org/10.1007/s10618-017-0544-8

30. Pandey, A.V., Krause, J., Felix, C., Boy, J., Bertini, E.: Towards understanding human similarity perception in the analysis of large sets of scatter plots. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3659–3669 (2016). https://doi.org/10.1145/2858036.2858155

31. Prolific (2014). https://www.prolific.co

32. Qualtrics (2005). https://www.qualtrics.com

33. Redhu, N., Thakur, Z.: Chapter 23 - network biology and applications. In: Singh, D.B., Pathak, R.K. (eds.) Bioinformatics, pp. 381–407. Academic Press (2022). https://doi.org/10.1016/B978-0-323-89775-4.00024-9

34. Royer, L., Reimann, M., Andreopoulos, B., Schroeder, M.: Unraveling protein networks with power graph analysis. PLoS Comput. Biol. **4**(7), 1–17 (2008). https://doi.org/10.1371/journal.pcbi.1000108

35. Sanfeliu, A., Fu, K.S.: A distance measure between attributed relational graphs for pattern recognition. IEEE Trans. Syst. Man Cybern. **SMC-13**(3), 353–362 (1983). https://doi.org/10.1109/TSMC.1983.6313167

36. Soni, U., Lu, Y., Hansen, B., Purchase, H.C., Kobourov, S., Maciejewski, R.: The perception of graph properties in graph layouts. CGF **37**(3), 169–181 (2018). https://doi.org/10.1111/cgf.13410

37. Stone, L., Simberloff, D., Artzy-Randrup, Y.: Network motifs and their origins. PLoS Comput. Biol. **15**(4), e1006749 (2019). https://doi.org/10.1371/journal.pcbi.1006749

38. Tarr, M.J., Pinker, S.: Mental rotation and orientation-dependence in shape recognition. Cogn. Psychol. **21**(2), 233–282 (1989). https://doi.org/10.1016/0010-0285(89)90009-1

39. Treisman, A.: Preattentive processing in vision. Comput. Vision, Graph. Image Process. **31**(2), 156–177 (1985). https://doi.org/10.1016/S0734-189X(85)80004-9

40. Tversky, A.: Features of similarity. Psychol. Rev. **84**(4), 327–352 (1977). https://doi.org/10.1037/0033-295X.84.4.327

41. Wallner, G., Pohl, M., von Landesberger, T., Ballweg, K.: Perception of differences in directed acyclic graphs: influence factors & cognitive strategies. In: Proceedings of the 31st European Conference on Cognitive Ergonomics, pp. 57–64 (2019). https://doi.org/10.1145/3335082.3335083

42. Wallner, G., Pohl, M., Graniczkowska, C., Ballweg, K., von Landesberger, T.: Influence of shape, density, and edge crossings on the perception of graph differences. In: Pietarinen, A.-V., Chapman, P., Bosveld-de Smet, L., Giardino, V., Corter, J., Linker, S. (eds.) Diagrams 2020. LNCS (LNAI), vol. 12169, pp. 348–356. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-54249-8_27

43. Wang, Y., et al.: Revisiting stress majorization as a unified framework for interactive constrained graph visualization. IEEE TVCG **24**(1), 489–499 (2018). https://doi.org/10.1109/TVCG.2017.2745919

44. Ware, C.: Visual Thinking for Information Design. Morgan Kaufmann (2021)

45. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences. Cambridge University Press (1994). https://doi.org/10.1017/CBO9780511815478

46. Wu, H.Y., Niedermann, B., Takahashi, S., Roberts, M.J., Nöllenburg, M.: A survey on transit map layout - from design, machine, and human perspectives. CGF **39**(3), 619–646 (2020). https://doi.org/10.1111/cgf.14030

47. Yu, J., Hu, Y., Yuan, X.: UNICON: a UNIform CONstraint based graph layout framework. In: 2022 IEEE 15th Pacific Visualization Symposium (PacificVis), pp. 61–70. IEEE (2022). https://doi.org/10.1109/PacificVis53943.2022.00015
48. Yuan, X., Che, L., Hu, Y., Zhang, X.: Intelligent graph layout using many users' input. IEEE TVCG **18**(12), 2699–2708 (2012). https://doi.org/10.1109/TVCG.2012.236
49. Zheng, J.X., Pawar, S., Goodman, D.F.: Graph drawing by stochastic gradient descent. IEEE TVCG **25**(9), 2738–2748 (2018). https://doi.org/10.1109/TVCG.2018.2859997