Evaluating the Effects of Size in LineSets

Dominique Tranquille
University of Brighton, UK
d.tranquille1@uni.brighton.ac.uk

Gem Stapleton
University of Brighton, UK
g.e.stapleton@brighton.ac.uk

Jim Burton
University of Brighton, UK
j.burton@brighton.ac.uk

Peter Rodgers
University of Kent, UK
p.j.rodgers@kent.ac.uk

ABSTRACT

LineSets represent information about sets by drawing one line for each set on an existing visualization of data items. This paper addresses the following question: *does manipulating the size of visual elements affect the comprehension of LineSets?* We empirically evaluated two types of size treatments applied to LineSets drawn on networks: varying set-line thickness, to reflect relative set cardinality, and varying node diameter, to reflect data items' relative degree of connectivity. The evaluation required participants to perform tasks that were thought to be aided by the size variations alongside tasks where no benefit was anticipated. Viewing comprehension through accuracy and time performance, we found that varying set-line thickness and node diameter significantly improves the effectiveness of LineSets. As a consequence, this research leads to the recommendation that LineSets vary sizes of lines and nodes.

KEYWORDS

Set visualization, LineSets, graphical properties

ACM Reference format:
Dominique Tranquille, Gem Stapleton, Jim Burton, and Peter Rodgers. 2017. Evaluating the Effects of Size in LineSets. In *Proceedings of VINCI '17, Bangkok, Thailand, August 14-16, 2017,* 8 pages.
DOI: 10.1145/3105971.3105978

1 INTRODUCTION

The rapid rise in the volume of data in recent years has led to the need for effective set visualization techniques, many of which have been devised [2, 7, 9, 11, 15, 16, 19]; see Alsallakh et al. [3] for an overview. Data items often lie in overlapping sets, where sets could represent brands, events, locations, organisations, products, services or even hash-tags. Our focus is a technique called LineSets, which visualizes sets by overlaying lines on top of an existing visualization of data items. An example is in Figure 1 which shows four sets, called England, France, Germany and Holland, and the nodes in the graph represent people. A node lying on a set-line indicates the person has visited the respective country. The black edges connecting nodes indicate the two people are friends; as we are visualizing social network data, the black edges are crucial to the information represented. As LineSets are an overlay technique, they can be applied to many different data sets.

In contrast to LineSets, many visualizations of sets, such as Euler diagrams, use closed curves instead of set-lines, exploiting overlapping regions [3]. When overlaying a network, the closed curves constrain the location of the data items and, thus, can compromise the layout of the network's edges. By contrast, when using Linesets the network is drawn first, with the lines subsequently overlaid. Whilst closed curves appear to be natural for representing sets, recent evidence suggests that using lines can be more effective [21]. This evidence leads us to posit that improving the design of LineSets could make them more effective than techniques based on closed curves. However, little work has been done on evaluating the impact of graphical properties on users' ability to understand the visualized set-data. Compared to other techniques that overlay group information on node-link diagrams, such as GMap and simple node colouring, Jianu et al.'s study found LineSets to be a viable alternative to GMap diagrams for group-only tasks, network-only tasks and group-network tasks [18]. Therefore, LineSets are a promising technique for overlaying group information on an existing visualization of data items. Given this and the extensive variety of applications of LineSets [2], it is important to understand how their graphical properties impact on task performance.

Alper et al.'s initial paper on LineSets focussed on exploring the potential of their new technique [2]. Their studies evaluated how to best draw the set-lines in order to connect elements. The results showed that LineSets outperformed BubbleSets in set membership

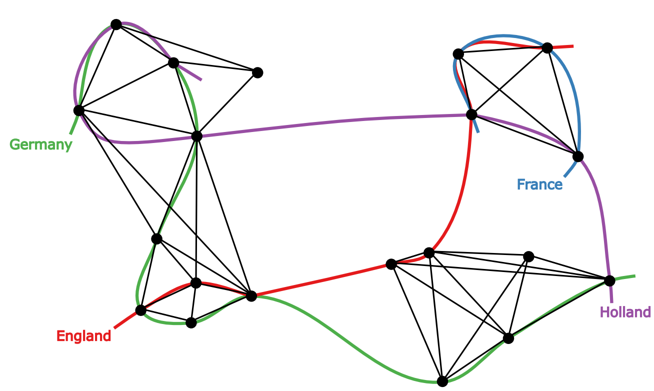

Figure 1: LineSet diagram presented in a default state.

VINCI '17, Bangkok, Thailand

978-1-4503-5292-5/17/08…$15.00
DOI: 10.1145/3105971.3105978

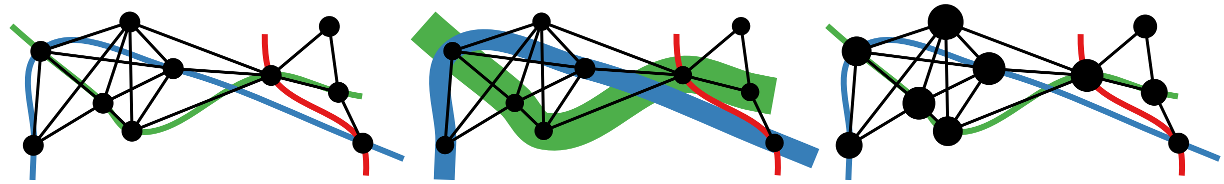

Figure 2: LineSets: standard (left), varying line thickness (middle), varying node diameters (right).

and intersection tasks and established two simple design guides: LineSets should be generated with set-lines that are as linear as possible and smooth.

However, there are a number of other graphical choices to be made when drawing LineSets; one of these choices is size. Bertin's Semiology of Graphics identifies size as one of the fundamental retinal variables to which we are known to be perceptually sensitive [4]. Treisman's hierarchy of features for feature coding classifies size as an important factor in order to facilitate visual search and improve element saliency [24]. Similarly, Healey recommends the use of size for ordinal and quantitative data classification [13]. For these reasons, it is important to evaluate element sizes in LineSets. Figure 2 shows visual differences between a standard LineSet visualization and ones that adopt varying set-line thicknesses and varying node diameters. Although there is extensive knowledge of the properties that humans are perceptually sensitive to, no work has been carried out in the specific context of social network visualization and how the implementation of changes to graphical properties might create unnecessary clutter or distraction that might impede the user.

This paper identifies whether varying set-line thicknesses and node diameters significantly improves the comprehension of LineSets drawn on networks. We conducted two empirical studies to establish whether either of these sizing treatments aids task performance. Sections 2 and 3 detail the experiment design and execution method common to both studies. Sections 4 (on line thickness) and 5 (on node diameter) provide details specific to each study and the respective results. The studies revealed significant differences between the sizing treatments. The study materials and collected data are available at
`http://readableproofs.org/dt-experiments-2017/`.

2 EXPERIMENT DESIGN

Each of the two empirical studies used a between-group design with two groups: one where the sizes were equal, the other where they varied. The size treatments had the following unique characteristics:

Set-line thickness experiment:

(1) *Varying Set-Line Thickness*: the thickness of the set-lines are relative to the cardinality of the represented sets, that is, how many nodes a line passes through
(2) *Equal Set-Line Thickness*: set-lines are all of an equal thickness (the current standard used by LineSets).

Node diameter experiment:

(1) *Varying node diameter*: the diameter of a node is relative to the number of edges connected to it,
(2) *Equal node diameter*: nodes are all of an equal diameter (the current standard used by LineSets).

Two dependent variables were recorded during each empirical study: the time required to answer the question and whether the question had been correctly answered. The independent variables consisted of the node diameter or line thickness treatment, data set size and question type. In each experiment, twenty-four LineSets were shown to the two groups of participants. Participants were asked to answer a question regarding the information shown in each LineSet diagram. If varying the size of elements impacts task performance then we would expect to find a significant difference in either time or accuracy between the two groups.

2.1 LineSet Diagram Drawing Conventions

To provide controlled variability, we drew 24 diagrams for each empirical study. There were 12 *Type 1* diagrams which represented four sets, as shown in Figure 3. The four sets comprised one set containing 10 to 15 data items, two sets containing 5 to 10 data items, and one set containing

We drew LineSet diagrams for the study using Inkscape. Manually drawing the diagrams allowed us to carefully control their layout features. All diagrams adhered to the same layout conventions and characteristics in order to remove unwanted variations between them. Each set-line was allocated a unique colour hue. This is the most common method used in LineSets and has been established as the most effective set-line colour treatment [23]. The colour palettes were derived from Brewer et al.'s work on the Munsell colour system [8] and Healey and Ware's work on using colours for labelling [12, 26].

To provide controlled variability, we drew 24 diagrams for each empirical study. There were 12 *Type 1* diagrams which represented four sets, as shown in Figure 3. The four sets comprised one set containing 10 to 15 data items, two sets containing 5 to 10 data items, and one set containing 1 to 5 data items. In addition, there were 20 data items and 40 edges between the data items. Exactly 11 data items were in multiple sets in Type 1 diagrams. The remaining 12 *Type 2* diagrams represented eight sets, as shown in Figure 4. The eight sets comprised two sets containing 15 to 20 data items, four sets containing 5 to 15 data items, and two set containing 1 to 5 data items. Nineteen data items were in multiple sets in Type 2 diagrams.

Each of the 24 diagrams was modified to reflect the graphical properties that were to be evaluated (varying set-line thickness or varying node diameter). This gave a total of 24 diagrams for each participant group which were semantically and syntactically identical except for the size treatments to which they were subjected.

2.2 Tasks for the Studies

The tasks performed by participants throughout both experiments fit, where appropriate, into Simonetto et al.'s group-level graph visualization taxonomy [22], which consists of group only tasks, group-node tasks, group-link tasks and group-network tasks. Because the experiments were intended to test the effects of a graphical alteration, the actual tasks selected reflected the hypotheses that were being tested. These tasks were:

- Extreme Set Size (ESS): identify the set of largest or smallest size (group-node tasks),
- Specific Set Size (SSS): count sets of a certain size (group-node tasks),
- Extreme Node Degree (END): identify the node of the highest or lowest degree and count the number of edges connected to that node,
- Specific Node Degree (SND): count how many nodes have a certain number of edges,
- Control (Co.): identify an element(s) in the diagram such that the task does not require set cardinality or node degree (group only and group-network tasks).

Although the END and SND style questions do not directly fit into the taxonomy, they were included so that participants had to identify a node of an extreme degree and count nodes with specified degrees; the END and SND tasks directly reflect the set-cardinality focused ESS and SSS questions. As such, these tasks were important as they could be aided by the varying node diameter treatment.

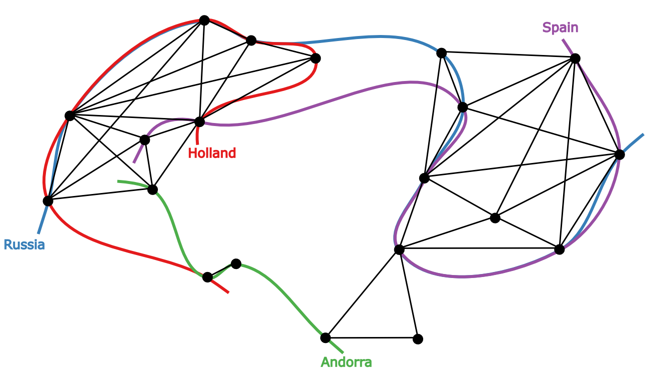

Figure 3: Example type 1.

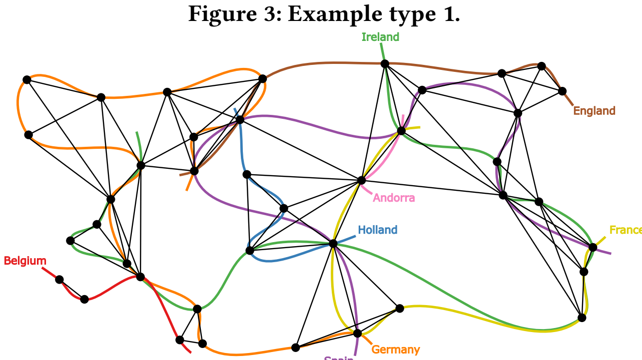

Figure 4: Example type 2.

The task types to evaluate the effects of varying set-line thickness were Extreme Set Size (ESS), Specific Set Size (SSS) tasks as well as Control (Co.) tasks. This was done to distinguish between questions where the thickness of the set-lines may be useful in finding the answer (ESS and SSS tasks) from tasks where the thickness of the set-line is irrelevant (Co., control tasks). The task types to evaluate node diameters were Extreme Node Degree (END), Specific Node Degree (SND) and Control (Co.) tasks. This allowed us to distinguish between questions where the size of the the node is potentially useful in finding the answer (Node Degree tasks) from tasks where the size of the node is irrelevant (Co. tasks).

3 EXPERIMENT EXECUTION

Each experiment session consisted of three phases; the first two were training phases and the third was where accuracy and time performance data were collected. A script was used to ensure each participant was treated equally throughout each session.

The first training phase was paper-based and introduced the participants to the concepts of LineSets and the task types. Additional material was added to the script in phase 1 of the session for participants in the varying set-line thickness and varying node diameter groups, who were informed about the semantics of the treatment to which they were exposed. Participants were shown two examples of each question; a first example was used to explain to the participant how to answer the question and the second example was for the participant to answer without assistance. If the answer was incorrect, the reason why was explained to the participant. Phase 2 consisted of computer-based training where the participant was introduced to the software tool that was used to collect performance data. They had to answer one of each question type before they could proceed to phase 3.

Performance data were collected in phase 3. A software tool developed for the purpose of conducting empirical studies [5–7] was used to collect our performance data, specifically time and accuracy data. The software randomised the order in which the questions were presented to participants, therefore limiting the possibilities of learning effects which could otherwise arise. Participants were presented with a LineSet diagram with a corresponding question and four possible answers, of which one was correct. The software enforced a two minute time limit per question to ensure the participant finished the session in a reasonable time.

4 SET-LINE THICKNESS

This experiment addressed the question *do set-line thickness variations affect the comprehension of LineSets?* Our primary motive to address this question is that we do not know if perceivable variations in the thickness of the set-lines in LineSets alters users' performance in different types of tasks. Existing experiments have already concluded that humans are perceptually sensitive to size variations [4, 17]. Despite this, it is yet to be discovered whether varying the thickness of the set-lines can improve LineSet visualizations or whether variations can facilitate the completion of a particular task type. Thus, our hypotheses were:

- The use of set-lines of varying thickness significantly improves task performance for ESS tasks and SSS tasks.

Specification	Type 1	Type 2
1		
2		
3		
4		
5		
6		

Table 1: Sample set-line thicknesses.

- The use of set-lines of varying thickness does not significantly improve task performance for Co. tasks.

It is unclear whether varying thickness could be detrimental to task performance when completing control tasks: for Co. tasks, the size variations are irrelevant and could act as a distractor.

4.1 Set-Line Treatments

Initially, all set-lines were of an equal thickness of 4 pixels, as shown in figure 5. There are no guidelines for how to best treat LineSets with varying set-line thicknesses, so we produced several samples with differing thickness ratios. In order to test the experiment's hypotheses, the chosen thicknesses had to provide a visually noticeable difference between each set-line whilst not cluttering the user's view. We calculated the thickness using the following formula:

$$LineThickness = \frac{(MaximumThickness - MinimumThickness)}{NumberofSets - 1}$$

Table 1 shows the thickness ratios that were applied to the set-lines that we considered prior to our experiment; these figures are scaled to 0.35 of their original sizes for space reasons. These thicknesses were obtained by adjusting the minimum thickness relative to the thickness of the edges (2px) and the maximum thickness relative to the diameter of the nodes (12px). As no guidelines currently exist, we used existing LineSets from previous work to apply our sample specifications to in order to determine a ratio that provided sufficiently distinguishable differences in thickness. Specification 5 was used to evaluate the effects of set-line thickness as we believed that it provided a sufficiently noticeable difference between the various sets without obscuring the user's view of other visual elements. Figures 5 and 6 show the visual differences of two semantically identical diagrams when one is subjected to set-lines of varying thickness.

A third treatment was considered where colour value variations was implemented in conjunction with set-line thickness to further emphasize the cardinality of the set-lines, Bertin argues that combining these two retinal variables diminishes the potential gains that they return when used separately. This is because the lightness and thinness of the smallest sets is highly likely to be overshadowed by the darkness and boldness of the biggest sets. Specifically in the context of LineSets, it has been found that treating set-lines with unique hues significantly improves user time performance in comparison the set-lines treated with varying values of a single hue to reflect set cardinality [23]. For these reasons, this treatment was excluded from this study.

4.2 Question Types

This study used tasks from Simonetto et al.'s group-level graph visualization taxonomy, as discussed in Section 2.2. The six specified question types were asked in the following format where italics denote the two variations of the same question type and bold show the criterion to satisfy:

Extreme Set Size Tasks
- ESS Max: e.g. which country has been visited by the *GREATEST* number of people?
- ESS Min: e.g. which country has been visited by the *SMALLEST* number of people?

Specific Set Size Tasks
- SSS Max: e.g. how many countries have been visited by **7** people or *MORE*?
- SSS Min: e.g. how many countries have been visited by **7** people or *FEWER*?

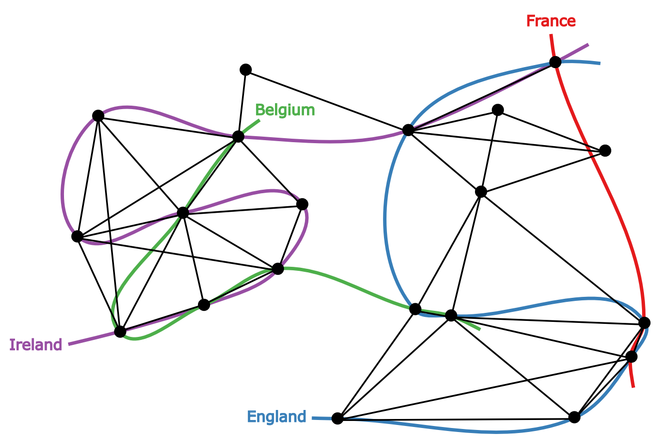

Figure 5: Type 1 diagram with set-lines of equal thickness.

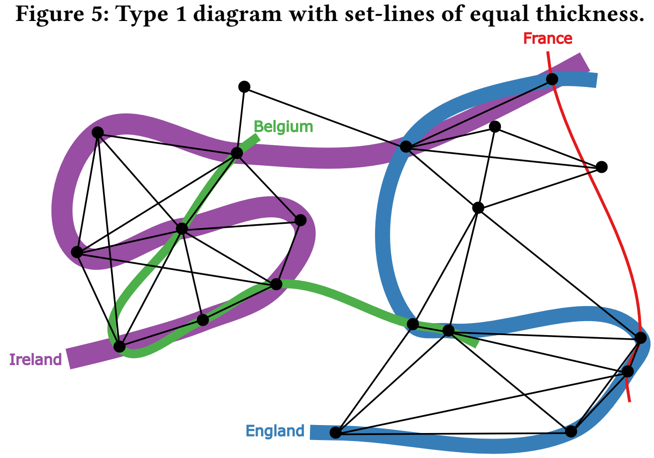

Figure 6: Type 1 diagram with set-lines of varying thickness.

Table 2: Summary of accuracy data and Kruskal-Wallis test for all question types and set-line thickness treatments.

	Equal	Varying	p-value	Ranking
Overall	$\frac{470}{600} = 78.3\%$	$\frac{478}{600} = 79.7\%$	0.740	N/A
ESS Max	$\frac{83}{100} = 83.0\%$	$\frac{99}{100} = 99.0\%$	0.000	V>E
ESS Min	$\frac{87}{100} = 87.0\%$	$\frac{83}{100} = 83.0\%$	0.611	N/A
SSS Max	$\frac{74}{100} = 74.0\%$	$\frac{72}{100} = 72.0\%$	0.760	N/A
SSS Min	$\frac{70}{100} = 70.0\%$	$\frac{64}{100} = 64.0\%$	0.348	N/A
Co. Max	$\frac{71}{100} = 71.0\%$	$\frac{72}{100} = 72.0\%$	0.590	N/A
Co. Min	$\frac{85}{100} = 85.0\%$	$\frac{88}{100} = 88.0\%$	0.458	N/A

Set Control Tasks

- Co. Max: e.g. how many countries have been visited by the person with the *MOST* friends?
- Co. Min: e.g. how many countries have been visited by the person with the *FEWEST* friends?

Each question was multiple choice. Four possible answers were included to reduce the number of correct answers that arise from guessing, without being excessive. Participants were asked to select the correct answer from four available options. Each of the six question styles was used twice for the Type 1 diagrams and twice for the Type 2 diagrams.

4.3 Results

Six participants were recruited for the Set-Line Thickness experiment pilot study. This uncovered a number of issues with the diagrams and questions. Firstly, one diagram was not correctly treated with varying thicknesses and a second diagram had an incorrect label. An additional diagram yielded a 17% accuracy rate. We believed this accuracy rate was because the target elements required for answering the question were placed too closely together, cluttering the participants' view. These three diagrams were modified accordingly.

We collected data from 50 participants (34 M, 16 F, average age 21, age range 18 to 43) in the main study; one participant chose not to disclose their age. No participants suffered from colour blindness. Each sample comprised 24 questions giving us a total of 1200 observations. As with other studies of a similar nature [1, 5–7], we considered accuracy to be more important than time in terms of a performance indicator. This is because the time to complete a task is redundant if the answer is ultimately wrong. Results were considered to be significant if $p \leqslant 0.05$.

We conducted a series of Kruskal-Wallis tests using the ranking each participant in terms of the number of correct answers they accrued. The analysis of correct answers is shown in Table 2; V and E abbreviate Varying and Equal respectively. The results indicate that *overall* people perform tasks no more accurately with LineSets treated with set-lines of varying thickness than with equal thickness treated diagrams; the overall accuracy rates were 78.3% and 79.7% respectively. The largest difference in accuracy was observed for ESS Max questions. Here a significant difference ($p = 0.000$ to 3d.p.) existed between varying thicknesses' accuracy rate of 99.0% and equal thicknesses' accuracy rate of 83.0%. Approximately 16 fewer correct answers were accrued by equal thickness diagrams per 100 questions in comparison to varying thickness diagrams for this type of question. No other significant differences were found. From these results we can suggest that set-lines of varying thickness can significantly improve accuracy for ESS Max type questions.

Table 3: Summary of mean times (st. dev.) and ANOVA results overall and by question type for set-line thickness.

Task Type	Skewness	Equal	Varying	p-value	Ranking
Overall	-0.014	36.07 (20.10)	30.90 (19.80)	0.007	V<E
ESS Max	-0.54	41.79 (20.78)	28.82 (18.09)	0.000	V<E
ESS Min	0.35	29.50 (19.16)	20.71 (14.14)	0.006	V<E
SSS Max	0.00	42.87 (18.99)	37.29 (21.17)	0.013	V<E
SSS Min	0.23	42.52 (23.04)	37.16 (23.13)	0.650	N/A
Co. Max	0.14	38.37 (19.45)	37.62 (20.76)	0.871	N/A
Co. Min	0.65	24.07 (17.18)	23.16 (15.05)	0.256	N/A

Table 3 summarises the (correct answer) mean times and standard deviations overall and for each question type. The analysis of time data is based on the time taken to provide a correct answer (shown in table 3), eliminating data where a correct answer was not provided. As the time data were not normal, a log transformation (base 10) was applied, reducing the skewness to within levels suitable for conducting a robust ANOVA. Participants from the varying thickness group had an overall mean task completion time of 30.088 seconds. This increased to 36.07 seconds for participants in the equal thickness group. Overall, varying set-line thickness allowed people to perform tasks significantly quicker than equal thickness set-lines ($p = 0.007$). The difference of approximately 6 seconds between the varying line thickness group and the equal line thickness group corresponds to about a 20% increase.

In terms of specific tasks, participants from the varying set-line thickness group were significantly faster than participants from the equal set-line thickness group for the first two question types, ESS Max ($p = 0.000$) and ESS Min ($p = 0.006$). Participants in the equal thickness group took approximately 13 seconds longer to complete ESS Max tasks and approximately 9 seconds to complete ESS Min tasks, equating to effects of about 45% and 42% respectively. SSS Max took approximately 6 seconds longer to complete, roughly 14% longer. In summary, our results suggest that overall and for ESS Max, ESS Min, and SSS Max tasks, varying set-line thickness effectively supports task performance, as hypothesized. It is surprising that the SSS Min tasks were not significantly aided by this treatment. Furthermore, no significant differences were revealed in the control tasks. The analysis allows us to suggest, therefore, that varying set-line thicknesses is beneficial for task performance.

5 NODE DIAMETER

The question *do node diameter variations affect the comprehension of LineSets*? is addressed in the second empirical study. Having established the effects of set-line thickness in section 4, this experiment considered the impact of varying the diameters of the displayed nodes in a proportional manner to represent cardinality. Our primary motive to address this question is that we do not know if perceivable variations in nodes diameter alters users performance in different types of tasks, our hypotheses were:

- The use of nodes of varying diameter significantly improves task performance for END tasks and SND tasks.
- The use of nodes of varying diameter does not significantly improve task performance when completing Co. tasks.

It is unclear whether varying diameters could be detrimental to task performance when completing control tasks: as with line thickness, for Co. tasks the size variations are irrelevant and could act as a distractor.

Spec.	Ratio	
1	1:0	
2	1:1.2	
3	1:2	

Table 4: Sample diameter ratios.

5.1 Node Treatments

Twenty-four diagrams were drawn using equal diameter nodes which then had their diameters varied to create 24 further LineSets diagrams. Nodes in the equal diameter treatment were subject to the standards and conventions we defined in section 2.1. The varying diameters of the nodes were calculated using this formula:

$$NodeDiameter = Min.NodeDiameter + (Edges * DiameterIncrease)$$

The minimum node diameter was 12 pixels; smaller nodes could be difficult to see as they may have been obscured by surrounding elements. We made several samples with differing diameter ratios, as illustrated in Table 4 and in Figures 7 and 8; these are scale to 0.3 of the original size for space reasons, it should be noted that the differences between the nodes are more clearly distinguishable at the sizes used in the diagrams shown in the study. As no guidelines currently exist, we used existing LineSets from previous work to apply our sample specifications to in order to determine a ratio that provided sufficiently distinguishable differences in diameter. These ratios were obtained by adjusting the nodes until the diagrams became too cluttered. Specification 3 was chosen as the nodes are sufficiently distinguishable from each other without the large nodes obscuring the view, allowing for a robust test of the hypotheses.

5.2 Question Types

Six question types were asked in the following format, where italics denote the two variations of the same question type and bold show the criterion to satisfy:

Extreme Node Degree Tasks

- END Most: e.g. how many friends does the person with the *MOST* friends have?
- END Few: e.g. how many friends does the person with the *FEWEST* friends have?

Specific Node Degree Tasks

- SND Most: e.g. how many people who own a **Tesla** have **5** friends or *MORE*?
- SND Few: e.g. how many people who own a **Mazda** have **3** friends or *FEWER*?

Node Control Tasks

- Co. Inter: e.g. how many **Saab** owners also own a **Tesla**?
- Co. Edge: e.g. how many **Fiat** owners are friends with **Mazda** owners?

Each question was multiple choice. Participants were asked to select the correct answer from four possible answers. Four options were included to reduce the number of correct answers that arise from guessing, without being excessive. Each of the six question styles was used twice for the Type 1 diagrams and twice for the Type 2 diagrams.

5.3 Results

Six participants were recruited for the node diameter experiment pilot study. Data collected from the pilot study revealed low accuracy rates for three questions. In one case this was due to an incorrect entry in the data collection software. In the second case, visual elements required for the answer were considered to be obscured; the diagram was modified. In the third case, the accuracy rate may

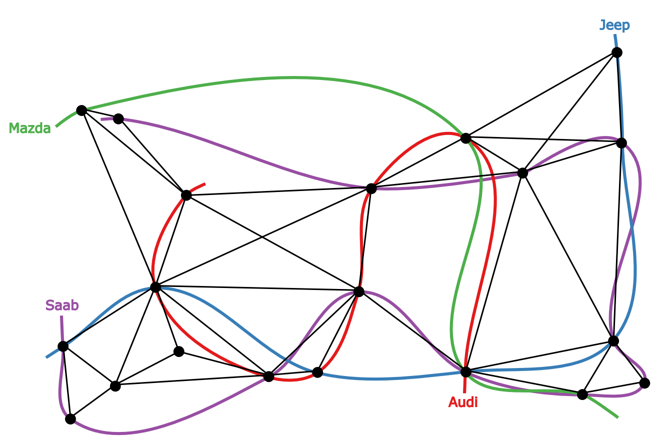

Figure 7: Type 1 diagram with nodes of equal size.

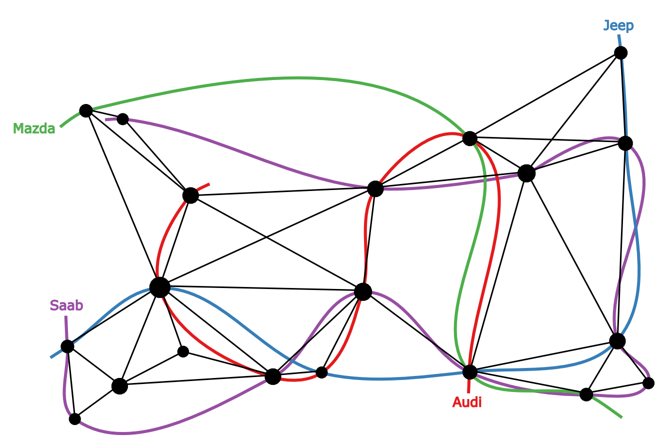

Figure 8: Type 1 diagram with nodes of varying size.

Table 5: Summary of accuracy data and Kruskal-Wallis test for all question types and node diameter treatments.

	Equal	Varying	p-value	Ranking
Overall	$\frac{615}{720} = 85.4\%$	$\frac{633}{720} = 87.9\%$	0.696	N/A
END Most	$\frac{114}{120} = 95.0\%$	$\frac{117}{120} = 97.5\%$	0.989	N/A
END Few	$\frac{116}{120} = 96.7\%$	$\frac{118}{120} = 98.3\%$	0.621	N/A
SND More	$\frac{113}{120} = 94.1\%$	$\frac{115}{120} = 95.8\%$	0.200	N/A
SND Few	$\frac{91}{120} = 75.8\%$	$\frac{95}{120} = 79.2\%$	0.782	N/A
Co. Inter.	$\frac{108}{120} = 90.0\%$	$\frac{112}{120} = 93.3\%$	0.528	N/A
Co. Edge	$\frac{73}{120} = 60.8\%$	$\frac{76}{120} = 63.3\%$	0.969	N/A

Table 6: Summary of mean times (st. dev.) and ANOVA results overall and by question type for node diameter.

Task Type	Skewness	Equal	Varying	p-value	Tukey Ranking
Overall	-0.07	39.39 (21.37)	34.54 (20.19)	0.009	V<E
END Most	-0.13	39.81 (15.94)	26.59 (11.57)	0.000	V<E
END Few	-0.05	26.49 (12.54)	20.97 (10.52)	0.001	V<E
SND More	-0.35	44.71 (20.14)	40.49 (19.87)	0.094	N/A
SND Few	-0.30	50.08 (23.98)	46.60 (22.32)	0.680	N/A
Co. Inter.	0.36	26.68 (13.26)	27.67 (13.46)	0.835	N/A
Co. Edge	-0.30	56.45 (25.27)	53.88 (22.01)	0.143	N/A

have been caused by the large number of elements that participants had to count in order find the answer; the question was simplified.

Data were collected from 60 participants (47 M, 13 F, average age 23, age range 18 to 45) in our main study; no participants suffered from colour blindness. Each sample comprised 24 questions, giving us a total of 1440 observations. Accuracy was considered to be more important than time in terms of a performance indicator, as is consistent with the set-line thickness experiment in section 4. Results were considered to be significant if $p \leqslant 0.05$.

The accuracy results are summarised in Table 5. The overall Kruskal-Wallis analysis of correct answers found no significant difference between LineSets treated with an equal node diameter and LineSets treated with nodes of varying size; the overall accuracy rates were 85.4% and 87.9% respectively. When the data were analysed by question type, there were no significant differences between the two treatments. Therefore we can suggest that manipulating the size of nodes does not have significant impact on user accuracy when using LineSets to answer questions.

The time results (for correct answers only) are summarised in Table 6. As the time data were not normal, a log transformation (base 10) was applied, reducing the skewness to within levels suitable for conducting a robust ANOVA. Overall, treating LineSets with nodes of varying size significantly improves user time performance ($p = 0.009$). Participants in the equal node diameter group took approximately 5 seconds longer to answer each question on average than participants from the varying node diameter group, which is roughly 14% longer.

In terms of the individual task types, the varying node diameter group yielded a significant time difference for END Few ($p = 0.001$) and END Most ($p = 0.000$) tasks. The biggest difference in time was found for END Most tasks where equal node diameter participants took, on average, 39.81 seconds to complete the questions in comparison to the varying node diameter group whose mean time was 26.59 seconds. This 13 second time difference equates to a 49.7% increase in time taken. For END Few tasks, the mean times for the equal and varying treatments were 26.49 seconds and 20.97 seconds respectively, yielding a 26% time increase for these tasks. No significant results were observed for the other task types. This analysis allows us to suggest that varying node diameters allow people to significantly reduce the time taken to identify the nodes of smallest and largest degree in LineSets but have no significant effect on other task types.

6 DISCUSSION AND IMPLICATIONS

The introduction highlighted that size variations are useful for improving ordered perception. This led to the hypothesis that varying sizes supports people with processing the visualized data more quickly or more accurately, at least for tasks that require size comparisons to be made. Our empirical studies support these hypotheses, providing evidence that the use of set-lines of varying thickness and nodes of varying diameter can, for certain tasks, significantly improve user performance. Moreover, for the remaining tasks, these size treatments did not have a significantly negative impact on task performance. Consequently we have provided evidence that, for the types of tasks in our study, varying line thickness and node diameter is not detrimental to, and sometimes beneficial for, task performance. In the discussion that follows, we seek to explain these results. Our focus is on time performance, as there was only one significant result for accuracy.

The reader is reminded that LineSets portray set membership using continuous lines that are overlaid on nodes or a node-link network. We make the following observations:

(1) Set membership is represented when nodes are situated on one or multiple lines. Manipulating the thickness of set-lines, as done for the LineSets used in our study, gives a visual indication of how many nodes belong to the set, relative to the other sets.
(2) Node connections are represented when two nodes are connected by an edge. Varying the diameter of the nodes, as done for the LineSets used in our study, will visually indicate to the user how many edges are connected to it, relative to the other nodes.

By varying size, we turn the problem of, say, finding the largest set (an ESS task) from a counting task to a *target detection* task where the user seeks to identify the thickest set-line; a target detection task is where a target element with a unique visual feature is identified by the user [14]. The results could, therefore, be explained by the fact that size is a preattentive property, which are those where the eye and the brain process an image in less than 250 milliseconds [14]: in our case, the visual target can be found preattentively when size is varied. The END tasks also become target detection tasks and all four (ESS Max, ESS Min, END Most, and END Few) lead to significant time improvements when size was exploited.

By contrast, seven of the remaining eight task types did not reveal significant time difference. Four of these eight tasks – SSS Max, SSS Min, SND More and SND Few – first required the identification of an element of a certain size and then size comparisons to be made; we call the first activity *identification* and the second part *comparison*. Based on the time results for ESS Max, ESS Min, END Most, and END Few, we would expect size to significantly help the comparison activity and, depending on the strategy employed could aid the identification activity (e.g. to eliminate smaller sets *with thinner lines*, if the elements are counted for a set with fewer than *n* elements, where *n* is given in the task). These results indicate that further studies are needed to reveal the extent to which size is beneficial for time performance and for which task types, beyond those considered here. Part of this research should seek to understand the strategies people use to solve tasks of this type to reveal whether the node diameters contribute to adopted problem solving strategies or whether other approaches are used.

Beyond LineSets, the results suggest that other set visualization techniques could benefit from exploiting size to convey cardinality and relative degree of connectivity. Set visualizations typically plot data items that belong to a set within an enclosed shape [2, 7, 9, 11, 15, 16, 19] but others also visualize sets with lines, such as linear diagrams [10, 25]. Where previous comparative evaluations of multiple techniques have revealed the relative strengths of certain techniques [18–20], exploiting size (and other features) has been largely overlooked, at least from the perspective of empirically evaluating the impact of design choices to 'optimize' any given technique. Such evaluations might reveal new strengths and weaknesses that could improve the way set visualizations are designed.

7 CONCLUSION

This paper set out to address the question *does manipulating the size of visual elements affect the comprehension of LineSets?* We selected two treatments based on perceptual theories on the use of size: varying set-line thickness, to reflect relative set cardinality, and varying node diameter, to reflect the data items' relative degree of connectivity. We performed two empirical experiments that compared these two sizing treatments against LineSets in their default state in order to evaluate their relative impact on accuracy and time performance. The results provide evidence that the use of varying set-line thicknesses and node diameters improves the understandability of LineSets. Taking into account both our accuracy analysis and time analysis, these variations do not impede user performance when completing control tasks and significantly improve users' performance for tasks that can exploit the varying size treatments. As a result of this, we suggest that LineSets should be treated with elements of varying size to best aid task performance.

Whilst we have focused on the use of varying the size of elements in LineSets, it is important to note that these diagrams were all drawn manually. This allowed the placement of nodes and set-lines to be controlled and modified. Future research should evaluate whether these findings are replicated in an automated layout environment with real-world data. It must also be established if these treatments interfere with each other when used in combination. Our expectation is that varying size will still be beneficial, as should be clear from the discussion in section 6. Adopting these varying size treatments should lead to an improved LineSets visualization technique.

REFERENCES

[1] J-W. Ahn, C. Plaisant, and B. Shneiderman. 2014. *A Task Taxonomy for Network Evolution Analysis*. IEEE Educational Activities Department, 365–376.
[2] B. Alper, N. Henry Riche, G. Ramos, and M. Czerwinski. 2011. *Design Study of LineSets, a Novel Set Visualization Technique*. Vol. 17. IEEE Educational Activities Department, 2259–2267.
[3] B.Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. 2014. Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. In *Eurographics Conference on Visualization*. The Eurographics Association, 1–21.
[4] J. Bertin. 1983. *Semiology of Graphics*. University of Wisconsin Press.
[5] A. Blake, G. Stapleton, P. Rodgers, L. Cheek, and J. Howse. 2012. *Does the Orientation of an Euler Diagram Affect User Comprehension?* Vol. 18. Springer, 185–190.
[6] A. Blake, G. Stapleton, P. Rodgers, L. Cheek, and J. Howse. 2014. The Impact of Shape on the Perception of Euler Diagrams. In *Diagrammatic Representation and Inference*. Springer, 123–137.
[7] A. Blake, G. Stapleton, P. Rodgers, and J. Howse. 2014. How Should We Use Colour in Euler Diagrams?. In *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*. ACM, 149–159.
[8] C. Brewer, M. Harrower, D. Heyman B., Sheesley, and A. Woodruff. 2013. Color Brewer. (2013).
[9] J. Ellson, E. R. Gansner, E. Koutsofios, S. C. North, and G. Woodhull. 2004. *Graphviz and Dynagraph — Static and Dynamic Graph Drawing Tools*. Springer, 127–148.
[10] B. Gottfried. 2014. *Set Space Diagrams*. Vol. 25. Academic Press, Inc., 518–532.
[11] L. Gou and X. Zhang. 2011. *TreeNetViz: Revealing Patterns of Networks over Tree Structures*. Vol. 17. IEEE Educational Activities Department, 2449–2459.
[12] C. Healey. 1996. Choosing Effective Colours for Data Visualization. In *Proceedings of the 7th Conference on Visualization '96*. IEEE Computer Society Press, 263–ff.
[13] C. Healey and J. Enns. 1998. Building perceptual textures to visualize multidimensional datasets. In *Visualization '98. Proceedings*. IEEE Computer Society Press, 111–118.
[14] C. H. Healey and J. T. Enns. 2012. *Attention and Visual Memory in Visualization and Computer Graphics*. Vol. 18. IEEE Educational Activities Department, 1170–1189.
[15] J. Heer and D. Boyd. 2005. Vizster: visualizing online social networks. In *Proceedings of the 2005 IEEE Symposium on Information Visualization*. IEEE Computer Society, 5–.
[16] N. Henry and J.D. Fekete. 2007. MatLink: Enhanced Matrix Visualization for Analyzing Social Networks. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*. Springer, 288–302.
[17] J. G. Hollands and I. Spence. 2001. *The Discrimination of Graphical Elements*. Vol. 15. John Wiley & Sons, Ltd., 413–431.
[18] R. Jianu, A. Rusu, Y. Hu, and D. Taggart. 2014. *How to Display Group Information on Node-Link Diagrams: An Evaluation*. Vol. 20. IEEE Educational Activities Department, 1530–1541.
[19] W. Meulemans, N. Henry Riche, B. Speckmann, B. Alper, and T. Dwyer. 2013. *KelpFusion: A Hybrid Set Visualization Technique*. Vol. 19. IEEE Educational Activities Department, 1846–1859.
[20] P. Rodgers, G. Stapleto, B. Alsallakh, L. Michallef, R. Baker, and S. Thompson. 2016. *A task-based evaluation of combined set and network visualization*. Vol. 367–368. Elsevier, 58–79.
[21] P. Rodgers, G. Stapleton, and P. Chapman. 2014. *Visualizing Sets with Linear Diagrams*. Vol. 22. ACM, 27:1–27:39.
[22] B. Saket, P. Simonetto, and S. G. Kobourov. 2014. *Group-Level Graph Visualization Taxonomy*. Vol. abs/1403.7421. The Eurographics Association.
[23] D. Tranquille, G. Stapleton, Jim Burton, and P. Rodgers. 2016. Evaluating the Effects of Colour in LineSets. In *Diagrammatic Representation and Inference: 9th International Conference*. Springer, 283–285.
[24] A. Treisman. 1988. Features and objects: The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section A* 40, 2 (1988), 201–237.
[25] G.W.F. von Leibniz. 1961. *Opuscules et Fragments inédits de Leibniz: extraits des manuscrits de la Bibliothéque royale de Hanovre par Louis Couturat*. Georg Olms Verlagsbuchhlandlung.
[26] C. Ware. 2004. *Information Visualization: Perception for Design* (2nd ed.). Morgan Kaufmann Publishers Inc., Chapter 6.1: Gestalt Laws, 189 – 205.