

# Exploring the limits of complexity: A survey of empirical studies on graph visualisation

Vahan Yoghoudjian<sup>a,\*</sup>, Daniel Archambault<sup>b</sup>, Stephan Diehl<sup>c</sup>, Tim Dwyer<sup>a</sup>,  
Karsten Klein<sup>d,1</sup>, Helen C. Purchase<sup>e</sup>, Hsiang-Yun Wu<sup>f</sup>

<sup>a</sup> Monash University, Australia

<sup>b</sup> Swansea University, United Kingdom

<sup>c</sup> University of Trier, Germany

<sup>d</sup> University of Konstanz, Germany

<sup>e</sup> University of Glasgow, United Kingdom

<sup>f</sup> University of TU Wien, Austria

## ARTICLE INFO

### Article history:

Received 13 December 2018

Accepted 25 December 2018

Available online 6 January 2019

### Keywords:

Graph visualisation  
Network visualisation  
node-link diagrams  
Cognitive scalability  
Evaluations  
Empirical studies

## ABSTRACT

For decades, researchers in information visualisation and graph drawing have focused on developing techniques for the layout and display of very large and complex networks. Experiments involving human participants have also explored the readability of different styles of layout and representations for such networks. In both bodies of literature, networks are frequently referred to as being 'large' or 'complex', yet these terms are relative. From a human-centred, experiment point-of-view, what constitutes 'large' (for example) depends on several factors, such as data complexity, visual complexity, and the technology used. In this paper, we survey the literature on human-centred experiments to understand how, in practice, different features and characteristics of node-link diagrams affect visual complexity.

© 2019 Zhejiang University and Zhejiang University Press. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

There has been much work done on designing algorithms that can efficiently scale to create pictures of very large graphs. However, what remains a more open question, is whether pictures of very large and complex networks require a mental effort that exceeds the capabilities of an average human brain.

Eick and Karr (2002) define the term *visual scalability* as the capability of visualisation tools to effectively display large data sets. They also discuss factors affecting visual scalability, like human perception, monitor resolution, visual metaphors, interactivity, data structures and algorithms, as well as the computational infrastructure. A more recent discussion of these, and similar factors, are presented by Jankun-Kelly et al. (2014). They also distinguish perceptual and cognitive scalability: "though elements may be perceivable, they may still exhaust cognitive resources".

In cognitive psychology, Miller's 'seven plus or minus two' (Miller, 1956) is commonly accepted as a rule-of-thumb for the limitation on peoples' working memory. Working memory is an example of one of the 'cognitive ceilings' that might affect peoples' ability to reason about large networks. Do working memory and other cognitive limitations have implications for the size and complexity of graphs that we should be trying to visualise?

Huang et al. (2009b) propose a framework for cognitive load in the context of graph visualisation. Based on results from cognitive psychology, they sketch a model that relates cognitive load during graph analysis tasks to mental effort and task performance in terms of response time and accuracy. They discuss a number of other factors affecting cognitive load: the domain (e.g. a highly technical and specific domain requires the user to relate their knowledge to the visual); the data itself (e.g. structure of the graph); the task (does it require deep understanding of the graph structure?); visual representation (does it follow best practice layout and design principles); demographic (e.g. the experience of the users); and time pressure. They report on one study of cognitive load that confirms some of these effects, but their model which suggests step changes in performance due to load, while compelling, is not fully validated.

In a more general evaluation of effect of display type on visualisation cognition, Yost and North (2006) demonstrated that more pixels make it possible to show more data without considerable loss of performance. Going from 2 to 32 mega pixels led to a 20-fold

\* Corresponding author.

E-mail addresses: [Vahan@Vahany.com](mailto:Vahan@Vahany.com) (V. Yoghoudjian), [d.w.archambault@swansea.ac.uk](mailto:d.w.archambault@swansea.ac.uk) (D. Archambault), [diehl@uni-trier.de](mailto:diehl@uni-trier.de) (S. Diehl), [Tim.Dwyer@monash.edu](mailto:Tim.Dwyer@monash.edu) (T. Dwyer), [karsten.klein@uni-konstanz.de](mailto:karsten.klein@uni-konstanz.de) (K. Klein), [Helen.Purchase@glasgow.ac.uk](mailto:Helen.Purchase@glasgow.ac.uk) (H.C. Purchase), [yun@cg.tuwien.ac.at](mailto:yun@cg.tuwien.ac.at) (H.-Y. Wu).

<sup>1</sup> Adjunct fellow of Monash University.

Peer review under responsibility of Zhejiang University and Zhejiang University Press.

increase in displayed data: the task completion times tripled, while accuracy only decreased from 95% to 92%. However, the data, visualisations and tasks considered (such as search and comparison) are relatively simple compared to networks and associated tasks involving understanding of connectivity. It cannot be assumed that such results carry over, or that display size is the only – or even most – significant limitation. The goal of this survey is to better understand these cognitive limitations. To differentiate this human aspect of scalability of node–link diagrams from technological or technique specific limitations, we use the term *cognitive scalability*.

This topic is important for our field, as such insights can guide the design of future techniques. For example, we are attempting to find tacit knowledge in past studies concerning the numbers of nodes and edges that are too difficult to work with in a single view. If we can establish such numbers, then it might suggest that we need to direct efforts away from algorithms and rendering techniques that can scale to huge numbers of network elements. Instead, for such large networks, we could focus on interactive ways to explore neighbourhoods (e.g. Dwyer et al. (2008) and van Ham and Perer (2009)) or abstractions (e.g. Abello et al. (2006), Archambault et al. (2008) and Dunne and Shneiderman (2013)) instead of attempting to display the full set of nodes and links.

Thus, we survey a large number of papers reporting empirical studies of node–link diagrams, being exhaustive within the corpora of core visualisation proceedings and journals. We aim to establish a consensus for definitions of adjectives like ‘large’ or ‘dense’ for node–link diagrams that are too complex to be easily comprehensible or useful for standard graph analysis tasks. We also provide an overview of the types of networks and tasks, as well as experimental design of these experiments.

In general, we find that the limits of scalability of the node–link network visualisation paradigm are rarely addressed directly. Rather, there seem to be tacit assumptions (or possibly unreported pilot findings) about what size node–link diagrams are useable for different tasks, and experiments stay within these bounds while testing specific techniques.

Our key findings are that only a small range of graph sizes and structures have been used in experimental evaluations of graph visualisation techniques, mostly limited to small and sparse graphs. In particular, three quarters of studies use graphs with 100 nodes and 200 edges or less and, the remaining studies test interactive techniques, such that only a small portion of the graph is shown on the screen at a time. These findings are discussed further throughout the paper and listed in full in the conclusion.

The paper is structured as follows: Section 2 discusses related surveys, primarily in graph visualisation; Section 3 outlines our scope, methodology, and describes our categorisation framework; then we present the results of our survey in Sections 4–6, followed by a discussion on trends in the network visualisation community in Section 7.

## 2. Related work

In the fields of graph drawing and visualisation, a number of surveys have considered scalability from different perspectives. In particular, there has been much discussion of the scalability of algorithms and computer hardware to compute node–link diagram layout. Such papers tacitly acknowledge that very ‘large’ and ‘dense’ graphs are difficult to read and hence propose interaction techniques to navigate aggregated graphs. They may even report on studies of the readability of networks using different layout, interaction or rendering techniques. Yet, rarely do they explicitly address the question of what is the largest (most complex) diagram that people can usefully comprehend.

Many past surveys characterise the techniques available for graph visualisation. The surveys of Herman et al. (2000) and von

Landesberger et al. (2011) are both of this type, focusing on techniques for graph visualisation and their strengths and weaknesses. Elmquist and Fekete (2010) characterise techniques in information visualisation that use hierarchical representations as a form of data abstraction. Recent surveys have also focused on specific areas of network visualisation including multi-faceted graph visualisation (Hadlak et al., 2015), group structures in graphs (Vehlow et al., 2015, 2017), matrix reordering techniques (Behrisch et al., 2016), edge bundling (Lhuillier et al., 2017), and networks in social media (Chen et al., 2017).

These surveys organise graph visualisation methods at the technique level, or specialise in a particular technique and present a survey of research in the area in-depth. These surveys do not focus on questions about how scalable these representations are from a human-centred perspective.

In the area of dynamic graphs, surveys have been conducted on dynamic networks (Beck et al., 2017) and dynamic data in information visualisation in general (Bach et al., 2017a). There have also been reviews focused on the human-centred effectiveness of animation, small multiples, and drawing stability (mental map preservation) by summarising experimental results and providing guidelines for visualisation designers. One such work by Archambault and Purchase (2013) summarises empirical results that relate to mental map preservation in dynamic graph drawing. In a later work (Archambault and Purchase, 2016a), based on the results of new studies, they review the conditions where animation and small multiples are effective and present new results for diagrams of low drawing stability. These papers focus on dynamic network visualisation and do not consider network visualisation in general. While providing a survey of human-centred effectiveness of visualisations to some degree, they do not consider cognitive scalability of the representations directly.

In this paper, we review evaluations of node–link visualisations of static and dynamic graphs. What is unique to our survey is that it examines cognitive scalability of node–link visualisations of graphs through the lens of human-centred experiments, to gain bounds on the sizes of graphs that have been displayed to the human while still usefully supporting analysis tasks. We seek to answer this question by surveying the literature of controlled experiments involving human participants to test node–link diagram representations of networks. Our summary information about the networks, techniques and tasks considered in these studies also presents an up-to-date snapshot of the evolution and state-of-the-art of controlled network visualisation evaluation.

## 3. Methodology

In this section, we clarify the scope of this survey, including the venues that we examined, and describe our categorisation framework. We have tried to be as systematic as possible in covering complete conference and journal venues with clearly defined constraints, as detailed below.

### 3.1. Scope of survey

To the best of our ability, we have sought to include in this survey all papers with a human-centred experiment (formal user study) where at least one of the conditions is a node–link representation. Both static and dynamic graph drawing studies were considered. For much of our analysis we focus on individual studies, where papers could contain multiple studies. Throughout the survey, ‘paper’ refers to the publication that presents the study, while ‘study’ refers to an individual experiment.

Our time range begins with the earliest formal human studies in network visualisation of which we are aware (Purchase et al., 1995) and ends on the 31st of March 2018. We consider the major

**Table 1**  
Venues considered in this survey.

Venue	First paper	Studies	Papers	References
<i>ACM CHI</i>	2006	16	13	Zhao et al. (2015), Dunne and Shneiderman (2013), Alper et al. (2014), Moscovich et al. (2009), Alper et al. (2013), Lee and Archambault (2016), Zhao et al. (2016), Song et al. (2010), Holten and van Wijk (2009), Ezaiza et al. (2016), Schöffel et al. (2016), Nekrasovski et al. (2006) and Chang et al. (2017)
<i>Diagrams</i>	2006	5	5	Yoon et al. (2006), Purchase and Samra (2008), Novick and Catley (2006), Chivers and Rodgers (2014) and Ware et al. (2008)
<i>EuroVis &amp; CGF</i>	2009	16	13	Okoe et al. (2014), Saket et al. (2016), Ghani et al. (2011), Okoe and Jianu (2015), Saket et al. (2015), Ghani et al. (2012), Ziemkiewicz and Kosara (2009), Shin et al. (2011), Archambault et al. (2010), Bae et al. (2017), Cypko et al. (2017), Tanahashi et al. (2016) and Didimo et al. (2018)
<i>GD</i>	1995	25	23	Hu and Shi (2014), Meulemans and Schulz (2015), Bridgeman and Tamassia (2000), Kobourov et al. (2014), Archambault et al., Zimmer and Kerren (2015), Burch et al. (2011b), Federico and Miksch (2016), Giacomo et al. (2013), Marner et al. (2014), Purchase et al. (2006), Huang et al. (2005), Archambault and Purchase (2012b), Barth et al. (2015), Purchase et al. (2012a), Blythe et al. (1995), Purchase et al. (2000), Purchase et al. (1995), Greffard et al. (2011), Purchase (1997b), Kindermann et al. (2017), Okoe et al. (2017) and Ballweg et al. (2017)
<i>IVJ</i>	2002	19	14	North et al. (2011), Dawson et al. (2015), Robertson et al. (2002), Ware et al. (2002), Farrugia and Quigley (2011), Keller et al. (2006), Huang et al. (2009b), Archambault and Purchase (2016b), Ghoniem et al. (2005), Ware and Bobrow (2005), Liu et al. (2017), Verspoor et al. (2018), Koylu and Guo (2017) and Hoque and Carenini (2018)
<i>PacificVis</i>	2008	7	6	Huang et al. (2009a), Holten et al. (2011), Huang et al. (2008), Archambault and Purchase (2012a), Burch et al. (2013) and Huang et al. (2016a)
<i>InfoVis &amp; TVCG</i>	2003	61	47	Archambault et al. (2011), Purchase et al. (2012b), Dwyer et al. (2009), Xu et al. (2012), Netzel et al. (2014), Bae and Watson (2011), Rufiange and McGuffin (2013), Borkin et al. (2013), Burch et al. (2011a), Wong et al. (2006), Bach et al. (2013), Kieffer et al. (2015), Jianu et al. (2014), Yuan et al. (2012), Marriott et al. (2012), Saket et al. (2014), van Ham and Rogowitz (2008), Cohe et al. (2015), Alper et al. (2011), Tennekes and de Jonge (2014), Lee et al. (2006a), Hlawatsch et al. (2014), Dwyer et al. (2013), Shi et al. (2015), Kwon et al. (2016), Guo et al. (2015), Huang et al. (2016b), Tan et al. (2007), Tu and Shen (2008), Giacomo et al. (2007), Henry et al. (2008), Ziemkiewicz and Kosara (2008), Kadaba et al. (2007), Wu et al. (2017), Cordeil et al. (2017), Yang et al. (2017a), Bach et al. (2017b), Srinivasan and Stasko (2018), Yang et al. (2017b), Kwon et al. (2018), Yeghoydjian et al. (2018), Walny et al. (2018), Zhao et al. (2018), Hu et al. (2018), Pienta et al. (2018), Ghoniem et al. (2004), Wong et al. (2005) and Wong et al. (2003)
Other		3	3	Huang and Huang (2010), Huang (2007) and Huang et al. (2010)
<b>Total</b>		<b>152</b>	<b>124</b>	

conferences and journals listed in Table 1, and examine all publications from this date, or the founding of, the conference/journal. This date is based on the above limit and accessibility of the venue. Well-known articles outside these venues are also included and listed under 'Other'.

Each author examined a venue that they were familiar with. Another author double checked the encoding, thus having at least two authors check each venue and encoding.

For most venues, we were able to read all titles and further examine the abstracts of papers that were relevant to our survey. The only exception was *ACM CHI*, where there were far too many papers: *CHI* accepts up to 600 papers in total each year. Instead, we found relevant papers at *CHI* using the *HCI Bibliography*.<sup>2</sup> A query of this database, limited to the *CHI* conference and using search terms '(network | graph) visuali\* study', returned 25 results, of which close inspection revealed 12 to be relevant, as detailed in Table 1.<sup>3</sup>

### 3.2. Categorisation framework

Information about each paper was collected and coded according to a number of criteria. Section 4 reports our findings on the sizes of graphs used in experiments. We identified the number of nodes and edges used within each study and computed the density of the graphs. If the study was on dynamic graphs, we counted the number of timeslices. We noted if this information was explicitly stated or whether it needed to be derived or inferred

(only nodes, only edges, or both). Exact numbers were sometimes unavailable, but we estimated the sizes based on figures of the stimuli (e.g. Huang et al. (2010) and Huang and Huang (2010)). If authors provided the algorithm and parameters used to generate the graphs, we computed the size based on this information (e.g. Holten and van Wijk (2009), Holten et al. (2011), Ware and Bobrow (2005) and Xu et al. (2012)).

Section 5 discusses other factors we found relating to scalability and, as such, is divided into three parts: HCI factors, graph drawing factors, and study design. The first part presents our findings on factors relative to human-computer interaction and scalability. We coded and described the types of tasks (Lee et al., 2006b) and interaction (Yi et al., 2007) used in each study. Application areas – if any – and the challenges they pose were also collected and are discussed in this section. The second part presents graph drawing factors related to scalability. We gathered information about the types of graphs, whether they were static or dynamic, and if they had attributes. We recorded information about the graph structure and noted if the data was real or generated. We also coded the layout algorithm used in the experiment. The third part discusses factors with respect to study design, including the number and nature of participants and whether the studies were within or between subject. In addition, we discuss the results of the study, and present the studies that are interesting in this data set.

Section 6 presents information about how the authors of experiments decide how large a graph should be presented to a participant. In particular, it discusses pilot studies and justification for using graphs of a certain size. In Section 7, we discuss the evolution of studies and their design over time in our community. Information about how studies have evolved and their venues is presented here also. The final sections of this paper include a discussion, recommendations for our community, and a conclusion.

<sup>2</sup> <http://hcibib.org/bs.cgi>.

<sup>3</sup> We also provide our curated bibliography as an online Tableau story <http://vahany.com/media/networkSize.html>.

#### 4. Basic measures of complexity

In this section, we consider measures of size that can be used to describe the graphs used in studies.

There are numerous measures that could be used when considering complexity of graph visualisations. However, we have found that in reports on graph study design and methodology, typically, few are considered. The number of nodes is the most commonly reported measure. The number of edges and/or density, on the other hand, is less frequently provided – although it is known to be a significant factor (Ghoniem et al., 2005). More edges, inevitably lead to more edge crossings, which is known to affect readability (Purchase, 1997a). When the data changes over time (i.e. dynamic graphs) additional measures, such as number of timeslices, become similarly important to gauge complexity.

Since our interest is primarily in findings regarding the scalability of network visualisation, we report on the distribution of graph sizes considered within studies. Table 2 summarises the minimum, maximum, average, median, and upper/lower quartiles for these metrics. While there is a large range across all these metrics, in each case the median is closer to the minimum. This skewed distribution implies that the majority of studies use small graphs.

In order to have a better understanding of the range of sizes of graphs used across and within user studies, we plotted for each study the number of nodes against the number of edges for both the smallest graph used in each study and the largest (Figs. 1, 2). Each minimum and maximum pair is connected by a link showing the range of graph sizes evaluated in that study. The circles are: hollow ○ if the number of nodes and edges was not mentioned by the authors; double enclosed ⊙ if only the number of nodes was mentioned; full ● if both number of nodes and edges were mentioned, or both metrics were otherwise apparent (e.g. if number of nodes  $n$  was given and the graph type was a tree, we assumed  $n - 1$  edges were present). In cases where the information was not mentioned we estimated the graph sizes from the figures (e.g. Zhao et al. (2016), Schöffel et al. (2016), Okoe and Jianu (2015), Ziemkiewicz and Kosara (2009), Bridgeman and Tamassia (2000), Burch et al. (2011a), Kieffer et al. (2015), van Ham and Rogowitz (2008), Tennekes and de Jonge (2014), Giacomo et al. (2007), Ziemkiewicz and Kosara (2008) and Tanahashi et al. (2016)). We also used the hue of the circles to differentiate between static ■ and dynamic ■ graphs.

Where precise sizes were not provided, we did our best to infer approximate sizes. For example, in some cases we were able to estimate the graph sizes from the figures (Zhao et al., 2016; Schöffel et al., 2016; Okoe and Jianu, 2015; Ziemkiewicz and Kosara, 2009; Bridgeman and Tamassia, 2000; Burch et al., 2011a; Kieffer et al., 2015; van Ham and Rogowitz, 2008; Tennekes and de Jonge, 2014; Giacomo et al., 2007; Ziemkiewicz and Kosara, 2008; Tanahashi et al., 2016).

The following subsections discuss each of the four measures of Table 2 in more detail.

##### 4.1. Number of nodes

The number of nodes in a graph is an important, but incomplete indicator, of complexity. It is clear that, for most tasks, difficulty is affected by introducing more nodes to a connected graph. We would assume that most experimenters pilot, or at least consider, graphs with different numbers of nodes to avoid tasks that are too trivial or impossible. Yet out of the 152 studies covered in this survey, 28 studies do not mention node count at all.

Among the studies that are included in our survey, 15 studies use graphs with more than 1000 nodes (Marner et al., 2014; Wong et al., 2006; Borkin et al., 2013; Nekrasovski et al., 2006; Dunne and Shneiderman, 2013; Archambault et al., 2010; Zimmer and

Kerren, 2015; Ware and Bobrow, 2005; Tu and Shen, 2008; Wu et al., 2017; Liu et al., 2017; Kwon et al., 2018; Yoghourdjian et al., 2018) and another nine that use graphs with more than 500 nodes (Moscovich et al., 2009; Lee and Archambault, 2016; Okoe et al., 2014; Jianu et al., 2014; Yuan et al., 2012; Purchase et al., 2012b; Yoghourdjian et al., 2018; Srinivasan and Stasko, 2018).

Among these studies, the majority aim is to evaluate tools that use interactive exploration to extract parts (e.g. neighbourhoods) of the graph (20/24 studies, 83%). Some of these studies evaluate aggregation techniques, thus they would require graphs with a large number of nodes, in order to highlight the benefits of compressing several nodes into fewer representations.

It is problematic to infer cognitive scalability of graph visualisation in the presence of interactivity because most of these studies do not ask the participants to perform the tasks on the whole graph. Rather, only a part of the graph is visible. When the authors do not report the precise number of nodes actually visible to the user, there is little that can be inferred about cognitive scalability. Eight of these 24 studies also use smaller networks with fewer than 100 nodes in their evaluations. This is shown by the long lines (Figs. 1, 2) that connect different graph sizes used by the same studies.

In order to understand the selection of number of nodes, we plotted a histogram of the number of nodes. Fig. 3 shows a number of spikes around specific numbers of nodes: 20, 50, and 100. The biggest spikes are at graphs with 50 nodes, which were used by 18 studies, followed by 20 and 100, used in 16 and 14 studies respectively. We believe that spikes at these round numbers suggest experimenters choose the number of nodes arbitrarily. These round numbers were not identified by empirical research and a formal study on ceiling and floor effects for graph cognitive scalability might lead to a better selection of number of nodes.

Nine additional studies use graphs with more than 200 nodes. Similar to the above, some of these studies only show subparts of the network (Ware et al., 2008; Cypko et al., 2017; Kwon et al., 2016; Okoe et al., 2017), while some evaluate tools that scale well with large networks (Ware et al., 2008; Cypko et al., 2017; Kwon et al., 2016; Holten et al., 2011; Giacomo et al., 2013; Burch et al., 2011a, 2013; Greffard et al., 2011). Four out of these nine studies have also used networks with less than 100 nodes.

To conclude, only 56 out of 152 studies use graphs with more than 100 nodes. Most of the studies that use a large number of nodes, use abstractions or aggregation to show only parts of the graph at a time, but do not report on the number of nodes seen at a given abstraction. In general, it is not clear why most researchers choose to use graphs with 100 nodes or less (121/152 studies, 80%). Some authors mention pilot studies where they discovered ceiling or floor effects, which could be seen as indications of cognitive scalability. We discuss these in Section 6.1; however to our knowledge, controlled studies were not conducted to verify or explain these effects.

##### 4.2. Number of edges

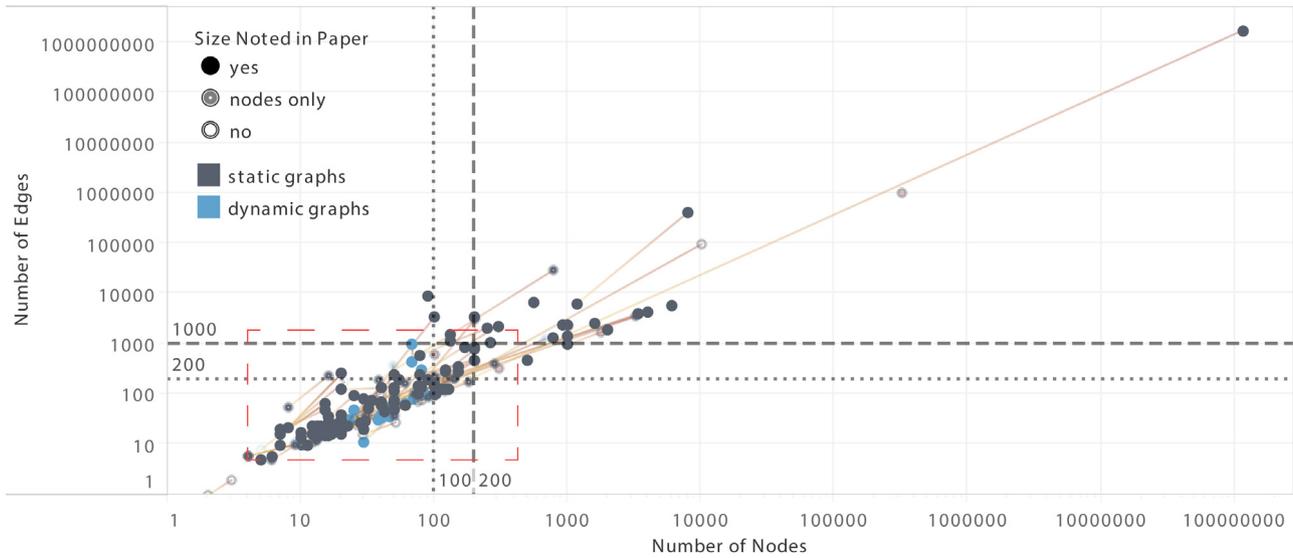
The number of edges in a graph is also an important indicator of complexity, especially in node-link diagrams, where the edges are drawn as lines. A large number of edges necessitate an increased number of crossings and overlap (collinear and therefore ambiguous lines). An excessive number of links often lead to what are known as 'hairball' visualisations. Among the 152 studies covered in this survey, 124 explicitly mention the number of nodes, while only 87 specify the number of edges.

There are only 28 studies that use graphs with 1000 edges or more. Most of these allow the participants to look at parts of the network instead of performing the task on the whole network (Nekrasovski et al., 2006; Dunne and Shneiderman, 2013; Archambault et al., 2010; Ware and Bobrow, 2005; Liu et al., 2017;

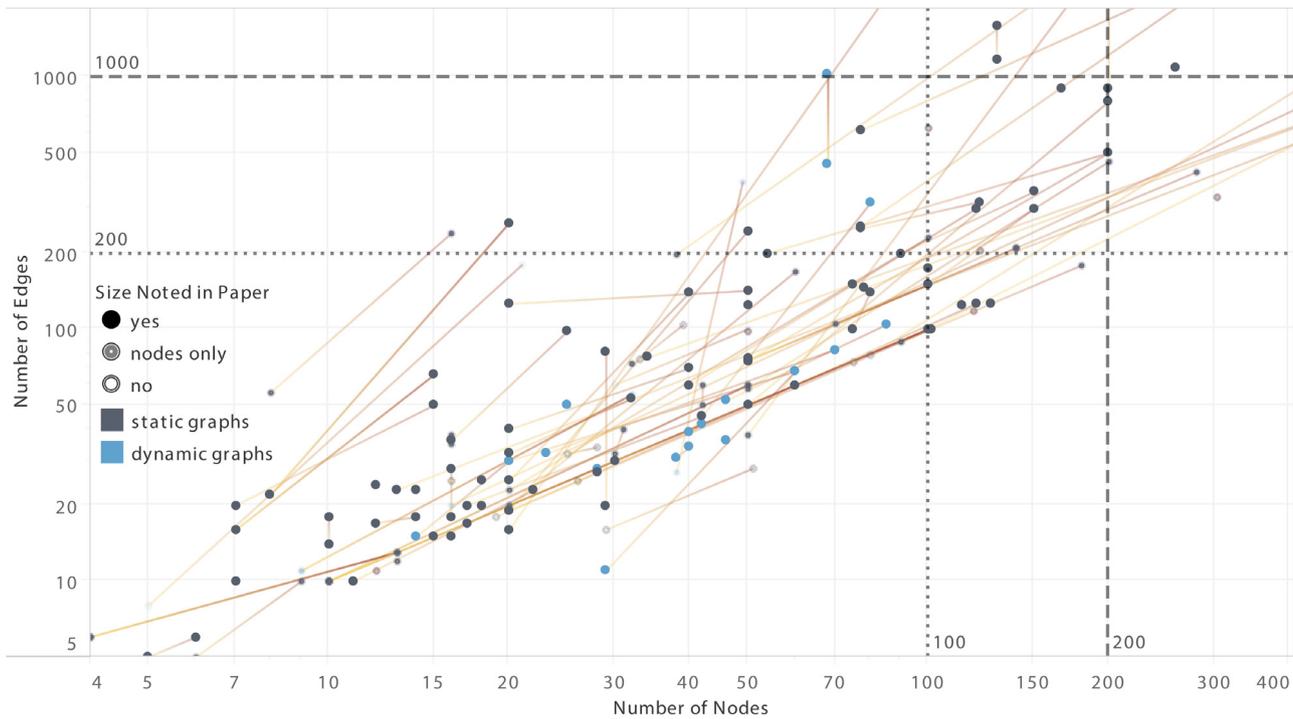
**Table 2**

Summary of graph sizes used in usability studies of graph visualisation. We used  $\frac{|E|}{|V|(|V|-1)}$  to calculate density, and  $\frac{|E|}{|V|}$  to calculate linear density.

Measure	Minimum	Maximum	Average	Median	Lower quartile	Upper quartile
Nodes	2	113 M.	457,242	49.5	18.5	116.5
Edges	1	1.8 B.	7.87 M.	73	23	242.5
Density	0.014	115.5	10	5	1.7	10
Linear density	0.38	102.8	3.5	1.5	1	2.4
Timeslices	2	15	6	6	3	7



**Fig. 1.** The size of graphs used in user studies. The x-axis shows the number of nodes, while the y-axis shows the number of edges. Both axes have a log scale. The grey circles represent static graphs, while the blue shows dynamic graphs.



**Fig. 2.** The section, marked by red dotted lines in Fig. 1, magnified for better readability.

Verspoor et al., 2018; Lee et al., 2006a; Okoe et al., 2014; Moscovich et al., 2009; Kwon et al., 2016; Zimmer and Kerren, 2015; Giacomo et al., 2013; Okoe et al., 2017; Tu and Shen, 2008; Yuan et al., 2012; Henry et al., 2008; Pienta et al., 2018; Wong et al., 2006; Srinivasan

and Stasko, 2018; Shi et al., 2015). Similar to studies that use a large number of nodes to highlight the benefits of aggregation or interaction methods, some studies use a large number of edges to show the benefits of edge compression, bundling, or highlighting

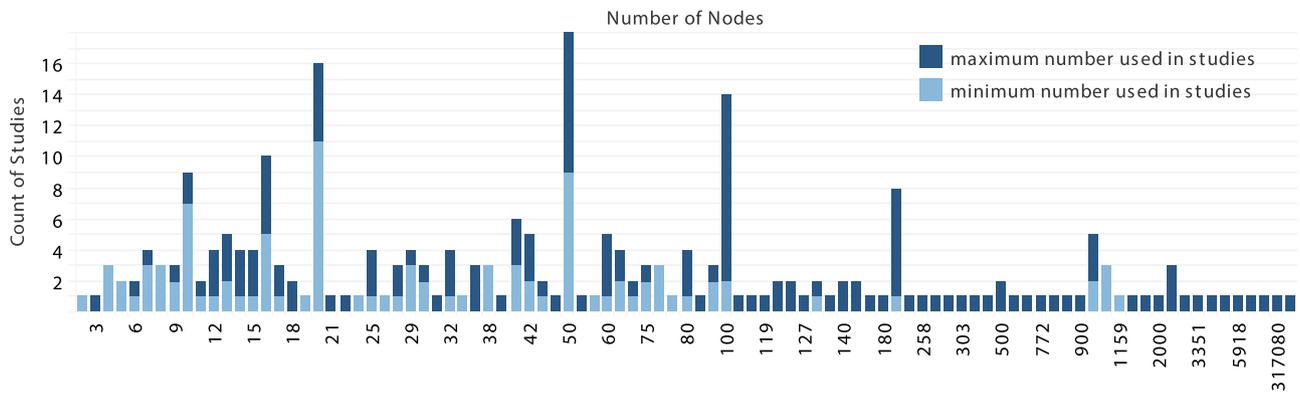


Fig. 3. Histogram of the minimum (light blue) and maximum (dark blue) number of nodes of graphs used in studies.

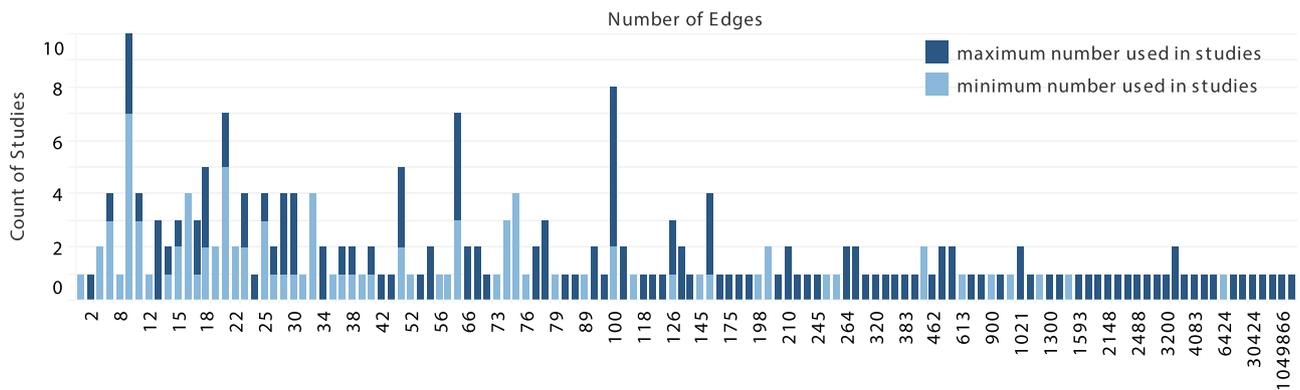


Fig. 4. Histogram of the minimum (light blue) and maximum (dark blue) number of edges of graphs used in studies.

techniques. Many evaluate tools or techniques, by design, scale well to handle graphs with a large number of edges. For example, a study by Giacomo et al. (2013) evaluates a technique that highlights edges in order to enhance the readability of graphs that have many edge crossings. Another category of studies that use a large number of edges, evaluate visualisations (e.g. adjacency matrices) that scale well with a large number of edges in comparison to node–link diagrams (Ghoniem et al., 2005; Bae and Watson, 2011; Henry et al., 2008; Yang et al., 2017b; Kwon et al., 2018; Okoe et al., 2017).

Again, just as we saw spikes in frequency at round numbers of nodes used for study graphs (Fig. 3), in Fig. 4 we see spikes particularly at 10, 20, 60, and 100 edges. However, the spikes are less pronounced: only ten, seven, seven, and eight studies, respectively. The smaller spikes for number of edges compared to number of nodes, tend to suggest that experimenters choose a specific number of nodes first, then adjust the density, presumably to control the difficulty level for their specific tasks. We found that 97 studies use graphs with less than 100 edges, while 75 studies use graphs with 100 edges or more.

In summary, we were surprised that about half of the studies do not report the number of edges. We would argue that without this information, graph evaluations are difficult to reproduce. The majority of studies use graphs with less than 1000 edges (126/152 studies, 83%). The 28 studies that use graphs with 1000 edges or more, either evaluate tools that require dense community structures, aim to show that node–link diagrams fail to perform well on graphs with a large number of edges, or highlight the benefits of aggregation and interaction techniques.

#### 4.3. Density

The previous section mentions studies that use graphs with a large number of edges. Nonetheless, most of these graphs have very low densities (relatively few edges to the number of nodes).

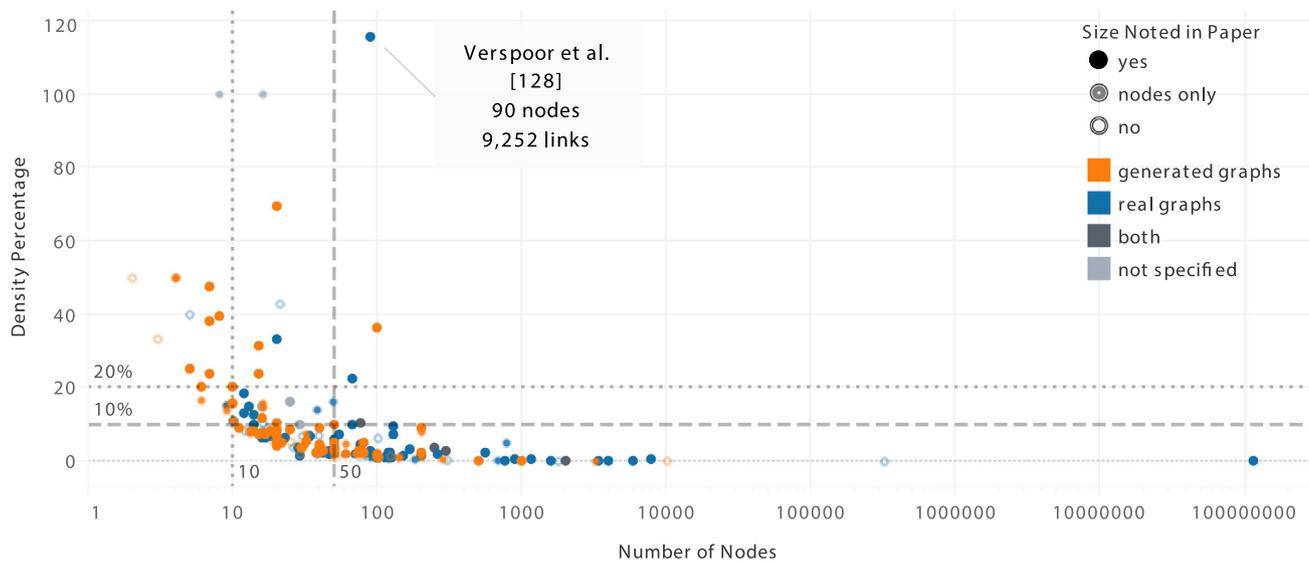
In this section we focus on studies that explicitly consider the effect of density on readability. We used  $\frac{|E|}{|V|(|V|-1)}$  to calculate density. This represents the ratio of the number of existing edges  $E$  to the number of all possible edges for the number of nodes  $V$ . For simplicity, we multiply this ratio by 100 to achieve percentages.

There are also other ways to derive density. The so-called *linear density*  $\frac{|E|}{|V|}$  is often used to compare the number of edges  $E$  to the number of nodes  $V$ . In this measure, tree-like graphs will have density close to 1. Ghoniem et al. (2004) suggest using *square-root density*  $\sqrt{\frac{|E|}{|V|^2}}$ , since its range is bounded to the interval  $[0, \frac{1}{\sqrt{2}}]$ . Any of these definitions can be used; the choice depends on utility (Melancon, 2006).

Fig. 5 shows the density of real and generated graphs used in studies. Most studies use graphs with densities of 50% or less. There are only three studies that use graphs with densities higher than 50% (Verspoor et al., 2018; Yang et al., 2017a; Hlawatsch et al., 2014). All three studies evaluate methods that are tailored to scale well for dense networks.

In fact, all studies that use graphs with ten nodes or more and a density of more than 20% evaluate matrix-like visualisations that scale well for dense graphs (Zhao et al., 2016; Keller et al., 2006; Ghoniem et al., 2005; Yang et al., 2017b) or evaluate edge-compression and edge-bundling tools (Dwyer et al., 2013; Bach et al., 2017b).

In summary, in the surveyed studies, dense graphs are mostly small and have less than ten nodes. The only studies that evaluate visualisations using dense graphs (>20% density) with more than



**Fig. 5.** The density of graphs used in the studies. Out of 152 studies, 129 studies use sparse graphs with densities of less than 20%, while 119 studies use graphs with less than 10% density.

a trivial number of nodes, tend to be those testing matrix diagrams or edge compression techniques. With the exception of these types of studies, all studies that use graphs with more than 50 nodes, choose to use sparse graphs with a density of less than 10%. 119 out of 152 studies (78%) use graphs with a density of less than 10%, while 129 out of 152 studies (85%) use graphs with less than 20% density.

#### 4.4. Number of timeslices

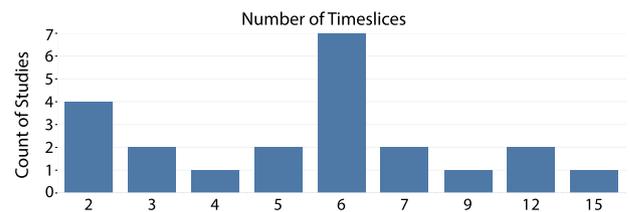
In the case of dynamic graphs or static graphs with dynamic attributes, and in addition to the number of nodes, and the number of edges or density, the number of timeslices is an important measure of cognitive scalability.

Among the 152 studies (described in 124 papers) covered in this survey, 22 use dynamic graphs. Compared to the reporting of size metrics discussed above, it seems dynamic graph evaluation papers are reasonably consistent about reporting the number of timeslices.

Fig. 6 shows a histogram of the number of timeslices. For previous metrics, we used the minimum and maximum values to derive the charts, but since there are only a few studies that use dynamic graphs, we included all the values. There is no visible difference between odd and even numbers. There is, however, a clearly visible spike at six timeslices. Seven studies, out of the total 22, use dynamic graphs with six timeslices.

We suggest that six timeslices are typically chosen as it is small enough for a small multiples representation of a dynamic graph to fit on a screen with each timeslice being at a reasonable scale. Farrugia and Quigley (2011) conduct two studies in order to compare animated displays to static ones. They used two graphs with six timeslices each. For the static view, they placed the six timeslices next to each other on a  $2 \times 3$  grid. Similarly, Archambault and Purchase (2016b) conduct a study that evaluates the effects of three factors—static versus animated presentation of dynamic attribute values, force-directed versus hierarchical layout of constant graph structure, and ‘with history’ versus ‘without history’ persistence, for displaying graphs with dynamic attributes. They use six timeslices, with each covering one-sixth of the screen.

Shi et al. (2015) conduct the only study that tests a variety of time slice counts. They demonstrate the scalability of various aggregation techniques against the more typical small multiples display of timeslices. They use four dynamic graphs with different



**Fig. 6.** Histogram of the number of timeslices of dynamic graphs used in the surveyed studies.

numbers of timeslices. They use graphs with 15, 12, 4, and 5 timeslices and 674, 109, 298, and 16 nodes respectively. For their small-multiples condition, they use either a 6 or 24-cell grid to make the full set of timeslices for each graph visible. Unlike the studies above, where the number of timeslices was chosen to fill the small multiples grid, in this study the odd numbers of timeslices would have left part of the screen unused.

Others choose the number of timeslices according to the norms of specific application areas. North et al. (2011) use a directed graph with 46 nodes, 36 edges, and 12 timeslices in their evaluation of three existing visualisation methods for dynamic graphs. They claim that this is a typical size of pathways used by Biologists.

In addition to the number of timeslices, the number of graph elements that change from one timeslice to another is important. Authors report these values in different ways. For example, some state the maximum number of changed elements (Archambault and Purchase, 2012b), while others report on the average (Purchase and Samra, 2008; Purchase et al., 2006).

In summary, the number of timeslices for dynamic graphs is often small. Most studies use less than ten timeslices (except for North et al. (2011); Shi et al. (2015)). Moreover, they often consider only a fixed number of timeslices (except Shi et al. (2015) who use four variations). Some studies pick the number of timeslices to best fit their visualisations on screen. Others choose what is common in the respective application area.

## 5. Other factors and scalability

In addition to the basic measures of the previous section, other factors and their interplay influence the cognitive scalability of graph visualisations. In the following sections, we discuss HCI, graph drawing, and study design factors.

## 5.1. HCI factors

Graph visualisations are often tailored to efficiently serve domain-specific tasks. Modern visualisation tools are equipped with heavy interaction. Such Human–Computer Interaction (HCI) factors play a key role in allowing graph visualisations to scale to larger data sets. They were rarely investigated in earlier visualisations, but have been widely discussed in recent times.

In this section, we discuss the interplay between different types of tasks, interaction techniques and application areas on one hand, and the basic measures of graph size on the other.

### 5.1.1. Tasks

To analyse and understand the trend, as well as the impacts of scalability in relation to tasks, we classified the various tasks used in studies into four main categories based on the taxonomy of Lee et al. (2006b). Fig. 7(a) shows a summary of tasks investigated in our survey: *Topology-based* tasks allow participants to detect node adjacency, accessibility, common connection, and connectivity. *Attribute-based* tasks support identification of nodes and links by the data attributes associated with them. *Browsing* tasks include following or revisiting a path. High-level or deliberately more abstract tasks were classified as *Overview* tasks. Note that we considered tasks that asked the participants to find the shortest path to be in the *Topology-based* category, even if they required some paths following. We categorised path following tasks as *Browsing*, only if participants were explicitly asked to follow a certain path.

Table 3 shows how the different tasks were used in the 152 studies (noting that several studies used more than one task). *Topology-based* tasks were the most common (47%), followed by *Attribute-based* (22%) and *Overview* (21%). Three quarters of the studies (114 studies) used tasks that are categorised as *Topology-based*.

This might be an indication that researchers perceive tasks concerned with understanding the structure of the graph as good evaluation measures for graph visualisation. However, the number of studies that use *Topology-based* decreases as graph size increases, while *Overview* tasks become more popular. They become equally popular (37%) in studies that use graphs with 300 nodes or more. For graphs with 1000 nodes or more, *Overview* tasks are used excessively (43%), while *Topology-based* tasks become less common (30%). We believe that this is expected, since *Overview* tasks do not require a detailed understanding of the graph, but deal with general properties and estimates.

In order to understand the combinations of tasks in each study, we plot these as composite nodes in Fig. 8. Six studies use only *Topology-based* tasks when using graphs with 500 nodes or more (Burch et al., 2013, 2011a; Nekrasovski et al., 2006; Okoe et al., 2014; Wong et al., 2006). The first three studies (Burch et al., 2013, 2011a; Nekrasovski et al., 2006) use graphs with a simpler structure, i.e. trees. Moreover, the trees they use have less than 1000 nodes. The other three studies (Wong et al., 2006; Nekrasovski et al., 2006; Pienta et al., 2018), which use graphs with more than 5000 nodes, allow participants to interact with the graphs and show only parts of the graph at a given time. This implies that tasks and interactions are mutually reinforcing each other in large graph visualisation.

Marner et al. (2014) use a large graph with 7885 nodes and 427,406 edges in their study. They project the graph on a wall-sized display and ask participants to untangle it until they achieve a better overview. We believe that even though their process allowed for an *Overview* task, it would have been almost impossible for the participants to perform more complex tasks, such as path-finding or counting triangles. Similarly, Kwon et al. (2018) use graphs of up to 113 million nodes and 1.8 billion edges in their

study, but only ask the participants to rate the similarities between the graphs. Note that among all tasks, *Topology-based* tasks have been widely used in most of the studies since the connectivity plays a key role for understanding graph structures. Path finding is the most common task (66 over 114 in our survey) among *Topology-based* tasks, and therefore can serve as the primary task for demonstrating the usability of graph visualisation. For example, following the investigation by Bae and Watson (2011), and Ghoniem et al. (2004), it is reasonable to use node–link diagrams for path-finding purposes when the graph is less than 200 nodes.

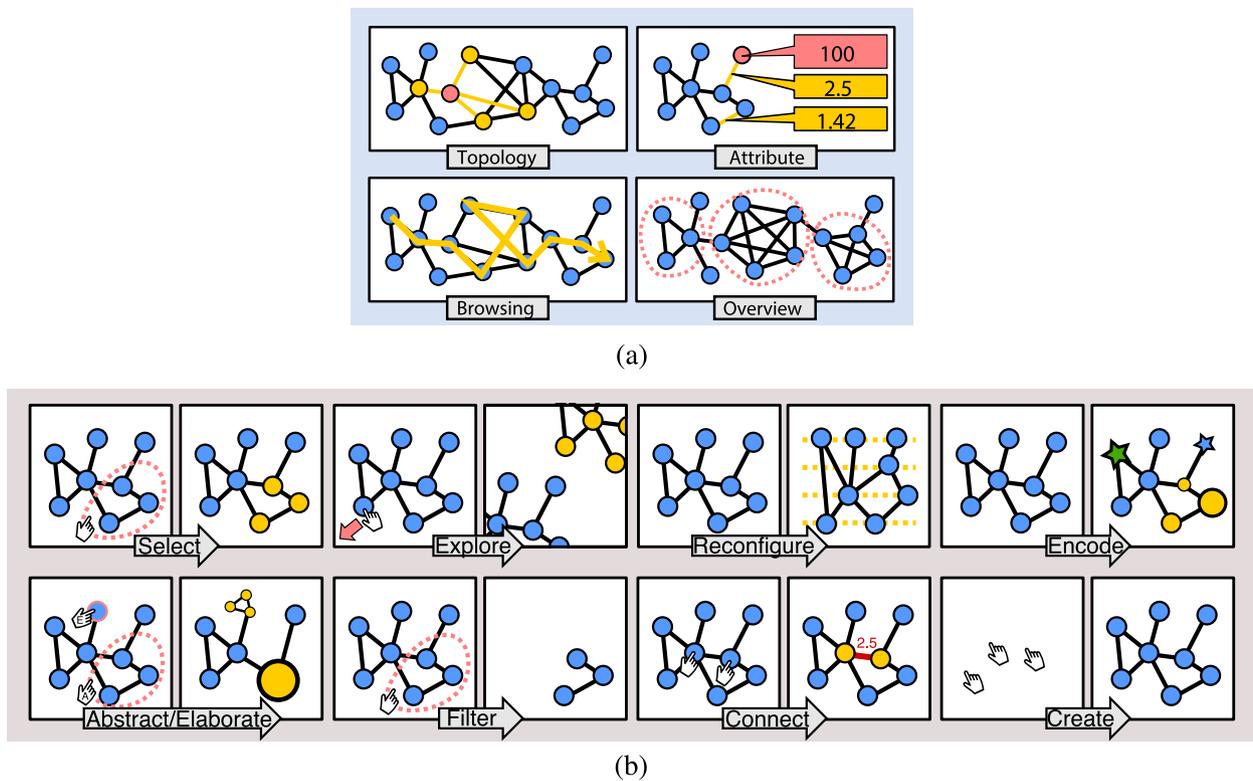
Another interesting study using *Topology-based* tasks is conducted by Moscovich et al. (2009). They use two graphs with 1000 nodes. The sparser one has 1485 edges, while the denser one has 2488 edges. We believe that they wanted a graph large enough to make the task difficult, especially when finding the immediate neighbours of a given node. The authors propose an interactive navigation technique *Bring-and-Go*, which aids this task by bringing all adjacent nodes closer to a selected node. This study demonstrates the complexity of these tasks on graphs with a large number of edges. Both error rates and performance times increase significantly from the sparser graph to the denser graph.

Some authors discuss the choice of graph sizes in relation to the results of their studies. Huang et al. (2009b) justify the increase in errors by explaining that human perception and cognitive systems become overburdened when dealing with large graphs, even with 25 nodes and 98 links. Okoe et al. (2014) use a graph with 900 nodes and 2500 edges. However, they mention a high error rate of more than 50%. They associate this to the difficulty of the tasks. They explain that the large number of edges voided the highlighting advantage of the evaluated technique. Wong et al. (2006) mention that they tried to use graphs that were not too complex, but their attempts were not successful for all the tasks.

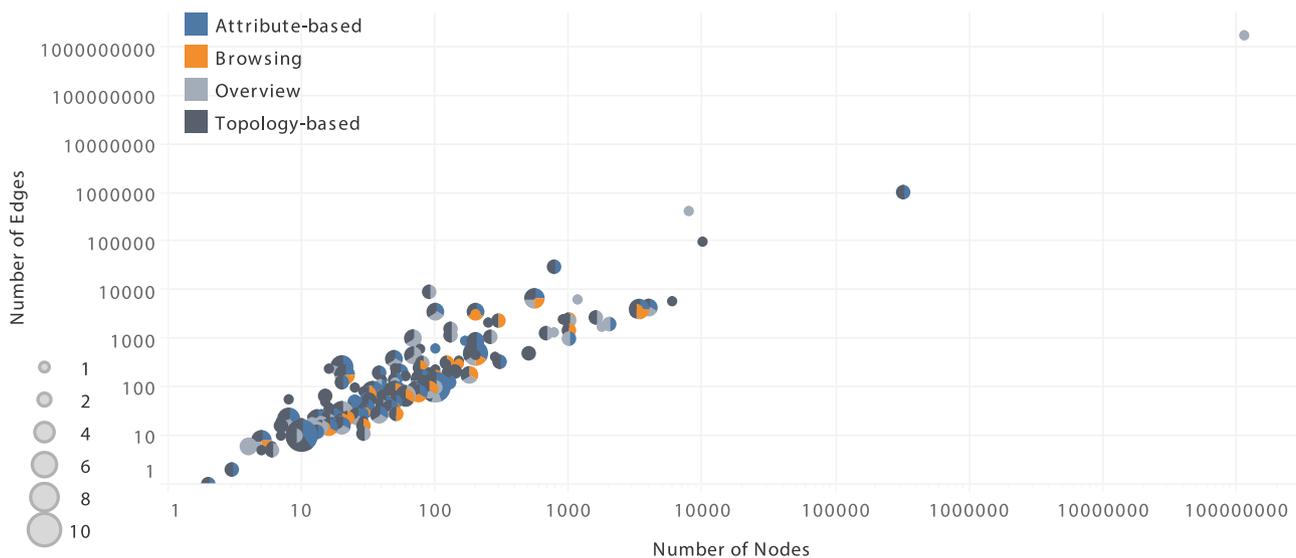
Only nine out of the 53 studies that require participants to perform *Attribute-based* tasks, use more than 200 nodes (Dunne and Shneiderman, 2013; Lee and Archambault, 2016; Archambault et al., 2010; Cypko et al., 2017; Borkin et al., 2013; Jianu et al., 2014; Srinivasan and Stasko, 2018; Pienta et al., 2018; Zimmer and Kerren, 2015). Two of these studies aggregate multiple nodes into singular representations: *motifs* (Dunne and Shneiderman, 2013) and *metanodes* (Archambault et al., 2010). It is natural to assume that *Attribute-based* tasks would be difficult on graphs with too many elements, since *Attribute-based* tasks are related to data attributes associated to nodes and links. While this is backed by our survey for number of nodes, the number of edges ranges between 10 and 1000 for most studies using *Attribute-based* tasks.

For *Browsing* tasks, we assumed that experimenters would use sparse graphs. This is backed by our findings as shown in Fig. 9. Also, in addition to the density of the graphs, studies that require performing *Browsing* tasks use graphs with few number of nodes ( $\leq 200$ ), with the exception of five outliers (Moscovich et al., 2009; Archambault et al., 2010; Ware and Bobrow, 2005; Kwon et al., 2016; Srinivasan and Stasko, 2018). Four of these outliers use interactive highlighting to assist the participants in performing the task (Moscovich et al., 2009; Ware and Bobrow, 2005; Kwon et al., 2016; Srinivasan and Stasko, 2018), while the fifth (Archambault et al., 2010) uses *metanodes* and *metaedges*, which are aggregations of multiple nodes and edges into singular representations. As seen in Fig. 8, *Browsing* tasks are only used on very sparse graphs ( $< 10\%$ ), except for one study by Zhao et al. (2016) that uses graphs with 40% and 42.9% densities. The latter allows participants to highlight specific nodes and their connections, thus showing small subsets of edges at a given time.

In summary, *Topology-based* tasks are the most common tasks to evaluate graph visualisation; nonetheless, the number of *Topology-based* tasks is relatively reduced when using graphs with more than 300 nodes. In such cases, *Overview* tasks become more



**Fig. 7.** Task and interaction taxonomies. Task taxonomy (a) proposed by Lee et al. (2006b), which mainly includes *Topology*, *Attribute*, *Browsing*, and *Overview* tasks. Interaction types (b) introduced by Yi et al. (2007), which include *Select*, *Explore*, *Reconfigure*, *Encode*, *Abstract/Elaborate*, *Filter*, and *Connect*. However, *Create* is newly added in our taxonomy.



**Fig. 8.** Types of different categories of tasks used in studies with relation to the number of nodes (on the x-axis) and the number of edges (on the y-axis). The areas of the circles reflect the count of studies that use similar values. Both axes have log scales.

common, especially when the tasks are assisted with multiple interaction techniques. Sparse graphs (<10%) with few nodes ( $\leq 200$ ) are used when asking the participants to perform *Browsing* tasks. Similarly, *Attribute-based* tasks are not common in studies that use graphs with more than 200 nodes.

### 5.1.2. Interaction

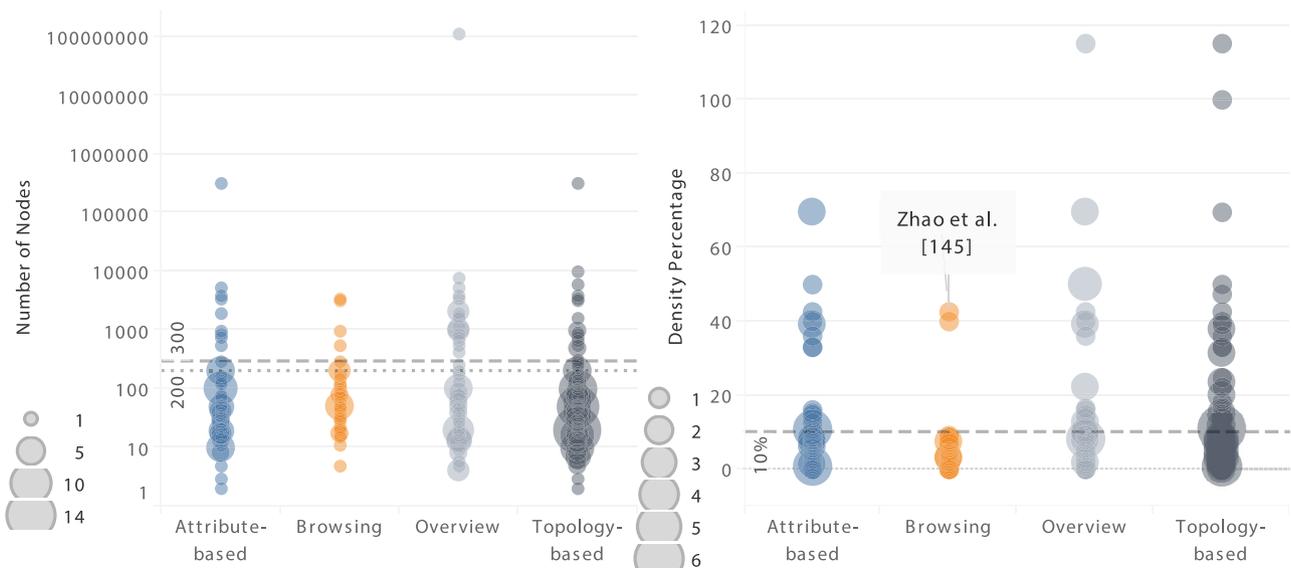
Early graph experiments did not make much use of interaction; their focus being on the interpretation of static graph drawing—often being simply presented on paper (e.g. Purchase (1997b),

Purchase et al. (2000) and Purchase et al. (1995)). In more recent years, several experiments have tested the worth of node-link diagrams using interactive systems that permit more extensive exploration of the relational information. In such cases, the use of interaction techniques is often a crucial component of the study design.

We used the interaction taxonomy of Yi et al. (2007) as a means of classifying the different types of interaction used in the experimental studies: *Select* allows users to mark objects on screen as interesting; *Explore* enables users to navigate a hidden subset

**Table 3**  
The type of tasks used within the studies.

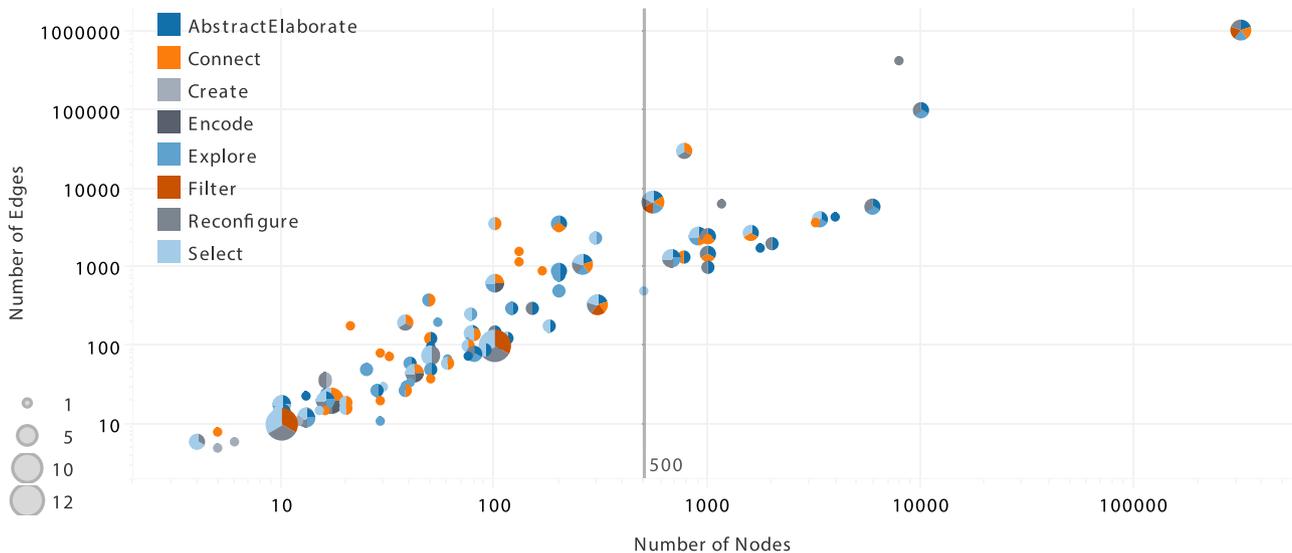
Task	References	# of studies
Attribute	Zhao et al. (2015), Dunne and Shneiderman (2013), Lee and Archambault (2016), Zhao et al. (2016), Song et al. (2010), Ezaiza et al. (2016), Schöffel et al. (2016), Tanahashi et al. (2016), Chivers and Rodgers (2014), Saket et al. (2016), Saket et al. (2015), Ghani et al. (2012), Shin et al. (2011), Archambault et al. (2010), Bridgeman and Tamassia (2000), Archambault et al., Zimmer and Kerren (2015), Federico and Miksch (2016), Huang et al. (2005), North et al. (2011), Robertson et al. (2002), Keller et al. (2006), Archambault and Purchase (2016b), Ghoniem et al. (2005), Wong et al. (2005), Borkin et al. (2013), Bach et al. (2013), Jianu et al. (2014), Saket et al. (2014), Tennekes and de Jonge (2014), Lee et al. (2006a), Hlawatsch et al. (2014), Guo et al. (2015), Huang et al. (2016b), Tan et al. (2007), Giacomo et al. (2007), Ziemkiewicz and Kosara (2008), Kadaba et al. (2007), Srinivasan and Stasko (2018), Zhao et al. (2018), Pienta et al. (2018), Koylu and Guo (2017); Hoque and Carenini (2018), Bae et al. (2017) and Cypko et al. (2017)	53
Browsing	Alper et al. (2014), Moscovich et al. (2009), Zhao et al. (2016), Song et al. (2010), Saket et al. (2016), Archambault et al. (2010), Archambault and Purchase (2012b), Barth et al. (2015), Ware and Bobrow (2005), Huang et al. (2009a), Huang et al. (2008), Dwyer et al. (2009), Netzel et al. (2014), Bae and Watson (2011), Wong et al. (2003), Cohe et al. (2015), Lee et al. (2006a), Kwon et al. (2016), Srinivasan and Stasko (2018), Zhao et al. (2018), Hu et al. (2018), Hoque and Carenini (2018) and Kindermann et al. (2017); Didimo et al. (2018)	26
Overview	Dunne and Shneiderman (2013), Alper et al. (2014), Alper et al. (2013), Zhao et al. (2016), Saket et al. (2015), Ziemkiewicz and Kosara (2009), Archambault et al. (2010), Archambault et al., Zimmer and Kerren (2015), Federico and Miksch (2016), Marnier et al. (2014), Purchase et al. (2006), Huang et al. (2005), Purchase et al. (2000), Greffard et al. (2011), North et al. (2011), Ghoniem et al. (2005), Ware and Bobrow (2005), Archambault and Purchase (2012a), Archambault et al. (2011), Netzel et al. (2014), Borkin et al. (2013), Bach et al. (2013), Kieffer et al. (2015), Jianu et al. (2014), Yuan et al. (2012), Tennekes and de Jonge (2014), Hlawatsch et al. (2014), Shi et al. (2015), Guo et al. (2015), Tu and Shen (2008), Henry et al. (2008), Wu et al. (2017), Srinivasan and Stasko (2018), Yang et al. (2017b), Kwon et al. (2018), Yoghourdjian et al. (2018), Liu et al. (2017), Verspoor et al. (2018), Koylu and Guo (2017), Okoe et al. (2017) and Ballweg et al. (2017)	52
Topology	Dunne and Shneiderman (2013), Alper et al. (2014), Moscovich et al. (2009), Alper et al. (2013), Lee and Archambault (2016), Zhao et al. (2016), Song et al. (2010), Holten and van Wijk (2009), Ezaiza et al. (2016), Nekrasovski et al. (2006), Huang (2007), Tanahashi et al. (2016), Yoon et al. (2006), Purchase and Samra (2008), Novick and Catley (2006), Chivers and Rodgers (2014), Ware et al. (2008), Okoe et al. (2014), Saket et al. (2016), Ghani et al. (2011), Okoe and Jianu (2015), Saket et al. (2015), Ziemkiewicz and Kosara (2009), Shin et al. (2011), Archambault et al. (2010), Hu and Shi (2014), Meulemans and Schulz (2015), Bridgeman and Tamassia (2000), Kobourov et al. (2014), Archambault et al., Burch et al. (2011b), Federico and Miksch (2016), Giacomo et al. (2013), Purchase et al. (2006), Huang et al. (2005), Purchase et al. (2012a), Blythe et al. (1995), Purchase et al. (1995), Purchase (1997b), North et al. (2011), Dawson et al. (2015), Robertson et al. (2002), Ware et al. (2002), Farrugia and Quigley (2011), Keller et al. (2006), Huang et al. (2009b), Ghoniem et al. (2005), Ware and Bobrow (2005), Huang et al. (2009a), Holten et al. (2011), Huang et al. (2008), Archambault and Purchase (2012a), Burch et al. (2013), Huang et al. (2016a), Archambault et al. (2011), Purchase et al. (2012b), Dwyer et al. (2009), Xu et al. (2012), Netzel et al. (2014), Rufiange and McGuffin (2013), Wong et al. (2005), Burch et al. (2011a), Wong et al. (2006), Kieffer et al. (2015), Jianu et al. (2014), Marriott et al. (2012), Saket et al. (2014), van Ham and Rogowitz (2008), Alper et al. (2011), Tennekes and de Jonge (2014), Lee et al. (2006a), Hlawatsch et al. (2014), Dwyer et al. (2013), Shi et al. (2015), Kwon et al. (2016), Tan et al. (2007), Henry et al. (2008), Ziemkiewicz and Kosara (2008), Kadaba et al. (2007), Huang and Huang (2010), Huang et al. (2010), Chang et al. (2017), Cordeil et al. (2017), Yang et al. (2017a), Bach et al. (2017b), Srinivasan and Stasko (2018), Yang et al. (2017b), Walny et al. (2018), Zhao et al. (2018), Hu et al. (2018), Pienta et al. (2018), Liu et al. (2017), Verspoor et al. (2018), Hoque and Carenini (2018), Kindermann et al. (2017), Okoe et al. (2017), Ballweg et al. (2017), Didimo et al. (2018), Bae et al. (2017) and Cypko et al. (2017)	114



**Fig. 9.** Types of different categories of tasks used in studies with relation to the number of nodes (on the left) and the density (on the right). The areas of the circles reflect the count of studies that use similar values.

of data; *Reconfigure* changes spatial arrangement; *Encode* allows users to transform data values to their preferred visual language (e.g. colour, size, or shapes); *Abstract/Elaborate* enables users to adjust the level of detail of a data representation; *Filter* allows

users to prevent the display of data fulfilling given conditions, and *Connect* allows the highlighting of relationships between data. For the purposes of this survey, we added an additional category: *Create*, which allows users to generate graph drawings.



**Fig. 10.** Types of different interaction used in studies with relation to the number of nodes and the number of edges. Both axes have log scales.

Of the 152 studies, 80 used no interaction at all. Table 4 shows how the remaining 72 used interaction (noting that several studies used more than one interaction technique). Of all the occasions when interaction techniques were used, *Abstract/Elaborate* was most common (20%), followed by *Select* (19%), *Explore* (16%) and *Reconfigure* (16%). Thus, there is no single category that dominates (unlike our finding with the task category).

Fig. 10 shows the different types of interaction used in the studies, with respect to graph size. We note that 25 of 60 studies (42%) use a single interaction type for those graphs with 500 nodes or fewer, while for those studies using graphs with more than 500 nodes, this ratio drops to 32% (6 out of 19 studies).

Tiny graphs (e.g. 4 nodes and 5 edges) are easy to understand without the support of interaction (see lower dashed lines shown in Fig. 11), although even studies that used graphs with as few as ten nodes and ten edges used interaction. The majority of studies that do not use any interaction use graphs with 100 nodes or less (70/80 studies, 88%) and 500 edges or less (71/80 studies, 89%) (upper dashed lines in Fig. 11). These bounds allow us to group the interaction techniques into two sets: techniques used on graphs with greater than 100 nodes and 500 edges: *Abstract/Elaborate*, *Connect*, *Explore*, *Reconfigure*, *Select*; and techniques used on graphs with fewer than 100 nodes and 500 edges: *Encode*, *Create*, and *Filter*. If we consider the degree of effort required by the user in applying these techniques, the first set can be considered relatively straight-forward—the techniques are usually tested by having participants apply the built-in functionality to look at the data in a different way until the answer is obvious. By contrast, *Encode* and *Create* require more effort on the part of the users, since (possibly creative) decisions need to be made.

Interaction tasks can come with a time trade-off, however, for example, finding labels inside an abstract node (which represents a set of aggregated nodes), could be more time consuming due to the interaction, than the task of finding labels outside of it, which is always completed faster (Dunne and Shneiderman, 2013). Care, therefore, needs to be taken in the interpretation of the timing data collected (especially if participants are allowed unlimited time to perform their task), since such data might include superfluous interaction.

To conclude, the combination of multiple interactions can improve the visualisation of large graphs, and this is more significant as graph size increases. Based on the studies we analysed, the floor effect of interaction lies at the boundary of nodes equal to 4 and edges equal to 5, with ceiling effect at 100 nodes and 500 edges.

### 5.1.3. Application areas

There are several domain-specific studies where the aim is enhanced understanding of the content, and so the tasks were clearly focused on the domain knowledge (e.g. Robertson et al. (2002); Novick and Catley (2006)). For example, Tanahashi et al. (2016) performed a comparative study of four different ways of presenting data for the purposes of introducing information visualisation to novices, where a graph drawing was one of the visualisation types. Their analysis not only considered the efficacy of the visualisations, but also looked at two different types of learning (active and passive), and two different teaching methods (top-down and bottom-up). They were therefore able to propose guidelines for writing effective information visualisation tutorials. North et al. (2011) compared two different types of evaluation (benchmark and insight) using three different visualisation alternatives depicting gene expression data, and all tasks were related to understanding of the data.

While some application areas have well defined and restricted characteristics for the networks under investigation, in others there can be a large range of impacting factors that potentially affect scalability. In the life sciences, a variety of network types are investigated, ranging in size from a few dozen to a few million nodes, and showing a similar variety in other network characteristics, e.g. density or diameter. In addition, the required tasks can differ significantly between use cases, e.g. from deciding reachability to the detection of dynamic patterns, which affects the limits of readability. RNA sequence graphs, like the ones used in Marner et al. (2014), can have up to several million nodes and edges, and dense local structures, but they often have a very sparse global structure, making the visual detection of so-called repeats (loops that indicate repetitive structures in RNA sequences), a feasible task. On the other hand, metabolic pathways like the ones used in Zimmer and Kerren (2015) are often planar or near-planar graphs with low local and global density, but require that the semantics associated with the metabolic flow is incorporated in the layout and the visual representation. While these pathways are parts of a large and complex network of metabolic reactions in an organism, the visual analysis is often restricted to such sparse sub-networks that have a specific functionality, e.g. the synthesis of a particular biomolecule.

### 5.2. Graph drawing factors

For several decades now, the field of Graph Drawing has led to the development of efficient algorithms for layout computation as

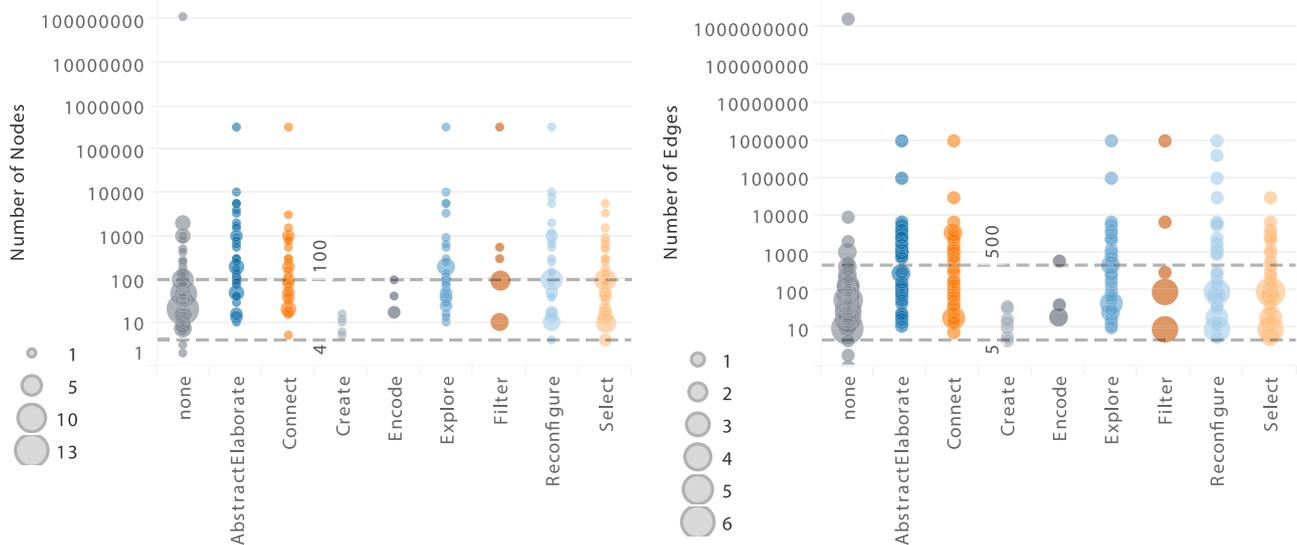


Fig. 11. Interaction types used in studies with relation to the number of nodes (on the left), and the number of edges (on the right).

Table 4

The type of interaction used within the studies.

Interaction	References	# of studies
None	Alper et al. (2014), Alper et al. (2013), Holten and van Wijk (2009), Schöffel et al. (2016), Huang (2007), Tanahashi et al. (2016), Yoon et al. (2006), Purchase and Samra (2008), Chivers and Rodgers (2014), Okoe and Jianu (2015), Ziemkiewicz and Kosara (2009), Hu and Shi (2014), Meulemans and Schulz (2015), Bridgeman and Tamassia (2000), Archambault et al., Burch et al. (2011b), Giacomo et al. (2013), Purchase et al. (2006), Huang et al. (2005), Archambault and Purchase (2012b), Purchase et al. (2012a), Blythe et al. (1995), Purchase et al. (2000), Purchase et al. (1995), Greffard et al. (2011), Purchase (1997b), North et al. (2011), Dawson et al. (2015), Ware et al. (2002), Farrugia and Quigley (2011), Keller et al. (2006), Huang et al. (2009b), Archambault and Purchase (2016b), Huang et al. (2009a), Holten et al. (2011), Huang et al. (2008), Archambault and Purchase (2012a), Burch et al. (2013), Huang et al. (2016a), Xu et al. (2012), Kieffer et al. (2015), Jianu et al. (2014), Alper et al. (2011), Tennekes and de Jonge (2014), Hlawatsch et al. (2014), Dwyer et al. (2013), Guo et al. (2015), Tan et al. (2007), Ziemkiewicz and Kosara (2008), Kadaba et al. (2007), Huang and Huang (2010), Huang et al. (2010), Wu et al. (2017), Yang et al. (2017a), Bach et al. (2017b), Yang et al. (2017b), Kwon et al. (2018), Yoghourdjian et al. (2018), Hu et al. (2018), Verspoor et al. (2018), Koylu and Guo (2017), Ballweg et al. (2017), Didimo et al. (2018) and Bae et al. (2017)	80
Abstract–Elaborate	Dunne and Shneiderman (2013), Alper et al. (2014), Moscovich et al. (2009), Song et al. (2010), Ezaiza et al. (2016), Nekrasovski et al. (2006), Ware et al. (2008), Okoe et al. (2014), Ghani et al. (2011), Shin et al. (2011), Archambault et al. (2010), Kobourov et al. (2014), Zimmer and Kerren (2015), Purchase et al. (2012b), Netzel et al. (2014), Wong et al. (2005), Wong et al. (2003), Borkin et al. (2013), Wong et al. (2006), Yuan et al. (2012), Saket et al. (2014), Cohe et al. (2015), Lee et al. (2006a), Shi et al. (2015), Tu and Shen (2008), Giacomo et al. (2007), Srinivasan and Stasko (2018), Pienta et al. (2018), Liu et al. (2017), Okoe et al. (2017) and Cypko et al. (2017)	32
Connect	Zhao et al. (2015), Moscovich et al. (2009), Lee and Archambault (2016), Zhao et al. (2016), Ezaiza et al. (2016), Ware et al. (2008), Okoe et al. (2014), Federico and Miksch (2016), Ghoniem et al. (2005), Ware and Bobrow (2005), Bae and Watson (2011), Yuan et al. (2012), Lee et al. (2006a), Huang et al. (2016b), Henry et al. (2008), Chang et al. (2017), Cordeil et al. (2017), Srinivasan and Stasko (2018), Walny et al. (2018), Zhao et al. (2018), Pienta et al. (2018), Liu et al. (2017), Hoque and Carenini (2018), Kindermann et al. (2017), Okoe et al. (2017) and Cypko et al. (2017)	29
Create	Novick and Catley (2006), Purchase et al. (2012b), Marriott et al. (2012) and van Ham and Rogowitz (2008)	4
Encode	Ezaiza et al. (2016), Huang et al. (2016b) and Zhao et al. (2018)	4
Explore	Ezaiza et al. (2016), Nekrasovski et al. (2006), Okoe et al. (2014), Saket et al. (2016), Ghani et al. (2011), Saket et al. (2015), Ghani et al. (2012), Shin et al. (2011), Archambault et al. (2010), Kobourov et al. (2014), Federico and Miksch (2016), Archambault et al. (2011), Purchase et al. (2012b), Ruffiang and McGuffin (2013), Wong et al. (2005), Borkin et al. (2013), Wong et al. (2006), Saket et al. (2014), Lee et al. (2006a), Shi et al. (2015), Kwon et al. (2016), Huang et al. (2016b), Srinivasan and Stasko (2018), Pienta et al. (2018) and Okoe et al. (2017)	26
Filter	Ezaiza et al. (2016), Robertson et al. (2002), Bach et al. (2013), Srinivasan and Stasko (2018), Pienta et al. (2018) and Cypko et al. (2017)	9
Reconfigure	Alper et al. (2014), Moscovich et al. (2009), Lee and Archambault (2016), Ezaiza et al. (2016), Nekrasovski et al. (2006), Zimmer and Kerren (2015), Marner et al. (2014), Robertson et al. (2002), Dwyer et al. (2009), Wong et al. (2005), Wong et al. (2006), Bach et al. (2013), Kieffer et al. (2015), van Ham and Rogowitz (2008), Shi et al. (2015), Srinivasan and Stasko (2018), Zhao et al. (2018), Pienta et al. (2018), Okoe et al. (2017) and Cypko et al. (2017)	25
Select	Lee and Archambault (2016), Okoe et al. (2014), Archambault et al. (2010), Barth et al. (2015), Robertson et al. (2002), Ghoniem et al. (2005), Purchase et al. (2012b), Dwyer et al. (2009), Netzel et al. (2014), Wong et al. (2005), Borkin et al. (2013), Burch et al. (2011a), Kieffer et al. (2015), Cohe et al. (2015), Shi et al. (2015), Kwon et al. (2016), Huang et al. (2016b), Cordeil et al. (2017), Srinivasan and Stasko (2018), Zhao et al. (2018), Liu et al. (2017), Kindermann et al. (2017), Okoe et al. (2017) and Cypko et al. (2017)	30

well as graph visualisation metaphors, and has also investigated the impact of the resulting visualisations on readability and task performance. For some years, a main focus has been on the computational complexity and scalability of algorithms, but since the development of methods that scale to several million nodes and

edges, the focus has shifted to the visual complexity and human interpretability of the resulting layouts (Eades et al., 2017).

As layout methods differ in computational scalability, in their performance on certain graph classes, and also in the features and characteristics of the resulting layouts, the interplay between

graph structure and layout method used in studies will strongly impact the limits of cognitive scalability. Comparisons between different methods are rare, as the selection is often motivated by real-world application requirements. While constraint-based methods can create high-quality layouts, which might be of interest for studying the limits of cognitive scalability, the methods do not scale well regarding the computational complexity. Thus, for large graph sizes, researchers have to resort to fast heuristics, e.g. multi-level force-based methods.

### 5.2.1. Graph type and structure

While the number of nodes and the density can give a first indication on the complexity of a graph with respect to cognitive scalability, a more fine-grained description of the structure is necessary to investigate its impact. Certain graph classes might result in layouts with minor quality, e.g. low diameter graphs like friendship networks can lead to hairball drawings when standard force-directed methods are applied. On the other hand, graphs with a globally sparse structure like the ones used by Marner et al. (2014) are well suited for untangling and structure identification tasks, even when the number of nodes is huge, as the global and local structures simply scale with the size.

A further distinction can be made regarding the use of directed and undirected graphs. While 103 of the studies used undirected graphs, only 33 of them used directed graphs, and 15 used trees. With the exception of the two near-tree sparse graphs from the *OnGrax* study (Zimmer and Kerren, 2015), and one graph from a study on directed edge representations (Holten et al., 2011), all graphs with more than 200 nodes are either undirected or trees. Some studies based their graph selection on real world examples. In some cases, the graphs were taken from specific application areas. Most commonly: social networks (Alper et al., 2014; Lee and Archambault, 2016; Ezaiza et al., 2016; Hu and Shi, 2014; Federico and Miksch, 2016; Huang et al., 2005; Blythe et al., 1995; Farrugia and Quigley, 2011; Huang et al., 2009b; Archambault and Purchase, 2016b), followed by co-authorship networks (Zhao et al., 2016; Bach et al., 2013; Alper et al., 2011; Shi et al., 2015; Henry et al., 2008) and biological networks (Nekrasovski et al., 2006; Zimmer and Kerren, 2015; Marner et al., 2014; North et al., 2011; Kieffer et al., 2015). (Shi et al., 2015) use four dynamic graphs from two domains—communication and co-authorship networks.

Other studies used extractions of real graphs, which were closely related to ones in practice. For example, North et al. (2011) used graphs that represented a subset of a biological data set, while others generate graphs with similar size to graphs that commonly appear in usage. Tan et al. (2007) use two trees with 127 nodes with the number of nodes chosen to be typical of tournament brackets considered in fantasy football leagues.

### 5.2.2. Layout method

The distribution of layout methods in the investigated studies shows the expected dominance of force-directed methods. Variants of this class of methods scale well computationally, and appear to be the preferred method used in a variety of publicly available graph visualisations and systems. Together, with the linear time tree layout methods, these methods are the only ones used for studies with more than one thousand nodes. More than 50% of the studies used a force-directed layout to draw the networks, followed by multiple types, indicated in around 18% of the studies. Note that our classification of layout methods specifies the initial layout of subjects presented with, and does not consider whether the subjects could manually, or by means of an algorithm, change the layout in an interactive interface.

While the practical computational scalability of many methods changed due to improved algorithms and implementations, the relative performance stayed the same with force-directed and tree

layouts being by far the fastest, and methods that require solvers, e.g. constraint-based methods, being slowest.

Borkin et al. (2013) present a radial-based tree layout to display file system provenance and motivate their choice, with the failure of node-link diagrams to show a high-level summary for the large-scale provenance data graphs. Their study results indicate that users were more efficient with the interactive radial layout representation than with an existing conventional node-link diagram tool.

The range of the basic graph size metrics for each of the layout methods fits the expectation. Manual layouts are only performed on graphs of 120 nodes or less, with the notable exception of the study on the collaborative graph visualisation system *OnGrax* (Zimmer and Kerren, 2015). In this system, however, an initial layout was given, which was manually created by domain experts, and the user could simply rearrange this layout manually.

### 5.3. General study design factors

Just over half the studies (92/152 studies, 61%) follow a typical design of asking participants to perform graph reading tasks under different conditions (the independent variables – often different layouts, different visualisations or different interaction techniques) and collecting response time and accuracy data as the dependent variables (with some also collecting preference choices). Most studies rely on these three dependent variables to measure the efficiency of the visualisations at hand.

With respect to graph size, there is no discernible difference in the sizes used for typical graph reading studies and the others—both categories have similar distribution of graph sizes.

Some experiments collected process data as the main dependent variable (e.g. Burch et al. (2011a) and Burch et al. (2013)). Others collected eye-tracking data (Yeghoydjian et al., 2018; Chang et al., 2017; Netzel et al., 2014; Ware et al., 2008). Wong et al. (2005) counted the number of mouse actions (click, move, zoom, pan, and turn) as a measure of extent of interaction with different forms of visualisation of labelled graphs, while Nekrasovski et al. (2006) looked particularly at the mouse drag action.

There were 129 studies that used a within-participants study design, 22 used between-participants, and one used a mixture of both. Within-participants studies are popular because they have the advantage of eliminating any effects relating to variability between the participants, and, while they may be subject to the learning effect, the effects of this are easily mitigated by appropriate randomisation. They do, however, tend to take longer than between-participants experiments, where the participants can have more time to work with only a selected few of the stimuli (rather than all of them, as it is in the within-participants case). Thus, the tasks for within-participants studies tend to be smaller and simpler than those used in between-participants studies.

We might expect that there would be more between-participants' experiments in recent years, since such experiments are more suitable for crowd-sourcing: within-participant experiments tend to take too long to be appropriate for crowd-sourced participants. However, this is not the case – there is a similar publication year profile for both categories of study.

## 6. Size rationale

In the previous sections we discussed how different metrics could affect scalability. This was mainly done based on what we could gather from the aim and the results of the studies. In Section 4, we reviewed the range of number of nodes and edges for graphs used in studies, and discovered that most studies using large numbers of nodes and edges want to highlight the advantages of using specific techniques. However, in some cases, the authors

explicitly mention the reasons for picking a specific number of nodes or edges. Some authors performed pilot studies which allowed them to determine the size of graphs that would best suit their evaluation. Others were based on the authors' experiences and understandings of the requirements. We discuss these further in the following sections.

### 6.1. Pilot studies

The papers in our survey rarely mention pilot studies that determine the impact of ceiling or floor on scalability. We believe that many conduct pilot studies do not mention them explicitly. In this section, we discuss selected few studies that provide the reasoning for their pilot studies and the decisions made with regards to factors that affect cognitive scalability.

Archambault et al. (2010) mention that the largest graph size for their study was determined by pilot studies. They start with small, medium and large sized graphs; however the pilot participants could not complete the tasks on the large graphs. Thus, instead of the large graph, they use a smaller graph. The largest graph they use has 3351 nodes and 4083 edges. Archambault and Purchase (2016b) use a pilot study to find a reasonable graph size. They use a maximum of 50 nodes and 100 edges. Similarly, Dawson et al. (2015) mention a pilot study to balance density and difficulty. They use graphs with 75 nodes and 150 edges.

Some design their pilots to find reasonable graph densities to allow the participants to finish the tasks in a restricted time limit. Kobourov et al. (2014) find 120 nodes and 2.5 density as maximum measures to complete the tasks under two minutes, while others use pilot studies in order to explore different layouts and graph structures for specific tasks; e.g. Netzel et al. (2014) aim at finding thresholds for assisting tasks of finding the longest link and biggest cluster.

Saket et al. (2014) mention a pilot study where they chose 50 nodes as minimum and 200 nodes as maximum. They also chose three density levels  $N$ ,  $2N$ , and  $4N$ . Borkin et al. (2013) conducted a pilot study to determine the boundaries between the easy and difficult tasks. They mention that graphs of 10s, 100s, 1000s, and 10,000s of nodes were compared. They classified trees with 42 to 346 nodes as easy, while trees with 1192 to 5480 nodes as hard. Similarly, Marriott et al. (2012) note that their pilot study showed that larger graphs beyond 6 nodes were too difficult to be memorised. Conversely, Xu et al. (2012) rely on a pilot study to choose 100 as the maximum number of nodes for their graphs. They further increase this to 200 in their second study, due to the lack of a ceiling effect in their first study.

### 6.2. What is small? what is large?

Testing the boundaries of readability/scalability may generally not be a primary goal for experimental studies. Instead, researchers might pick a size range that they consider acceptable in order not to have scalability as a confounding factor in their results. In addition, technical limitations, (e.g. screen size and resolution, and also typical requirements from application areas, like characteristics of occurring networks of interest), might play an important role in the determination of graph sizes, but are often not reported explicitly and will also change over the years. Furthermore, there might be standard benchmark sets used, or simply graphs picked based on availability, instead of using graphs that allow one to investigate scalability effects.

Some studies do not mention pilots, but justify their choices of graph size as an attempt to meet a particular requirement for their studies. Archambault et al. (2011) use real graphs, with the largest having 60 nodes and 68 edges. They explain that they chose two data sets each consisting of graphs with realistic size and structure.

Sometimes, the rationale is task-oriented. For example, Kieffer et al. (2015) justify their use of small graphs to allow the participants to manually draw the graphs in a reasonable amount of time. Similarly, Purchase et al. (2012b) use two graphs with 10 nodes and 11 and 18 edges, to make it manageable for the task of drawing the networks. Blythe et al. (1995) note that they use a small graph in order not to overwhelm the participants with the amount of information. They use a small graph with 12 nodes and 24 edges.

Others, such as Hlawatsch et al. (2014), justify using small graphs to avoid the need for interaction. They use ten graphs with eight nodes and 22 to 40 edges. They also use ten graphs with 20 nodes and 147 to 264 edges. Similarly, Alper et al. (2011) chose not to vary the graph size drastically in order to avoid the requirement of zooming.

Zhao et al. (2015) justify their size selection to fit the diagrams to screen. They use graphs with 167 nodes and 902 edges in their evaluation of a visualisation called *MatrixWave*. Kadaba et al. (2007) mention that they needed graphs that were small enough to be memorisable. They use a daisy-structured graph with 11 nodes. In a second study they use smaller graphs with 3 nodes and 2 edges. Holten and van Wijk (2009) explain that they wanted to generate graphs with an adequate number of vertices, without causing a large amount of visual clutter caused by an excessively high edge density.

In contrast, some studies use large graphs in order to highlight the benefits and improvements of some techniques with respect to scalability. Lee et al. (2006a) justify their selection of graphs with 200 nodes as complex graphs. They also state that 200 nodes are considered to be an upper bound for currently studied food webs. Huang et al. (2016a) mention that they choose graphs with a density ranging from 10% to 20%, and their drawings have the same crossing ratio of 40% in order to have reasonably complex graphs. The largest graph they use has 50 nodes and 245 edges. Dwyer et al. (2009) state that the graphs they used were larger in size (50 nodes), than ones (17 nodes) used in another study which inspired their work, while Okoe et al. (2017) justify their selection of a graph with 258 nodes and 1090 edges as larger than previously used graphs, yet sufficiently small to be evaluated in a browser.

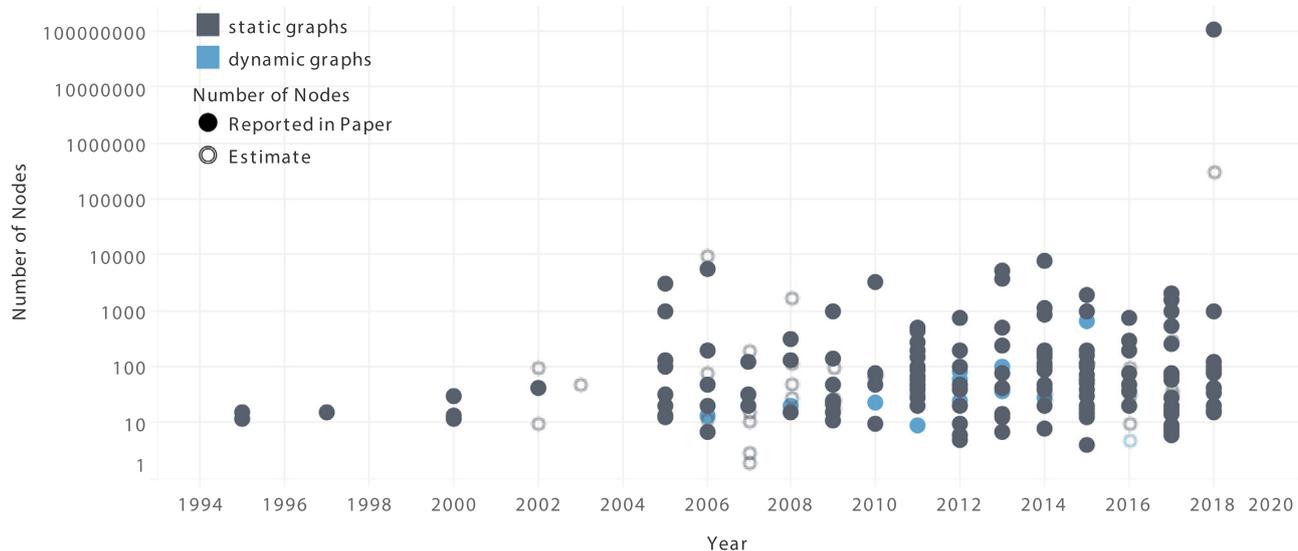
To conclude, some authors provide a rationale for their selection of graphs with a specific range of node and edge counts. These often resemble our discussions and hypotheses, nonetheless, we wanted to keep a clear separation between what constitutes our opinions and the rationale provided by the experimenters.

## 7. Research trends

By analysing our survey data from a historical perspective, we tried to gain some insights about the development of the research community, including trends with respect to graph sizes, participants, tasks, and interaction types used in the studies.

Of the studies surveyed in this paper, 40% were published in TVCG/InfoVis, 16% at Graph Drawing, 13% in the Information Visualisation Journal, and another 11% at CHI. Historically, the first seven studies (Blythe et al., 1995; Purchase et al., 1995; Purchase, 1997b; Bridgeman and Tamassia, 2000; Purchase et al., 2000) were all published at Graph Drawing between 1995 and 2000. The first study (Robertson et al., 2002) in the Information Visualisation Journal appeared in 2002. The first one (Wong et al., 2003) in TVCG/InfoVis was in 2003, whereas the first one (Nekrasovski et al., 2006) at CHI in 2006, and the first one (Ziemkiewicz and Kosara, 2009) at EuroVis in 2009.

The seminal papers (Blythe et al., 1995; Purchase et al., 1995) of 1995 focused on static graphs and were followed by another ten papers (with a total of 16 studies) on static graphs. The first study on dynamic graphs (Purchase et al., 2006) was published in 2006, more than ten years after the first one on static graphs. As we cover



**Fig. 12.** The number of nodes of static versus dynamic graphs used in studies for the period from 1995 to 2016.

a period of 24 years from 1995 to 2018 in this paper, it is reasonable to compare the first and second half of the studied period, with the first half ending in 2006, and the second half starting in 2007. Whilst during the first half, the average publication frequency was around two studies per year, it considerably increased to 11 studies per year for the second half. Of these studies, on average, 9 focused on static and 2 on dynamic graphs.

Fig. 12 shows the number of nodes of static and dynamic graphs used in studies in different years. While for static graphs the number of nodes increased from below 20 in 1995 to several thousand ten years later, the graph size of dynamic graphs stayed 100 or below (with one exception (Shi et al., 2015) in 2015).

Since we observed a considerable increase in the graph size for static graphs, we expected that researchers would also try to recruit more participants for their studies. It turns out that there was actually a big difference, if we compare the first and the second half of our studied period. In the first half the median number of participants, the number of studies that used static graphs, was 14. In the second half it increased to 21. Interestingly, a closer look at the first half of the studied period reveals a dramatic drop of the median number of participants. It decreased from 75 for the original 7 studies (Blythe et al., 1995; Purchase et al., 1995; Purchase, 1997b; Bridgeman and Tamassia, 2000; Purchase et al., 2000) (all published at GD) of the period from 1995 to 2000 to 9 for the subsequent 11 studies (Robertson et al., 2002; Ware et al., 2002; Wong et al., 2003; Huang et al., 2005; Ghoniem et al., 2005; Ware and Bobrow, 2005; Wong et al., 2005) from 2001 to 2005 (mostly published in the Information Visualisation Journal).

## 8. Summaries of findings and discussion

### 8.1. Author personal experiences

Several of the authors of this paper have conducted experimental studies included in this survey; we therefore have our own insights into the problem of choosing an appropriately-sized graph for an empirical study. In many cases (e.g. (Purchase and Samra, 2008; Purchase et al., 2000, 1995; Archambault et al., 2010)), the original graph chosen was found to be too large during piloting, and scaled down. This was not because it was thought that the participants would not be able to complete the task correctly, but simply because it was necessary to keep the tasks short in duration so that the experiment did not take too long, especially when using a within-participants' study design. Even in cases where

real-world data was used (e.g. Archambault and Purchase (2016b, 2012a)), these graphs were filtered to make them of an appropriate manageable size for within-participants experiments. On the one occasion when a complex UML diagram was used in a between-participant study, the variability between subjects was so great that no sensible results were achieved: one of the many times when data was simply discarded.

Another common way we selected our graph sizes (and, in some cases, graph structure) was with reference to published work, particularly if the study was deliberately building on prior experiments — even our own. The size and structure of the graphs used in Dwyer et al. (2009) and Purchase et al. (2012b) were based on those used by van Ham and Rogowitz (2008); the number of slices in the dynamic graph used in Archambault and Purchase (2012b) was based on the number used in our own prior work Archambault et al. (2011). Evidence that the choice of size of graph has proven successful in prior work is a clear pointer as to an appropriate size to choose for experiments that are similar in scope.

A further way to define graph characteristics for experimental studies is based on the requirements from application areas. In Marner et al. (2014), the graphs were provided by biologists that used the graphs in a previous publication to analyse sequencing data by applying network analysis and visualisation. The graph set was chosen to exploit and test the affordances of a wall-sized display.

### 8.2. Recommendations for experimenters

Here we present our recommendations, based on our findings, for future design and reporting of studies.

We recommend that authors of study papers report precisely the number of nodes and edges used in studies, as well as the number of timeslices for dynamic graphs. In the case of interactivity, such as semantic zooming or neighbourhood browsing, the number of nodes visible on screen should also be clearly recorded to differentiate between scalability due to interaction versus cognitive and perceptual scalability. Similarly, studies that use aggregation to collapse parts of the network into less numerous glyphs, should report on the number of glyphs visible.

Authors of papers describing studies on network analysis tasks should explicitly state how they controlled the size and density of the network data tested. This needs not entail extra work. Simple recording of observations during piloting could be greatly informative to the kind of meta-analysis we have performed in this

survey. Alternatively, if the choice of size is arbitrary or based on previous studies, explicitly stating the provenance of the choice can help readers to better reason about size effects.

We would call for studies that explicitly test scalability of network visualisation by testing more than a couple of different graph sizes to search for ceiling and floor effects and controlling for other variables. Moreover, we call for the reporting of whatever ceiling and floor effects that are found.

## 9. Conclusions

### 9.1. Key findings on graph properties

- There are some clear 'default' numbers of nodes and edges used by most studies. At a maximum, most studies use graphs with 100 nodes, followed by 50 nodes and 100 edges, while their smallest graphs have 20, followed by 50 and 10 nodes, and 10 edges. The round numbers used by most studies suggest the choice of size to test is somewhat arbitrary, as opposed to being based on empirical evidence of cognitive limitations.
- 80% of studies (121 / 152 studies) use graphs with 100 nodes or less, while only 37% (56 / 152 studies) use graphs with more than 100 nodes.
- 74% of studies (113 / 152 studies) use graphs with 200 edges or less, while only 34% (52 / 152 studies) use graphs with more than 200 edges.
- 70% of studies (23 / 33 studies) that use graphs with more than 200 nodes use interaction and aggregation techniques to show only parts of the network to the participants at a given time.
- Most of the studies that use graphs with more than 1000 edges evaluate tools that are intended to be able to cope with a substantially large number of edges, or are oriented towards performing well for networks of specific structure, e.g. densely connected communities.
- Only 12% (18 / 152 studies) of the studies surveyed use graphs with a density of more than 20%.
- Studies that use graphs with more than 20% density either use small graphs (< 10 nodes), evaluate matrix representations, or evaluate edge bundling and compression techniques.
- 32% of studies (7 / 22 studies) that use dynamic graphs use six timeslices. This is often to best fit small multiples representations to screen.

### 9.2. Key Findings on Task Types

- The most common type of tasks is *Topology-based* (114 / 152 studies, 75%), however for graphs with 1000 nodes or more, *Overview* tasks prevail (13 / 18 studies, 72%).
- *Attribute-based* tasks are commonly used for graphs with 200 nodes or less (47 / 53 studies, 89%) and less than 10% density (42 / 53 studies, 79%).
- *Browsing* tasks are commonly used for graphs with 200 nodes or less (24 / 26 studies, 92%) and less than 10% density (25 / 26 studies, 96%).

### 9.3. Key Findings on Interaction Types

- Studies with larger graphs (> 500 nodes) tend to use multiple types of interaction, while using a single type of interaction is more common in studies with smaller graphs ( $\leq$  500 nodes).
- The majority of studies with no interaction used graphs with 100 nodes or less (70 / 80 studies, 88%) and less than 10% density (57 / 80 studies, 71%).

**Table 5**

Four categories of graph size based on number of nodes.

	# of nodes	# of studies	# of studies with no interaction	# of studies with overview tasks
Small	$\leq 20$	62 (41%)	39 (63%)	19 (31%)
Medium	[21, 50]	50 (33%)	31 (62%)	12 (24%)
Large	[51, 200]	56 (37%)	27 (48%)	16 (29%)
v. large	>200	33 (22%)	10 (30%)	19 (58%)

**Table 6**

Four categories of graphs based on linear density.

	Linear density	# of studies	# of studies with no interaction	# of studies with overview tasks
Tree-like & disconnected	[0, 1.0]	49 (32%)	20 (41%)	15 (31%)
Sparse	[1.01, 2.0]	67 (44%)	40 (60%)	20 (30%)
Dense	[2.01, 4.0]	31 (20%)	18 (58%)	7 (23%)
v. dense	>4.0	26 (17%)	12 (46%)	13 (50%)

- The interaction types of *Abstract/Elaborate*, *Connect*, *Explore*, *Reconfigure*, and *Select* were used on larger graphs, while *Create*, *Encode*, and *Filter* were used on smaller graphs ( $\leq$  100 nodes and < 10% density).

### 9.4. Discussion

While there has been a significant focus on computational scalability of node–link diagrams layout and rendering, in a race to visualise the largest networks, it seems researchers have understudied human cognitive limitations in understanding such diagrams. A better knowledge of cognitive scalability in this regard would have several benefits. In tools that allow the user to interactively explore a large network through neighbourhood or aggregated views, tool developers could more intelligently control the number of elements in these views. Furthermore, if our community could give clear and informed guidance to users of graph visualisation it would help them to select the right tool for their purpose. For example, in creating figures for papers, biologists reporting on the interactions of particular proteins may be better off showing a focused neighbourhood around those specific proteins rather than providing a hairball. Similarly, if the users were experimenters, they would choose their corpora, design their tasks and pick layout methods based on these well-defined limits.

Thus, the aim of this survey was to explore factors that would affect cognitive scalability via reviewing existing empirical studies that have used node–link diagrams. We have noticed that controlled experiments tend to focus on graph data sets of size within a fairly limited window (tens to a few hundred nodes and low density). We also discovered that even though the most common type of tasks performed on a network, in general, is related to the topology of the network, overview tasks become more popular for larger networks. Similarly tasks related to detailed attributes associated to the nodes or edges and browsing are not common on networks with more than a couple hundred nodes. With regards to interaction, studies with large graphs tend to allow for more than one type of interaction. Furthermore, we discovered that some interaction types, such as *Create*, *Encode* and *Filter* are only used on small graphs, while others, e.g., *Abstract/Elaborate*, *Connect*, *Explore*, *Reconfigure*, and *Select* are also used on large graphs.

This survey has also helped identify some weaknesses in the design and reporting of empirical studies that use node–link diagrams. For example, several studies do not report on the sizes of the graphs used, while others do not report on the number of elements visible to the participants at a given time. We provide a

list of recommendations to overcome these weaknesses in future studies.

A motivation for this work was to identify or validate terms that are used to categorise ranges for graph size. Table 5 presents four categories with respect to number of nodes. According to the surveyed studies, there are clear cuts at 20, 50, and 200 nodes. We categorise these into ranges that represent small, medium, large and very large graphs. Following significant thresholds of linear density, which have been identified by Melancon (2006), we categorise sparse, dense, and very dense graphs in Table 6. Hopefully this breakdown gives future researchers a clear motivation for selecting different graph sizes for their studies. For example, there are only seven studies that use dense graphs with overview tasks.

Our findings, indicate a threshold at 200 nodes and 10% density. This threshold is respected by empirical studies that include tasks requiring a detailed analysis of the network. Nonetheless, we believe that this threshold is a result of the expert intuition of the researchers, rather than empirical research. A controlled study is needed to validate and refine this threshold.

## Acknowledgments

This survey began as part of a working group output of the NII Shonan Seminar No. 2015-1 *Big Graph Drawing: Metrics and Methods*, and we would like to thank this seminar series for the role it played in this survey. We would like to thank Tamara Munzner for her ideas and feedback at this seminar which helped focus the topic of this paper. The second author would like to thank EPSRC First Grant EP/N005724/1. The last author would like to thank the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 747985. This work was supported by the Australian Research Council Discovery Project grant DP140100077.

## References

- Abello, J., van Ham, F., Krishnan, N., 2006. ASK-GraphView: A large scale graph visualization system. *IEEE Trans. Vis. Comput. Graphics* 12 (5), 669–676.
- Alper, B., Bach, B., Henry Riche, N., Isenberg, T., Fekete, J.-D., 2013. Weighted graph comparison techniques for brain connectivity analysis. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*. ACM, New York, NY, USA, pp. 483–492.
- Alper, B.E., Henry Riche, N., Hollerer, T., 2014. Structuring the space: A study on enriching node-link diagrams with visual references. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*. ACM, New York, NY, USA, pp. 1825–1834.
- Alper, B., Höllerer, T., Kuchera-Morin, J., Forbes, A., 2011. Stereoscopic highlighting: 2d graph visualization on stereo displays. *IEEE Trans. Vis. Comput. Graphics* 17 (12), 2325–2333.
- Archambault, D., Munzner, T., Auber, D., 2008. Grouseflocks: Steerable exploration of graph hierarchy space. *IEEE Trans. Vis. Comput. Graphics* 14 (4), 900–913.
- Archambault, D., Purchase, H.C., 2012a. The mental map and memorability in dynamic graphs. In *Visualization Symposium (PacificVis)*, 2012 IEEE Pacific, pp. 89–96.
- Archambault, D., Purchase, H.C., 2012b. Mental map preservation helps user orientation in dynamic graphs. In *Graph Drawing - 20th International Symposium, GD 2012*, Redmond, WA, USA, September 2012, 19–21, Revised Selected Papers, pp. 475–486.
- Archambault, D., Purchase, H.C., 2013. The map in the mental map: experimental results in dynamic graph drawing. *Int. J. Hum-Comput. Stud.* 71 (11), 1044–1055.
- Archambault, D., Purchase, H.C., 2016a. Can animation support the visualisation of dynamic graphs? *Inform. Sci.* 330, 495–509, SI: Visual Info Communication.
- Archambault, D., Purchase, H.C., 2016b. On the effective visualisation of dynamic attribute cascades. *Inf. Vis.* 15 (1), 51–63.
- Archambault, D., Purchase, H.C., Pinaud, B., 2010. Difference map readability for dynamic graphs. In *Graph Drawing - 18th International Symposium, GD 2010*, Konstanz, Germany, September (2010) 21–24. Revised Selected Papers, pp. 50–61.
- Archambault, D., Purchase, H.C., Pinaud, B., 2010. The readability of path-preserving clusterings of graphs. *Comput. Graph. Forum* 29 (3), 1173–1182.
- Archambault, D., Purchase, H.C., Pinaud, B., 2011. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Trans. Vis. Comput. Graphics* 17 (4), 539–552.
- Bach, B., Dragicevic, P., Archambault, D., Hurter, C., Carpendale, S., 2017a. A descriptive framework for temporal data visualizations based on generalized space-time cubes. *Comput. Graph. Forum* 36 (6), 36–61.
- Bach, B., Pietriga, E., Fekete, J.-D., 2013. Graphdiaries: animated transitions and temporal navigation for dynamic networks. *IEEE Trans. Vis. Comput. Graphics* 20 (5), 740–754.
- Bach, B., Riche, N.H., Hurter, C., Marriott, K., Dwyer, T., 2017b. Towards unambiguous edge bundling: investigating confluent drawings for network visualization. *IEEE Trans. Vis. Comput. Graphics* 23 (1), 541–550.
- Bae, J., Helldin, T., Riveiro, M., 2017. Understanding indirect causal relationships in node-link graphs. *Comput. Graph. Forum* 36, 411–421, Wiley Online Library.
- Bae, J., Watson, B., 2011. Developing and evaluating quilts for the depiction of large layered graphs. *IEEE Trans. Vis. Comput. Graphics* 17 (12), 2268–2275.
- Ballweg, K., Pohl, M., Wallner, G., von Landesberger, T., 2017. Visual similarity perception of directed acyclic graphs: a study on influencing factors. *arXiv preprint arXiv:1709.01007*.
- Barth, L., Gemsa, A., Niedermann, B., Nöllenburg, M., 2015. On the readability of boundary labeling. In: *Graph Drawing and Network Visualization - 23rd International Symposium, GD 2015*, Los Angeles, CA, USA, September (2015) 24–26, Revised Selected Papers, pp. 515–527.
- Beck, F., Burch, M., Diehl, S., Weiskopf, D., 2017. A taxonomy and survey of dynamic graph visualization. *Comput. Graph. Forum* 36 (1), 133–159.
- Behrisch, M., Bach, B., Henry Riche, N., Schreck, T., Fekete, J.-D., 2016. Matrix reordering methods for table and network visualization. *Comput. Graph. Forum* 35 (3), 693–716.
- Blythe, J., McGrath, C., Krackhardt, D., 1995. The effect of graph layout on inference from social network data. In: *Graph Drawing, Symposium on Graph Drawing, GD '95*, Passau, Germany, September 1995, 20–22, Proceedings, pp. 40–51.
- Borkin, M.A., Yeh, C.S., Boyd, M., Macko, P., Gajos, K.Z., Seltzer, M., Pfister, H., 2013. Evaluation of filesystem provenance visualization tools. *IEEE Trans. Vis. Comput. Graphics* 19 (12), 2476–2485.
- Bridgeman, S.S., Tamassia, R., 2000. A user study in similarity measures for graph drawing. In: *Graph Drawing, 8th International Symposium, GD 2000*, Colonial Williamsburg, VA, USA, September 2000, 20–23, Proceedings, 2000, pp. 19–30.
- Burch, M., Andrienko, G., Andrienko, N., Höferlin, M., Raschke, M., Weiskopf, D., 2013. Visual task solution strategies in tree diagrams. In: *Proceedings of the 6th IEEE Pacific Visualization Symposium (PacificVis 2013)*, pp. 169–176.
- Burch, M., Heinrich, J., Konevtsova, N., Höferlin, M., Weiskopf, D., 2011a. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Trans. Vis. Comput. Graphics* 17 (12), 2440–2448.
- Burch, M., Vehlow, C., Konevtsova, N., Weiskopf, D., 2011b. Evaluating partially drawn links for directed graph edges. In: *Graph Drawing - 19th International Symposium, GD 2011*, Eindhoven, The Netherlands, September (2011) 21–23, Revised Selected Papers, 2011, pp. 226–237.
- Chang, C., Bach, B., Dwyer, T., Marriott, K., 2017. Evaluating perceptually complementary views for network exploration tasks. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, pp. 1397–1407.
- Chen, S., Lin, L., Yuan, X., 2017. Social media visual analytics. *Comput. Graph. Forum* 36 (3), 563–587.
- Chivers, D., Rodgers, P., 2014. Octilinear force-directed layout with mental map preservation for schematic diagrams. In: *Diagrammatic Representation and Inference - 8th International Conference, Diagrams 2014*, Melbourne, VIC, Australia, July 28 - August 1, 2014. Proceedings, pp. 1–8.
- Cohe, A., Liutkus, B., Bailly, G., Eagan, J., Lecolinet, E., 2015. Schemelens: a content-aware vector-based fisheye technique for navigating large systems diagrams. *IEEE Trans. Vis. Comput. Graphics* 22 (1), 330–338.
- Cordeil, M., Dwyer, T., Klein, K., Laha, B., Marriott, K., Thomas, B.H., 2017. Immersive collaborative analysis of network connectivity: cave-style or head-mounted display? *IEEE Trans. Vis. Comput. Graphics* 23 (1), 441–450.
- Cypko, M., Wojdziak, J., Stoehr, M., Kirchner, B., Preim, B., Dietz, A., Lemke, H., Oeltze-Jafra, S., 2017. Visual verification of cancer staging for therapy decision support. *Comput. Graph. Forum* 36, 109–120, Wiley Online Library.
- Dawson, J.Q., Munzner, T., McGrenere, J., 2015. A search-set model of path tracing in graphs. *Inf. Vis.* 14 (4), 308–338.
- Didimo, W., Kornaropoulos, E.M., Montecchiani, F., Tollis, I.G., 2018. A visualization framework and user studies for overloaded orthogonal drawings. *Comput. Graph. Forum* 37, 288–300, Wiley Online Library.
- Dunne, C., Shneiderman, B., 2013. 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 '13*. ACM, New York, NY, USA, pp. 3247–3256.
- Dwyer, T., Lee, B., Fisher, D., Quinn, K.I., Isenberg, P., Robertson, G., North, C., 2009. A comparison of user-generated and automatic graph layouts. *IEEE Trans. Vis. Comput. Graphics* 15 (6), 961–968.
- Dwyer, T., Marriott, K., Schreiber, F., Stuckey, P.J., Woodward, M., Wybrow, M., 2008. Exploration of networks using overview+detail with constraint-based cooperative layout. *IEEE Trans. Vis. Comput. Graphics* 14 (6), 1293–1300.

- Dwyer, T., Riche, N.H., Marriott, K., Mears, C., 2013. Edge compression techniques for visualization of dense directed graphs. *IEEE Trans. Vis. Comput. Graphics* 19 (12), 2596–2605.
- Eades, P., Hong, S., Nguyen, A., Klein, K., 2017. Shape-based quality metrics for large graph visualization. *J. Graph Algorithms Appl.* 21 (1), 29–53.
- Eick, S.G., Karr, A.F., 2002. Visual scalability. *J. Comput. Graph. Statist.* 11 (1), 22–43.
- Elmqvist, N., Fekete, J.-D., 2010. Hierarchical aggregation for information visualization: overview, techniques, and design guidelines. *IEEE Trans. Vis. Comput. Graphics* 16 (3), 439–454.
- Ezaiza, H., Humayoun, S.R., AlTarawneh, R., Ebert, A., 2016. Person-vis: Visualizing personal social networks (ego networks). In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 1222–1228.
- Farrugia, M., Quigley, A., 2011. Effective temporal graph layout: a comparative study of animation versus static display methods. *Inf. Vis.* 10 (1), 47–64.
- Federico, P., Miksch, S., 2016. Evaluation of two interaction techniques for visualization of dynamic graphs. In *Graph Drawing and Network Visualization - 24th International Symposium, GD 2016, Athens, Greece, September (2016) 19-21, Revised Selected Papers*, pp. 557–571.
- Ghani, S., Elmqvist, N., Yi, J.S., 2012. Perception of animated node-link diagrams for dynamic graphs. *Comput. Graph. Forum* 31 (3), 1205–1214.
- Ghani, S., Riche, N.H., Elmqvist, N., 2011. Dynamic insets for context-aware graph navigation. *Comput. Graph. Forum* 30 (3), 861–870.
- Ghoniem, M., Fekete, J.-D., Castagliola, P., 2004. A comparison of the readability of graphs using node-link and matrix-based representations. In: *Information Visualization, 2004 INFOVIS 2004 IEEE Symposium on*. Ieee, pp. 17–24.
- Ghoniem, M., Fekete, J.-D., Castagliola, P., 2005. On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. *Inf. Vis.* 4 (2), 114–135.
- Giacomo, E.D., Didimo, W., Grilli, L., Liotta, G., 2007. Graph visualization techniques for web clustering engines. *IEEE Trans. Vis. Comput. Graphics* 13 (2), 294–304.
- Giacomo, E.D., Didimo, W., Liotta, G., Montecchiani, F., Tollis, I.G., 2013. Exploring complex drawings via edge stratification. In: *Graph Drawing - 21st International Symposium, GD 2013, Bordeaux, France, September (2013) 23-25, Revised Selected Papers*, pp. 304–315.
- Greffard, N., Picarougne, F., Kuntz, P., 2011. Visual community detection: An evaluation of 2d, 3d perspective and 3d stereoscopic displays. In: *Graph Drawing - 19th International Symposium, GD 2011, Eindhoven, The Netherlands, September (2011) 21-23, Revised Selected Papers*, pp. 215–225.
- Guo, H., Huang, J., Laidlaw, D.H., 2015. Representing uncertainty in graph edges: an evaluation of paired visual variables. *IEEE Trans. Vis. Comput. Graphics* 21 (10), 1173–1186.
- Hadlak, S., Schumann, H., Schulz, H.-J., 2015. A survey of multi-faceted graph visualization. In *Eurographics Conference on Visualization (EuroVis) - STARS*.
- Henry, N., Bezerianos, A., Fekete, J., 2008. Improving the readability of clustered social networks using node duplication. *IEEE Trans. Vis. Comput. Graphics* 14 (6), 1317–1324.
- Herman, I., Melançon, G., Marshall, M.S., 2000. Graph visualization and navigation in information visualization: A survey. *IEEE Trans. Vis. Comput. Graphics* 6 (1), 24–43.
- Hlawatsch, M., Burch, M., Weiskopf, D., 2014. Visual adjacency lists for dynamic graphs. *IEEE Trans. Vis. Comput. Graphics* 20 (11), 1590–1603.
- Holten, D., Isenberg, P., van Wijk, J.J., Fekete, J.-D., 2011. An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *Visualization Symposium (PacificVis), 2011 IEEE Pacific*, pp. 195–202.
- Holten, D., van Wijk, J.J., 2009. A user study on visualizing directed edges in graphs. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 2299–2308.
- Hoque, E., Carenini, G., 2018. Interactive topic hierarchy revision for exploring a collection of online conversations. *Inf. Vis.* 1473871618757228.
- Hu, Y., Shi, L., 2014. A coloring algorithm for disambiguating graph and map drawings. In *Graph Drawing - 22nd International Symposium, GD 2014, Würzburg, Germany, September (2014) 24-26, Revised Selected Papers*, pp. 89–100.
- Hu, Y., Shi, L., Liu, Q., 2018. A coloring algorithm for disambiguating graph and map drawings.
- Huang, W., 2007. Using eye tracking to investigate graph layout effects. In: *Visualization, 2007 APVIS'07 2007 6th International Asia-Pacific Symposium on*. IEEE, pp. 97–100.
- Huang, W., Eades, P., Hong, S.-H., 2009a. A graph reading behavior: Geodesic-path tendency. In *Proceedings of the 2th IEEE Pacific Visualization Symposium (PacificVis 2009)*, pp. 137–144.
- Huang, W., Eades, P., Hong, S.-H., 2009b. Measuring effectiveness of graph visualizations: a cognitive load perspective. *Inf. Vis.* 8 (3), 139–152.
- Huang, W., Eades, P., Hong, S.-H., Duh, H.B.-L., 2016a. Effects of curves on graph perception. In: *Pacific Visualization Symposium (PacificVis), 2016 IEEE*. IEEE, pp. 199–203.
- Huang, W., Eades, P., Hong, S.-H., Lin, C.-C., 2010. Improving force-directed graph drawings by making compromises between aesthetics. In: *Visual Languages and Human-Centric Computing (VL/HCC), 2010 IEEE Symposium on*. IEEE, pp. 176–183.
- Huang, W., Hong, S.-H., Eades, P., 2005. Layout effects on sociogram perception. In: *Graph Drawing, 13th International Symposium, GD 2005, Limerick, Ireland, September 2005, 12–14, Revised Papers*, pp. 262–273.
- Huang, W., Hong, S.-H., Eades, P., 2008. Effects of crossing angles. In: *Visualization Symposium, 2008 PacificVIS '08 IEEE Pacific*, pp. 41–46.
- Huang, W., Huang, M.L., 2010. Exploring the relative importance of crossing number and crossing angle. In *2010 International Symposium on Visual Information Communication, VINCI '10, Beijing, China - September (2010) 28-29*, p. 10.
- Huang, X., Zhao, Y., Ma, C., Yang, J., Ye, X., Zhang, C., 2016b. Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE Trans. Vis. Comput. Graphics* 22 (1), 160–169.
- Jankun-Kelly, T., Dwyer, T., Holten, D., Hurter, C., Nöllenburg, M., Weaver, C., Xu, K., 2014. Scalability considerations for multivariate graph visualization. In: *Kerren, A., Purchase, H.C., Ward, M.O. (Eds.), Multivariate Network Visualization LNCS 8380*. Springer, pp. 207–235, Chapter 10.
- Jianu, R., Rusu, A., Hu, Y., Taggart, D., 2014. How to display group information on node-link diagrams: an evaluation. *IEEE Trans. Vis. Comput. Graphics* 20 (11), 1530–1541.
- Kadaba, N.R., Irani, P.P., Leboe, J., 2007. Visualizing causal semantics using animations. *IEEE Trans. Vis. Comput. Graphics* 13 (6), 1254–1261.
- Keller, R., Eckert, C.M., Clarkson, P.J., 2006. Matrices or node-link diagrams: which visual representation is better for visualising connectivity models? *Inf. Vis.* 5 (1), 62–76.
- Kieffer, S., Dwyer, T., Marriott, K., Wybrow, M., 2015. Hola: human-like orthogonal network layout. *IEEE Trans. Vis. Comput. Graphics* 22 (1), 349–358.
- Kindermann, P., Meulemans, W., Schulz, A., 2017. Experimental analysis of the accessibility of drawings with few segments. *arXiv preprint arXiv:1708.09815*.
- Kobourov, S.G., Pupyrev, S., Saket, B., 2014. Are crossings important for drawing large graphs? In *Graph Drawing - 22nd International Symposium, GD 2014, Würzburg, Germany, September (2014) 24-26, Revised Selected Papers*, pp. 234–245.
- Koylu, C., Guo, D., 2017. Design and evaluation of line symbolizations for origin-destination flow maps. *Inf. Vis.* 16 (4), 309–331.
- Kwon, O.-H., Crnovrsanin, T., Ma, K.-L., 2018. What would a graph look like in this layout? a machine learning approach to large graph visualization. *IEEE Trans. Vis. Comput. Graphics* 24 (1), 478–488.
- Kwon, O., Muelder, C., Lee, K., Ma, K., 2016. A study of layout, rendering, and interaction methods for immersive graph visualization. *IEEE Trans. Vis. Comput. Graphics* 22 (7), 1802–1815.
- Lee, A., Archambault, D., 2016. Communities found by users – not algorithms: Comparing human and algorithmically generated communities. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. ACM, New York, NY, USA, pp. 2396–2400.
- Lee, B., Parr, C.S., Plaisant, C., Bederson, B.B., Veksler, V.D., Gray, W.D., Kotfila, C., 2006a. Treeplus: Interactive exploration of networks with enhanced tree layouts. *IEEE Trans. Vis. Comput. Graphics* 12 (6), 1414–1426.
- Lee, B., Plaisant, C., Parr, C.S., Fekete, J.-D., Henry, N., 2006b. Task taxonomy for graph visualization. In: *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*. ACM, pp. 1–5.
- Lhuillier, A., Hurter, C., Telea, A., 2017. State of the art in edge and trail bundling techniques. *Comput. Graph. Forum* 36 (3), 619–645.
- Liu, D., Guo, F., Deng, B., Qu, H., Wu, Y., 2017. egocomp: A node-link-based technique for visual comparison of ego-networks. *Inf. Vis.* 16 (3), 179–189.
- Marner, M.R., Smith, R.T., Thomas, B.H., Klein, K., Eades, P., Hong, S., 2014. GION: interactively untangling large graphs on wall-sized displays. In *Graph Drawing - 22nd International Symposium, GD 2014, Würzburg, Germany, September (2014) 24-26, Revised Selected Papers*, pp. 113–124.
- Marriott, K., Purchase, H., Wybrow, M., Goncu, C., 2012. Memorability of visual features in network diagrams. *IEEE Trans. Vis. Comput. Graphics* 18 (12), 2477–2485.
- Melançon, G., 2006. Just how dense are dense graphs in the real world?: a methodological note. In: *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*. ACM, pp. 1–7.
- Meulemans, W., Schulz, A., 2015. A tale of two communities: Assessing homophily in node-link diagrams. In: *Graph Drawing and Network Visualization - 23rd International Symposium, GD 2015, Los Angeles, CA, USA, September (2015) 24-26, Revised Selected Papers*, pp. 489–501.
- Miller, G.A., 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63 (2), 81.
- Moscovich, T., Chevalier, F., Henry, N., Pietriga, E., Fekete, J.-D., 2009. Topology-aware navigation in large networks. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09*. ACM, New York, NY, USA, pp. 2319–2328.
- Nekrasovskii, D., Bodnar, A., McGrenere, J., Guimbretière, F., Munzner, T., 2006. An evaluation of pan & zoom and rubber sheet navigation with and without an overview. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 11–20.
- Netzel, R., Burch, M., Weiskopf, D., 2014. Comparative eye tracking study on node-link visualizations of trajectories. *IEEE Trans. Vis. Comput. Graphics* 20 (12), 2221–2230.

- North, C., Saraiya, P., Duca, K., 2011. A comparison of benchmark task and insight evaluation methods for information visualization. *Inf. Vis.* 10 (3), 162–181.
- Novick, L.R., Catley, K.M., 2006. Interpreting hierarchical structure: Evidence from cladograms in biology. In: *Diagrammatic Representation and Inference, 4th International Conference, Diagrams 2006, Stanford, CA, USA, June (2006) 28–30, Proceedings*, pp. 176–180.
- Okoe, M., Alam, S.S., Jianu, R., 2014. A gaze-enabled graph visualization to improve graph reading tasks. *Comput. Graph. Forum* 33 (3), 251–260.
- Okoe, M., Jianu, R., 2015. Graphunit: evaluating interactive graph visualizations using crowdsourcing. *Comput. Graph. Forum* 34 (3), 451–460.
- Okoe, M., Jianu, R., Kobourov, S., 2017. Revisited network representations. In: *25th Symposium on Graph Drawing (GD)*.
- Pienta, R., Hohman, F., Endert, A., Tamersoy, A., Roundy, K., Gates, C., Navathe, S., Chau, D.H., 2018. Vigor: interactive visual exploration of graph query results. *IEEE Trans. Vis. Comput. Graph.* 24 (1), 215–225.
- Purchase, H., 1997a. Which aesthetic has the greatest effect on human understanding?. In: *International Symposium on Graph Drawing, Springer*, pp. 248–261.
- Purchase, H.C., 1997b. Which aesthetic has the greatest effect on human understanding?. In: *Graph Drawing, 5th International Symposium, GD '97, Rome, Italy, September (1997) 18–20, Proceedings*, pp. 248–261.
- Purchase, H.C., Aldler, J., Carrington, D.A., 2000. User preference of graph layout aesthetics: A UML study. In: *Graph Drawing, 8th International Symposium, GD 2000, Colonial Williamsburg, VA, USA, September (2000) 20–23, Proceedings*, pp. 5–18.
- Purchase, H.C., Cohen, R.F., James, M.I., 1995. Validating graph drawing aesthetics. In: *Graph Drawing, Symposium on Graph Drawing, GD '95, Passau, Germany, September (1995) 20–22, Proceedings*, pp. 435–446.
- Purchase, H.C., Hamer, J., Nöllenburg, M., Kobourov, S.G., 2012a. On the usability of lombardi graph drawings. In: *Graph Drawing - 20th International Symposium, GD 2012, Redmond, WA, USA, September (2012) 19–21, Revised Selected Papers*, pp. 451–462.
- Purchase, H.C., Hoggan, E.E., Görg, C., 2006. How important is the mental map? - an empirical investigation of a dynamic graph layout algorithm. In: *Graph Drawing, 14th International Symposium, GD 2006, Karlsruhe, Germany, September (2006) 18–20 Revised Papers*, pp. 184–195.
- Purchase, H.C., Pilcher, C., Plimmer, B., 2012b. Graph drawing aesthetics - created by users, not algorithms. *IEEE Trans. Vis. Comput. Graphics* 18 (1), 81–92.
- Purchase, H.C., Samra, A., 2008. Extremes are better: Investigating mental map preservation in dynamic graphs. In: *Diagrammatic Representation and Inference, 5th International Conference, Diagrams 2008, Herrsching, Germany, September (2008) 19–21 Proceedings*, pp. 60–73.
- Robertson, G., Cameron, K., Czerwinski, M., Robbins, D., 2002. Animated visualization of multiple intersecting hierarchies. *Inf. Vis.* 1 (1), 50–65.
- Rufiange, S., McGuffin, M.J., 2013. Diffani: visualizing dynamic graphs with a hybrid of difference maps and animation. *IEEE Trans. Vis. Comput. Graphics* 9 (12), 2556–2565.
- Saket, B., Scheidegger, C., Kobourov, S.G., 2016. Comparing node-link and node-link-group visualizations from an enjoyment perspective. *Comput. Graph. Forum* 35 (3), 41–50.
- Saket, B., Scheidegger, C., Kobourov, S.G., Börner, K., 2015. Map-based visualizations increase recall accuracy of data. *Comput. Graph. Forum* 34 (3), 441–450.
- Saket, B., Simonetto, P., Kobourov, S., Börner, K., 2014. Node, node-link, and node-link-group diagrams: an evaluation. *IEEE Trans. Vis. Comput. Graphics* 20 (12), 2231–2240.
- Schöffel, S., Schwank, J., Stärz, J., Ebert, A., 2016. Multivariate networks: A novel edge visualization approach for graph-based visual analysis tasks. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, ACM*, pp. 2292–2298.
- Shi, L., Wang, C., Wen, Z., Qu, H., Lin, C., Liao, Q., 2015. 1.5d egocentric dynamic network visualization. *IEEE Trans. Vis. Comput. Graphics* 21 (5), 624–637.
- Shin, H., Park, G., Han, J., 2011. Tablorer - an interactive tree visualization system for tablet pcs. *Comput. Graph. Forum* 30 (3), 1131–1140.
- Song, H., Kim, B., Lee, B., Seo, J., 2010. A comparative evaluation on tree visualization methods for hierarchical structures with large fan-outs. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM*, pp. 223–232.
- Srinivasan, A., Stasko, J., 2018. Orko: facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE Trans. Vis. Comput. Graphics* 24 (1), 511–521.
- Tan, D.S., Smith, G., Lee, B., Robertson, G.G., 2007. Adaptivtree: Adaptive tree visualization for tournament-style brackets. *IEEE Trans. Vis. Comput. Graphics* 13 (6), 1113–1120.
- Tanahashi, Y., Leaf, N., Ma, K., 2016. A study on designing effective introductory materials for information visualization. *Comput. Graph. Forum* 35 (7), 117–126.
- Tennekes, M., de Jonge, E., 2014. Tree colors: color schemes for tree-structured data. *IEEE Trans. Vis. Comput. Graphics* 20 (12), 2072–2081.
- Tu, Y., Shen, H., 2008. Balloon focus: a seamless multi-focus+context method for treemaps. *IEEE Trans. Vis. Comput. Graphics* 14 (6), 1157–1164.
- van Ham, F., Perer, A., 2009. Search, show context, expand on demand: Supporting large graph exploration with degree-of-interest. *IEEE Trans. Vis. Comput. Graphics* 15 (6), 953–960.
- van Ham, F., Rogowitz, B.E., 2008. Perceptual organization in user-generated graph layouts. *IEEE Trans. Vis. Comput. Graphics* 14 (6), 1333–1339.
- Vehlow, C., Beck, F., Weiskopf, D., 2015. The state of the art in visualizing group structures in graphs. In: *Eurographics Conference on Visualization (EuroVis) - STARS*.
- Vehlow, C., Beck, F., Weiskopf, D., 2017. Visualizing group structures in graphs: A survey. *Comput. Graph. Forum* 36, 201–225. Wiley Online Library.
- Verspoor, K., Ofoghi, B., Robles Granda, M., 2018. Commviz: visualization of semantic patterns in large social communication networks. *Inf. Vis.* 17 (1), 66–88.
- von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J.J., Fekete, J.-D., Fellner, D.W., 2011. Visual analysis of large graphs. *Comput. Graph. Forum* 30 (6), 1719–1749.
- Walny, J., Huron, S., Perin, C., Wun, T., Pusch, R., Carpendale, S., 2018. Active reading of visualizations. *IEEE Trans. Vis. Comput. Graphics* 24 (1), 770–780.
- Ware, C., Bobrow, R., 2005. Supporting visual queries on medium-sized node-link diagrams. *Inf. Vis.* 4 (1), 49–58.
- Ware, C., Gilman, A.T., Bobrow, R.J., 2008. Visual thinking with an interactive diagram. In: *Diagrammatic Representation and Inference, 5th International Conference, Diagrams 2008, Herrsching, Germany, September (2008) 19–21 Proceedings*, pp. 118–126.
- Ware, C., Purchase, H.C., Colpoys, L., McGill, M., 2002. Cognitive measurements of graph aesthetics. *Inf. Vis.* 1 (2), 103–110.
- Wong, N., Carpendale, M.S.T., Greenberg, S., 2003. Edgelens: An interactive method for managing edge congestion in graphs. In *9th IEEE Symposium on Information Visualization (InfoVis 2003)*, 20–21 2003, Seattle, WA, USA, pp. 51–58.
- Wong, P.C., Foote, H., Jr, G.C., Mackey, P., Perrine, K., 2006. Graph signatures for visual analytics. *IEEE Trans. Vis. Comput. Graphics* 12 (6), 1399–1413.
- Wong, P.C., Mackey, P., Perrine, K., Eagan, J., Foote, H., Thomas, J., 2005. Dynamic visualization of graphs with extended labels. In *IEEE Symposium on Information Visualization (InfoVis 2005)*, 23–25 2005, Minneapolis, MN, USA, p. 10.
- Wu, Y., Cao, N., Archambault, D., Shen, Q., Qu, H., Cui, W., 2017. Evaluation of graph sampling: a visualization perspective. *IEEE Trans. Vis. Comput. Graphics* 23 (1), 401–410.
- Xu, K., Rooney, C., Passmore, P., Ham, D.-H., Nguyen, P.H., 2012. A user study on curved edges in graph visualization. *IEEE Trans. Vis. Comput. Graphics* 18 (12), 2449–2456.
- Yang, Y., Dwyer, T., Goodwin, S., Marriott, K., 2017a. Many-to-many geographically-embedded flow visualisation: an evaluation. *IEEE Trans. Vis. Comput. Graphics* 23 (1), 411–420.
- Yang, X., Shi, L., Daianu, M., Tong, H., Liu, Q., Thompson, P., 2017b. Blockwise human brain network visual comparison using nodetrix representation. *IEEE Trans. Vis. Comput. Graphics* 23 (1), 181–190.
- Yi, J.S., ah Kang, Y., Stasko, J., 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Vis. Comput. Graphics* 13 (6), 1224–1231.
- Yeghordjian, V., Dwyer, T., Klein, K., Marriott, K., Wybrow, M., 2018. Graph thumbnails: Identifying and comparing multiple graphs at a glance. *IEEE Trans. Vis. Comput. Graphics*.
- Yoon, D., Narayanan, N.H., Lee, S., Kwon, O., 2006. Exploring the effect of animation and progressive revealing on diagrammatic problem solving. In *Diagrammatic Representation and Inference, 4th International Conference, Diagrams 2006, Stanford, CA, USA, June 2006, 28–30, Proceedings*, pp. 226–240.
- Yost, B., North, C., 2006. The perceptual scalability of visualization. *IEEE Trans. Vis. Comput. Graphics* 12 (5), 837–844.
- Yuan, X., Che, L., Hu, Y., Zhang, X., 2012. Intelligent graph layout using many users' input. *IEEE Trans. Vis. Comput. Graphics* 18 (12), 2699–2708.
- Zhao, J., Glueck, M., Chevalier, F., Wu, Y., Khan, A., 2016. Egocentric analysis of dynamic networks with egolines. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM, New York, NY, USA*, pp. 5003–5014.
- Zhao, J., Glueck, M., Isenberg, P., Chevalier, F., Khan, A., 2018. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE Trans. Vis. Comput. Graphics* 24 (1), 340–350.
- Zhao, J., Liu, Z., Dontcheva, M., Hertzmann, A., Wilson, A., 2015. Matrixwave: Visual comparison of event sequence data. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA*, pp. 259–268.
- Ziemkiewicz, C., Kosara, R., 2008. The shaping of information by visual metaphors. *IEEE Trans. Vis. Comput. Graphics* 14 (6), 1269–1276.
- Ziemkiewicz, C., Kosara, R., 2009. Preconceptions and individual differences in understanding visual metaphors. *Comput. Graph. Forum* 28 (3), 911–918.
- Zimmer, B., Kerren, A., 2015. Displaying user behavior in the collaborative graph visualization system ongrax. In *Graph Drawing and Network Visualization - 23rd International Symposium, GD 2015, Los Angeles, CA, USA, September (2015) 24–26, Revised Selected Papers*, pp. 247–259.