Systematic Application of Circle-Similar Shapes to Visualize Database-Homogeneity in a Big Data Environment

Verena Lechner(✉), Karl-Heinz Weidmann, and Isabella Hämmerle

Research Center for User Centered Technologies, University of Applied Sciences Vorarlberg, Hochschulstrasse 1, 6850 Dornbirn, Austria
uct@fhv.at

Abstract. Creating a useful visualization of database-homogeneity in a big data environment contains the task to present the underlying data in a perceivable way to the human observer. This paper provides a deep insight into the results of a research project commissioned by Crate Technology GmbH, where the before mentioned task was tackled by using a modified method of radar charts. A novel visualization tool to observe database-homogeneity was created with the aim to integrate it into the UI of the massive scalable elastic SQL Data store developed by Crate. The results of the formative usability test of this tool showed that the principle is understood intuitively by almost all test subjects already during the first task without any briefing about how to identify the problematic state of the single nodes of the database cluster. The results showed that this tool is expedient to visualize database-metadata in a big data environment.

Keywords: Database visualization · Big data · Form perception · Shape perception · Radar chart · Visualization · User interface · Formative usability test · Homogeneity

1 Introduction

Visualizing homogeneity in the area of database-metadata in a big data environment holds several challenges, as for example the amount of data to be displayed is not only huge but also changes during time [1, 2], therefore a good visualization should offer these data in an easily perceivable way to the human observer. For the research project with Crate Technology GmbH, who developed an elastic SQL Data Store that is massively scalable [3], the task was to develop such a visualization combining these already mentioned challenges with the precondition, that the observer doesn't necessarily has to be a highly professional database specialist. Moreover, as the databases of Crate are massively scalable, a broad variety of number of nodes has to be included into the visualization. This paper will first provide an insight into the development of a new visualization tool for database homogeneity. Then, an overview is given about the results of a formative usability test with the resulting tool that has been carried out at the Research Institute for User Centered Technologies (UCT). One of the main objectives was to explore whether the potential users are able to use the developed

A. Marcus (Ed.): DUXU 2016, Part III, LNCS 9748, pp. 167–179, 2016.
DOI: 10.1007/978-3-319-40406-6_16

visualization system intuitively, i.e. without prior instruction, for the observation of the database-homogeneity and to find potential improvable aspects.

2 Structure of the Visualization

2.1 Data to Be Represented

The visualization to be developed should enable the users to observe the database-homogeneity, for which we used a combination of four types of database metadata as the basic data set. Basically, database metadata can be defined as “data about database data”, so it’s not the data in the database itself, database metadata rather has the task to describe the database [4]. The four types of database metadata used for this tool are the average load for the last minute[1], the used disk (data size)[2], the used heap[3] and the number of shards[4]. These four data types are considered per node, the number of which may vary considerably, depending on the configuration of the database setup. As the visualization should enable an overview about the database-homogeneity, it’s not expedient to visualize the absolute values of each node, but the difference of these four data types of each node to the completely homogeneous cluster – so it’s the difference to the average of all nodes, which defines a relative measure.

2.2 Basic Concept of the Visualization

For the development of the visualization tool of metadata that should be applicable not only for highly professional database specialists, basic principles of human perception of shapes [5–9] were combined with the information visualization method of radar charts [10, 11]. The four types of database metadata we used for this visualization are available on the level of a single node. Therefore the basic element of the visualization is the node itself, which is (in its ideal state) represented as a circle[5]. Invisible in the overview mode of the visualization are the four circular arranged axes of the radar chart that lie behind the apparent circle-similar shape. In contrast to the radar chart first described by Georg Mayr in 1877 [10], the data points on the circular-arranged axes

[1] The arithmetic mean of all ethernet traffic (in bps) calculated over the last minute per node.

[2] The quotient of secondary-storage (disk) usage (blocks) in relation to the total available secondary-storage per node.

[3] The quotient of heap storage usage (bytes) in relation to the total available heap storage per node.

[4] The number of database shards (a horizontal partition of data in a database) per node.

[5] The reason why the form of a circle was chosen to represent a node in its ideal state has been investigated during an earlier stage of the research project. A detailed description of the form perception principles with regard to this project can be found in the proceedings of the HCI2015 [12]. Furthermore, an online web-survey with 399 test persons has been carried out in March 2015 in order to investigate the differences between the perception of polygons, circles and circle-similar shapes. The results of this survey as well as an analysis of comparable tools on the market have been published in January 2016 [13].

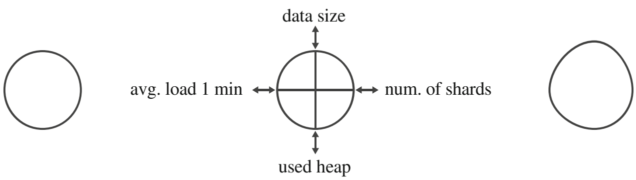

Fig. 1. Principle of the representation of a single node. 1. node in an ideal state (left circle); 2. with the axes labeling (middle circle); 3. inhomogeneous node (right deformed circle).

aren't connected with straight lines (resulting in a polygon) but with curves, that create a regular circle if the data points have the same distance from the center (which is the case when it has the ideal value) and a deformed circle-shape in the other case (cf. Fig. 1). So every node of the visualized database is represented as such a shape, the axes represent the four values "average load", "used disk", "used heap" and "number of shards". The ideal value of these metrics (observing cluster homogeneity) has been determined by Crate as the average of the single values of these metrics of all nodes. Therefore, if the observer of the database visualization perceives all the nodes as regular circles (and thus can assume that the cluster homogeneity is in the ideal state), depends on whether the values of these four axes of all nodes are the same. As soon as the values of one or more nodes deviate from the average, the forms of all other nodes of the cluster are modified as well to a lesser extent. Throughout this paper, only the nodes that deviate at least 15 % from the regular circle are stated as "inhomogeneous nodes".

The single nodes of a database running on Crate are arranged in a dark blue filled circle, which represents a whole cluster. The space between each node is 1,1 times larger than the radius of the ideal circle. That ensures that even if the maximum deviation of the data point on the axes is achieved with two adjacent nodes (50 % of the radius of the ideal circle), the outer edges don't touch each other. In order to make it easier for the user to interpret the deformations of the single circles, a legend is provided on the upper right corner. A white tooltip-box on single nodes shows the exact values of the four axes of the selected node as well as the exact average value of the whole cluster. Clicking on one of the nodes triggers a zoom into the selected node and shows also the axes (labeled with the underlying data type, the exact values of the enlarged node as well as the average value) and the ideal circle to compare the curves (cf. Fig. 2).

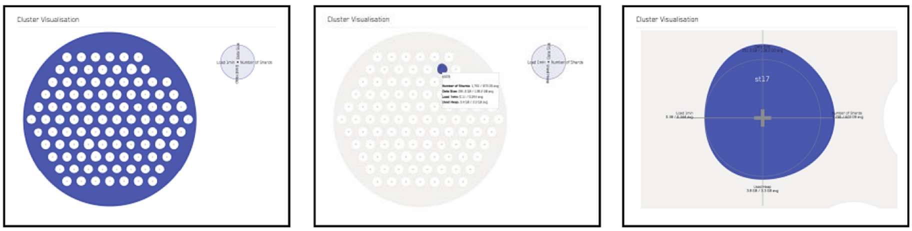

Fig. 2. Overview of the interface in its basic state with 100 nodes (left picture); in the mouse-over mode (middle picture); while zoomed in to a single node (right picture).

3 Testing of the Prototype

3.1 Objectives

The main aim of the developed visualization tool is to enable the end users to intuitively decide if the cluster running on Crate is in a homogeneous state and if not, to identifiy (not necessarily the whole number, but at least some of) those nodes that are inhomogeneous. Thus the observer may decide what problem could lie behind the deformation of the nodes and try to solve it. Therefore, the main objective of the testing was to evaluate if the test participants recognize that the shape of the nodes represents the status of a node and if they are capable to identify them in different settings with cluster sizes. The result should show if it's expedient to visualize database homogeneity through shapes as described earlier in this paper. Furthermore, questions such as which strategy was applied to scan through the clusters or whether the direction in which the circular shape is deformed (either inwards or outwards the center of the circle) have any impact on the simplicity of detection should be answered.

3.2 Method

Formative Usability Test. Due to the fact that the interface is still under development, carrying out a formative usability test was the chosen method to scrutinize it for potential points for improvement [14]. As the visual elements shown on the screen needed to be perceived with the same attributes (such as size, color, position), the test has been carried out under repeatable conditions between 14 to 20 October 2015 in the UCT in Dornbirn, Austria. Moreover, relevant data (as described later) wouldn't be possible to gather if the tests would have been carried out locally independent, as the technical systems have to be installed in a fix setup. Recording the comments of the test participants given through using the thinking aloud approach is also simplified by having a unified locality of the test.

Test Participants. The test has been carried out with 12 (graduated and undergraduated) computer science students of the University of Applied Sciences in Vorarlberg (who according to Crate are potential users with at least the minimum of required prior knowledge). Three test participants were studying in the fifth semester of the bachelor of computer science and nine participants were students of the master of computer science. The age of the test persons ranged between 22 and 56 (median 32). Two of them were women, thus the gender ratio is unbalanced. But considering that according to current statistics of the U.S. and the U.K. the number of women working in the IT sector is much lower than the number of men[6], therefore the unbalanced gender ratio within the group of test participants seems appropriate.

[6] According to the last update of the National Center for Women & Information Technology women held only 26 % of the computing occupations in the U.S. in 2014, taking into account 500 U.S. based companies [15]. With view on Europe, in the UK the percentage of women being employed as IT and Telecommunications Professionals was at 12.2 in 2013 [16].

Structure and Procedure. The test consisted of 12 different tasks the participants had to process. Each task showed a cluster that contained 8, 50 or 100 nodes with 1 to 10 inhomogeneous nodes. The testing was structured into two parts, the first offered only static pictures of the cluster as shown in Fig. 2 to solve the eight given tasks. In the second part, the test users had to interact with the visualization (cf. Fig. 2) and therefore got additional information and other views of the cluster.

As shown in Tables 1 and 2, the tasks 2, 6, 7 and 8 (static part of the test) contained the same number of nodes in a cluster and the same number of inhomogeneous nodes with the same percentage of deformation as tasks 9, 10, 11 and 12 (interactive part). Besides the fact, that in the second part of the test the earlier described interactive elements have been offered, the only difference between these two groups of tasks was the positioning of the modified nodes. Thereby it's important to emphasize that the changes of the positions have been altered only very slightly, that on the one hand the positions can't be memorized, but on the other hand the single tasks still stay comparable.

Before each task, the test participants got a written task description on the screen, instructing them to identify the node/s in a problematic state. For tasks 1, 2 and 9 they additionally had to tell the test leader the reason why they thought the selected nodes are in a problematic state. (How the problematic state of a node is represented and subsequently how the test persons should define the nodes that are in a problematic state wasn't explained at all.) After task 10 and 11 of the interactive part, an online questionnaire had to be filled out.

Used Software and Measurement Parameters. Five methods of data collection have been used for the testing:

- As a basis for all the analyses the screen (including the mouse positions) has been recorded during the whole test sessions.
- In order to gain precise data about where the test participants were looking at, the eye movements have been recorded with an eye tracking system.
- To collect the data of the think-aloud method the test participants have been filmed.
- Further information about the subjective evaluation of the participants has been gathered with a short online questionnaire. The questions concerned the easiness to identify the inhomogeneous nodes, the helpfulness of the arrangement of the nodes, the similarity of the inhomogeneous nodes to the homogeneous ones, the number of nodes displayed and if the size of the nodes enabled a comfortable perception. The questions had to be rated on a 5-point semantic differential (−2 to +2). Participants also had to answer how the ideal values of the single axes of a node have been developed (two answer options were given).

Data Analysis. The data of all 12 participants were included in the statistical analysis. The amount of errors made was determined by measuring firstly how many deformed circles in each task were not identified as inhomogeneous nodes (hereinafter "error type 1") and secondly how many almost regular circles were classified mistakenly as inhomogeneous nodes (hereinafter "error type 2"). The relation of identified inhomogeneous nodes to the amount of inhomogeneous nodes in each task was also calculated and is displayed in percentage. The deformations of the inhomogeneous nodes were

Table 1. Overview of the tasks of the static part of the test

Task numbers	Number of nodes in the cluster	Number of inhomogeneous nodes	Position of the inhomogeneous node/s	Percentage of deformation in comparison with the radius of the ideal value
task 1	8	1	left horizontal axis: value below average	30 %
task 2	8	1	left horizontal axis: value above average	40 %
task 3	8	2	bottom vertical axis: value above average	20 %
task 4	8	2	bottom vertical axis: value below average	20 %
task 5	8	3	bottom vertical axis: value above average	20 %
task 6	50	6	3 nodes: right horizontal and upper vertical axis: value above average; 3 nodes: right horizontal and upper vertical axis: value below average	25 %
task 7	100	10	5 nodes: right horizontal and upper vertical axis: value above average; 5 nodes: right horizontal and upper vertical axis: value below average	25 %
task 8	100	2	right horizontal and upper vertical axis: value below average	50 %

Table 2. Overview of the tasks of the interactive part of the test

Task numbers	Number of nodes in the cluster	Number of inhomogeneous nodes	Position of the inhomogeneous node/s	Percentage of deformation in comparison with the radius of the ideal value
task 9	8	1	left horizontal axis: value above average	40 %
task 10	50	6	3 nodes: right horizontal and upper vertical axis: value above average; 3 nodes: right horizontal and upper vertical axis: value below average	25 %
task 11	100	10	5 nodes: right horizontal and upper vertical axis: value above average; 5 nodes: right horizontal and upper vertical axis: value below average	25 %
task 12	100	2	right horizontal and upper vertical axis: value below average	50 %

classified in hereafter so called "too small nodes", which included the nodes that had at least one value of the four axes below the average, and in hereafter so called "too big nodes", which had at least one value of the four axes above the average of the whole cluster (cf. Tables 1 and 2).

To identify the strategy used to search for the inhomogeneous nodes eye tracking data were analyzed. The results lead us to the two categorizations: The test participants used either a "chaotic" or "line-based" strategy (cf. Fig. 3). The category "chaotic" contained all participants who didn't use an explicit line-based strategy. The categorisation was done for each task with at least 50 nodes.

The answers of the questionnaire were transformed to values from 1 to 5, where 1 represented the positive end and 5 the negative end.

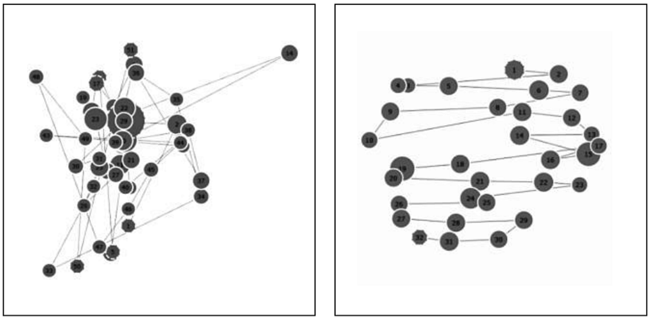

Fig. 3. Exemplary eye tracking data: chaotic (left square) and line-based strategy (right square)

Not all tasks were included in every analysis, since the percentage of deformation of the problematic nodes wasn't the same for all clusters as well as the ratio of inhomogeneous nodes to the total amount of nodes in a cluster differed (cf. Tables 1 and 2). Which tasks are included is therefore stated for each step of analysis.

Statistical Analysis. Descriptive statistics was used to describe the amount and types of errors the participants made. Independent Student's *t*-tests were calculated to determine if there are any differences between the cluster sizes (only 50 or 100 nodes) regarding the subjective evaluation of the visualization system with the questionnaire.

A Pearson's chi-square test was conducted to test for a relationship between the size of the cluster and the method used to scan the cluster. Furthermore, Pearson's chi-square tests were calculated to test if there is a relationship between the cluster size and the amount of identified inhomogeneous nodes as well as if there is any relationship between the size of the inhomogeneous node (too small or too big) and the probability that the problematic node is identified.

The significance level was fixed at the 5 % level. The statistical analysis was performed using IBM SPSS Statistics Data Editor Version 23.0.

3.3 Results

Correct Identification of the Inhomogeneous Nodes. The average percentage of all identified inhomogeneous nodes (ratio of detected inhomogeneous nodes to the amount of inhomogeneous nodes per task) and of all too small and too big inhomogeneous nodes were calculated once for all tasks and once for the tasks 6, 7, 10 and 11. If all tasks were included in the analysis the participants identified 86 % of all inhomogeneous nodes compared to 75 %, if just the tasks 6, 7, 10 and 11 were considered. If the type of deformation was regarded separately it can be seen that the percentage of identified inhomogeneous too big nodes was 90 % (for all tasks) resp. 87 % compared to 77 % resp. 62 % for the nodes that were too small.

After the first, the second and the ninth task, the test persons were asked to describe orally, why they thought that the nodes they selected in the previous task are in a problematic state. As shown in Table 3, the answers were classified in three categories. For the right answers two categories were defined: If the test participants named the axis (as described in the legend in the right upper corner) where the deformation took place, the answer was right because it was the most precise answer they could give (named as "right answer type 1" in Table 3). The second type of right answers contains all answers where the test participants only named the deformation of the single circles representing a node as the indicating parameter for the problematic state (named as "right answer type 2" in Table 3). If a test participant didn't name one of the previous reasons, the answer was wrong (named as "wrong answer" in Table 3).

Table 3. Oral answers to the question why the test persons thought that the before selected nodes are in a problematic state.

Answer	Num. of persons that selected that answer after task 1 (n = 12)	Num. of persons that selected that answer after task 2 (n = 10)	Num. of persons that selected that answer after task 9 (n = 11)
right answer type 1	3	2	8
right answer type 2	8	8	3
wrong answer	1	0	0

Pearson's chi-square test revealed a significant association between the type of deformation and whether the nodes were identified or not, χ^2 (1) = 30.76, $p < .001$ as well as for the size of a cluster and whether the nodes were identified or not, χ^2 (1) = 3.88, $p < .05$. Based on the odds ratio, the odds of nodes to be identified as inhomogeneous nodes were 3.92 times higher if they were too big than if they were too small and 1.63 times higher if they were in a cluster with 50 nodes than if they were in a cluster with 100 nodes. Tasks included in the analysis were task number 6, 7, 10 and 11 since they had the same amount of too big and too small nodes within each task and all nodes had the same degree of deformation.

Error Rates. The average of the absolute value of errors of all 12 participants of type 1 and type 2 per task is displayed in Table 4. These results have been calculated once for all tasks as well as just for tasks with a cluster size of at least 50 nodes and a uniform deformation of 25 % (tasks 6, 7, 10, 11).

Table 4. Comparison of the error rate of error type 1 and type 2 of two different task groups

Tasks included in the analysis	Error type 1 per task	Error type 2 per task	Error type 1 per person per task	Error type 2 per person per task
all tasks	9.8	8.5	0.8	0.7
tasks 6, 7, 10, 11	25.5	9.8	2.1	0.8

Table 5. Absolute values of used search strategies per cluster size (n = number of tasks)

Cluster size	Chaotic	Line-based
50 nodes (n = 2)	21	3
100 nodes (n = 4)	27	21

As can be seen in Table 4 the error rate of type 1 increased from 0.8 per person per task to 2.1 if not all tasks but just the more complex ones (min. 50 nodes, deformation value 25 %) were included. This rise in mistakes wasn't observed for errors of type 2 (0.7 to 0.8).

Half of the participants (6 persons) answered the question (of the questionnaire) how the ideal values of the single axes have been developed correctly (through the average of all nodes displayed at the same time) and five persons had chosen the wrong answer (through a fixed value). One person thought that none of the two answers is correct.

Strategy Used to Scan Through the Clusters. Table 5 shows how often the two scanning strategies (chaotic or line-based) were used for different cluster sizes. All tasks that had a cluster size of at least 50 nodes were included in this analysis (cf. Table 5).

Pearson's chi-square test to conduct if there is a relationship between the size of the cluster and the strategy used to scan a cluster was significant, χ^2 (1) = 7.03, p < .01. Based on the odds ratio, the odds of using a chaotic search strategy were 5.4 times higher for 50 node clusters than for 100 node clusters.

Table 6. Means, standard errors and p-values of the *t*-tests to compare different questions depending on the cluster size

Question (categorisation of the answer)	Cluster size	*M*	*SE*	*p*
It was difficult /easy for me to detect the problematic nodes. (1 = easy, 5 = difficult)	50 nodes	2.83	0.83	1
	100 nodes	2.83	0.83	
The presentation of the single nodes was too small /too big for a comfortable perception. (1 = too big, 5 = too small)	50 nodes	3.12	0.58	.058
	100 nodes	3.92	1.16	
The presentation showed too many /too less nodes to compare. (1 = too less, 5 = too many)	50 nodes	3.83	0.94	.65
	100 nodes	4.00	0.85	
The problematic nodes have been too similar to the not problematic ones /the difference was large enough to easily distinguish. (1 = large enough, 5 = too similar)	50 nodes	3.50	0.67	.67
	100 nodes	3.67	1.15	

Results of the Subjective Evaluation of the Visualization System. The *t*-tests revealed only a significant difference between 50 node clusters and 100 node clusters for the question, if the arrangement of the nodes was confusing or helpful, $t(22) = -2.11$, $p < .05$. On average, participants rated the arrangement of the 50 node clusters less confusing ($M = 2.75$, $SE = 0.22$) than for the 100 node clusters ($M = 3.42$, $SE = 0.23$). The means, standard errors and *p*-values for the other questions can be seen in Table 6.

4 Discussion

The results showed that all participants but one understood the visualization system intuitively already in the first task in terms of recognizing that the shape of the nodes represents the status of a node. After the 9^{th} task eight participants answered the open question why they thought that the nodes they have selected in the previous task were in a problematic state totally correct (answers referred to the axis of the nodes). The other three participants just mentioned the deformation of the nodes, which is also a right answer, but not as precise as the other one. On the other side the question how the ideal values of the single axes have been developed was just answered right by six participants. This indicates, that the system can be operated by users even if they don't know how the ideal values of the nodes are developed, but recognizing the deformation of the nodes is sufficient.

If all tasks were considered, 86 % of all inhomogeneous nodes were identified to be in a problematic state but just 75 % if only the more complex tasks (at least 50 nodes and more than two inhomogeneous nodes per cluster) were included in the analysis. Interestingly was the result, that the type of deformation (too small or too big) seems to be important for the identification of the inhomogeneous nodes, since 90 % (87 % for the complex tasks) of the nodes that were too big were identified compared to just 77 % (62 % for the complex tasks) of the nodes that were too small. This indicates that - at least for our specific setting (regarding distance between the nodes, colours and display size of the nodes) - too big nodes are easier to identify than too small nodes.

The analysis of the error types also revealed impressive results. If all tasks were considered in the error rate analysis each participant didn't identify 0.8 inhomogeneous nodes per task as problematic ones (error type 1) and rated 0.7 of the circular shaped nodes mistakenly as problematic ones (error type 2). If just the complex tasks were considered, the amount of not identified inhomogeneous nodes rose from 0.7 to 2.1 per task. This high increase of errors wasn't observed for the circular shaped nodes that were rated as problematic ones (increase from 0.7 to 0.8). Based on this results it can be concluded that with increasing complexity more inhomogeneous nodes are overlooked but not more homogeneous nodes are rated mistakenly as inhomogeneous.

For the subjective evaluation where the participants rated different aspects of the visualization we expected some differences depending on the cluster size. The results showed just a significant difference between the cluster sizes for the arrangement of the nodes. The smaller cluster was rated less confusing. The easiness to identify homogeneous nodes was rated exactly the same for 50 and 100 nodes, it was neither perceived as very easy nor as very difficult. The problematic nodes were neither rated as too similar to the problematic ones to easily identify them nor was the difference rated as large enough. The participants rated the nodes as rather too small for a comfortable perception and as rather showing too many nodes per cluster. On the one hand it seems like the participants weren't really satisfied, but also not extremely dissatisfied with the visualization. But on the other hand, the dissatisfaction didn't increase significantly from 50 nodes to 100 nodes, even not for the statement if the visualization showed too many nodes. The question arises, how the ratings would be if the clusters were even bigger.

The analysis of the search strategy showed that a line-based search strategy is more often applied for clusters with 100 nodes than for clusters with just 50 nodes. This indicates that with a rising number of nodes the participants started to use a strategic approach, probably because with a higher number of nodes the chance to overlook an inhomogeneous node is bigger. The question arises, if a visualization tool that is developed to give a quick overview of the cluster homogeneity should be scanned line-based. It has to be said that the arrangement of the nodes in the clusters supported a line-based strategy, since the nodes were arranged not totally chaotic but in horizontal lines. It would be interesting, if a line-based strategy would also be used if this isn't the case. Further research is necessary to gain better insights which search strategy should be applied to fulfil the task best and what kind of node arrangement best supports that the users apply this particular strategy.

5 Conclusion

The test of the prototype implementation confirmed the usefulness of the selected approach. Participants didn't have problems to understand the developed visualization tool intuitively without any explanation and performed at a reasonable well level for all selected tasks. They performed even better compared to their subjective judgement. Certainly, there is room for improvement of the presentation of the visualization e.g. the users rated the nodes as rather too small for a comfortable perception. Further research is necessary to confirm our findings and to see if our findings are also applicable for bigger clusters.

References

1. Cox, M., Ellsworth, D.: Managing big data for scientific visualization. In: Siggraph 1997, Computer Graphics Annual Conference Series, Course 4, Exploring Gigabyte Datasets in Real-Time: Algorithms, Data Management, and Time-Critical Design. Assn for Computing Machinery, Los Angeles (1997)
2. Illingworth, J.A., Lippstreu, M., Deprez-Sims, A.-S.: Big data in talent selection and assessment. In: Big Data at Work. The Data Science Revolution and Organizational Psychology. p. 368. Routledge Taylor & Francis Group, New York (2016)
3. CRATE Technology GmbH: Crate. Your Elastic Data Store. https://crate.io/
4. Parsian, M.: JDBC Metadata, MySQL, and Oracle Recipes: A Problem-Solution Approach. Apress, Berkeley (2006)

5. Bernhard, T.: Wahrnehmung der Schönheit in der Architektur. epubli, Berlin (2012)
6. Naumann, A.: Raimer Jochims: FarbFormBeziehungen: Anschauliche Bedingungen seiner Identitätskonzeption. Königshausen & Neumann, Würzburg (2005)
7. Arnheim, R.: Art and visual perception: a psychology of the creative eye. In: 50th Anniversary Printing. Original edition: Art and Visual Perception: A Psychology of the Creative Eye. Universitiy of California Press, Berkeley and Los Angeles 1954. University of California Press (2004)
8. Pizlo, Z.: 3D Shape: Its Unique Place in Visual Perception. MIT Press, Massachusetts (2008)
9. Alexander, K.: Kompendium der visuellen Information und Kommunikation. Springer-Verlag, Berlin (2013)
10. Mayr, G.: Die Gesetzmäßigkeit im Gesellschaftsleben. Statistische Studien. Oldenbourg, München (1877)
11. Takezawa, K.: Guidebook to R Graphics Using Microsoft Windows. Wiley, Hoboken (2013)
12. Lechner, V., Weidmann, K.-H.: Visualizing database-performance through shape, reflecting the development opportunities of radar charts. In: HCI International 2015 Conference Proceedings. Springer, Los Angeles (2015)
13. Lechner, V., Weidmann, K.-H., Mayer, H.O.: Visualisierung der Datenbank-Performance durch die methodische Verknüpfung von Grundlagen der Formwahrnehmung mit den Prinzipien des Radardiagramms. In: design2product. Beiträge zur empirischen Designforschung. Band 5 - Methoden der Designforschung, pp. 148–168. Bucher, Hohenems (2015)
14. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1994)
15. National Center for Women & Information Technology: Women in IT: The Facts Infographic. https://www.ncwit.org/resources/women-it-facts-infographic-2015-update
16. Baker, C., Cracknell, R.: Women in Public life, the Professions and the Boardroom. House of Commons Library, London (2014)