Configuration Finder: A Tidy Visual Interface for Effective Faceted Search

P. Riehmann, A. Schädler, J. Harder, J. Herpel, and B. Froehlich

Bauhaus-Universität Weimar, Germany

Abstract

We present an interactive visualization aiding users in making informed decisions about large product data sets consisting of quantitative and categorical attributes. Our approach tries to overcome common problems between parallel attribute axes, for instance limited horizontal space or clutter, by introducing novel visual concepts such as proxy axes, fusion axes, and hybrids of set-based and individual axis connections. A proxy axis represents a group of semantically related attributes, which can be interactively explored and seamlessly integrated into the display. Fusion axes allow users to reduce the number of axes by merging categorical+categorical or categorical+quantitative attribute axes. Set-based or individual connections between axis pairs are chosen according to the involved attribute types. The pilot study and expert reviews showed that these novel concepts are understood, considered to be very useful and favored over up-to-date webshop interfaces.

CCS Concepts

• ***Human-centered computing*** → ***Visualization techniques; Information visualization; Visualization theory, concepts and paradigms;***

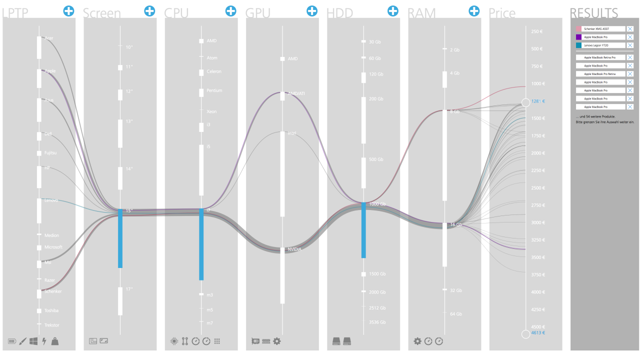

Figure 1: *Connections between axes provide continuation and alignment. Between the axis RAM and the continuous axis Price individual connections are shown whereas ribbons (set-based connections) are used between the other axes. Potential candidates are currently selected in the result list on the right and highlighted by different colors. Their individual paths are drawn on top of the respective ribbons. Proxy axes reveal the represented set of axes on demand by clicking on their respective ⊕ icon (see Figure 2).*

1. Introduction

An analysis of webshop interfaces with respect to their adherence to Hearst's guidelines on search interfaces [Hea09] revealed that most follow these guidelines only partially and provide very limited means for effective faceted search. Generally missing is information on the availability of certain attribute combinations. This can only be seen once a value for the first attribute is selected, since only then the remaining available values for the second attribute are shown. If the desired combination is not among the remaining ones, one has to go back to the selection of the first attribute or try to select the second attribute first etc. So, finding a certain combination can become a frustrating trial and error process.

Parallel Coordinate Plots [Ins09] (Heinrich [HW13] provides a thorough survey) in their most basic form show all available attribute combinations at once but suffer from clutter and overplotting when depicting numerous products and their many attributes. Various approaches tried to overcome these issues including Siirtola [Sii00] by grouping multiple lines by logical operations, Wong [WCG03] by an edge lens that bends lines to reveal underlying structures, Graham et al. [GK03] by using quadratic and cubic curves instead of straight lines, Peng et al. [PWR04] by axes reordering techniques, and Kauer [KK18] with Bifocal Parallel Coordinates. Yet all these techniques only treat symptoms of an already cluttered display.

Instead, we focus on limiting the number of attributes and attribute relations that need to be considered at any time. Our system starts with an overview by displaying a small initial set of proxy axes and important regular axes which show the distribution of attribute values across all products. A proxy axis represents a group

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

DOI: 10.2312/evs.20201042

delivered by EUROGRAPHICS DIGITAL LIBRARY
www.eg.org diglib.eg.org

of semantically related attributes which can be exposed on demand. Attribute relationships, encoding the available configurations of the shown products, are added once the user starts to narrow down the search space. They are visualized as ribbons or individual line connections depending on the types of adjacent axes. The axis connections are independently formed for all adjacent axis pairs to limit the number of connections and to provide a smooth branching and merging of connections across the attribute axes (Figure 1). To reduce the number of separate axes and connections, fusion axes can be set up by merging two axes into a single one. We explored the usability of these novel concepts in a user study and expert reviews.

2. Related Work

Product data attributes are mainly categorical. Bendix and Kosara introduced Parallel Sets for exploring categorical data [KBH06]. Quantitative relationships are shown as ribbons passing from one category on one axis to another category on the next. A similar older tool without connections is Parallel Bargrams (or EZ-Chooser) [WLHS01]. Riehmann [ROF12] provided a product selection interface as enhanced PCPs for mixed (categorical and continuous attributes) data. Yet, each product is still drawn as an individual path. In this regard Tuor et al. [TEL18] presented an evaluation of variants of representing such mixed data.

Besides PCP derivatives early papers as Dynamic Queries and Home Finder [Shn94, WS92] were suited for finding real-estates. More general tools for analyzing multi-attribute data were the Influence Explorer [TSDS95] and the Attribute Explorer [ST98], both providing multiple histograms combined with interactive range selections. Later on, Bautista et al. [BC06] presented redesigned Value Charts (originally [CL04]). Subsequently, Conati [CCH*14] tested “the Impact of User Characteristics and Different Layouts on an Interactive Visualization for Decision Making.” A study by Yi [Yi08] proved the usefulness of a visual decision tool for choosing a nursing home. Icon-based visualizations are an alternative for multivariate data (see Chernoff Faces [Che73]). Spence and Parr [SP91] used miniature house icons (as direct metaphorical icons) to encode attributes of real estates. Similarly, Riehmann [RMF14] introduced comic-like characters (for children between four and eight) whose shape and features depend on the main ingredients of food products. The number of attributes that could depicted by a glyph is limited, though.

A number of versatile tools for an in-depth analysis of tabular data were proposed, e.g. Bertifier [PDF14], SMARTExplore [BBS*18], LineUp [GLG*13] or Taggle [FGS*20]. However, their powerful abilities come with a certain operational and visual complexity of numerous small visual elements, which contradicts with our approach of visual reduction.

3. Visual Concept

Our design is aimed at reducing the number of visual elements from the beginning, allowing the user to become familiar with the available attributes and attribute values without being distracted by the connections between the attribute axes (see video). As soon as the first value is chosen (indicating the user feels familiar enough to start investigating), the connections appear immediately and provide quantitative relations between attribute values.

3.1. Semantic Attribute Hierarchies with Proxy Axes

Is it really necessary to show all axes at once and all the time to make an informed decision? Most datasets contain attributes more important to a decision than others. We already tried to reduce the number of axes by ranking and interacting with a separate axis repository [ROF12]. Yet, we did not take advantage of semantic relations. Many of the attributes are related to one another or even subordinated. For example, processors or graphics cards have a marketing name as well as detailed specifics such as speed, number of cores, etc. Such attributes belong together and can be organized in groups or even in shallow hierarchies. For that we introduced proxy axes as representatives of an attribute group, which consists of semantically related attributes considered to be not as important as the one represented by a proxy axis (see Figure 2).

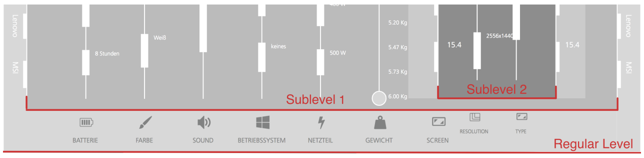

Figure 2: *The laptop proxy axis is opened (first sublevel) which also contains the Screen axis that is opened as the second sublevel.*

A single proxy axis can be opened for revealing all represented (sub) axes. The opened axis is split vertically in half by a sliding horizontal animation. This way an open proxy axis shows connections (ribbons) to the (proxy) axes left and right as well as to its sub axes inside. The inner space containing the sub axes can be clearly discriminated by its background. The maximum number of axes on a FullHD screen is about 10 for guaranteeing proper ribbon shapes. Each of them could stash and reveal up to 8 sub axes (along with their split outer proxy axis framing them), which also could contain sub axes themselves resulting in 640 axes for three levels.

Attribute icons (designed for the dataset and positioned below their respective sub axis) provide a visual hint of the contained sub axes even while the proxy axis is closed and sub axes are not visible (see Figure 1. Every time an axis is opened the icons are being moved to their respective sub axes (see Figure 2).

Selection works similarly for a proxy axis and a sub axis. Clicking on a value area filters the ribbons accordingly for all possible connections (regardless being inside or outside a proxy axis). Selections on sub axes persist, no matter being open or closed. Reopening an axis recreates the state exactly as it was before which makes it ready to proceed or altering the selections previously made. On a closed proxy axis the sub axis icons indicate the attributes chosen inside. Hovering over a highlighted icon reveals the picked value.

3.2. Set-based and Individual Connections

Parallel Sets [KBH06] already reduce screen elements by considering sets instead of individual items or products. Sets are represented as bars on each axis and connected by ribbons. But, for a

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

larger number of attributes and in particular if quantitative ones are included, this might quickly result in a large number of tiny ribbons. Therefore, we decided to compute axis connections only on the basis of pairs of adjacent axes, since parallel axes are best suited for showing the relationships between two consecutively displayed attributes. For two categorical axes attribute combinations are shown by set-based connections using ribbons as with Parallel Sets. If one of two adjacent axes is quantitative we use line-based connections that represent individual products as with Parallel Coordinates. When connections switch from ribbon to line-based, i.e. two categorical axes are followed by a quantitative axis or vice versa we make sure that vertical ribbon positions and ribbon widths on one side of the central axis match the positions and accumulated widths of the individual lines on the other side (Gestalt Law of continuation).

In order to reduce the number of individual line connections, we could bin quantitative attributes such that only a few values remain. For some attributes this is naturally the case (e.g. memory of 4GB, 8GB, etc.) but if the value distribution is spread across the entire quantitative axis, individual lines might still be the preferred option. Furthermore, in certain situations, it is necessary to make an individual product or a small number of items stand out against the rest. Selecting particular items in the result list (on the right) for closer inspection creates individually colored lines that are overlaid in front of the existing ribbons to reveal their attribute values across all axes. Determining reasonable crossing points on the bars of a categorical axis is crucial since the individual curves should be displayed within the boundaries of their respective ribbons.

3.3. Fusion Axes

To be even more space efficient, axes can be combined; either pairs of quantitative+categorical or categorical+categorical axes (Figure 3) by mapping quantitative or discrete value ranges onto the discrete categorical values represented as bars on the second axis. A quantitative attribute range is encoded by a color gradient whereas for categorical values the bars are further split for each new subset.

The color gradients on the bars of a categorical axis share a common color range, mapping from the lowest value to the highest value for a particular attribute. This allows the user to recognize particular relations regarding the categorical values and their distribution across the quantitative range. Applying a quantitative attribute such as price to all axes reveals additional information about the distribution of price ranges within the value bars. In our notebook examples, the i7 processors and the 64 GByte bars are the darkest and thus most expensive ones.

Due to the distribution of the actual quantitative values in the range, a linear color gradient would be incorrect. In order to provide a realistic visual distribution of the color mappings each value bar's gradient has to be constructed individually in a following pixel-based manner. We know how many items (products/laptops in our case) are represented by the height of a value bar and thus how many items one horizontal line of pixels represents. For each single pixel we average these n values of the merged quantitative attribute and map the resulting value by a global color mapping. For the pixel lines representing the lowest and the highest value (positioned at the top and bottom of a bar) we do not average; we simply choose the lowest (respective highest) value for color mapping in order to represent the entire range depicted by an individual value bar. We draw a bar as well as sub bars always at least with 2 pixels height for being able to hover above and perform a selection. This approach limits the total number of bars and sub bars that can be displayed. To avoid discontinuities (see Figure 4) we opted against overlaying ribbons when selecting a subrange or sub bar and simply treat those as regular selections that narrow down the result set further.

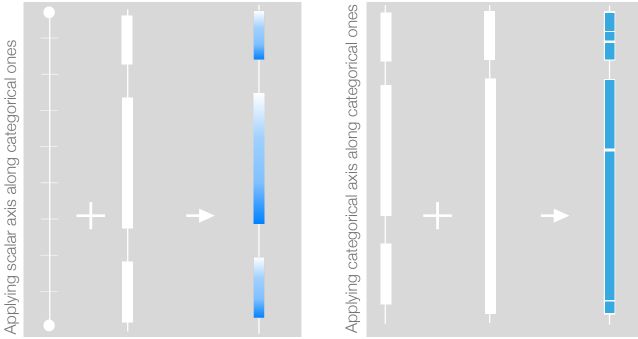

Figure 3: *Merging a quantitative axis with a categorical one results in a new axis (left), resembling the categorical axis, but displaying the quantitative value information on its bars as a pixel-based color gradient expressing the quantitative values belonging only to the categorical values of the respective categorical bar. Merging two categorical axes (right) maps the categorical value range of one axis to the bars of the other axis, resulting in sub bars. The heights of the sub bars are mapped to the number of records/items sharing both attribute values.*

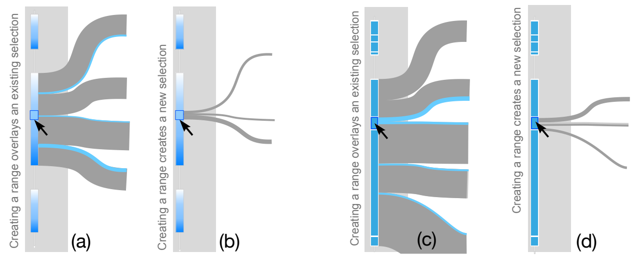

Figure 4: *Selecting ranges on merged axes. (a) Overlaying the ribbons when selecting on a gradient would display discontinuities between the selected range and the highlights needed for individual ribbons. (b) Thus, we opted to brush a desired range as a new selection and show the new ribbons vertically aligned to the selected range. (c) The same problem appears for selecting a sub bar and thus a subset of two merged categorical axes. So, selecting a sub bar (d) also presents the new ribbons aligned.*

3.4. Result List

While narrowing down a product set the result list (Figure 1, right side) shows the top ten products. Potential purchase candidates can be highlighted with a distinctive color in the list as well as when

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

drawn as paths across all axes (see also Subsection 3.2 and Figure 1). In order to maintain candidates with different attribute values, they are resistant to any further changes and will be kept in the list and in the main view even if their values are deselected.

4. Pilot Study and Expert Reviews

We surveyed popular webshops and reimplemented the interface most compliant with Hearst's guidelines [Hea09] and tested it against our system (without fusion axes). Additionally, we performed reviews with six experts working in webshop design and development (with fusion axes).

The 8 participants (5 female and 3 male, between 21 and 31 between) performed seven tasks with both interfaces in a within-subject, counter-balanced design. The participants, students from the humanities (without any knowledge of PCPs, etc.), were introduced to both systems in detail followed by self-performed sample tasks and a 5 minute play phase per interface. The reimplemented webshop interface presented the results as a list of 10 items per page with detailed attribute values (see Figure 5, only the top two items). We opted against a matrix-like gallery just showing pictures (as it might be useful for fashion articles) which could contain more items but would require an additional click or hovering to reveal the actual attribute values. Four tasks were destined to find a particular model, two to find a balance (optimum) between two attributes, and one to look for a notebook for themselves. CSUQ (Computer System Usability Questionnaire [Lew95]) was used instead of SUS (System Usability Scale [Bro13]) due to its more specific questions.

Our analysis showed that a familiar webshop interface (the best according to Hearst's guidelines) does not offer any advantage regarding task completion times over the previously unfamiliar and briefly-introduced configuration finder. Task completion times for two of the tasks showed significant differences ($p < 0.05$) in a two-sided t-test for our system. Moreover, of the two described task types the consideration (balancing) type show a statistically significant difference ($p < 0.05$) in favor of new system. CSUQ measured medium values for the web interface, but in contrast, excellent values for our system. No participant's CSUQ score for our system was worse than 2.27 (on a scale from 1 to 7, the lower the better, see Figure 5). The web interface was perceived as more confusing. In particular, participants were annoyed by the fact that the web interface had to be scrolled more often to see how the result list had changed after setting filters differently. In contrast, our system was described as "easy to use and not overloaded ... showing efficiently what was selected." The majority of the participants stated being more confident having found the correct results with the configuration finder and that it appeared to be more trustworthy. The results of the CSUQ reveal how well the visual concepts of our system had been understood, especially, regarding the distinction of product attributes shown by proxy axes and sub axes. Six persons fully agreed "It was easy for me to find the categories I was looking for", whereas the rest mostly agreed to this point.

Aside from regular user tests, we asked six experts (3 per session) that are with one of the very first companies providing commercial webshop systems. As with our regular participants a brief demonstration sufficed to enable our experts to perform example

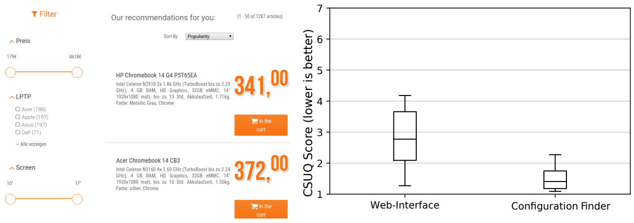

Figure 5: *Fully functional webshop reimplementation (left). The CSUQ scores give a clear impression (lower is better, right).*

tasks as well as narrowing down their own laptop requirements or interests. The reviews turned quickly into a market analysis in order to find typical relations and missing configurations nobody offers. Our experts also saw potential as in-house application to compare a company's portfolio against the market or for optimizing purchases. They especially liked the fusion axes for combining price information with any categorical attribute.

5. Conclusion, Discussion, and Future Work

Our system generalizes concepts from Parallel Sets and PCPs. Proxy axes represent a set of hierarchically organized related attributes for providing a minimal set of attributes at startup. They literally split in half to show the represented axes seamlessly integrated into the visual display. Fusion axes represent a pair of categorical axes or a pair of a categorical and a quantitative axis. A fusion axis shows all combinations of attribute values of the two fused axes. We effectively reduce clutter and provide a tidy display by forming axis connections independently between adjacent axis pairs. Connections across multiple axes for individual products are provided only on demand. Our pilot study and expert reviews show that users understand these new concepts, feel more confident in their decisions and appreciate the visualization of available attribute combinations. Nevertheless, further studies are necessary to reveal in detail the effects of our techniques and design decisions.

In this paper, we focused on product data with mainly categorical attributes and only a few quantitative ones. The concepts of fusion axes and proxy axes can be adapted for datasets from other domains with similar properties. Currently, proxy axes and fusion axes are manually curated. However, (semi-)automatic approaches for deriving proxy hierarchies would be desirable but seem highly domain dependent. Furthermore, the (sub) axes represented by a proxy axis should be obvious to the user which we tried to achieve by manually designing appropriate icons. They would be difficult to create automatically.

While our web-based implementation renders with Javascript solely on the client and already scales to thousands of products, even larger datasets would require aggregation techniques such as clustering very similar products into a single one. Nevertheless, we believe that the concepts presented in this paper and implemented in the Configuration Finder are an important step to establish visualizations based on parallel axes in future faceted search interfaces.

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

References

[BBS*18] Blumenschein M., Behrisch M., Schmid S., Butscher S., Wahl D. R., Villinger K., Renner B., Reiterer H., Keim D. A.: Smartexplore: Simplifying high-dimensional data analysis through a table-based visual analytics approach. In *2018 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2018), pp. 36–47. 2

[BC06] Bautista J., Carenini G.: An integrated task-based framework for the design and evaluation of visualizations to support preferential choice. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (2006), AVI '06, ACM, pp. 217–224. 2

[Bro13] Brooke J.: Sus: A retrospective. *J. Usability Studies 8*, 2 (Feb. 2013), 29–40. 4

[CCH*14] Conati C., Carenini G., Hoque E., Steichen B., Toker D.: Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Comput. Graph. Forum 33*, 3 (2014), 371–380. 2

[Che73] Chernoff H.: The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association 68*, 342 (1973), 361–368. 2

[CL04] Carenini G., Loyd J.: Valuecharts: Analyzing linear models expressing preferences and evaluations. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (New York, NY, USA, 2004), AVI '04, ACM, pp. 150–157. 2

[FGS*20] Furmanova K., Gratzl S., Stitz H., Zichner T., Jaresova M., Lex A., Streit M.: Taggle: Combining overview and details in tabular data visualizations. *Information Visualization 19*, 2 (2020), 114–136. 2

[GK03] Graham M., Kennedy J.: Using curves to enhance parallel coordinate visualisations. In *Proc.of the Seventh International Conference on Information Visualization* (2003), pp. 10–16. 1

[GLG*13] Gratzl S., Lex A., Gehlenborg N., Pfister H., Streit M.: Lineup: Visual analysis of multi-attribute rankings. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (Dec 2013), 2277–2286. 2

[Hea09] Hearst M.: *Search User Interfaces*. Cambridge University Press, 2009. 1, 4

[HW13] Heinrich J., Weiskopf D.: State of the Art of Parallel Coordinates. In *Eurographics 2013 - State of the Art Reports* (2013), Sbert M., Szirmay-Kalos L., (Eds.), The Eurographics Association. 1

[Ins09] Inselberg A.: *Parallel Coordinates, Visual Multidimensional Geometry and Its Applications*. Springer, 2009. 1

[KBH06] Kosara R., Bendix F., Hauser H.: Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics 12*, 4 (July 2006), 558–568. 2

[KK18] Kaur G., Karki B. B.: Bifocal parallel coordinates plot for multivariate data visualization. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: IVAPP,* (2018), INSTICC, SciTePress, pp. 176–183. 1

[Lew95] Lewis J. R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *Int. J. Hum.-Comput. Interact. 7*, 1 (Jan. 1995), 57–78. 4

[PDF14] Perin C., Dragicevic P., Fekete J.: Revisiting bertin matrices: New interactions for crafting tabular visualizations. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2082–2091. 2

[PWR04] Peng W., Ward M. O., Rundensteiner E. A.: Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proc.* (2004), INFOVIS'04, pp. 89–96. 1

[RMF14] Riehmann P., Möbus W., Froehlich B.: Visualizing food ingredients for children by utilizing glyph-based characters. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (New York, NY, USA, 2014), AVI '14, ACM, pp. 133–136. 2

[ROF12] Riehmann P., Opolka J., Froehlich B.: The product explorer: Decision making with ease. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (New York, NY, USA, 2012), AVI '12, ACM, pp. 423–432. 2

[Shn94] Shneiderman B.: Dynamic queries for visual information seeking. *IEEE Software 11* (1994), 70–77. 2

[Sii00] Siirtola H.: Direct manipulation of parallel coordinates. In *CHI '00 extended abstracts on Human factors in computing systems* (2000), CHI EA '00, pp. 119–120. 1

[SP91] Spence R., Parr M.: Cognitive assessment of alternatives. *Interacting with Computers 3*, 3 (1991), 270 – 282. 2

[ST98] Spence R., Tweedie L.: The attribute explorer: information synthesis via exploration. interacting with. *Computers 11* (1998), 137–146. 2

[TEL18] Tuor R., Evéquoz F., Lalanne D.: Parallel bubbles - evaluation of three techniques for representing mixed categorical and continuous data in parallel coordinates. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: IVAPP,* (2018), INSTICC, SciTePress, pp. 252–263. 2

[TSDS95] Tweedie L., Spence B., Dawkes H., Su H.: The influence explorer. In *Conference Companion on Human Factors in Computing Systems* (1995), CHI '95, ACM, pp. 129–130. 2

[WCG03] Wong N., Carpendale S., Greenberg S.: Edgelens: an interactive method for managing edge congestion in graphs. In *Proc.* (2003), INFOVIS'03, pp. 51–58. 1

[WLHS01] Wittenburg K., Lanning T., Heinrichs M., Stanton M.: Parallel bargrams for consumer-based information exploration and choice. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2001), UIST '01, ACM, pp. 51–60. 2

[WS92] Williamson C., Shneiderman B.: The dynamic homefinder: Evaluating dynamic queries in a real-estate information exploration system. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (1992), SIGIR '92, ACM, pp. 338–346. 2

[Yi08] Yi J. S.: *Visualized decision making: development and application of information visualization techniques to improve decision quality of nursing home choice*. Doctoral thesis, Georgia Institute of Technology, 2008. URL: https://smartech.gatech.edu/bitstream/handle/1853/24662/yi_jisoo_200808_phd.pdf. 2

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.