Tutorial on Comparative Visualization: Interactive Designs and Algorithms Depending on Data and Tasks

T. von Landesberger, K. Ballweg*
TU Darmstadt, Germany

Hans-Jörg Schulz†
Aarhus University, Denmark

Natalie Kerracher‡
Edinburgh Napier University, UK

Margit Pohl§
TU Wien, Austria

ABSTRACT

Data comparison in various domains can be effectively supported by visual analytics solutions combining interactive visualization and algorithmic analysis. The design of such solutions should match the comparison problem at hand: the input data and the task specification. This requires several choices from algorithm to visual design and interaction. Such design choices also need to consider human perception capabilities. Our tutorial presents how the differences in data and task characterizations influence visual-analytical solution designs. We will first present a conceptual framework, which defines a set of dimensions along which the comparison problem is defined. We then show how this specification influences the comparative solution design both in theory and using real world examples. Our tutorial provides visualization designers with a means to systematize domain problem analysis and to learn which algorithms, visual designs and interactions to use when, also taking into consideration human perception and cognition capabilities. The tutorial is held at a beginner to an intermediate level.

1 TUTORIAL MOTIVATION

Comparison is an important data analysis task in many domains such as biology, finance, transportation or medicine. Comparison tasks can be effectively supported by a combination of interactive visualization and algorithmic data analysis [19, 20]. The comparative systems needs to match the comparison problem at hand: input data and the task [4, 27, 31, 35]. The solutions need to suitably combine visual design, interaction and algorithmic analysis, taking into account the limited user's cognition capabilities. This is difficult, as many choices need to be made when designing comparison solutions. Which algorithms to use? Which similarity function to use? Which visualization type should be used? Which visual encoding should be used? How to support comparison by interaction? Therefore, visualization designers need a comprehensive characterization of comparison problems for developing suitable visual comparison systems. This tutorial addresses this need and offers a detailed description of comparison problems (data and tasks) together with their implications for the comparison solutions.

We will present both theoretic guidelines and practical examples of best practices. Our tutorial builds upon visualization literature on task and data characterizations, comparison visualization techniques such as [1, 4, 15, 19–21, 24, 25, 27, 35, 38] as well as our extensive experience with designing visual analytics solutions for comparative tasks across various application domains such as finance [45], medicine [42], biology [13, 23, 28], transportation [44], meteorology [43], and perception studies [9, 46].

2 TUTORIAL ORGANIZATION

The tutorial is organized along the visual comparison process. It starts with the comparison problem description as a combination of input data and task. Then, comparison operations turn input data into outputs. This operation can be done algorithmically, visually or in a combined way. The outputs are shown to the user in an interactive visualization. The visual design should include interaction and should consider user's cognitive capabilities. The planned schedule follows this overall outline:

(Introduction – ca. 10-15 minutes)

Part 1: Specification of Comparison Problems: Data and Tasks We will present a set of conceptual dimensions along which the comparison data and task can be characterized. (ca. 30 minutes, N. Kerracher + T. von Landesberger)

Part 2: Algorithmic Comparison: We will present computational methods for comparing and relating data items ranging from matching approaches, via classification methods, to clustering algorithms. (ca. 45 minutes, H.-J. Schulz)

(Coffee Break – ca. 30 minutes)

Part 3: Visual Design and Interaction: We will present types and real examples of visual designs and interaction techniques fitting to various comparison problem specifications (ca. 45–60 minutes, T. von Landesberger)

Part 4: Perception and Cognition: We will present the perceptual and cognitive mechanisms relevant for visual comparison, and underline them with insights and guidelines derived from studies on visual comparison. (ca. 30 minutes, M. Pohl + K. Ballweg)

(Wrap-Up and Q&A – ca. 10–15 minutes)

2.1 Part 1: Comparison Problems: Data and Tasks

Data and task characterization is needed to describe the comparison problem to be solved. Proper data and task characterization are a prerequisite for finding suitable visual comparison solutions. For example, financial analysts need to analyze the implications of a financial crisis by determining the change of financial institutions and their connections in a financial network after the crisis, to the reference network before the crisis, e.g., 2010 to 2005.

Therefore, we will first present important dimensions for specifying comparison problems that can be used across application domains. We will include many examples of specifications along these dimensions. In the financial example, these are: 1) the comparison purpose (i.e., what is the target output) and 2) the comparison input, on which the task is performed [4]. The comparison purpose is to determine change. This change is determined by comparing the *existence* of *nodes and edges* in the *network of 2010 to the network of 2005*. Thus, the comparison input consists of a) what is compared – i.e., nodes and edges of the two networks; b) what is compared to what – i.e., 2010 and 2005 networks; and c) according to which aspect – i.e., their existence.

We will explain each dimension in detail both conceptually and using real examples. This set of dimensions will then be used in the next sections for characterizing suitable solutions.

2.2 Part 2: Algorithmic Comparison

Data comparison is often supported by algorithmic means. In these cases, the visualization's responsibility is not so much to allow for visual comparison, but for the interactive steering of the used algorithm and for the interactive exploration of its result. We can

*e-mail: office@gris.tu-darmstadt.de
†e-mail: hjschulz@cs.au.dk
‡e-mail: N.Kerracher@napier.ac.uk
§e-mail: margit@igw.tuwien.ac.at

subdivide algorithmic comparison approaches by the cardinality of the involved data.

1-to-1 Comparison: Matching aims to map one set of data onto another set of data as closely as possible. Common instances are graph matching, pattern matching, and string matching – both exact and inexact – that have applications from de-duplication of data records [12] to detecting errors in crowd-sourced data [22].

1-to-Many Comparison: Classification is the process of assigning one data item a suitable class out of a finite number of predefined classes. Common examples are deciding for a pattern of incoming network connections whether these are malicious or not, and part of speech tagging of words in a sentence. Algorithmic means to do that are decision trees [40] and the *k*-nearest neighbor algorithm [36], for example.

Many-to-Many Comparison: Clustering compares all data items to all others to establish groups of similar data items. Clustering can be done bottom-up by successively agglomerating data items into larger groups, or top-down by successively partitioning a single large group that encompasses all data items. Clustering is used in a variety of application domains from biomedicine [26] to the analysis of time-oriented data [5].

This part will give a comprehensive overview of and introduction to these three flavors of algorithmic comparison, explain their commonalities and differences, their respective uses and outcomes, as well as potential difficulties and pitfalls in their use.

2.3 Part 3: Interactive Visualization Design

This section will present the implications of comparison problem specification (see Part 1) on the visual-interactive design of comparative visualizations both as means of data comparison and as showing comparison result of algorithmic calculation (see Part 2).

There are many comparative visualizations, see surveys [2–4, 10, 21, 24, 25, 30, 41]. The main categories independent of the input data are [19, 20]: *juxtaposition* (showing compared data in separate coordinate space), *superposition* (overlaying the data in the same coordinate space), and *explicit representation* of relationships. These basic representations are often combined or have variants such as superposition in space (e.g., overlay) or superposition in time (e.g., animation, interchangeable display). These categories are, however, not discriminative enough for providing detailed design guidelines. For example, juxtaposition has several variants, even for one data type and often require mixed types such as juxtaposition with explicit encoding. We will present more detailed discrimination of the designs according to the specification of the comparison purpose and of the comparison inputs.

Moreover, interaction supports comparative visualizations, e.g., by bringing closer the items to be compared. Specific interaction techniques have been proposed e.g., [32, 39]. The choice of the interaction technique should support the comparison goal. We will present examples of such interaction techniques.

We will first present the main categories of comparative visualizations and their concrete design specifications as implications of data and task characterization. We will present real examples of the techniques as well as discuss their advantages and disadvantages. This will provide the audience guidelines for choosing suitable visual design for the task at hand.

2.4 Part 4: Perception and Cognition

Respecting perception and cognition for visual data comparison is important because multiple studies have shown that the human similarity perception strongly diverges from mathematical measures [11, 17, 33, 34]. Furthermore, studies have shown that there are factors tricking the human perception and cognition to over-/underestimate the data items similarity [8].

We will first present relevant models explaining the perceptual and cognitive mechanism relevant for visual data comparison. Second, we present insights and guidelines resulting from studies on visual data comparison for various data types. The models in the first half build up the knowledge base to better understand the studies presented in the second half of this part. We are convinced that the tutorial attendees will be able to design improved comparative visualizations by employing this knowledge base about the relevant perception and cognition processes.

Models We will address 1-to-1, 1-to-many and many-to-many comparisons. Psychological investigations concerning similarity perception mostly address 1-to-1 comparisons [29]. We discuss several different formal approaches to model similarity perception from the literature.

Practical guidelines We present insights and guidelines resulting from studies on visual data comparison for various data types. This includes the following body of work: studies on dynamic graphs by Bach et al. [7] and Ghani et al. [18], study on perception of graph differences by Archambault et al. [6] and Bridgeman et al. [14], many-to-many comparison of graphs and scatterplots [9, 33], the influence of contours on perceiving data similarity in star glyphs [16] and evaluation of comparison tasks in various bar charts by Srinivasan et al. [37].

3 Tutorial Material and Learning Goals

The participants will be provided with tutorial notes including an extensive literature list on the subject as well as additional material such as demos and videos.

Participants can expect to gain a comprehensive overview of the state-of-the-art on comparison in data visualization. Through the selected example scenarios, systems, and techniques that are highlighted and discussed throughout the tutorial, the participants will also get a grasp on how to utilize the various concepts on the different levels in practice. Finally, the overview of the perceptual issues sets our tutorial apart from mere method-oriented courses, as this additional angle on the subject will allow participants to design and evaluate visualization solutions for comparison tasks and help to avoid common mistakes.

4 Tutorial Organizers

Tatiana von Landesberger is a group leader in Interactive Graphics Systems Group, Technische Universitt Darmstadt. She focuses on visual analytics of complex and large data in various applications. In 2017, she finished her habilitation on the topic of Visual Data Comparison. She has lead several projects developing novel techniques for visual data comparison in various applications. Together with H.-J. Schulz and D. Baur, she held a tutorial at VIS 2014 on interaction for data visualization.

Kathrin Ballweg is a PhD Student in Interactive Graphics Systems Group, Technische Universitt Darmstadt. She focuses on visual comparison of graphs in node-link diagrams.

Hans-Jörg Schulz is an associate professor at Aarhus University in Denmark. His research focuses on Visual Analytics solutions for structured data with a strong emphasis on hierarchical structures and networks. He has previously held a tutorial at VIS 2012 on "Showing Relationships in Data" with M. Streit and A. Lex, and he was involved in two VIS tutorials in 2013 and 2014 on drawing large graphs and interaction for data visualization.

Natalie Kerracher is a research fellow at Edinburgh Napier University. She completed her Ph.D. degree within the Centre for Algorithms, Visualisation and Evolving Systems in 2017. She has published widely in information visualization and visual analytics with a focus on task taxonomies and network visualization.

Margit Pohl studied psychology and computer science. She is associate professor at the Vienna University of Technology. She is leader of the Human-Centered Visualization group. Her main reserach areas are cognition and reasoning in visualization and visual analytics and Human-Computer Interaction.

References

[1] J.-w. Ahn, C. Plaisant, and B. Shneiderman. A task taxonomy for network evolution analysis. *TVCG*, 20(3):365–376, 2014.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011.

[3] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. Visualizing sets and set-typed data: State-of-the-art and future challenges. In *EuroVis – STAR Papers*, pages 1–21, 2014.

[4] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual analytics of movement*. Springer Science & Business Media, 2013.

[5] N. Andrienko and G. Andrienko. State transition graphs for semantic analysis of movement behaviours. *Information Visualization*, 17(1):41–65, 2018.

[6] D. Archambault, H. C. Purchase, and B. Pinaud. Difference map readability for dynamic graphs. In *Proc. of GraphDrawing'10*, pages 50–61. Springer, 2011.

[7] B. Bach, E. Pietriga, and J.-D. Fekete. GraphDiaries: Animated transitions and temporal navigation for dynamic networks. *TVCG*, 20(5):740–754, 2014.

[8] K. Ballweg, M. Pohl, G. Wallner, and T. von Landesberger. Visual Similarity Perception of Directed Acyclic Graphs: A Study on Influencing Factors. In *Graph Drawing and Network Visualization*, pages 241–255, Cham, 2018.

[9] K. Ballweg, M. Pohl, G. Wallner, and T. von Landesberger. Perception of animated node-link diagrams for dynamic graphs. *JGAA*, in press.

[10] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. In *EuroVis –STARs*. Eurographics, 2014.

[11] A. Bernstein, E. Kaufmann, C. Bürki, and M. Klein. How similar is it? towards personalized similarity measures in ontologies. In *Wirtschaftsinformatik 2005*, pages 1347–1366. 2005.

[12] M. Bilgic, L. Licamele, L. Getoor, and B. Shneiderman. D-Dupe: An interactive tool for entity resolution in social networks. In *Proc. of VAST'06*, pages 43–50. IEEE, 2006.

[13] S. Bremm, T. von Landesberger, M. Heß, T. Schreck, P. Weil, and K. Hamacher. Interactive visual comparison of multiple trees. In *IEEE VAST*, pages 31–40. IEEE, 2011.

[14] S. Bridgeman and R. Tamassia. Difference metrics for interactive orthogonal graph drawing algorithms. In *Graph Drawing*, pages 57–71. Springer Berlin Heidelberg, 1998.

[15] M. Burch. The dynamic graph wall: visualizing evolving graphs with multiple visual metaphors. *J. Visualization*, pages 1–9, 2016.

[16] J. Fuchs, P. Isenberg, A. Bezerianos, F. Fischer, and E. Bertini. The influence of contour on similarity perception of star glyphs. *IEEE Trans. Vis. Comput. Graphics*, 20(12):2251–2260, 2014.

[17] X. Gao, B. Xiao, D. Tao, and X. Li. A survey of graph edit distance. *Pattern Analysis and applications*, 13(1):113–129, 2010.

[18] S. Ghani, N. Elmqvist, and J. S. Yi. Perception of animated node-link diagrams for dynamic graphs. *Comput. Graph. Forum*, 31(3pt3):1205–1214, 2012.

[19] M. Gleicher. Considerations for visualizing comparison. *IEEE TVCG*, 24(1):413–423, 2017.

[20] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[21] M. Graham and J. Kennedy. A survey of multiple tree visualisation. *Information Visualization*, 9(4):235–252, 2010.

[22] M. Hascoët and P. Dragicevic. Interactive graph matching and visual comparison of graphs and clustered graphs. In *Proc. of AVI'12*, pages 522–529. ACM, 2012.

[23] M. Hess, S. Bremm, S. Weissgraeber, K. Hamacher, M. Goesele, J. Wiemeyer, and T. von Landesberger. Visual exploration of parameter influence on phylogenetic trees. *IEEE CG&A*, 34(2):48–56, 2014.

[24] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE TVCG*, 19(3):495–513, 2013.

[25] J. Kehrer, H. Piringer, W. Berger, and M. E. Gröller. A model for structure-based comparison of many categories in small-multiple displays. *IEEE TVCG*, 19(12):2287–2296, 2013.

[26] M. Kern, A. Lex, N. Gehlenborg, and C. R. Johnson. Interactive visual exploration and refinement of cluster assignments. *BMC Bioinformatics*, 18(406), 2017.

[27] N. Kerracher, J. Kennedy, and K. Chalmers. A task taxonomy for temporal graph visualisation. *IEEE TVCG*, 21(10):1160–1172, 2015.

[28] N. Kerracher, J. Kennedy, K. Chalmers, and M. Graham. Visual techniques to support exploratory analysis of temporal graph data. In *EuroVis-Short Papers*, pages 1–21, 2015.

[29] R. l. Goldstone and J. Y. Son. Similarity. In K. J. Holyoak and R. G. Morrison, editors, *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, 2005.

[30] L. McNabb and R. S. Laramee. Survey of surveys (sos)-mapping the landscape of survey papers in information visualization. In *CGF*, volume 36, pages 589–617. Wiley, 2017.

[31] T. Munzner. A nested model for visualization design and validation. *IEEE TVCG*, 15(6):921–928, 2009.

[32] M. Novotný and H. Hauser. Similarity brushing for exploring multidimensional relations. In *WSCG*, 2006.

[33] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *CHI*, pages 3659–3669, 2016.

[34] E. Pekalska and R. P. W. Duin. *The dissimilarity representation for pattern recognition: Foundations and applications*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2005.

[35] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *TVCG*, 19(12):2366–2375, 2013.

[36] C. Seifert and E. Lex. A novel visualization approach for data-mining-related classification. In *Proc. of IV'09*, pages 490–495. IEEE, 2009.

[37] A. Srinivasan, M. Brehmer, B. Lee, and S. Drucker. Whats the difference?: Evaluating variants of multi-series bar charts for visual comparison tasks. pages 1–10. ACM, January 2018.

[38] C. Tominski. CompaRing: Reducing Costs of Visual Comparison. In *EuroVis – SP*, pages 137–141. Eurographics Association, 2016.

[39] C. Tominski, C. Forsell, and J. Johansson. Interaction support for visual comparison inspired by natural behavior. *IEEE TVCG*, 18(12):2719–2728, 2012.

[40] S. van den Elzen and J. J. van Wijk. BaobabView: Interactive construction and analysis of decision trees. In *Proc. of VAST'11*, pages 151–160. IEEE, 2011.

[41] C. Vehlow, F. Beck, and D. Weiskopf. The state of the art in visualizing group structures in graphs. In *EuroVis – STARs*, pages 21–40, 2015.

[42] T. von Landesberger, D. Basgier, and M. Becker. Comparative local quality assessment for 3D medical image segmentation with focus on statistical shape model-based algorithms. *IEEE TVCG*, 2015.

[43] T. von Landesberger, S. Bremm, N. Andrienko, G. Andrienko, and M. Tekušová. Visual analytics methods for categoric spatio-temporal data. In *IEEE VAST*, pages 183–192, Oct 2012.

[44] T. von Landesberger, F. Brodkorb, and P. Roskosch. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatia-temporal graphs and clustering. *IEEE TVCG*, 22(1):11–20, 2016.

[45] T. von Landesberger, S. Diel, S. Bremm, and D. W. Fellner. Visual analysis of contagion in networks. *Information Visualization*, 14(2):93–110, 2015.

[46] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *Proc. of InfoVis'05*, pages 157–164, 2005.