Final remarks

We consider the optimization approach and the greedy heuristic as two complementary approaches, which we plan to implement and compare in experiments. Concerning the rendering of an Euler diagram based on its dual graph our hope is that we can re-use existing methods and software. A crucial question that needs further discussions, however, is how the uncertainties introduced with our simplification methods can be visualized.

References

1 Jean Flower and John Howse. Generating Euler diagrams. In *Proc. 2nd International Conference on Diagrammatic Representation and Inference (Diagrams 2002)*, pages 61–75, 2002.

2 Peter Rodgers, Leishi Zhang, and Andrew Fish. General Euler diagram generation. In *Proc. 5th International Conference on Diagrammatic Representation and Inference (Diagrams 2008)*, pages 13–27, 2008.

3 Peter Rottmann, Markus Wallinger, Annika Bonerath, Sven Gedicke, Martin Nöllenburg, and Jan-Henrik Haunert. MosaicSets: Embedding set systems into grid graphs. *IEEE Transactions on Visualisation and Computer Graphics*, 29(1):875–885, 2023.

4 Paolo Simonetto, Daniel Archambault, and Carlos Scheidegger. A simple approach for boundary improvement of Euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):678–687, 2016.

5 Paolo Simonetto and David Auber. Visualise undrawable Euler diagrams. In *Proc. 12th International Conference Information Visualisation*, pages 594–599, 2008.

6 Paolo Simonetto and David Auber. An heuristic for the construction of intersection graphs. In *Proc. 13th International Conference Information Visualisation*, pages 673–678, 2009.

7 Paolo Simonetto, David Auber, and Daniel Archambault. Fully automatic visualisation of overlapping sets. In *Proc. 11th Eurographics / IEEE – VGTC Symposium on Visualization (EuroVis'09)*, pages 967–974, 2009.

8 Youjia Zhou, Archit Rathore, Emilie Purvine, and Bei Wang. Topological simplifications of hypergraphs. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

4.3 Types of Uncertainty in Set Visualization

Christian Tominski (Universität Rostock, DE), Michael Behrisch (Utrecht University, NL), Susanne Bleisch (FH Nordwestschweiz – Muttenz, CH), Sara Irina Fabrikant (Universität Zürich, CH), Eva Mayr (Donau-Universität Krems, AT), Silvia Miksch (TU Wien, AT), and Helen C. Purchase (Monash University – Clayton, AU)

Set visualization deals with visual methods to support people understand and make sense of sets, their elements, and relations thereof. Existing methods such as Euler diagrams, Venn diagrams, and bi-partite node-link representations focus on communicating set memberships, their cardinality, and their possible intersections. However, designing visual representations of uncertain sets appears to be challenging. This is mainly due to the fact that not only the data D themselves need to be encoded visually, but also the information about their uncertainty U needs to be communicated to a reader. Above all, set visualization users must be able to extract all the encoded information (about the data and their uncertainty) from the visualization, which can be formulated abstractly as a pipeline:

$$(D, U) \xrightarrow{m} V \xrightarrow{i} (D', U').$$

The visualization designer defines a mapping m of data D and uncertainty U to create a visual representation V. Through an interpretation i of the visual representation V, human observers extract their own versions of data D' and uncertainty information U'. The scientific challenge is to understand the cognitive process of i and to devise mappings m so that ideally $D = D'$ and $U = U'$ for all human observers. The congruence of D and D', as well as U and U', can serve as a guiding principle for the visualization of uncertain data.

While set visualizations themselves are an active research frontier there are far fewer research activities in the understanding of the implications of uncertainty for set visualization. In the first place, it is still unclear how uncertainty is defined in the context of set-type data. Only if we know what types of uncertainty are relevant for set type data can we design expressive visual representations of uncertain sets. Therefore, we conceptualized uncertainty in the context of set visualization by examining (a) which aspects of set-type data might be affected by uncertainty, and (b) which characteristics of uncertainty might influence the visualization design.

Undeniably, uncertainty bears the notion of something being known, unknown, vague, and/or containing varying accuracy. So, the starting point of our discussion centered around specifying what is known and what is unknown. In a perfect world, we know the data and we assume that they are accurate. For set-type data this means that we know for certain all elements, all existing sets, and the set membership of each element. There are also associated data attributes we know with certainty, for example, set size as an important derived set attribute. There may be further data attributes given for elements or sets for which we know their data values with certainty (e.g., the number of female members of a team). Given these data characteristics (D), the visualization of set-type data is primarily concerned with communicating (i) set membership, (ii) set properties, and (iii) associated data attributes. An overview of suitable visualization methods for the cases where set characteristics are certain is available in [1].

While we might believe to know things accurately in a perfect world, in the real world, however, there is certainty about uncertainty (U) surrounding us. Just take the weather predictions, for example, and the often heard statement "There is a 70% chance of rain tomorrow" on your favorite weather app. We thus asked ourselves, how much do we actually know about data uncertainty? In a perfect world, we know that there exists no uncertainty at all, which we denote as $U = 0$. In the real world, however, one can distinguish two scenarios. First, we know that there is uncertainty, but we cannot tell accurately where it is, what it is, or how much of it exists. In other words, we know for a fact that uncertainty is present in our data, but no further details. We denote this as $U > 0$. In the second scenario, we not only know that uncertainty exists in our data, but we also know with certainty where, what, and how much of it exists in our data. For the sake of simplicity, we denote this as $U = p$. The letter p is a strong simplification of what could be known about the uncertainty in our dataset. Depending on the given data characteristics we are interested in, p can take different forms. When set membership of an element a and a set X is certain, one can say either $a \in X$ or $a \notin X$. Under uncertainty, p might denote a probability of a being a member of X, $P(a, X) = p$, which is a notation known from fuzzy sets. In this case, p can be understood as a plain probability value. Yet, we could also say that p denotes a more complex probability

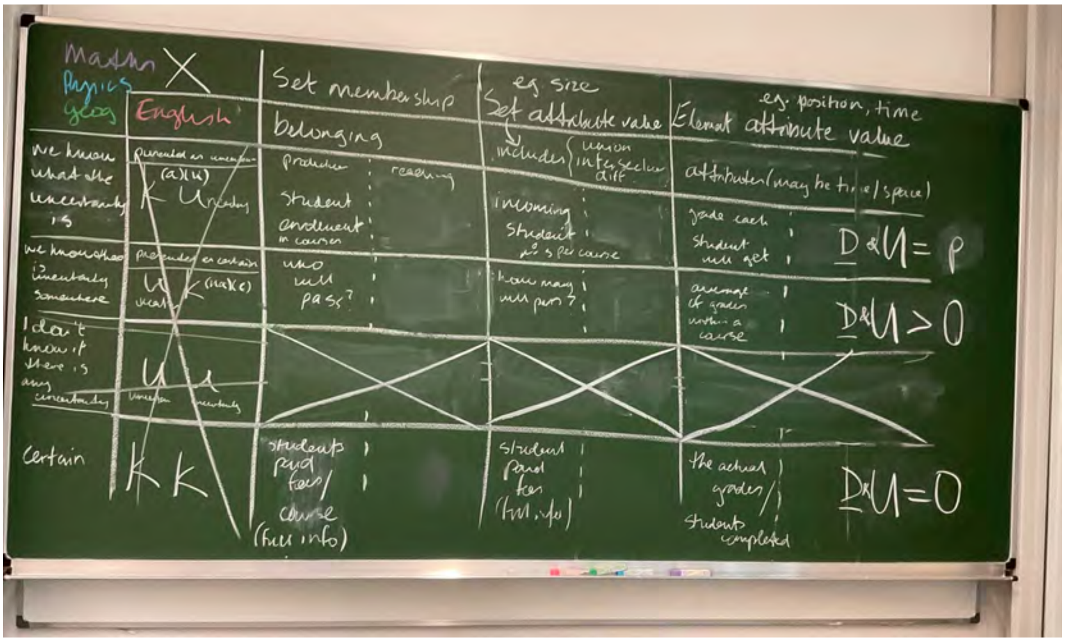

Figure 7 Conceptual framework sketched as a table with columns and rows representing data characteristics D and types of uncertainty U, respectively. The individual columns were discussed in separate subgroups.

distribution (e.g., $p = \mathcal{N}(\mu, \sigma^2)$) based on which set membership is decided. Also, in relation to the data attributes of elements or sets, we may understand p as the probability value of an attribute taking a particular data value. The same holds for the notion of p being a probability distribution. Additionally, it is common for uncertain attribute values to specify them via a range of possible values, in which case $p = [l, u]$ is some interval with a lower and upper bound of l and u.

Overall, the discussion of the characteristics of set data D and the types of uncertainty U led us to a conceptual framework of uncertainty in set visualization. In terms of D, the framework distinguishes: set membership, set attributes, and element attributes. Related to U, we use the different plausible types of (un)certainty: certainty ($U = 0$), uncertainty as a binary fact ($U > 0$), and uncertainty as quantifiable measure ($U = p$). We captured the framework in a table whose columns and rows respectively represent D and U, as shown in fig. 7.

Based on this conceptual framework, we then systematically discussed possible visualization designs to illustrate examples and highlight challenges of integrating uncertainty in set visualizations. Three subgroups were formed, each working on a selected data characteristic (i.e., table column). As a baseline, each subgroup used the simple case of a visual representation with zero uncertainty ($U = 0$). The group that dealt with set membership worked with bi-partite node-link and matrix representations, which were gradually expanded to include unknown uncertainty ($U > 0$) and known uncertainty ($U = p$) by varying the visual encoding of links and matrix cells as indicated in fig. 8.

Set attributes turned out to be particularly challenging to visualize when uncertainty is involved. The reason for this is that derived data attributes depend by definition on other data characteristics, which also can include varying levels of uncertainty. For example, set size depends on set memberships. Leaving this particular challenge for future work, the

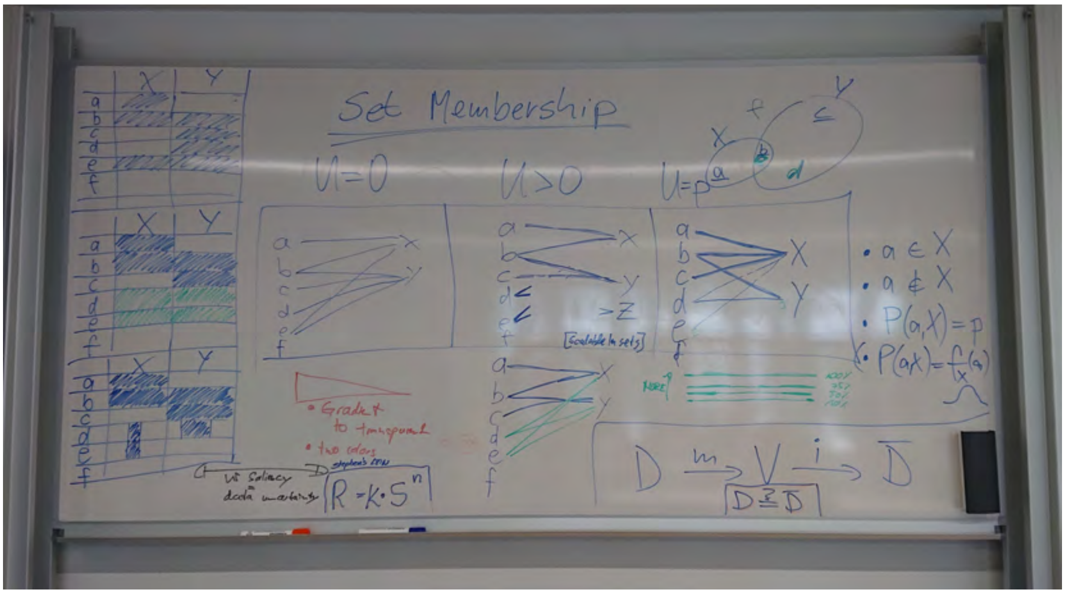

Figure 8 Integrating uncertainty into visual representations of set membership. Two alternatives were sketched: bi-partite node-link representation with uncertainty encoded on the links (center) and matrix visualization with uncertainty visualized in the matrix cells.

subgroup designed and discussed visual representations where set attributes do not depend on other factors. They came up with node-link-style representations as shown in fig. 9. Sets are represented as bigger nodes being linked to their belonging set elements, which are depicted as smaller nodes. The set attributes are shown as pie charts within the bigger nodes, where color hue indicates certain attribute values and hatching marks uncertain set elements. The same encoding is applied to the attributes of the individual set elements on the smaller nodes.

Finally, one subgroup worked on visualizing uncertain element attributes. Their focus was not so much on coming up with new designs, but to review the existing knowledge about general uncertainty visualization. For example, cartography has a long history in working with uncertain data, but also the visualization community studied this topic in detail. Particularly, the works by Alan MacEachren et al. [5, 6], Kristin Potter et al. [7, 8, 2], Amit Jena et al. [4], and Theresia Gschwandtner et al. [3] offer profound insight into how uncertain data values can be encoded visually, and to what degree humans can interpret and understand the depicted information. With these general considerations, the table of the developed conceptual framework could be filled completely. Based on the intense and productive discussions centered on the conceptual framework for set visualization and uncertainty, we drafted an outline for a journal article that will summarize key results of the research conducted at the Dagstuhl-Seminar. Our planned article will also include a synthesis of recommendations to be considered when designing visualizations for uncertain set data and an outline of future research directions.

This working group consisted of (in alphabetical order) Michael Behrisch, Susanne Bleisch, Sarah Fabrikant, Eva Mayr, Silvia Miksch, Helen Purchase, and Christian Tominski (see fig. 10). Helen Purchase headed the group. Christian Tominski drafted this report. All members of the team contributed significantly to the discussions, provided feedback and edited this report, and will be co-authors of the planned journal article.

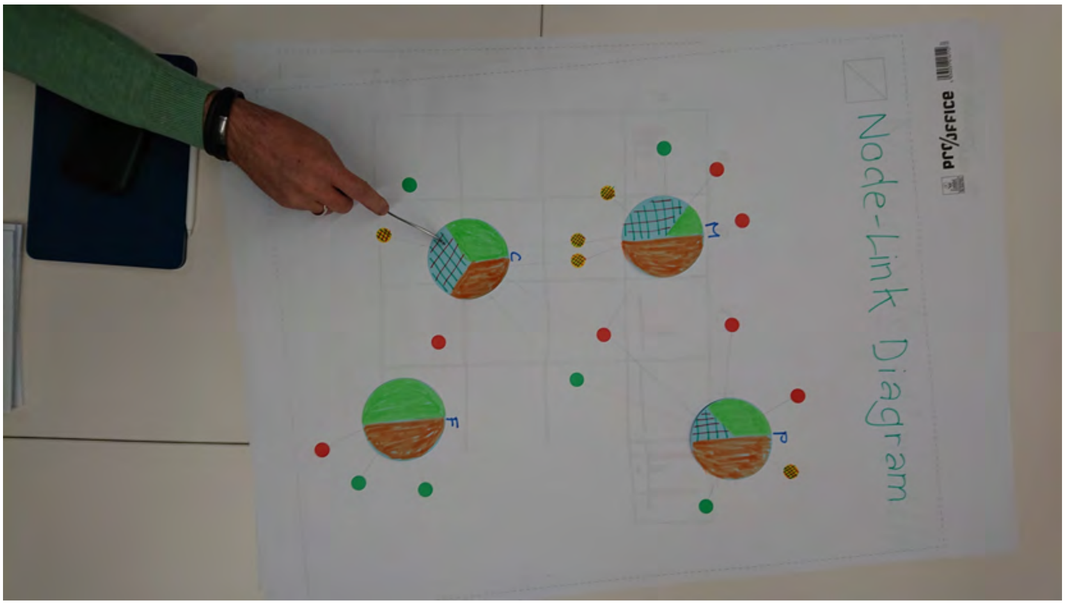

Figure 9 Node-link depiction where larger nodes visualize attributes of sets by color hue (certain) or by hatching (uncertain) within a pie chart and smaller nodes denote set elements using the same visual variables.

Figure 10 Members of the working group (from left to right): Helen Purchase (lead), Susanne Bleisch, Christian Tominski, Eva Mayr, Silvia Miksch, Sarah Fabrikant, and Michael Behrisch.

References

1 Bilal Alsallakh, Luana Micallef, Wolfgang Aigner, Helwig Hauser, Silvia Miksch, and Peter J. Rodgers. The State-of-the-Art of Set Visualization. *Comput. Graph. Forum*, 35(1):234–260, 2016. doi:10.1111/cgf.12722.

2 Georges-Pierre Bonneau, Hans-Christian Hege, Chris R. Johnson, Manuel M. Oliveira, Kristin Potter, Penny Rheingans, and Thomas Schultz. Overview and State-of-the-Art of Uncertainty Visualization. In Charles D. Hansen, Min Chen, Christopher R. Johnson, Arie E. Kaufman, and Hans Hagen, editors, *Scientific Visualization*, Mathematics and Visualization, pages 3–27. Springer, 2014. doi:10.1007/978-1-4471-6497-5_1.

3 Theresia Gschwandtner, Markus Bögl, Paolo Federico, and Silvia Miksch. Visual Encodings of Temporal Uncertainty: A Comparative User Study. *IEEE Trans. Vis. Comput. Graph.*, 22(1):539–548, 2016. doi:10.1109/TVCG.2015.2467752.

4 Amit Jena, Ulrich Engelke, Tim Dwyer, Venkatesh Raiamanickam, and Cécile Paris. Uncertainty Visualisation: An Interactive Visual Survey. In *2020 IEEE Pacific Visualization Symposium, PacificVis 2020, Tianjin, China, June 3-5, 2020*, pages 201–205. IEEE, 2020. doi:10.1109/PacificVis48177.2020.1014.

5 Alan M. MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, Robert Murray, Mark Gahegan, and Elisabeth Hetzler. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005. doi:10.1559/1523040054738936.

6 Alan M. MacEachren, Robert E. Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. Visual Semiotics & Uncertainty Visualization: An Empirical Study. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2496–2505, 2012. doi:10.1109/TVCG.2012.279.

7 Kristin Potter. *The Visualization of Uncertainty.* PhD thesis, University of Utah, USA, 2010.

8 Kristin Potter, Joe Kniss, Richard F. Riesenfeld, and Chris R. Johnson. Visualizing Summary Statistics and Uncertainty. *Comput. Graph. Forum*, 29(3):823–832, 2010. doi:10.1111/j.1467-8659.2009.01677.x.

4.4 Set Size Visualization with Dependent and Independent Uncertainties

Nathan Van Beusekom (TU Eindhoven, NL), Steven Chaplick (Maastricht University, NL), Amy Griffin (RMIT University – Melbourne, AU), Martin Krzywinski (BC Cancer Research Centre – Vancouver, CA), Marc van Kreveld (Utrecht University, NL), and Hsiang-Yun Wu (FH – St. Pölten , AT)

Problem Setting

We discuss the visualization of set sizes and their uncertainties. We distinguish between independent and dependent uncertainty in set sizes, where the latter refers to elements that are certainly present, but it is unknown to which set they belong, among two (or more) possibilities. We present three options to visualize sets sizes and their uncertainties. For each, we discuss when to use them and what their advantages and disadvantages are.