Workshop on Visualization for Communication (co-located with IEEE VIS), Berlin, Germany, October 2018

PetalVis - Floral Visualization for Communicating Set Operations

Ayush Kumar*
Stony Brook University, New York, USA
Michael Burch†
Eindhoven University of Technology
Dina Kurbanismailova
University of Reutlingen
Uwe Kloos
University of Reutlingen
Klaus Mueller‡
Stony Brook University, New York, USA

(a) (b) (c)

Figure 1: Floral visualizations with selected color coded elements: (a) Intersection of several sets. (b) Symmetric difference of several sets. (c) Union of several sets.

Abstract

Communicating set operations like intersections, unions, or symmetric differences of several sets by a visual representation is a challenging task when the number of involved sets increases. Moreover, it is difficult to provide an intuitive, readable, and understandable visual metaphor with the goal to convince the spectators about the attractiveness and easy-to-use visual depiction in order to solve a certain kind of communication task. To address the communication and exploration of a larger number of sets and operations based on them, we introduce PetalVis making use of a floral visual metaphor. Moreover, we address the perceptual limitations of traditional Venn diagrams by exploiting the visual encoding of the set operations and selections by different color schemes focusing on making the diagrams even more informative as well as more aesthetically pleasing. However, although not each individual set operation result is uniquely represented in an area, our approach can show those by color codings. We illustrate the usefulness of the novel diagram type by varying several parameters like number of sets involved or selection of several set operation results.

Index Terms: Human-centered computing—Visualization—Visualization techniques; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Communicating set operations is an important task when it comes to illustrating certain combinations of element or object groups. For example, showing the group of developers who worked at several modules in a software system together could be done by a visual depiction of the modules connected to the involved developers [3]. However, if the number of modules and developers increases, we wish to get a more scalable variant than the traditionally existing set visualizations [1].

For example, Venn diagrams were initially used by John Venn around 1880 [13, 14] and considered to be one of the most traditional methods for visualizing sets [1]. They are that intuitive and useful that they are taught to young children in school already, when it comes to explaining and communicating the combination of different sets by set operations [8].

Moreover, ever since the introduction of Venn diagrams, they were used as visual tools for representing set properties, typically drawn as closed curves in circular, elliptical, or polygonal shapes. A Venn diagram consists of multiple overlapping closed curves, usually circles, each representing a set. The curves are overlapping, building a representative area of each of the possible subset combinations out of n sets, hence showing containment, exclusion, and intersection properties.

Visually communicating these set operations and relations is very important for gaining insights into the underlying set data. On the negative side, we can find an increasingly challenge to visualize, but also to interpret Venn diagrams when the number of underlying basic sets increases while also the aesthetics is typically not well exploited to make the diagrams attractive for a larger population.

In this paper, we introduce PetalVis which is a floral visualization, additionally making use of color coding to indicate the results of combinations (see Figure 1). These aesthetically appealing diagrams are scaling to larger set numbers than traditional Venn diagrams that are typically build from 2 or 3 sets (sometimes a bit more) used in mathematical education to introduce elementary set theory to young children or even students. However, although we cannot indicate all sets and set operations in unique closed curves, highlighting and color coding might be used to enhance the visual outcomes.

*e-mail: aykumar@cs.stonybrook.edu
†e-mail: m.burch@tue.nl
‡e-mail: mueller@cs.stonybrook.edu

We illustrate our novel idea by means of varying parameters and color codings. Although we think that our approach produces aesthetically appealing diagrams we are also aware of the fact that this is still work in progress, i.e., we are trying to improve the visual encoding of the set operations in a better way, for example, by also taking into account the number of contained elements into the region sizes or color codings.

(a) (b) (c)

Figure 2: Different set operations and visual encodings: (a) Intersection of 4 sets. (b) Symmetric difference of 4 sets. (c) Union of 4 sets.

2 Related Work

Alsallakh et al. [1] surveyed set visualizations by categorizing them into several classes such as Euler and Venn diagrams [10], Euler diagram variants [9], overlays like in BubblesSets [4], node-link diagrams [12], matrix-based approaches [5], and aggregation-based techniques [2].

Our approach makes use of overlapping regions and consequently falls into the category of Euler and Venn diagrams while using differently sized and angularly oriented shapes based on flower petals, i.e., transformed elliptical shapes. By following this visual encoding strategy our approach comes close to the work of Micallef and Rodgers [6, 7] using ellipses and different color codings for the enclosed regions.

From a general perspective we see benefits of our region-based Venn diagram approach due to its reduced visual clutter effect [11] that is present in line-based diagrams to which node-link diagrams belong. Moreover, matrix- and aggregation-based techniques make it hard to identify set operations for larger set numbers due to a difficulty in communicating the correspondence between the encoded subsets. On the other hand, overlays like BubbleSets [4] demand for algorithmic concepts to compute a good layout and are perceptually difficult to communicate when consisting of a larger number of sets.

If the focus is on aesthetic and symmetric approaches we see drawbacks in the traditional techniques involving many sets, i.e., the symmetry effect is often lost and hence, also the aesthetics criterion cannot be followed that easily anymore. In particular, in line-based diagrams many crossing lines and hairball effects can lead to a bad quality and the aesthetics suffers from that. Consequently, we follow the visual encoding of a number of sets in a natural floral visualization enhanced by color codings.

3 Visual Design of PetalVis

We base our visual depiction of the sets and their set operations on similar concepts like in Venn diagrams. In those diagrams each set is represented by a region inside a closed curve. We also follow this principle and try to focus on communication of set relationships by providing an aesthetical representation based on a floral visual metaphor.

3.1 Design Criteria

To reach our goal, we follow the principles of readability, easy selections, and aesthetics.

- **Readability:** We assume that flowers are easy to read and intuitive, i.e., the understanding and communication of set operations becomes an easier task than if a more traditional way of visualization is used.

- **Selection:** Using region-based representations for the result of the set operations produces representative areas that can be interacted with more easily than line-based diagrams consisting of lots of overlaps and visual clutter [11]. This means they can, for example, be selected by mouse or eye gaze-based interactions.

- **Aesthetics:** Following patterns from nature to make diagrams more aesthetically appealing is a good way to make them more attractive for a larger population of users. Following this principle we make use of a floral visual metaphor combined with color codings and differently large petal shapes.

Transforming a number of n sets into such a region-based set visualization is easily done by placing individual petals either clockwise or counter-clockwise around a circle center. Turning them by a certain angle in a well-defined angular direction and following principles of symmetry while also varying the petal sizes leads to the final outcome of the PetalVis (see Figure 2).

3.2 Communication Patterns

When communicating information to human observers it is important to understand the visual patterns that a diagram is able to reflect. In our case of PetalVis we focus on several set-related visual patterns that region-based set diagrams support.

- **Subset pattern** $A \subset B$**:** A subset can be detected by inspecting enclosed regions and subregions. The enclosed subregion representing set A is a subset of the larger region representing set B.

- **Intersection pattern** $A \cap B$**:** Overlapping subregions indicate a combination of them and hence, reflect intersections. To understand which subsets are involved we have to follow the surrounding lines to complete the representing sets.

- **Union pattern** $A \cup B$**:** Combinations of subregions, i.e., the summed up regions build the representative regions for unions. Those can be observed by inspecting all involved representative regions in the combined region.

- **Complement pattern** $A \setminus B$**:** Inspecting a region without an overlapping subregion gives a representative subregion of the complement, i.e., all elements located in A but not in B.

- **Number of involved sets:** If more sets are involved in an intersection, this can be reflected in the color coding of the corresponding subregions. This visual effect perceptually strengthens the communication of the number of set pattern.

It may be noted that a communication of visual patterns can be even more efficient if it is supported by interaction techniques, for example, by selecting and highlighting certain representative subregions. Extending such a region-based diagram to an area-proportional shape diagram, i.e., a scaled Venn diagram is generally possible, but has an impact on the general shape and hence, visual aesthetics and symmetries.

Figure 3: Selecting an element and highlighting the involved subsets.

3.3 Interactions

At the moment we have included a few interactions with the goal to further get insights from the otherwise static plot (see Figure 3). In our perspective, it is important to start with an overview visualization to communicate the general idea of the set combinations, but after this first impression one should be able to ask dynamically for more insights:

- **Number of involved sets:** Changing the number of involved sets has an impact on the appearance of the visualization. It may be noted that the enclosed regions will become smaller and smaller the more sets are involved. This phenomenon is also the case for traditional Venn diagrams.
- **Selection of diagram regions:** Apart from inspecting the static diagram, the user can click on regions to, for example, see which sets are involved. This interaction technique makes the understanding and communication of set operations much easier.
- **Color coding:** Indicating the number of involved sets can be visually encoded in the color coding of each individual region. This visualization strategy makes the diagrams more aesthetically appealing following a similar principle as in nature in which petals are differently shaded due to different light conditions.

3.4 Varying Parameters

There is a list of parameters that can be varied like the number of sets, the set operations, and the color codings indicating set operations or supporting the communication of the visual patterns.

- **Number of involved sets:** Increasing the number of sets leads to many subset regions, too many to easily communicate the corresponding visual patterns. In such a situation, interaction techniques are important means to understand the set operations. Figure 2 shows three operation scenarios of 4 sets in which color coding is already used for indicating the differences and for enhancing the visual appearance and aesthetics. If the number of sets is much larger we end up in situations like the three ones reflected in Figure 4.
- **Set operations:** Figures 1, 2, and 4 each illustrate three different set operation scenarios which are the intersection, the symmetric difference, and the union of either 4 or many sets. As we can see the set operations are supported by color codings to further indicate the corresponding outcomes of the operations.
- **Color coding:** Color coding is useful to indicate the set overlap number (see Figure 1), e.g., by changes in color hue reflected in the greenish colors in Figure 1.

3.5 Discussion and Limitations

Although we generated an aesthetically appealing diagram type for representing set operations we are aware of the fact that our approach has several limitations.

For example, if too many sets are combined this can result in very tiny regions that may be difficult to select by mouse interaction. Hence, a more dynamic approach with a zooming lens would be suitable to first look into such small regions.

Moreover, the tracking of individual region separation lines becomes more and more difficult if more sets are involved. This is a general drawback of these region based diagrams, although this task is important to reliably solve the task of detecting which subsets are meant and of which sets and set operations they consist.

4 Conclusion

In this paper, we have introduced a non-traditional visualization technique of Venn diagrams resembling flower petals to perform set operations such as unions, intersections, or symmetric differences. In our visualization we can visually communicate any number of sets and naively interact with them in order to communicate the underlying set operations and combinations by highlighting certain selected subregions. Color codings are used to visually enhance the set combinations allowing an easier tracking of the corresponding sets. For future work we are planning to add more interaction techniques like clicking on a certain region and then showing the combinations of all the involved sets and set operations in a common expression. Moreover, we will extend the color coding by also showing the number of involved set elements in each of the displayed regions. For evaluation purposes we will conduct an eye tracking experiment to figure out for which set number human observers get problems to interpret the diagrams and where the visual attention was paid over time.

Acknowledgments

This research was supported by NSF grant IIS 1527200 & MSIT (Ministry of Science and ICT), Korea, under the ICTCCP (IITP-2017-R0346-16-1007) supervised by the IITP.

References

[1] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. J. Rodgers. The state-of-the-art of set visualization. *Computer Graphics Forum*, 35(1):234–260, 2016.

[2] F. Bendix, R. Kosara, and H. Hauser. Parallel sets: Visual analysis of categorical data. In *IEEE Symposium on Information Visualization*, p. 18, 2005.

[3] M. Burch, T. Munz, F. Beck, and D. Weiskopf. Visualizing work processes in software engineering with developer rivers. In *3rd IEEE Working Conference on Software Visualization, VISSOFT*, pp. 116–124, 2015.

[4] C. Collins, G. Penn, and M. S. T. Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1009–1016, 2009.

[5] B. H. Kim, B. Lee, and J. Seo. Visualizing set concordance with permutation matrices and fan diagrams. *Interacting with Computers*, 19(5-6):630–643, 2007.

[6] L. Micallef and P. Rodgers. Computing the region areas of euler diagrams drawn with three ellipses. In *Joint Proceedings of the Fourth International Workshop on Euler Diagrams and the First International Workshop on Graph Visualization in Practice co-located with Diagrams*, pp. 1–15, 2014.

[7] L. Micallef and P. Rodgers. eulerAPE: Drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One*, 9(7), 2014.

[8] J. E. Moore. What makes these mammals? The art of sorting: Using Venn diagrams to learn science process skills. *Journal of Science Activities: Classroom Projects and Curriculum Ideas*, 39(4), 2003.

(a) (b) (c)

Figure 4: Set operations and visual encodings of many more sets: (a) Intersection of many sets. (b) Symmetric difference of many sets. (c) Union of many sets.

[9] N. H. Riche and T. Dwyer. Untangling Euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1090–1099, 2010.

[10] P. Rodgers, L. Zhang, and A. Fish. General Euler diagram generation. In *Diagrammatic Representation and Inference, 5th International Conference*, pp. 13–27, 2008.

[11] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: a measure of display clutter. In *Proceedings of the 2005 Conference on Human Factors in Computing Systems*, pp. 761–770, 2005.

[12] J. T. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[13] J. Venn. On the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(59):1–18, 1880.

[14] J. Venn. On the employment of geometrical diagrams for the sensible representations of logical propositions. In *In Proceedings of the Cambridge Philosophical Society*, vol. 4, pp. 47–59, 1880.