EuroVis Workshop on Visual Analytics (2020)
K. Vrotsou and C. Turkay (Editors)

Dual Radial Set

Krešimir Matković[1] Denis Gračanin[2], Matea Bardun[3], Rainer Splechtna[1], and Helwig Hauser[4]

[1]VRVis Research Center, Vienna, Austria; [2]Virginia Tech, Blacksburg, Virginia, USA
[3]AVL-AST d.o.o, Zagreb, Croatia; [4]University of Bergen, Bergen, Norway

Abstract

Set-typed data visualizations require novel interactive representations, especially when visualizing multiple set-typed data attributes. The challenge is how to effectively analyze relations between data elements from different set-typed attributes. We build on Radial Set view to support simultaneous visualization of two set-typed attributes. The main contributions include: Dual Radial Set view that supports simultaneous visualization of two groups of sets; an extension of Radial Set view that can display empty sets; and two new view configurations, the equal sector size and the relative-size scaled sectors. The two new view configurations also can be applied to the original Radial Set view. We conducted an informal evaluation using a movies data set as a case study. The evaluation results demonstrate the advantages of the proposed approach.

CCS Concepts

• ***Human-centered computing*** → ***Visualization techniques; Visualization toolkits;***

1. Introduction

The data sets explored in visualization, statistics or data science usually contain numerical (discrete and continuous) and categorical (ordinal and nominal) data. However, when analyzing data, analysts often need to represent data as time series, graphs/networks, or sets. The set-typed data are traditionally visualized using Euler [EHdCmdC02] or Venn [M.A80] diagrams. Such representations work well for relatively small number of sets. For more complex analysis and for a larger number of sets, visualization community has developed several novel, interactive representations.

We extend Radial Set view (Alsallakh et al. [AAMH13]) to simultaneously visualize two set-typed attributes and analyze cross set relations between elements of different sets. We use as an example a movies data set where each movie is classified based on multiple genres and awards. Representing the genres and awards attributes as set-typed attributes, allow us to explore correlations between specific sets of genres and sets of awards. Such exploration is not possible using the state of the art techniques. Arguably, we could use the same technique twice in a coordinated multiple views setup. However, an integrative view has many advantages. The proposed view integrates dual sets and shows correlations in a more intuitive way. We also extend the original Radial Set view to include empty sets — a concept regularly used in mathematics and, interestingly, often neglected or ignored in many visualization techniques. Finally, we also describe two view configurations, the relative-size scaled view and the equal sector-size view. These configurations are also applicable to the original Radial Set view to better suit analysis tasks.

The main contributions of this paper can be summarized as follows: **(1)** Dual Radial Set view, a new view that supports simultaneous visualization of two set-typed attributes. **(2)** An extension of Radial Set view that supports depicting of empty sets, and **(3)** Two configurations of Dual Radial Set view, the relative-size scaled sector sizes configuration, and the equal sector-sizes configuration. The two new configurations can be used for Radial Set view, too. These three contributions make it possible to analyze complex cross-set relations in en efficient and intuitive way. We informally evaluate the contributions using a well known movies data set [IMD].

2. Related Work

The set-typed data visualization is designed to effectively analyze relations between data elements and attributes from different sets when the number of sets is large. Traditional approaches, such as Euler [EHdCmdC02] and Venn [M.A80] diagrams, do not scale well. A complex collection of intersecting sets can be simplified as a strict hierarchy that is easier to manage and visualize [RD10].

Set'o'gram [FMH08] uses a frequency-based representation of set-typed data, in a way related to mosaic plots [HK81], histograms, and pixel-based approaches (liner visual structures). This approach provides a middle-ground between a high-level set representation and a low-level, individual element representation. Radial Set view [AAMH13] is based on the same idea but uses a radial layout to analyze set memberships for large sets. Frequency-based representations support fast search for and analysis of overlaps between the sets, as well as related overlaps for other attributes.

UpSet [LGS*14] uses a matrix layout to represent associated

DOI: 10.2312/eurova.20201080

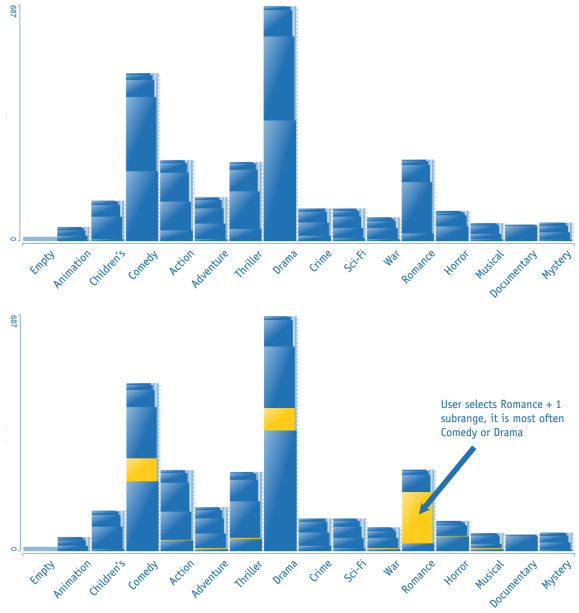

Figure 1: *An example of set'o'gram [FMH08] view displaying movie genres. **(Top)** The view shows a distribution of movie genres. Most of the movies belong to the drama genre. Since a movie can belong to more than one genre, the sub-sectors in each bin show how many movies belong to a single genre (the lowest, full width sub-sector), the genre plus one more, plus two more, etc. Some movie genres appear more often alone (drama) than others (e.g., romance). **(Bottom)** The user hovers over a sub-sector (romance plus one) and determines that drama and comedy genres are the most frequent plus one genres.*

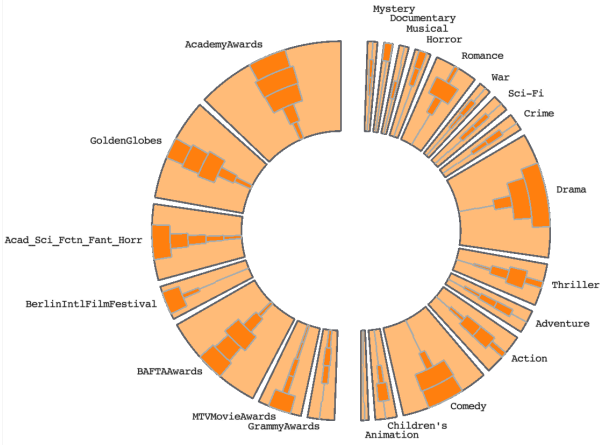

Figure 2: *The proposed Dual Radial Set view makes it possible to depict two set-typed attributes simultaneously. The view shows movie awards on the left and movie genres on the right. The idea is similar to the set'o'gram idea. The arc lengths of annulus sectors show the frequency of each genre (compare with bin heights in Figure 1). The bins in each sector show frequencies of genre or award alone, then plus one, plus two, etc. The user can configure the number of bins.*

data, including the number of elements in the aggregates (based on groupings and queries) and intersections. PowerSet [AR17] uses treemaps [Shn09] for a scalable, comprehensive overview of non-empty intersections in a collection of sets. OnSet [SMDS14] provides a specialized techniques for interactive, scalable visualization of large-scale binary set data. Alsallakh et al. provide a systematic overview of state-of-the-art techniques for visualizing set relations [AMA*14]. We build on Radial Set and set'o'gram to support simultaneous analysis of two set-typed attributes simultaneously.

3. Set Typed Data and Multiple Sets

It is natural to group together similar objects and to show relations between the groups (sets), such as union or intersection. When dealing with various data driven tasks in everyday life we often think in terms of sets. For example, if we want to classify fruits based on their colors, we can create sets of different colors. Oranges belong to the orange color set but apples belong to multiple color sets. If we further classify fruits based on vitamins that fruits contain, we can extend our sets or we can add another group of sets, vitamins (vitamin A, vitamin B, vitamin C, etc.). An orange, now belongs to the orange color set, as well as to the vitamin B and vitamin C sets. Splitting the sets into these two groups seems logical and intuitive, something we often to when dealing with set-typed data. Interestingly, the set-typed data visualization neglects this grouping of sets and focuses on a single group of sets only.

We use the movies data set which contains 1590 movies [IMD]. Movie attributes include year, IMDb rating (1 to 5), number of awards, and number of genres. In addition, there are two groups of sets, genres and awards. Each movie can belong to sets from the genres group (comedy, war, drama, ...), and to sets from the awards group (academy award, golden bear, ...). We can study relation between the two groups of sets. For example, we can examine if movies with an academy award and two more awards are more likely to be of a certain genre(s). Answering such questions is naturally supported by proposed Dual Radial Set view.

4. Dual Radial Set

The basic visualization idea in Dual Radial Set view builds on the set visualization method introduced by Freiler et al. [FMH08]. The main goal is to show how many elements each set has, as well as to show how many elements belong to other sets at the same time. Figure 1a shows set'o'gram for movie genres of the movies data set. The overall height of each bin shows how many movies belong to each genre. We see that drama genre is the largest set, followed by comedy and romance genres. There are some members of the empty set — movies with no genre specified.

The subdivision of bins shows how many movies have single genre (the widest subdivision in the bottom of each bin), how many have a genre plus one more, plus two more, etc. While drama and comedy genres have While a relatively large portion of movies with single genre are drama and comedy movies, very few of single

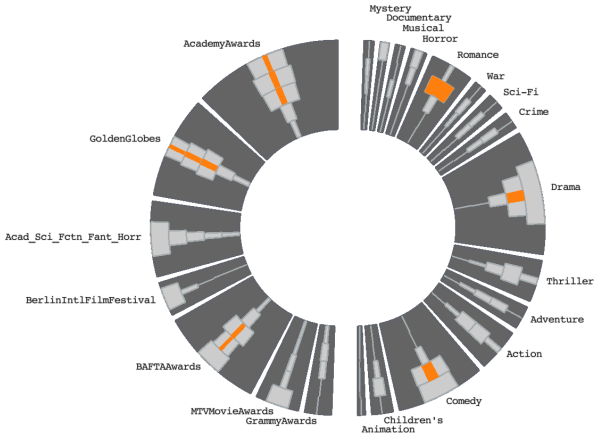

Figure 3: *The user clicks on the romance plus one genre movies and sees that the plus one genre is most often drama or comedy genre for romance movies. The romance plus one genre movies received Academy, Golden Globe, and BAFTA awards.*

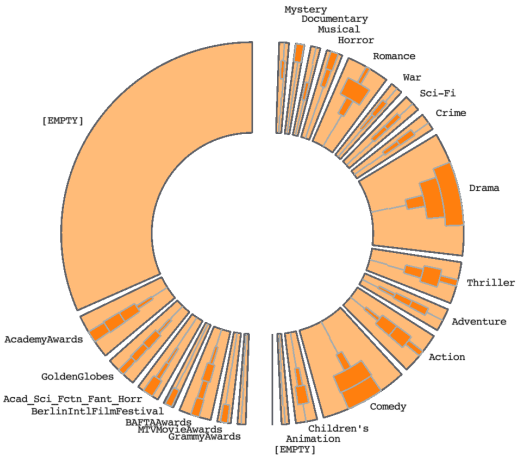

Figure 4: *Dual Radial Set view supports visualization of empty sets. We see that more than half of the movies did not get an award, and that there are some movies without assigned genre.*

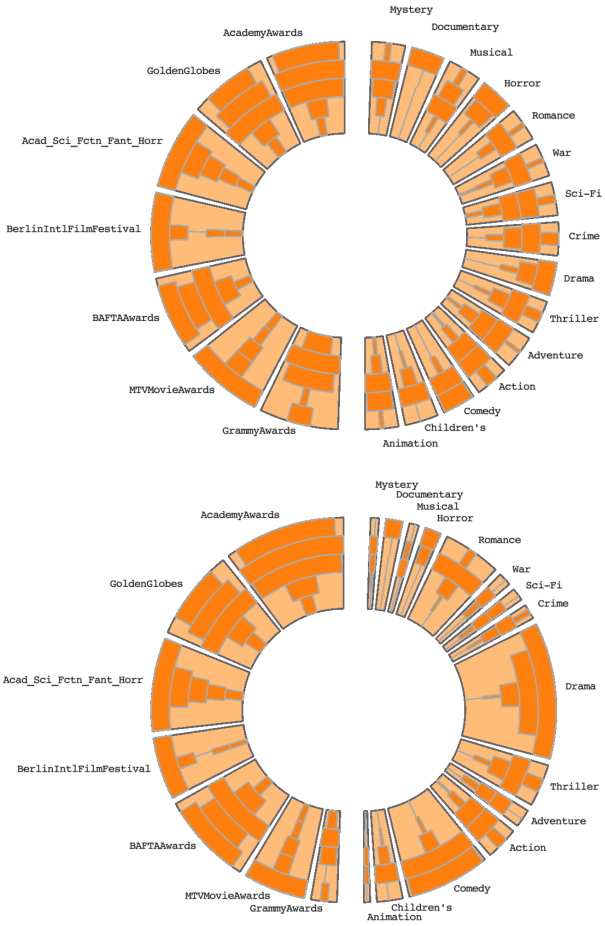

Figure 5: *Two display configurations support additional analysis tasks.* ***(top)*** *The equal sector size configuration depicts all sectors with the same arc length. In this way it is easier to explore relations within a sector.* ***(bottom)*** *The relative-size scaled configuration has the arc lengths scaled according to the longest sub-sector, not according to the total count of movies in a sector. If there is an outlier sub-sector with many movies, the whole sector is highlighted.*

genre movies are romances. If we would like to see the genres that are the most often combined with romance genre (movies with two genres), we can simply hover over the second subdivision, and related second subdivision parts from other genres are highlighted. Figure 1 shows that if romance movie has one more genre, it is most often drama or comedy genre. There are some other genres (yellow stripes), but the number of corresponding movies is small.

In addition to genres, movies also have the movie awards attribute. Instead of using two set'o'grams, or two Radial Set views, we introduce a new view, Dual Radial Set view. Figure 2 shows the new view. The right half of the circle shows movie genres. It shows the same information as set'o'gram in Figure 1. The arc length of each annulus sector corresponds to the number of movies. We see that drama genre has the longest arc followed by comedy and romance genres. The sub-sectors in each sector show the number of movies having single genre, two genres, three genres, and so on. Dual Radial Set view in Figure 2 is configured so that the innermost sector shows all movies with five or more genres. The left part of the view shows the awards. If we click on the romance plus one genre sector (Figure 3) we see the most frequent combinations (drama and comedy genres), just as in set'o'gram. In addition, we see the awards distribution for these movies on the left. The selected movies have Academy Awards, Golden Globes, and BAFTA Awards in various combinations.

Dual Radial Set view supports empty sets. Figure 4 shows empty sets for movie genres and awards. As expected, in the data set there are many more movies without an award than without a genre.

Figure 6: ***(a)*** *A possible configuration from an analysis session.* ***(b)*** *Dual Radial Set view for the same data as in* ***(a)*** *except that the single genre movies are brushed.* ***(c)*** *Movies with at least one award brushed, just a few single genre movies got an award.*

Finally, besides the standard configuration (with and without empty sets), Dual Radial Set view also supports different sector scaling possibilities tailored for specific analysis tasks. If the user wants to compare what is happening in each sector, the standard configuration gives much better insight into large sectors. In order to support task of individual sector exploration, we added the equal sector size configuration. Figure 5 top shows this configuration. All sectors have equally sized arcs. The overall distribution is not shown but a single sector exploration is much easier now. It is also possible to scale sectors based on the longest sub-sector (relative-size scaled configuration), and not on the largest total count of movies in a sector (Figure 5 bottom). Note how the drama sector becomes larger due to a single large sub-sector, and the romance sector becomes narrower due to two large sub-sectors.

5. Informal Evaluation

We have integrated Dual Radial Set view in a coordinated multiple views system [MFGH08] to show brushed data from other views, and to to create brushes which are then composed using Boolean operations. Besides Dual Radial Set view, we deploy standard views such as histogram or scatter plot. We also compute descriptive statistics and display them using box-plots and statistics-enhanced scatter plot, as proposed by Radoš et al. [RSM*16]. Figure 6a shows a possible configuration. The views in the left column show additional attributes: year is depicted in the top histogram, number of genres in the middle histogram, and rating and awards count are shown in the scatter plot in the bottom of the left part of the screen. Dual Radial Set view is positioned centrally. On the right, a box plot with statistics of three numerical attributes, and a scatter plot with statistics of brushed data are shown.

The user brushes movies with high ratings in the lower left scatter plot and refines the brush by selecting movies with two genres in the histogram. The table in the lower part of the screen shows all details for the selected movies. It is easy to move the brush and select only single genre movies (Figure 6b). The awards changed, there are no Grammy awards, for single genre movies. Finally, we brush the movies with at least one award (Figure 6c). Only drama and comedy genres got at least one award as single genre movies (all other outer sub-sectors are empty).

6. Conclusions and Future Work

Dealing with sets is practically omnipresent in today's data driven world. In this paper we introduce a novel view to visualize two set-typed attributes simultaneously. By depicting two attributes simultaneously, our approach makes it possible to explore and analyze cross set relations. This is just the beginning of our research and we foresee many possible improvements. First, inclusion of more than two groups is a logical next step. Secondly, we plan to formally evaluate our contributions with a user study. Thirdly, we have identified already several application domains that can benefit from cross set analysis (e.g., medicine and nutrition). We plan to consult with the domain experts and provide novel solutions according to their needs. Finally, Dual Radial Set view builds on Radial Set view and the set'o'gram paradigm. We will research how other approaches in set visualization can be extended to support multiple set-typed attributes. We also plan to improve interactions and use the free space in the middle of Dual Radial Set view to display additional information.

7. Acknowledgments

We would like to thank Michael Beham from VRVis for numerous discussions and his valuable feedback. VRVis is funded by BMK, BMDW, Styria, SFG and Vienna Business Agency in the scope of COMET - Competence Centers for Excellent Technologies (854174) which is managed by FFG. This work was supported, in part, by a grant from the Virginia Tech Institute for Creativity, Arts, and Technology.

References

[AAMH13] ALSALLAKH B., AIGNER W., MIKSCH S., HAUSER H.: Radial sets: Interactive visual analysis of large overlapping sets. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis) 19* (12/2013 2013), 2496–2505. URL: http://publik.tuwien.ac.at/files/PubDat_219617.pdf, doi:10.1109/TVCG.2013.184. 1

[AMA*14] ALSALLAKH B., MICALLEF L., AIGNER W., HAUSER H., MIKSCH S., RODGERS P.: Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. In *EuroVis - STARs* (2014), Borgo R., Maciejewski R., Viola I., (Eds.), The Eurographics Association. doi:10.2312/eurovisstar.20141170. 2

[AR17] ALSALLAKH B., REN L.: Powerset: A comprehensive visualization of set intersections. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (Jan 2017), 361–370. doi:10.1109/TVCG.2016.2598496. 2

[EHdCmdC02] EULER L., HUNTER H., DE CARITAT MARQUIS DE CONDORCET J.: *Letters of Euler to a German Princess, on Different Subjects in Physics and Philosophy*. No. Bd. 1 in Letters of Euler to a German Princess, on Different Subjects in Physics and Philosophy. Murray and Highley, J. Cutrell, Vernor and Hood, Longman and Rees, Wynn and Scholey, G. Cawthorn, J. Harding, and J. Mawman, 1802. URL: https://books.google.at/books?id=_1oIAAAAIAAJ. 1

[FMH08] FREILER W., MATKOVIC K., HAUSER H.: Interactive visual analysis of set-typed data. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov 2008), 1340–1347. doi:10.1109/TVCG.2008.144. 1, 2

[HK81] HARTIGAN J. A., KLEINER B.: Mosaics for contingency tables. In *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface* (New York, NY, 1981), Eddy W. F., (Ed.), Springer US, pp. 268–273. 1

[IMD] IMDB: IMDb datasets. https://www.imdb.com/interfaces/. [Last accessed 28 February 2020]. 1, 2

[LGS*14] LEX A., GEHLENBORG N., STROBELT H., VUILLEMOT R., PFISTER H.: Upset: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 1983–1992. doi:10.1109/TVCG.2014.2346248. 1

[M.A80] M.A. J. V.: I. on the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 10*, 59 (1880), 1–18. URL: https://doi.org/10.1080/14786448008626877, arXiv:https://doi.org/10.1080/14786448008626877, doi:10.1080/14786448008626877. 1

[MFGH08] MATKOVIĆ K., FREILER W., GRAČANIN D., HAUSER H.: ComVis: A coordinated multiple views system for prototyping new visualization technology. In *Proceedings of the 12th International Conference on Information Visualisation (IV '08)* (9-11 July 2008), pp. 215–220. 4

[RD10] RICHE N. H., DWYER T.: Untangling euler diagrams. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov 2010), 1090–1099. doi:10.1109/TVCG.2010.210. 1

[RSM*16] RADOŠ S., SPLECHTNA R., MATKOVIĆ K., ĐURAS M., GRÖLLER E., HAUSER H.: Towards quantitative visual analytics with structured brushing and linked statistics. *Computer Graphics Forum 35*, 3 (2016), 251–260. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12901, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12901, doi:10.1111/cgf.12901. 4

[Shn09] SHNEIDERMAN B.: Treemaps for space-constrained visualization of hierarchies. Web document, 2009. URL: http://www.cs.umd.edu/hcil/treemap-history/index.shtml. 2

[SMDS14] SADANA R., MAJOR T., DOVE A., STASKO J.: Onset: A visualization technique for large-scale binary set data. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 1993–2002. doi:10.1109/TVCG.2014.2346249. 2