Review of *UpSet: Visualization of Intersecting Sets*

Grace Yi Chen

This article introduced in detail about a new intersecting sets visualization tool called UpSet, which visualizes set intersections in a matrix layout and could do intersections aggregations, sorting and queries. A set is a collection of distinct elements that share some common characteristics. It is very interesting to analyze how these meaningful characteristics co-occur in a dataset. Euler and Venn diagrams are the most common visualization method for sets and intersections. However, they are inefficient to use when the number of sets is large. UpSet could be appropriate to use here because it could visualize both the attributes and sets intersections together. UpSet utilizes the exclusive intersections and it decomposes the sets into all possible set intersections. Then, it allows users to analyze these intersections individually or as aggregates using symbols and colors. The authors also validated UpSet based on real-life cases and demonstrate its advantages.

I think the authors explained the advantages of UpSet in visualizing and exploring set intersections well. Although I am not familiar with set visualization tools, I get an idea about this new design and know how it could be applied in real world cases. Set visualization is related to multi-dimensional data visualization. UpSet is a visualization tool with a matrix layout and the columns of this matrix correspond to the sets and the rows correspond to intersections. UpSet could do queries in set space and element space, sorting by degree or cardinality and highlight by clicking the cardinality bars. Users can explore and know the number of elements in the aggregates and intersections, as well as additional summary statistics from subset using scatter plots and histograms. The authors showed one application of using UpSet in comparing the performance of several different tools identifying SNV. I think UpSet could be used to find the intersection of DE genes in different diseases and see if there are any common genes that could be targeted for common treatment.

Questions:

1. The authors mentioned that Figure 1 shows, that action movies are watched more often on average than comedies. However, I see the cardinality number for Comedy is larger than Action movies. I am wondering whether I understand it correctly of the definition of cardinality?
2. In Figure 11 (a), the purple tick in the fourth row is very tiny and hard to identify. I am wondering is there any better ways to identify this row?