Aggregating Hypergraphs by Node Attributes

David Trye[0000−0001−7626−3251], Mark Apperley[0000−0003−1588−1595], and David Bainbridge[0000−0002−0603−5728]

University of Waikato, Hamilton, New Zealand
dgt12@students.waikato.ac.nz, {mapperle,davidb}@waikato.ac.nz

PAOHVis [2, 5] displays hypergraphs [1, 3] in a matrix where rows represent nodes (dots) and columns represent hyperedges (vertical lines). We propose extensions to PAOHVis for leveraging repeated hyperedges in non-simple hypergraphs, and displaying multiple node attributes. This is accomplished through two aggregation functions: *count-based*, which targets low-level detail, and *binary*, for high-level overview. In doing so, we introduce a domain-agnostic framework for consolidating hypergraphs by one or more categorical node attributes.

Preliminary results indicate that these enhancements provide a clearer picture of overall patterns and distributions of hypergraph data. Consider Fig. 1, which illustrates the different aggregation levels applied to a fictional co-authorship dataset. There are 12 nodes (people) and 17 hyperedges (papers), 13 of which are distinct. Nodes are coloured and subsequently consolidated by the gender of the author. The legend summarises node/category frequencies for the respective hypergraph. Additional categorical node attributes (e.g. affiliation, position and field) can be displayed and aggregated at the same time, provided these are mapped to different visual channels (e.g. shape, outline and texture). Unless they are strongly correlated, aggregating a larger number of attributes greatly reduces the number of identical hyperedges, resulting in a less compact visualisation. Thus, it may be more fruitful to aggregate hypergraphs by each attribute in turn, rather than attempting to visualise all attributes at once.

Count-Based Aggregation. This kind of aggregation shows, for each hyperedge, the exact number of nodes per category. Hyperedges with the corresponding number of nodes in each category (e.g. all papers authored by exactly two men and one woman) are combined. The original size of each hyperedge is preserved and nodes are stacked as tightly as possible, from the top row downwards, in descending order of overall category frequency. This layout facilitates comparisons of hyperedge size, which can be difficult to assess in non-aggregated hypergraphs. The original nodes (people) can no longer be reliably identified, since the same node in a repeated hyperedge may represent a different person across separate instances.

Count-based aggregation is useful for tasks relating to category frequency and overall set size. The middle panel of Fig. 1. shows that all papers have between two and four authors, which was not so apparent in the non-aggregated chart (left panel), due to the different line lengths. It is also easier to see that papers tend to have more male than female authors, but that the paper with the most authors of the same gender is written by four women (and no men).

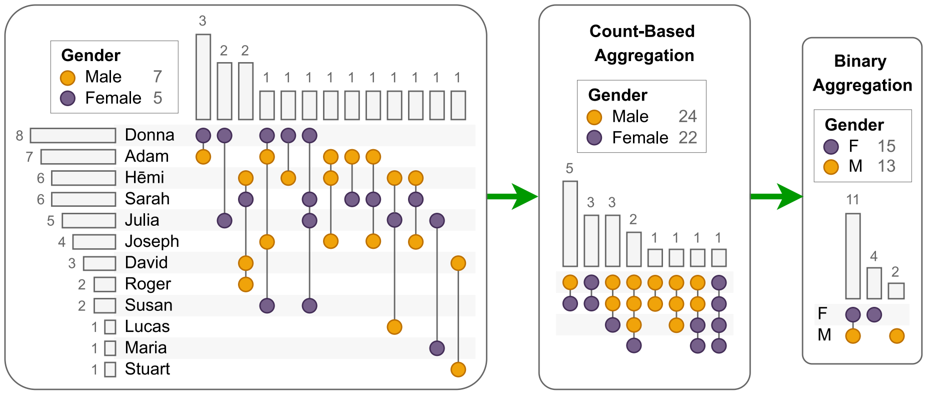

Fig. 1. Different levels of aggregation for a single node attribute (gender).

Binary Aggregation. Hypergraphs can be further aggregated by collapsing each category with multiple occurrences in a hyperedge into a single node. The bar chart then shows the number of hyperedges that contain at least one node from *precisely* the corresponding categories. While this has been partially implemented in PAOHVis, it is not currently possible to consolidate identical hyperedges, which is essential for obtaining a quick overview of hypergraphs that are very dense, especially since (certain) hyperedges are likely to elicit higher counts, given the smaller number of possible category combinations. If an attribute has more than two categories, the data can be aggregated even further, so that all hyperedges are *flattened* into pairwise combinations.

Binary aggregation helps analysts to see how many distinct categories tend to occur in a hyperedge (e.g. do all categories occur together or only some?) and whether particular combinations of categories are dominant. The right-most panel of Fig. 1 shows that, while papers tend to have more male authors, there are more papers authored solely by women (four) than by men (two).

In conclusion, building on PAOHVis, we advocate the consolidation of any repeated hyperedges and the encoding of their frequency in an aligned bar chart above each hyperedge. The result is visually similar to *UpSet* [4] but functionally different, with bar height denoting hyperedge multiplicity rather than set intersection size. This economises horizontal space, while also drawing attention to the distribution of recurrent hyperedges, especially when sorted by frequency.

Aggregation by node attributes is useful in situations where it is less important to know precisely which entities occur in relationships and more important to understand what *kinds* of entities they tend to be (e.g. to investigate a possible gender bias or to see how many papers have female-only or male-only authors). As the level of aggregation increases, more information about the original nodes and hyperedges is lost, in order to reveal more general patterns. It may be beneficial to view all levels of aggregation in conjunction, rather than in isolation.

References

1. Berge, C.: Graphs and hypergraphs (1973)
2. Buono, P., Valdivia, P.: Applications of dynamic hypergraph visualization. In: Proceedings of the 2022 International Conference on Advanced Visual Interfaces. AVI 2022, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3531073.3534495
3. Fischer, M.T., Frings, A., Keim, D.A., Seebacher, D.: Towards a survey on static and dynamic hypergraph visualizations. IEEE Transactions on Visualization and Computer Graphics **2**(3), 81–85 (2021). https://doi.org/10.1109/VIS49827.2021.9623305
4. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., Pfister, H.: Upset: Visualization of intersecting sets. IEEE Transactions on Visualization and Computer Graphics **20**(12), 1983–1992 (2014). https://doi.org/10.1109/TVCG.2014.2346248
5. Valdivia, P., Buono, P., Plaisant, C., Dufournaud, N., Fekete, J.D.: Analyzing dynamic hypergraphs with parallel aggregated ordered hypergraph visualization. IEEE Transactions on Visualization and Computer Graphics **27**(1), 1–13 (2021). https://doi.org/10.1109/TVCG.2019.2933196