Aggregating Hypergraphs by Node Attributes

David Trye, Mark Apperley & David Bainbridge | University of Waikato, Hamilton, New Zealand
dgt12@students.waikato.ac.nz

Background

- A *hypergraph* [1] is a generalisation of a graph in which (hyper)edges may connect any number of nodes; hypergraphs can be used to model complex, multi-entity relationships and have diverse applications
- *PAOHVis* visualises hypergraphs by mapping nodes to parallel, horizontal bars and depicting hyperedges as vertical lines [2]

Problem

- Few techniques exist for directly incorporating node attribute data into hypergraph visualisations
- When dealing with large datasets, it is often necessary to use operations like aggregation to reduce visual complexity and better utilise the available space

Contribution

- Building on the design of *PAOHVis* [2], we advocate simplifying hypergraphs by consolidating identical hyperedges and encoding their frequency in a horizontally-aligned bar chart
- Multiple categorical node attributes can be displayed in any hypergraph using visual channels such as colour, shape, size and outline
- We propose a domain-agnostic framework for aggregating hypergraphs by one or more categorical node attributes, distinguishing between *no aggregation*, *count-based* and *binary* functions.
- *No aggregation* corresponds to the default PAOHVis layout (with the potential addition of a bar chart)
- *Count-based aggregation* collapses hyperedges with the same category counts, and is useful for tasks relating to category frequency and hyperedge length
- *Binary aggregation* collapses multiple occurrences of each category into a single node, and is useful for detecting the presence of and comparing groups of categories

Co-Authorship Example: Nodes = people, hyperedges = papers, attribute = gender (fictional dataset)

Input: Matrix encoding attribute data for nodes (columns) and hyperedges (rows)

ID	Donna	Adam	Hēmi	Sarah	Julia	Joe	David	Roger	Susan	Lucas	Maria	Stuart	All Cat's	F&M	F-only	M-only
h1	F	M											F,M	1	0	0
h2	F	M											F,M	1	0	0
h3	F	M											F,M	1	0	0
h4	F				F								F,F	0	1	0
h5	F				F								F,F	0	1	0
h6			M	F			M	M					F,M,M,M	1	0	0
h7			M	F			M	M					F,M,M,M	1	0	0
h8	F	M				M			F				F,F,M,M	1	0	0
h9	F		M										F,M	1	0	0
h10	F			F	F				F				F,F,F,F	0	1	0
h11		M	M			M							M,M,M	0	0	1
h12		M		F									F,M	1	0	0
h13		M		F		M							F,M,M	1	0	0
h14			M		F					M			F,M,M	1	0	0
h15			M	F		M							F,M,M	1	0	0
h16					F						F		F,F	0	1	0
h17							M					M	M,M	0	0	1
	8	7	6	6	5	4	3	2	2	1	1	1		11	4	2

Identical hyperedges (rows) are always consolidated

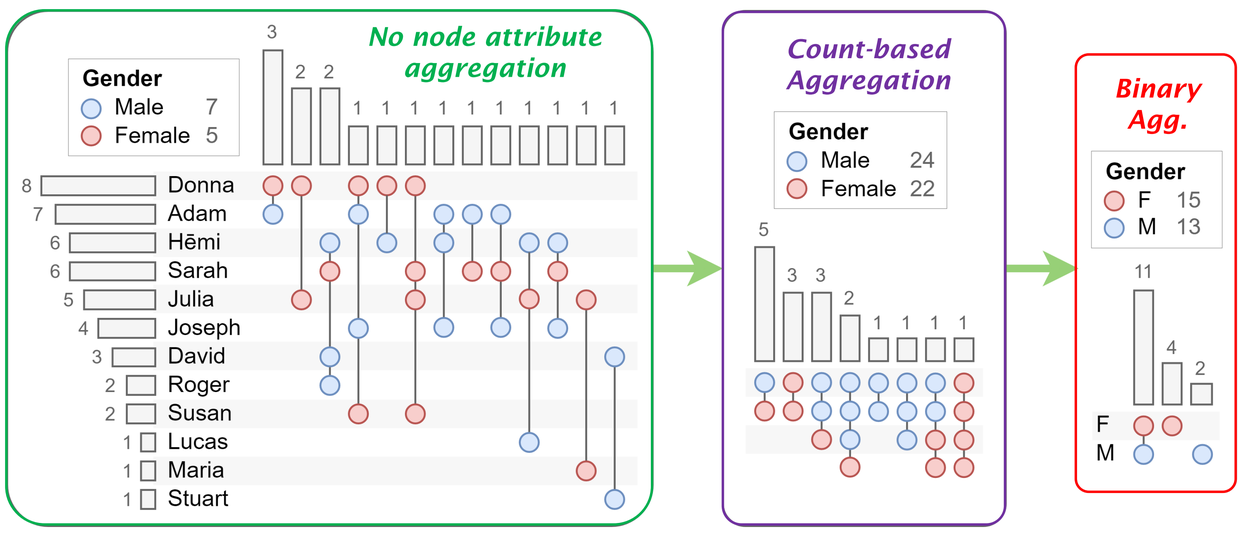
For count-based aggregation, we compute the frequency of each combination of categories

Non-aggregated legend shows number of blue and pink columns in matrix (M=7, F=5)

Sum all M's and F's to get the numbers in the count-based legend (M=24, F=22)

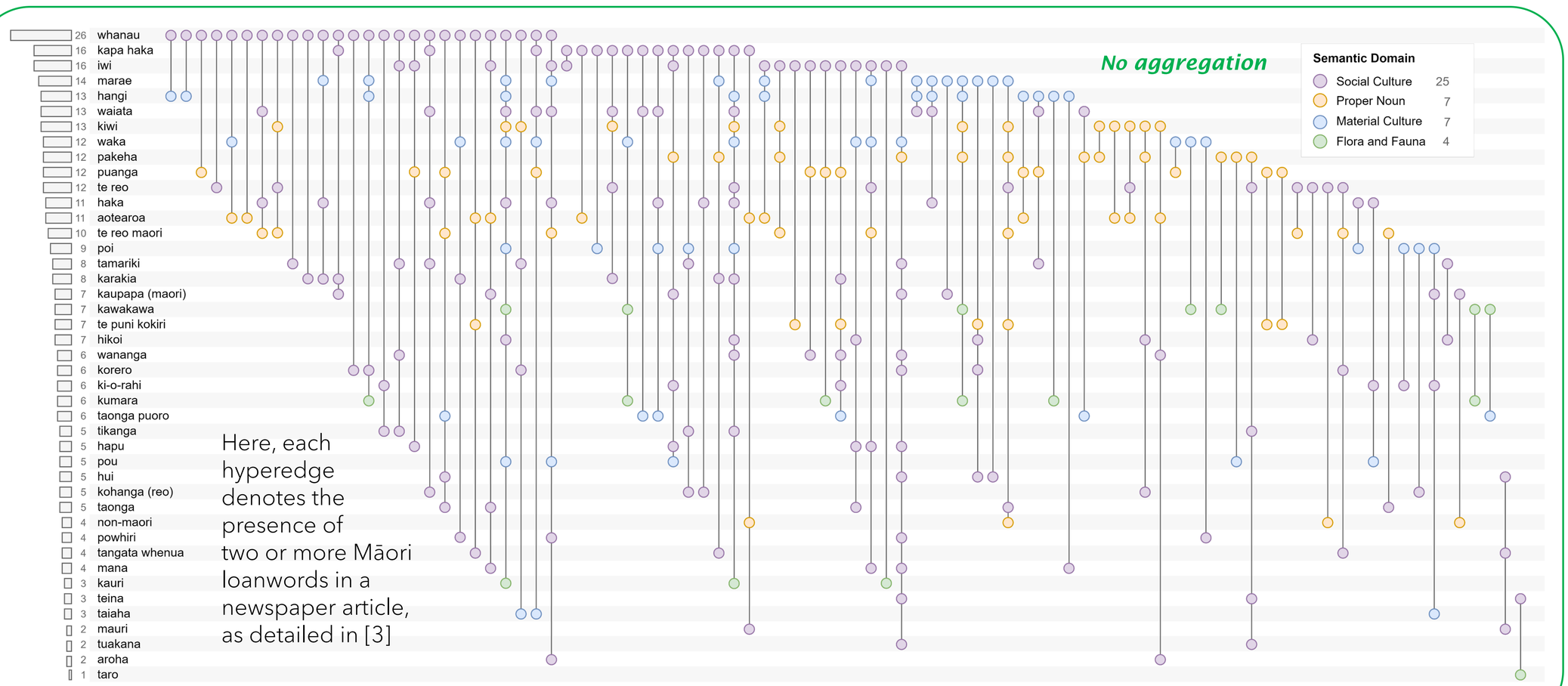
11+4=15 (F)
11+2=13 (M)

Output: Hypergraph visualisations reflecting different levels of aggregation



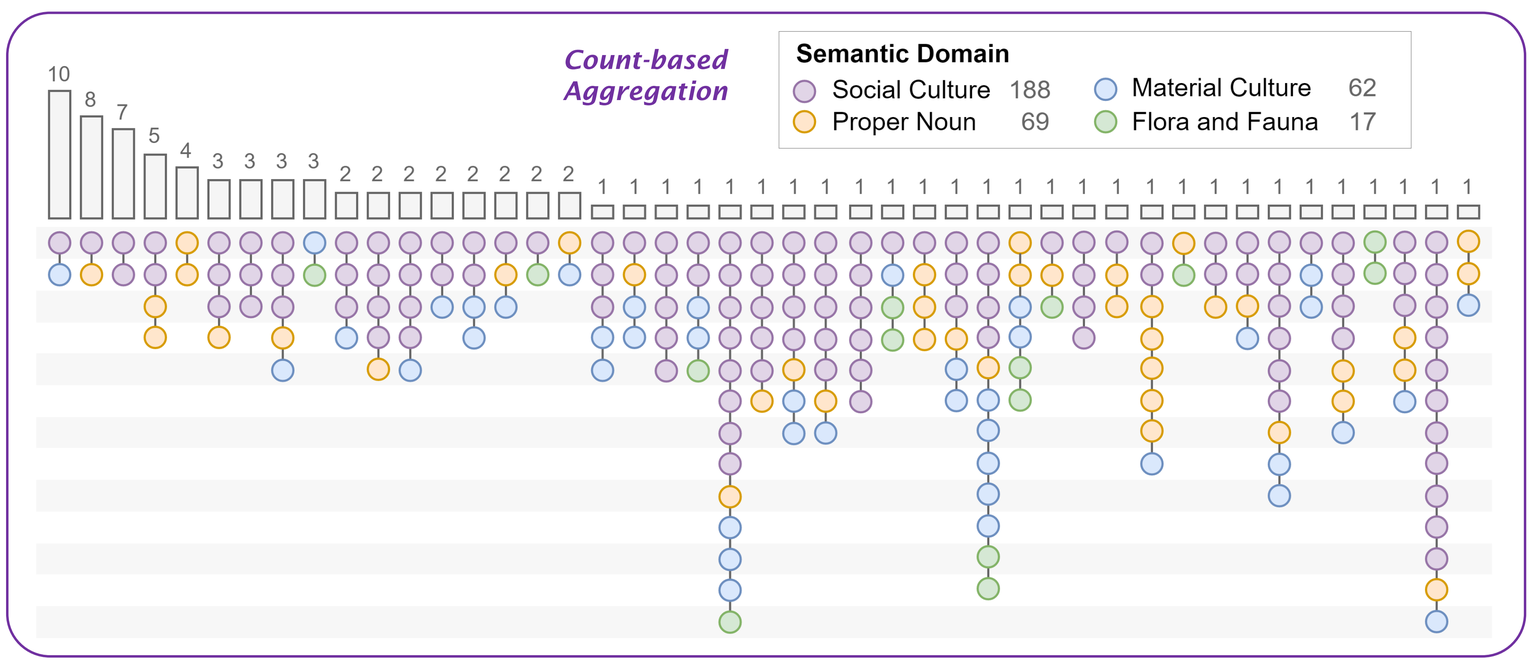
Insights: Most papers are authored by a mixture of men and women. All papers have between two and four authors. There are up to 3 male authors and 4 female authors per paper. Papers tend to have more male than female authors, but there are more female-only papers (4) than male-only ones (2).

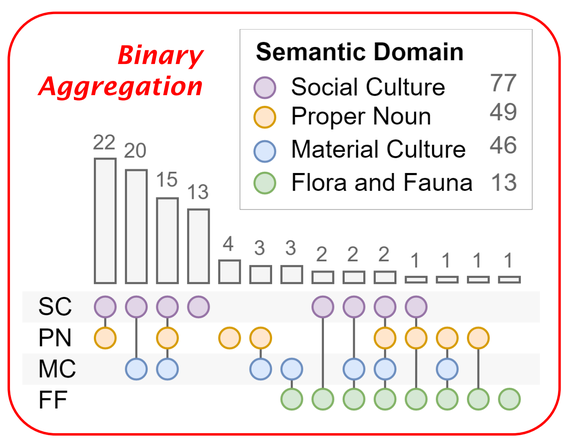
Loanword Co-occurrence Example [3]: Nodes = Māori loanwords, hyperedges = NZ newspaper articles, attribute =semantic domain



Insights:

- Social culture loans are the most prolific category, not only appearing in the most hyperedges (articles) but also having more instances within those hyperedges
- Among repeated hyperedges (for count-based aggregation), there is *never* more than a single flora and fauna term, and *rarely* more than a single material culture loan
- Unique hyperedges tend to be much larger, with more loans per category
- Looking at binary aggregation, there are only two hyperedges containing nodes from all four categories
- Despite the high frequency of social culture loans, most hyperedges do not contain *only* social culture terms, being accompanied by at least one loan from another category





References
[1] Berge, C.: Graphs and hypergraphs (1973)
[2] 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)
[3] Trye, D., Calude, A.S., Keegan, T.T., Falconer, J.: When loanwords are not lone words: Using networks and hypergraphs to explore Māori loanwords in New Zealand English. International Journal of Corpus Linguistics (2022)