An Investigation into the Representational Suitability of Tree Visualizations

Omkar S. Chekuri*
The University of Oklahoma

Chris Weaver†
The University of Oklahoma

ABSTRACT

In this poster we present work in progress to develop a more comprehensive understanding of the capabilities of common hierarchical visualization types for representing various structural properties of tree data. Our primary goal is to better inform designers and domain experts about which hierarchical visualization techniques are well suited (or not) for representing particular aspects of tree structure. We describe a design space of the data characteristics of various hierarchical structures and the visual channels of the different visual representation techniques use to represent them, systematically assess first the possibility then the suitability of each technique to represent each kind of structure, then identify important patterns across techniques and structures that both confirm current design wisdom and offer more grounded guidance when selecting a technique for visualizing particular structures of interest.

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

1 INTRODUCTION

The visualization community continues to show substantial interest in diverse techniques for visualizing hierarchical data. In addition to the wide variety of hierarchical visualization techniques now commonly in use, *multiple tree* visualizations have been developed to show different data characteristics in multiple different visualizations [1,7], and *hybrid visualization* techniques have been developed to support viewing of multiple different structural relationships and data attributes simultaneously [3,9]. Taxonomies have been developed [4] to understand the characteristics of hierarchical visualization techniques, and user studies [6,8] have been done to find the strengths and weaknesses of particular techniques. Visualization design builds on both theoretical foundations (e.g. [2]) and long-standing practical knowledge of how visual perceptual channels can effectively represent data attributes in visualizations. There is much less knowledge about how those channels are (or are not) effective for representing aspects of data structure, particularly for non-tabular data structures, including hierarchical data.

Hierarchical data varies across data domains and data set instances in terms of both the structural relationships and data attribute characteristics present in the data. Effective perception of both structure and attributes are highly influenced by the choice of visualization technique as well as the visual encoding mappings employed within them. By helping visualization designers assess the suitability of candidate tree visualization techniques in ways that consider multiple structures and attributes, we hope to facilitate more effective and efficient creation of visualization designs for hierarchical data analysis. As work in progress, we studied common tree visualization techniques, examining their ability and suitability to effectively represent different aspects of tree structure [and data attributes]. Here we summarize our results for a sample set of the studied visualization techniques and data properties most commonly encountered in hierarchical data, with fuller sets to be presented on the poster itself.

2 MODELING SUITABILITY FOR REPRESENTING STRUCTURE

To model the design space of tree visual representations, we surveyed the visualization literature, identified different types of tree visualization techniques, and selected a set of the 20 most common. We categorized the techniques based on their general characteristics of connection, containment, alignment, and adjacency as derived from variations of categorizations [5,9]. A sample of three categories and nine techniques in them are shown as column groupings and sub-groupings in Fig. 1. (The poster itself also includes *top-down dendrogram*, *left-right node-link*, *left-right dendrogram*, *radial dendrogram*, *indented treemap style*, *multi directional node-link*, *voronoi treemap*, *radial treemap*, *fan chart*, as well as the *2D grid* and *adjacency matrix* techniques in the adjacency category.)

We also identified the different tree substructures that are involved in visual data exploration and analysis in the surveyed visualization applications. We first categorized the substructures by their general topological character, then sub-categorized them into the different kinds of *properties* determined by the structure and/or attribute features of significant interest in each application's data. Each property type can be structural, statistical, or some specific kind of information type associated with pertaining tree elements. A sample of four structure types and 13 properties are shown as row groupings and sub-groupings in Fig. 1. (The poster includes eight additional properties and the additional structures *nodes*, *paths*, and *bipaths*.)

We assessed each visualization technique first to determine whether it is *possible* to visually represent each property. We then looked at each possible combination to determine the general *suitability* of the technique for visually representing the property. Pairs of columns in figure Fig. 1 summarize our determinations for each technique. The *possible* column states whether or not ('yes' or 'no') the visual channels used to depict the tree structure in the visual representation technique is capable of encoding the particular property. The *suitability* column states how effective those visual channels are in representing the property. Determination of effectiveness takes into account several qualitative criteria, described in the list below, resulting in values of **very high**, **high**, **medium**, **low**, **very low**, or **not at all**. Suitability can be a single value, a range of values, or even multiple distinct values when different examples suggest effectiveness can be quite divergent under different circumstances. For instance, for *top-down node-link* techniques we observe that the general suitability to show categorical data amongst siblings is *high* when the number of siblings is very low but rapidly falls off to *low* as the number of siblings increases. We chose the values of the qualitative scale for *suitability* by considering the following criteria, factored in according to their apparent relative importance:

- ***Visual Channel Support*** The ability of a particular technique to visually encode a data property. For instance, circular nodes in node-link techniques are limited in the range of values that can be encoded as area without obscuring layout structure.

- ***Visual Channel Accuracy*** The accuracy with which a user can perceive the represented data attribute. We based our analysis on Mackinlay's perceptual rankings by data type [2].

*e-mail: omkar.chekuri@ou.edu
†e-mail: cweaver@ou.edu

				Connection								Containment						Alignment			
				Top-Down Node-Link		Radial Node-Link		Indented Outline		Rings Layout		One Dimensional Treemap		Circular Treemap		Squarified Treemap		Radial Sunburst		Icicle Plot	
Tree Structure	Property	Description	Example	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability
Tree	Chronological	The chronological information associated with the tree from root node to leaf nodes.	Chronological nature of family tree in a geneological hierarchy.	yes	high	yes	very high	no	not at all	no	not at all	no	not at all	no	very low or not at all	no	not at all	yes	low to medium	yes	low to medium
	Diverging	The ability of a tree to start with fewer nodes and grow in size into many nodes.	The diverging of a family in a geneological hierarchy.	yes	high	yes	very high	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all	yes	low	yes	low
	Converging	The ability of a tree to start with many nodes and shrink in size into fewer nodes.	The converging of teams that compete for a championship in sports tournament data.	yes	medium	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all
Level	Ordinal	The distinct levels of the tree that begin with the root node and increases in number with each additional connected edge.	The stages of progression of a sports tournament in a tournament hierarchy.	yes	very high	yes	high	yes	very high	yes	low	yes	very low or medium	yes	medium or high	yes	low	yes	very high	yes	very high
Parent- Child Relationship	Part - Whole	The relationship in which the child nodes make up parts of the parent node and combining them will constitute a parent node.	Continents comprising of countries in a geographical hierarchy.	yes	medium	yes	low	yes	very high	yes	high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high
	Association	The relationship in which the child node and a parent node are related by a relation that may cease to exist at a later point in time.	The relationship between a manager and an employee in an organizational hierarchy.	yes	very high	yes	very high	yes	high	yes	high	yes	very low	no	medium	yes	very low	yes	low to medium	yes	low to medium
	Weighted	The importance in terms of the value parent-child relationship carries in addition to the existance of a relationships in a tree.	The period time an employee works under a manager in an organizational hierarchy.	yes	low or not at all	yes	low or not at all	yes	very low	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all
	Categorical	The category of the parent-child relationship in addition to the existance of relationship in a tree.	The categories of paternal and maternal relationships in a geneological hierarchy.	yes	low or not at all	yes	low or not at all	yes	very low	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all	no	not at all
Siblings	Ordinal	The ordinality of the siblings arising from the ordinal data associated with the sibling nodes under a parent node in a tree.	The ordering of files within a folder in a file hierarchy.	yes	very high	yes	low to medium	yes	very high	no	not at all	yes	very high	no	not at all	yes	very low or not at All	yes	low to medium	yes	very high
	Nominal	The nominality of the siblings arising from the nominal data of the sibling nodes under a parent node in a tree.	The naming of the players in a team in a sports tournament hierarchy.	yes	low to medium	yes	medium	yes	very high	yes	very low	yes	low or very low	yes	very low or medium or not at all	yes	low to medium	yes	low to medium	yes	low to medium
	Quantitative	The quantitative nature of the siblings arising from the quantitative data of the sibling nodes under a parent node in a tree.	The size of the files in a folder in file hierarchy.	yes	low or not at all	no	not at all	no	not at all	no	not at all	yes	very low or not at all	yes	very low	yes	low to very low	yes	high	yes	high
	Interval	The interval nature of the data attributes of siblings arising from the interval data of the sibling nodes under a parent node in a tree.	The clustering of different nodes into different intervals in a dendrogram.	yes	low	yes	very low	yes	low	no	not at all	yes	low	no	not at all	yes	very low	yes	low	yes	low
	Categorical	The categorical nature of the siblings rising from the categorical data of the sibling nodes under a parent node in a tree.	Different genders of employees in an organizational hierarchy.	yes	low or high	yes	very low	no	not at all	yes	medium	no	not at all	no	not at all	no	not at all	yes	very low	yes	very low

Figure 1: The suitability of various types of tree visualization to visually represent some noteworthy properties of tree structures.

- ***Unintended Artifacts*** Visual channels can add unintended visual artifacts that are not present in the data. For instance, horizontal positioning of nodes in a node-link representation implies an ordering that may or may not exist in the data.
- ***Layout Limitations*** The choice of layout influences how one perceives data properties. For instance, order is typically easier to perceive in a linear layout than a radial layout. The angle and spacing between nodes in a node-link diagram can influence ability to perceive path and branching structures.
- ***Scalability*** Some visual channels become perceptually ineffective to encodea property as data scales, while others channels scale well to some extent. For example, treemap layouts are typically limited to showing at most a few levels at a time.

3 Results

The *suitability* columns in Fig. 1 suggest that there is quite a bit of variation in the visual representation techniques in terms of their ability to encode the properties. For instance, the ability to show properties is *yes* for all the properties for *top-down node-link* techniques, but *no* for most of properties for *circular treemap* techniques, and the support provided by the other techniques vary in between.

If we observe the rows in Fig. 1 for each property, the *possible* columns suggest that some properties can be encoded in most techniques we considered, while other properties are possible in only some. The *suitability* columns suggest that while some techniques like *top-down node-link* have the ability to encode most properties, only some properties have *very high* or *high* suitability. This suggests that some techniques are general purpose and can be used in most cases with reasonable suitability, whereas others have limited *very high to high* suitability that suggest more specialized purposes.

The suitability of techniques in the *alignment* category appear much more scattered than those in *containment*. This suggests that the model can provide not only guidance on the choice of a technique to show a desired property, but also how one can go about picking a technique when it needs to support multiple properties at once.

Overall, our study showed that there is quite a bit of difference in terms of the sets of properties that known tree visualization types can show, yet taken as a whole the existing tree visualization design space is clearly quite capable of representing at least the most common and important aspects of hierarchical structures.

4 Conclusions

We developed and applied a conceptual framework to assess the suitability of a wide variety of tree visualization techniques for visually representing important structural properties of tree data. Our results reveal rich patterns of similarities and differences in the suitability of the techniques across the various properties, and suggest opportunities for applying the techniques individually or collectively to diverse exploration and analysis needs. Moving forward we plan to apply knowledge of suitability to study the design of suites of interactive operations for editing trees directly in visualizations, and use the results to inform the design of tasks in user studies of interactive tree visualization techniques.

References

[1] M. Graham and J. Kennedy. A survey of multiple tree visualisation. *Information Visualization*, 9(4):235–252, 2010.

[2] J. Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.

[3] P. Neumann, S. Schlechtweg, S. Carpendale, et al. Arctrees: visualizing relations in hierarchical data. In *EuroVis*, pp. 53–60, 2005.

[4] W. Scheibel, M. Trapp, D. Limberger, and J. Döllner. A taxonomy of treemap visualization techniques. In *VISIGRAPP (3: IVAPP)*, pp. 273–280, 2020.

[5] H.-J. Schulz, S. Hadlak, and H. Schumann. The design space of implicit hierarchy visualization: A survey. *IEEE transactions on visualization and computer graphics*, 17(4):393–411, 2010.

[6] J. Stasko, R. Catrambone, M. Guzdial, and K. McDonald. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International journal of human-computer studies*, 53(5):663–694, 2000.

[7] S. T. Teoh. A study on multiple views for tree visualization. In *Visualization and Data Analysis 2007*, vol. 6495, pp. 99–110. SPIE, 2007.

[8] Y. Wang, S. T. Teoh, and K.-L. Ma. Evaluating the effectiveness of tree visualization systems for knowledge discovery. In *EuroVis*, vol. 6, pp. 67–74, 2006.

[9] S. Zhao, M. J. McGuffin, and M. H. Chignell. Elastic hierarchies: Combining treemaps and node-link diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 57–64, 2005.