An Investigation into the Representational Suitability of Tree Visualizations

Omkar Chekuri
omkar.chekuri@ou.edu
School of Computer Science
The University of Oklahoma

Chris Weaver
cweaver@ou.edu
School of Computer Science
The University of Oklahoma

Background

Visualization design builds on theoretical foundations (e.g. [1]) and long-standing practical knowledge of how visual channels can effectively represent data *attributes*, but there is much less knowledge about the effectiveness of those channels for representing aspects of data *structure*, particularly for trees.

How does one choose a visual representation technique? Hierarchical data in each domain has specific data characteristics that are more prominent than other domains.

What makes a hierarchical visual representation effective? The ability to show the data characteristics of the hierarchy in a manner that lets the user perceive the intended data accurately.

What can a tree show? A tree structure is composed of many structural elements each capable of showing different data types. Each of these different data types can show various properties that are specific to domain.

Structures of Interest

Trees contains many substructures of interest. Common structures we considered include:

Node: An item in the hierarchy.

Edge (parent child relationship): A relation between a parent node and a child node.

Siblings: The child nodes of a common parent.

Path: A sequence of edges connecting an ancestor and a descendent.

Bi-path: A path connecting two descendent nodes through a common ancestor.

Level: Nodes of the same path length to the root.

Tree: The entire hierarchy of nodes and edges.

Techniques + Structures

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 [2, 3].

We also identified the different tree substructures 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

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.

The figure below summarizes our determinations for each combination of technique and property.

Structure	Property	Description	Example	Connection																Containment								Alignment						Adjacency			
				TopDown NodeLink		TopDown Dendrogram		LeftRight NodeLink (Space Tree)		LeftRight Dendrogram		Radial NodeLink		Radial Dendrogram		Indented TreeMap Style Hybrid		Indented Outline		Multi Directional NodeLink		Rings		One Dimensional TreeMap (ArcTrees)		Circular TreeMap		Squarified TreeMap		Voronoi TreeMap		Nested Pie Tree		Fanchart		*Sunburst*		*Icicle Plot*		2D Grid Layout		Adjacency Matrix	
				Bohanec.2007		McGuirl.et.al.2020		Plaisant.et.al.2002		Knott.et.al.2019		Sheth.et.al.2003		Takahashi.et.al.2019		Gachet.et.al.2002		Husken.at.el.2007		[Tableau Community]		Teoh.et.al.2002		Neumann.et.al.2005		Wetzel.et.al.2004 as implemented in D3.js		Shneiderman.et.al.2001		Brath.et.al.2015		H.Schulz.et.al.2013		Draper.GM.et.al.2008		*Stasko.el.ot.2000*		*Woodburn.et.al.2019*		Brath.et.al.2015		Zhang.et.al.2019	
				Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability	Possible	Suitability
Tree	Chronological tree	The chronological information associated with the tree from root node to leaf nodes.	Chronological nature of family tree in a genealogical hierarchy.	yes	high	yes	very high	yes	very high or none	yes	very high	yes	very high	yes	very high	no	none	yes	high	yes	very low	no	none	no	none	no	very low or none	no	none	no	none	no	none	yes	medium to high	yes	medium to high	yes	high	no	none	yes	very low
	Diverging tree	The ability of a tree to start with fewer nodes and grow in size into many nodes.	The diverging of a family in a genealogical hierarchy.	yes	high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very low	yes	very high	yes	very high	yes	medium to high	yes	medium to high	yes	medium to high	yes	medium	yes	very high	yes	high	yes	very high	no	none	yes	very high	yes	very low	yes	very low
	Converging tree	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.	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	yes	very high	yes	medium	no	none	no	none	no	none	no	none	no	none	yes	high	no	none	no	none	no	none	yes	very low
	Spatial tree	The ability of the tree to be laid out spatially on 2D space such that nodes can convey spatial data.	Displaying geographical data by placing the nodes on the map.	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	yes	high	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none
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	medium	yes	very high or medium	yes	medium	yes	high	yes	medium	yes	medium	yes	very high	yes	medium to high	yes	low	yes	very low or medium	yes	medium or high	yes	low	yes	low to very low	yes	very low to low	yes	very high	yes	very high	yes	very high	yes	low to very high	yes	medium or none
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	high or none	yes	low	yes	high	yes	low	yes	high	yes	low	yes	very high	yes	low	yes	high	yes	very high	yes	very high	yes	very high	yes	medium	yes	high	yes	medium or none	yes	very high	yes	very high	no	high	yes	none
	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	no	none	yes	very high	no	none	yes	very high	no	none	yes	very high	yes	high	yes	very high	yes	high	yes	very low	no	medium	yes	very low	yes	very low	yes	very low	yes	very high or none	yes	low to medium	yes	low to medium	yes	high or none	yes	none
	Weighted	The importance in terms of the value parent-child relationship carries in addition to the existence of a relationships in a tree.	The period time an employee works under a manager in an organizational hierarchy.	yes	low or none	no	none	yes	low or none	no	none	yes	low or none	no	none	yes	very low	yes	very low	yes	very low	yes	medium	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	yes	very high
	Categorical	The category of the parent-child relationship in addition to the existence of relationship in a tree.	The categories of paternal and maternal relationships in a genealogical hierarchy.	yes	low or none	no	none	yes	low or none	no	none	yes	low or none	no	none	yes	very low	yes	very low	yes	very low	yes	low to medium	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	yes	very high
	Ordinal	The ordinality arising from the parent-child relationship in a tree.	The ordering of nodes from ancestors to descendants in a genealogical tree.	yes	very high	yes	high or none	yes	very high	yes	high or none	yes	very high	yes	high or none	no	none	no	none	yes	high	yes	high	no	none	no	none	yes	medium to high	yes	low to very low	no	medium to high	yes	very high	yes	very high	yes	very high	yes	medium	no	very high or none
	Grouped	The grouping of all the parent-child relations under a single parent.	All the relations between a folder and the files it it in a file hierarchy represent the same kind of relations and these relations can be grouped.	no	none	yes	very high	no	none	yes	very high	no	none	yes	very high	yes	very high	yes	very high	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none
Siblings	Ordinal	The ordinality of the siblings arising from the ordinal data associated with the sibling nodes.	The ordering of files within a folder in a file hierarchy.	yes	very high	yes	high	yes	very high	yes	very high	yes	low to medium	yes	very low	yes	very high	yes	very high	yes	high	no	none	yes	very high	no	none	yes	very low or none	no	none	yes	high	no	low to medium	yes	low to medium	yes	very high	yes	high or none	yes	very high or none
	Nominal	The nominalist of the siblings arising from the nominal data of the sibling nodes.	The naming of the players in a team in a sports tournament hierarchy.	yes	low to medium	yes	very high or low	yes	low	yes	very high or low	yes	medium	yes	very high or low	yes	medium to high	yes	very high	yes	very high	yes	very low	yes	low or very low	yes	very low or medium or none	yes	low to medium	yes	very low	yes	low to very high or none	yes	low to very high or none	yes	low to very high or none	yes	very high	yes	very high
	Quantitative	The quantitative nature of the siblings arising from the quantitative data of the sibling nodes.	The encoding of size of the files in a folder in file hierarchy.	yes	low or none	yes	low	yes	low or none	yes	low	no	none	yes	very low	no	none	no	none	yes	very high	no	none	yes	very low or none	yes	very low	yes	very low to low	yes	none	yes	very low or none	no	none	yes	very low to medium or none	yes	very low to medium or none	no	none	no	none
	Interval	The interval nature of the data attributes of siblings arising from the interval data of the sibling nodes.	The clustering of different nodes into different intervals in a dendrogram.	yes	low	yes	very high	yes	medium	yes	very high	yes	very low	yes	very high	no	very low	yes	low	no	very low	no	none	yes	low	no	none	yes	very low	no	none	yes	low	no	none	yes	low	yes	low	no	none	no	none
	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	no	none	yes	low or high	no	none	yes	very low	no	none	no	none	no	none	yes	low	yes	medium	yes	medium to high	no	none	no	none	no	none	no	none	yes	very low	yes	very low	no	none	no	none
	Identical	The ability to encode all the identical siblings in a tree in a same manner.	The existence of all files of same size and name under a folder in a same manner.	yes	high	yes	high	yes	high	yes	very high	yes	high	yes	very high	yes	high	yes	high	yes	very high	yes	low	yes	very high	yes	very high	yes	very high	no	none	no	none	yes	very high	yes	very high	yes	high	yes	medium	yes	very high
Path	Unidirectional	A sequence of edges connecting an ancestor and a descendent.	The path of the file in any folder in a file hierarchy.	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	high	yes	high	yes	high	yes	medium	yes	very low	yes	high	yes	very high	yes	very high	yes	very high	yes	high or none	yes	very low
Bi-path	Multidirectional	A path connecting two descendent nodes through a common ancestor.	The relationship by which two distinct cousins in a family tree are connected.	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	high	yes	high	yes	high	yes	medium to high	yes	high	yes	very low	yes	very low	yes	medium	yes	very high	yes	very high	yes	very high	yes	high or none	yes	very low
Node	Mono entity	The ability to encode each node as a single independent entity in a tree.	The encoding of each person in an organization as an single independent entity in an organization.	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high	yes	very high
	Multi entity	The ability to encode multiple entities as a single node in a tree.	The encoding of father and mother as a single parent node in a genealogical tree.	yes	high or none	yes	very high or none	yes	high or none	yes	very high or none	yes	high or none	yes	very high or none	no	none	no	none	yes	none	no	none	no	none	no	low	no	none	no	none	no	none	no	none	no	none	no	none	no	none	no	none

Suitability

We define *suitability* as how effective visual channels are for representing a structural property. We determined values of suitability on a qualitative scale (from very high to none) by considering the following criteria, factored in according to their apparent relative importance:

Visual Channel Support: The ability of the structure of a tree to encode a data property. It taking the values of **yes** if a property is supported and **no** incase the property is not supported.

Visual Channel Accuracy: The ability of the visual channel to represent the property of data in a manner to be accurately perceived by the users. We based part of our analysis on the perceptual ranking for a particular data type based on [1].

Unintended artifacts: The visual channels sometimes introduce artifacts in the visual representation that are not present in the data.

Layout limitations: The choice of layout influences how we perceive the data property. For instance, perceiving ordering is easier in linear layout compared to radial layout.

Scalability: The ability of a visual channel to effectively encode the property even as data scales.

Observations

Applying the suitability criteria resulted in the populated table above. It reveals quite a bit of variation in the suitability of different tree visualization techniques to encode various structural properties of trees.

For instance, **Multi Directional NodeLink** representations are suitable for most properties whereas **Circular TreeMaps** are suitable for few, and other techniques are in between.

Looking at rows, the **Possible** column suggests that some properties like **Path** can be encoded in most techniques, while other properties like **Weighted** (under **Parent-Child Relationship**) are possible in only some techniques. The **Suitability** column suggests that while some techniques have the ability to encode most properties, suitability is very high or high for only some of them.

Some techniques appear to be general purpose, of reasonable suitability in most cases, while other techniques are more special purpose with very high or high suitability for particular circumstances.

Overall, there is substantial variation in the aspects of structure that techniques can show well and consequently the tree visualization space has rich and complex potential for supporting visual exploration and analysis of hierarchical data structures.

Future Work and Potential Impacts

- Identify different data characteristics supported by various structural elements in hierarchies.
- Develop a framework to assess the ability of individual tree visualization techniques to represent structural properties of trees.
- Analyze similarities and differences in how different kinds of tree visualizations support visual representation of tree structures.
- Study how design of interactive operations for editing trees is informed by a better understanding of suitability.

References:

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

[2] 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.

[3] 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. IEEE, 2005.