Facilitating Graph Interpretation via Interactive Hierarchical Edges

Thomas S. McTavish
Center for Digital Data, Analytics & Adaptive Learning
Pearson
tom.mctavish@pearson.com

ABSTRACT
Graphs visualizations can become difficult to interpret when they fail to highlight patterns. Additionally, the data to be visualized may be hierarchical in nature. Therefore, graphs with hierarchical data need to offer means of telescoping that collapse or expand subgraphs while aggregating their data. In this paper, we demonstrate an interactive hierarchical edge graph on book prerequisite data, which can be generalized to a variety of hierarchical data. We illustrate the importance of ordering nodes (when possible) and coloring by various features. We then demonstrate various ways of performing exploratory data analysis by delivering various pieces of information on mouseovers and utilizing telescoping and filtering.

Keywords
Hierarchical edge bundling, prerequisite relationships

1. INTRODUCTION
When graphs contain many nodes and edges – especially different types of nodes and edges – they can quickly become difficult to visually interpret [7]. The common term is "hairball" as nodes and edges jumble into a tangled morass that occlude any meaningful patterns. Force-directed graphs operate to keep nodes with strong edges closer and nodes with weak or absent edges further apart [2]. This layout can aid in some contexts, but frequently exacerbates the hairball phenomenon. There are two striking visualization designs by Krzywinski and colleagues that aim at revealing interpretable patterns in graphs. At the core of each is at least one meaningful axis on which to align nodes. The first is *Circos*, which arranges sorted nodes along a circle [5]. Nodes are often displayed as arcs along the circle and edges between the arcs are visualized as chords or ribbons that cut through the middle of the circle. Circos has been used in over 500 publications, many related to large-scale genomic data. By arranging nodes along one axis in a circle, Circos easily discriminates nearby and distant edges. The widths of the nodes (length of the arc) can carry meaning and so can the width of the chord between connected arcs. Nodes and edges can also be colored to highlight features such as the node type, the source, and the target. It is also common to display many node features such as histograms of different measures within an arc, for example, Figure 3 in [6].

The other design by Krzywinski is *hive plots* [4]. Hive plots are comprised of multiple axes, each radiating from an inner ring. A given node may exist on one or more axes, aligned along the axis in some meaningful way. For example, an axis might sort nodes by different graph features such as a node's *closeness* – the average distance between a node and all others reachable from it. By placing nodes on various axes, a representation of where a node resides along some feature is captured. When edges are added, it may bring out relationships between adjacently-placed features. For example, anti-correlations of two features compared side-by-side will have many criss-crossed edges. In short, ordering nodes in some meaningful way(s) permits Circos and hive plots to better reveal patterns. Circos and hive plots, however, do not capture hierarchical relationships very well.

Hierarchical edge bundling is a visualization technique on hierarchical data that skews edges toward their parent nodes, which may be invisible in the graph [3]. The visual effect is that edges are channeled into larger, striking swaths while avoiding the direct clutter of the parent nodes. Any topology can be employed, but simpler geometric structures are most commonly used.

In this paper, we demonstrate an interactive hybrid of Circos plots with hierarchical edge bundling on mathematical book prerequisites. The books are structured hierarchically as a table of contents with chapters, sections, objectives, and exercises. Prerequisites map between objectives. The goal was to provide a means of highlighting prerequisites at the various levels, to call out important objectives, and also reveal holes. Through coloring, it is simple to discriminate chapters or to highlight nodes by features such as learner interaction frequency. Through telescoping, it is straightforward to determine those prerequisites that map across chapters, within a chapter, and within sections. Through filtering it is possible to display nodes and edges by their degree. Collectively, by aligning a curriculum along a circle, we demonstrate how this template can be used for displaying various relationships and features of hierarchical, educational data.

2. METHODS

Two higher education math books were selected that contained a table of contents and prerequisites as mapped by content matter experts. Interactivity data came from students, largely from the U.S., who were enrolled in courses spanning Fall semester 2012 through 2013 that used these books and the accompanying Pearson MathXL® homework system. All data was translated into JSON format for use in a web browser. The graph and its interactivity functionality was programmed using D3.js [1].

3. DEMONSTRATION

Figure 1 shows a screen shot of the graph and user controls. Displayed is a developmental math book with chapters starting at 12 o'clock and progressing clockwise. Nodes are colored by chapter and have ample spacing to easily discriminate them. Most nodes displayed are at the objective level. Within a chapter, slight separations between the nodes delineate the sections. Edges within a section are shown as little arcs. Edges within the chapter have a larger arc, and edges across chapters bend so that they bundle near to where a chapter node would be. We see various features at a glance. For example, the online appendix has no pre- or post-requisites across chapters. This is because it is shared across several books and is independent from this book.

Chapters 2, 11, and 13 are displayed at the chapter level hiding all of their section and objective nodes, whereas chapters 4 and 7 are at the section level. Chapters can be shown or hidden in the column of checkboxes on the right. For example, some appendix items have been removed from this display. The radio buttons correspond to the level of the hierarchy to display.

In Figure 1, the user has centered the mouse over the Chapter 2 node. The color of the node is green, so bold green edges reveal other chapters to which Chapter 2 is a prerequisite. Also shown are bold orange lines from Chapter 1 objectives coming into Chapter 2. The text of these pre- and post-requisites are listed on the left. We see at a glance that while Chapter 2 is prerequisite to Chapter 3, it links to other sections and objectives in a punctate fashion, completely ignoring the middle chapters of the book.

Edges are colored by their outgoing node color. Of course, they can be colored by their incoming node color or other feature. We have also colored nodes by their degree, highlighting critical objectives and important chapters. We have also colored nodes by performance measures such as the frequency of user interactions. Coloring can be selected through the pulldown menu on the top-left.

This utility also has some filtering capabilities to show/hide edges within sections, within chapters, and across chapters. Nodes can also be filtered out their degree or feature by which they are colored. Similarly, edges can be filtered out if they are below a weight threshold.

3.1 Limitations

While this visualization works well with book prerequisites, making graphs interactive as we have demonstrated, limits the quantity of nodes and edges because they have to be large enough to be selectable. Additionally, while edge bundling facilitates an interpretation of convergence, it also makes it difficult to select or hover over any individual edge for information. As presented, more than 3000 edges begins to be problematic. Similarly, when there are over 500 nodes along the circle, it can become difficult to select a node of interest with a mouse.

3.2 Next steps

This plot and circos plots have only one axis. Avenues to explore include reordering nodes along this axis by different features. Alternatively, hive plots could be extended with the ideas presented here, where various axes could utilize hierarchical data by swapping child nodes with aggregated parent nodes.

4. CONCLUSION

It is difficult to interpret graphs without an adequate visualization. In this work, we demonstrated a template that can be used on hierarchical data aligned along the axis of a circle. At a glance, it can reveal a lot of features, but through filtering, telescoping, and interactivity, exploratory data analysis can be performed to reveal features at various scales. As a template, it is quite useful for contrasting several graphs, or alternatively, illuminating various features within a general structure. For example, in our case using prerequisite data, nodes and edges might be colored by difficulty, fraction correct, time-on-task, or other measures of students interacting with these book objectives. Furthermore, this technique can be generically applied to other datasets.

5. REFERENCES

[1] M. Bostock, V. Ogievetsky, and J. Heer. D^3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, Dec. 2011.

[2] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, Nov. 1991.

[3] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, Sept. 2006.

[4] M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra. Hive plots—rational approach to visualizing networks. *Briefings in Bioinformatics*, 13(5):627–644, Sept. 2012. PMID: 22155641.

[5] M. Krzywinski, J. Schein, İ. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, Sept. 2009. PMID: 19541911.

[6] E. S. Mace, S. Tai, E. K. Gilding, Y. Li, P. J. Prentis, L. Bian, B. C. Campbell, W. Hu, D. J. Innes, X. Han, A. Cruickshank, C. Dai, C. Frère, H. Zhang, C. H. Hunt, X. Wang, T. Shatte, M. Wang, Z. Su, J. Li, X. Lin, I. D. Godwin, D. R. Jordan, and J. Wang. Whole-genome sequencing reveals untapped genetic potential in africa's indigenous cereal crop sorghum. *Nature Communications*, 4, Aug. 2013.

[7] D. Merico, D. Gfeller, and G. D. Bader. How to visually interpret biological data using networks. *Nature Biotechnology*, 27(10):921–924, Oct. 2009.

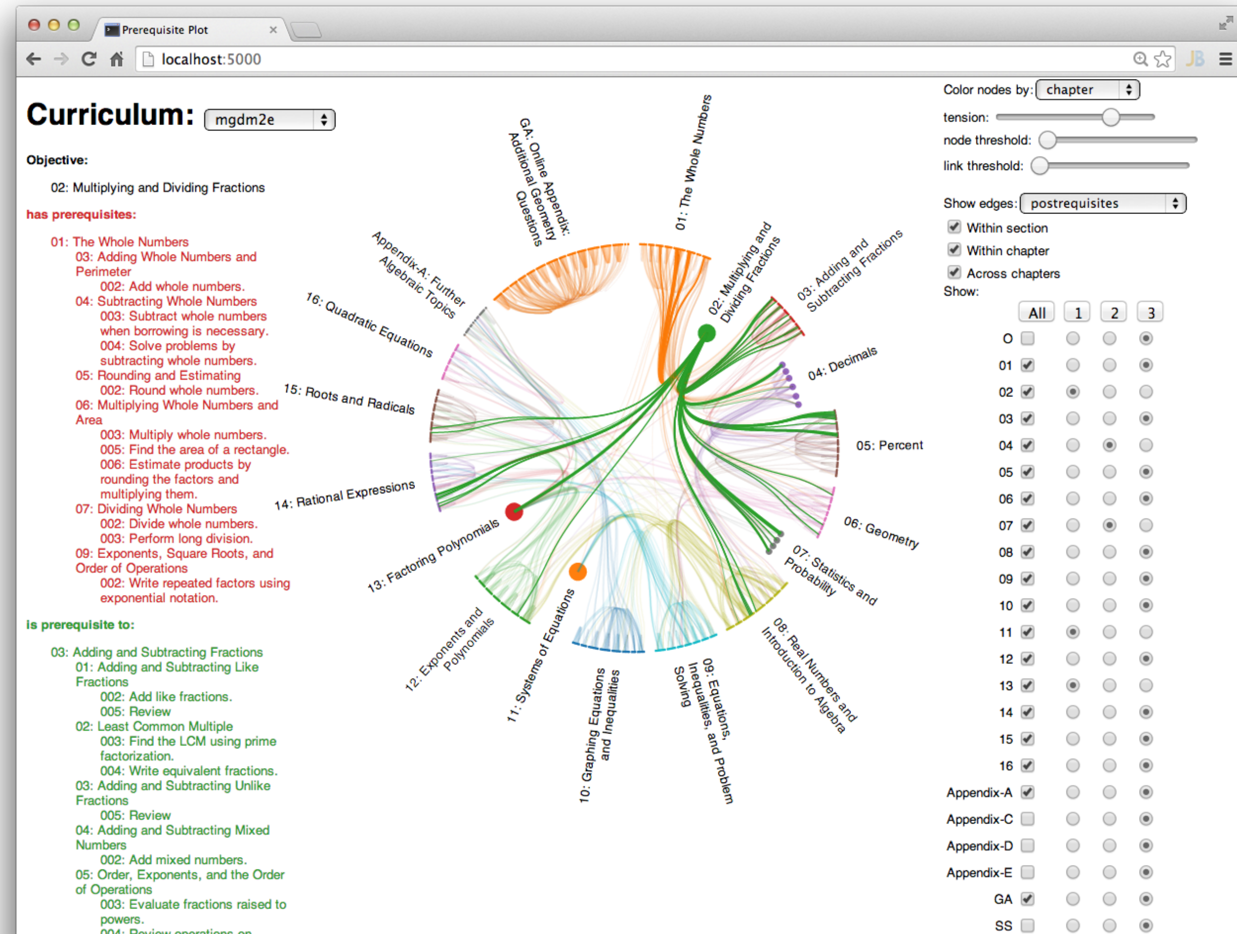

Figure 1: Screen shot of the interactive, hierarchical edge bundling graph.