Transportation Research Part C 51 (2015) 167–179

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Discovering temporal changes in hierarchical transportation data: Visual analytics & text reporting tools

John Alexis Guerra-Gómez [a,1], Michael L. Pack [b,*], Catherine Plaisant [c,2], Ben Shneiderman [d,3]

[a] University of Maryland, College Park, Center for Advanced Transportation Technology Laboratory, Human–Computer Interaction Lab, United States
[b] University of Maryland, College Park, Center for Advanced Transportation Technology Laboratory, Human–Computer Interaction Lab, Department of Civil & Environmental Engineering, United States
[c] University of Maryland, College Park, Human–Computer Interaction Lab, United States
[d] University of Maryland, College Park, Human–Computer Interaction Lab, Department of Computer Science, United States

ARTICLE INFO

Article history:
Received 14 November 2014
Received in revised form 25 November 2014
Accepted 25 November 2014
Available online 26 December 2014

Keywords:
Information visualization
Visual analytics
Change visualization
Tree visualization
Difference visualization

ABSTRACT

Analyzing important changes to massive transportation datasets like national bottleneck statistics, passenger data for domestic flights, airline maintenance budgets, or even publication data from the Transportation Research Record can be extremely complex. These types of datasets are often grouped by attributes in a tree structure hierarchy. The parent–child relationships of these hierarchical datasets allow for unique analytical opportunities, including the ability to track changes in the dataset at different levels of granularity, over time or between versions. For example, analysts can use hierarchies to uncover changes in the patterns of passengers flying in the United States over the last ten years, breaking down the data by states, cities, airports, and number of passengers. Exploring changes in travel patterns over time can help carriers make better decisions regarding their operations and long-range planning.

This paper describes TreeVersity2, a web-based data comparison tool that provides users with information visualization techniques to find what has changed in a dataset over time. TreeVersity2 enables users to explore data that can be inherently hierarchical or not (by categorizing them by their attributes). An interactive textual reporting tool complements the visual exploration when the amount of data is very large. The results of two case studies conducted with transportation domain experts along with the results of an exit questionnaire are also described. TreeVersity2 preloaded with several demo datasets can be found at (http://treeversity.cattlab.umd.edu) along with several example videos.

1. Introduction

Hierarchies help us organize and understand information. Examples of hierarchies include the evolutionary tree of species, the federal budget, and business organizational charts. Three examples are shown in Fig. 1. Many visualizations have been designed to represent, navigate, and analyze tree structures using node link representations (Plaisant et al., 1998),

* Corresponding author. Tel.: +1 301 405 0722.
E-mail addresses: john.guerra@gmail.com (J.A. Guerra-Gómez), packml@umd.edu (M.L. Pack), plaisant@cs.umd.edu (C. Plaisant), ben@cs.umd.edu (B. Shneiderman).
[1] Tel.: +1 740 591 0435.
[2] Tel.: +1 301 405 2768.
[3] Tel.: +1 301 405 2680.

http://dx.doi.org/10.1016/j.trc.2014.11.007
0968-090X/

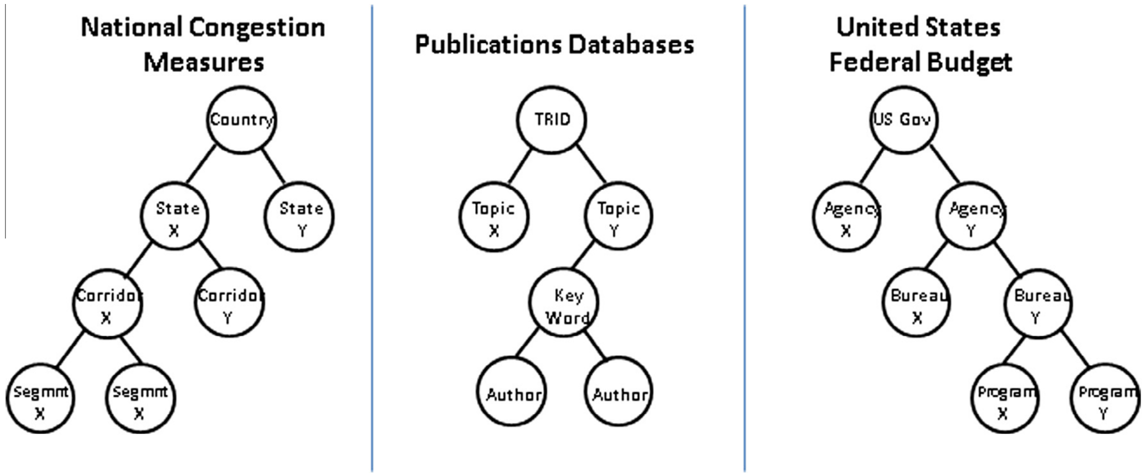

Fig. 1. Example of datasets viewed as hierarchies.

treemaps (Shneiderman, 1992), radial representations (Fisher et al., 2001) or icicle trees (Kruskal and Landwehr, 1983). Such techniques are now used in scientific or general-audience publications and websites to show transportation data. Our previous work (Guerra Gómez et al., 2013a) presented TreeVersity2 a visual comparison tool that introduced novel visualization techniques for finding changes in datasets over time looking at them as trees. This paper demonstrates how this novel visualization technique can be used by analysts to visually compare trees that change over time and thereby derive insights from transportation data.

Information visualization (InfoVis) techniques have been used to analyze a variety of transportation data (Beecham and Wood, 2014; Hughes, 2004; Pack et al., 2005; Pack, 2010; Wang et al., 2014) including airline passenger data (Hurter et al., 2014) and road congestion data (Cheng et al., 2013). The analysis of trends requires users to answer questions such as: how and where has congestion changed both nationally and locally? What new topics are emerging in our scientific publications? How has the department of transportation budget changed over the last ten years? How have airline maintenance budgets shifted from year to year, and have those changes affected safety? Those questions are challenging and require sifting through large amounts of data to evaluate multiple aspects of change, at multiple levels of aggregations. To address this challenge, a partnership between the Center of Advanced Transportation Technologies and the Human–Computer Interaction lead to the development of TreeVersity2 (Guerra Gómez et al., 2013b). TreeVersity2 uses visualization, an approach that complements alternative techniques used by others working with similar data, e.g. (Zografos and Madas, 2006) for airline passenger data or (Kamga and Yazıcı, 2014; Liu and Fei, 2010) for traffic congestion.

This paper first describes TreeVersity2 and illustrates how it allows transportation analysts to explore their data, then discusses the performance and implementation issues we encountered and how they were addressed – which will inform the development of future web-based visualization systems. Finally, two case studies conducted with transportation domain experts are described in detail, as well as the results of an exit questionnaire (based on a total of ten case studies).

2. Related work and definitions

While a significant amount of research has been conducted to visualize (Kruskal and Landwehr, 1983; Lamping, 1996; Robertson et al., 1991; Shneiderman, 1992) and explore (Card and Nation, 2002; Heer and Card, 2004; Parr et al., 2004; Plaisant et al., 1998) **single** trees, the problem of comparing **two** trees is significantly harder. We have identified and classified the following types of tree comparison (Fig. 2):

- **Type 0**: topological differences between two trees where the nodes only contain a label. Example: finding created, moved and removed topics between two versions of the TRB research publications hierarchical categorization system, without looking for number of papers.
- **Type 1**: positive and negative changes in leaf node values with aggregated values in the interior nodes (i.e. trees that can be visualized with a treemap (Shneiderman, 1992)) and no changes in topology. Example: comparing the change on the number of workers on the traffic agencies of the country, grouped by Agencies, Counties and Regions, given that no Agencies are created or removed.
- **Type 2**: positive and negative changes in leaves and interior node values with no changes in topology. Example: comparing the salaries in the Department of Transportation's organizational chart between two years, when no reorganization has occurred.