Event HyperGraph for analyzing large collections of text

Sam Yu-Te Lee*
University of California, Davis

Kwan-Liu Ma†
University of California, Davis

ABSTRACT

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

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

1 THIS IS MY INTRODUCTION

[1] is the best!

2 RELATED WORKS

3 METHODOLOGY

3.1 Preprocessing

The Methodology can work for any unstructured dataset

3.1.1 Summarization

Chatgpt for summarization

3.1.2 Document Embedding

OpenAI's embedding API

3.1.3 Major Participant Extraction

Chatgpt for major participant extraction and another model for entity linking

3.2 Modeling

3.2.1 Hierarchical HyperGraph Clustering

We organize the data into a hypergraph: nodes, hyperedges.

The hypergraph is clustered based on semantic and connectivity similarity: dual, Ravasz algorithm

Chatgpt to assign topics to each cluster

4 VISUALIZATION

4.1 Space Filling Curves

Introduce Gosper curve and generalized Hilbert curve, and how they are used for large graph layout

*e-mail: ytlee@ucdavis.edu
†e-mail: klma@ucdavis.edu

4.2 SFC for HyperGraph

Using the Gosper curve to layout the article graph

Concatenating four generalized Hilbert curve to layout the entity graph on the peripheral

4.3 Spacing Strategy

4.4 Border Approximation

4.5 Edge Bundling

ACKNOWLEDGMENTS

The authors wish to thank A, B, and C. This work was supported in part by a grant from XYZ.

REFERENCES

[1] G. Grinstein, D. Keim, and M. Ward. Information visualization, visual data mining, and its application to drug design. IEEE Visualization Course #1 Notes, October 2002.