

Cluster Visualization Techniques for Time-Varying Document Collections

Eric M. Cabral

Universidade de São Paulo (USP)
Instituto de Ciências Matemáticas e de Computação (ICMC)
Laboratório de Visualização, Imagens e Computação Gráfica (VICG)

cabral.eric@usp.br

June 22, 2018

Abstract

- Data clustering changes its structures over time and through user interactions
- Interaction with the data makes the visualization more effective and useful
 - A static view can show only one aspect of a dataset [Munzner and Maguire (2015)]
 - User-driven
- Lack* of specific time-varying data clustering visualization techniques
- General document domain

Keywords: interaction techniques, document visualization, text mining, visual analytics, time-varying

- 1 Introduction
- 2 Literature Review
- 3 Conclusions

Table of Contents

1 Introduction

2 Literature Review

3 Conclusions

Introduction – Dimensionality Reduction

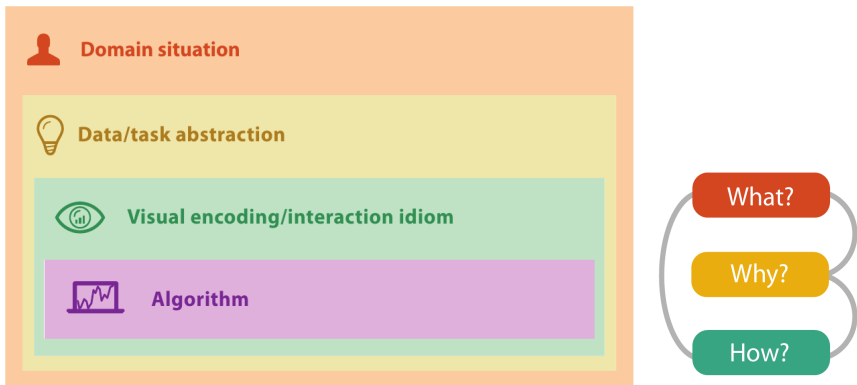
$$f : \mathbb{R}^m \rightarrow \mathbb{R}^n$$

$$m > n$$

- Dimensionality reduction (DR) techniques transform, or project, high-dimensional datasets into low-dimensional ones which can be more easily analyzed, interpreted, and visualized
- DR keep the data structure
 - Outliers
 - Clusters
- Information loss

Introduction – Visualization

Figure: Visualization design project



Fonte: Munzner and Maguire (2015)

Following Munzner's visualization design project framework:

- **Domain:** Time-varying document collections
- **Task and data abstraction:** Document clustering and state variation over time
- **Visual encoding/interaction idiom:** Yet to be specified
- **Algorithm:** Yet to be implemented

Three main interaction concerns:

- Individual and mass data tracking
- Data behaviour over time
- User interactions Tracking (events)*

Table of Contents

1 Introduction

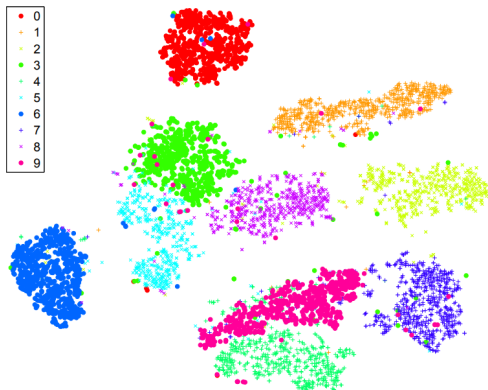
2 Literature Review

- Clustering Visualization
- Documents Visualization
- Time-varying Visualization

3 Conclusions

Clustering Visualization – t-SNE (2008)

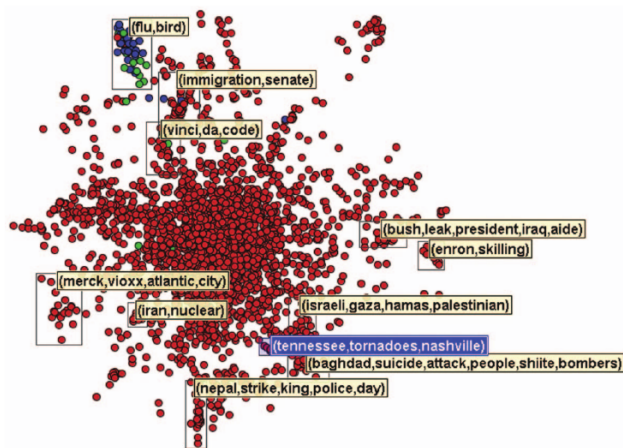
Figure: t-Distributed Stochastic Neighbor Embedding



Fonte: Van Der Maaten and Hinton (2008)

Clustering Visualization – LSP (2008)

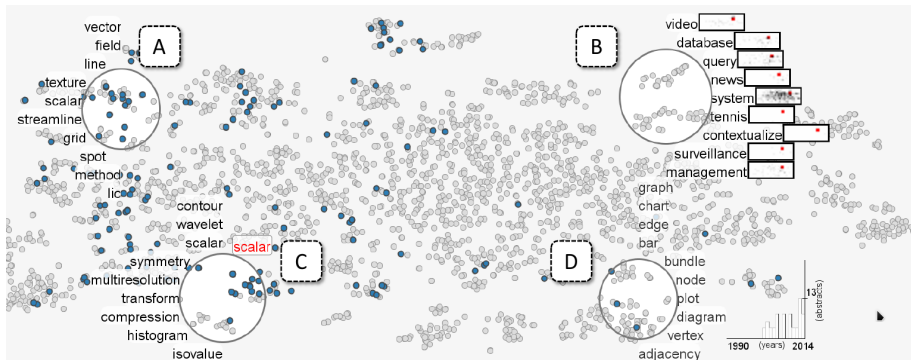
Figure: Least Square Projection



Fonte: Paulovich and Minghim (2008)

Document Visualization – DocuCompass (2016)

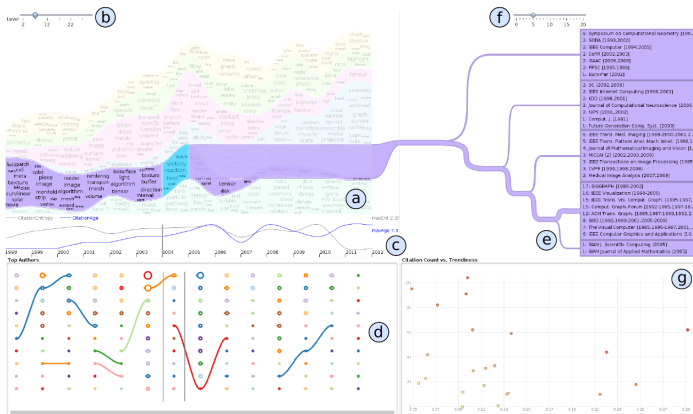
Figure: DocuCompass Overview



Fonte: Heimerl et al. (2016a)

Document Visualization – CiteRivers (2016)

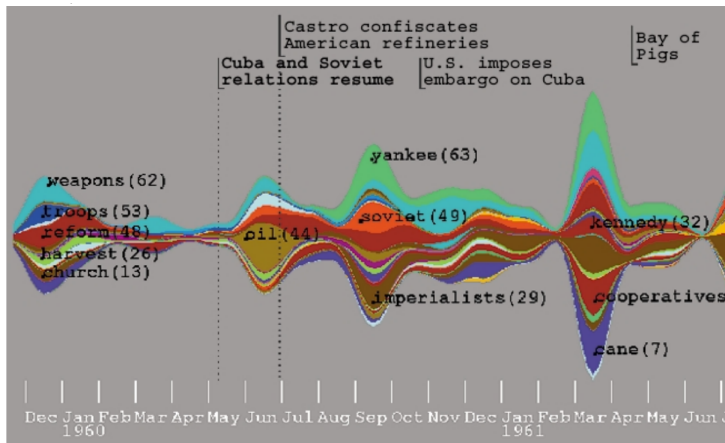
Figure: DocuCompass Overview



Fonte: Heimerl et al. (2016b)

Time-varying Visualization – ThemeRiver(2002)

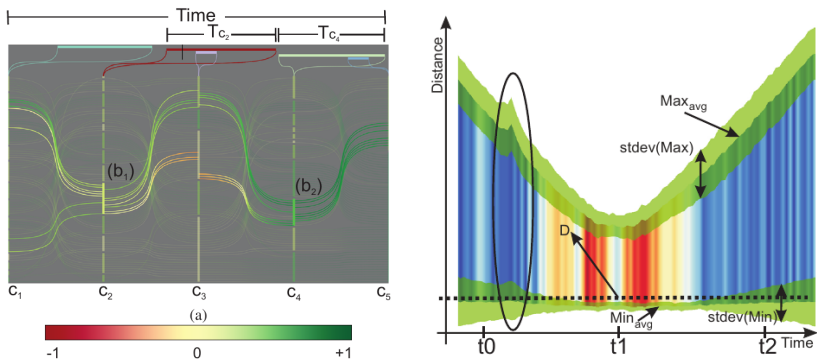
Figure: ThemeRiver – Stack graph



Fonte: Havre et al.

Time-varying Visualization – Turkey (2011)

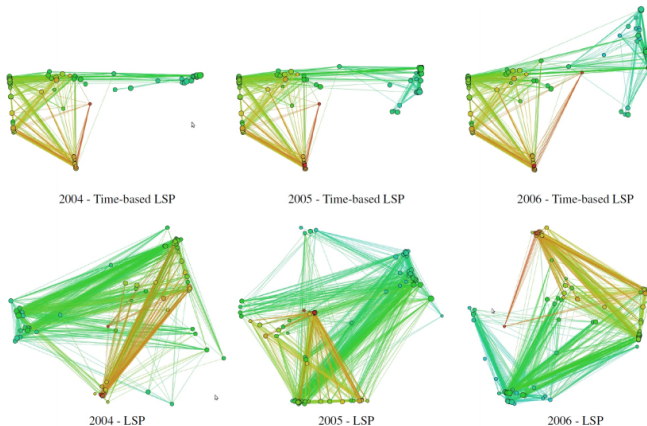
Figure: Cluster view and Temporal Signatures



Fonte: Turkey et al. (2011)

Time-varying Visualization – t-LSP (2011)

Figure: Time-based Least Square Projection



Fonte: Alencar (2012)

Table of Contents

1 Introduction

2 Literature Review

3 Conclusions

Conclusions

- The analysis is very limited due to the dinanimity nature of time-varying data collections
- There are many time-varying document collections visualization techniques, but none was able to track user interaction events
- This research lacks literature coverage to establish the state of the art in time-varying document collections visualization

- T. Munzner and Maguire, *Visualization analysis & design*, 2015, no. September.
- L. Van Der Maaten and G. Hinton, “Visualizing Data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- F. V. Paulovich and R. Minghim, “HiPP: A novel hierarchical point placement strategy and its application to the exploration of document collections,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1229–1236, 2008.
- F. Heimerl, Q. Han, S. Koch, and T. Ertl, “CiteRivers: Visual Analytics of Citation Patterns,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 190–199, 2016.
- , “CiteRivers: Visual Analytics of Citation Patterns,” *IEEE Transactions on Visualization and Computer Graphics*, 2016.

- S. Havre, B. Hetzler, and L. Nowell, “ThemeRiver: visualizing theme changes over time,” in *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*.
- C. Turkay, J. Parulek, N. Reuter, and H. Hauser, “Interactive visual analysis of temporal cluster structures,” *Computer Graphics Forum*, vol. 30, no. 3, pp. 711–720, 2011.
- A. B. Alencar, “Visualização da evolução temporal de coleções de artigos científicos,” p. 159, 2012.