

TDAview

An online visualization tool for topological data analysis

Kieran Walsh¹, Mircea A. Voineagu², Fatemeh Vafaei¹, Irina Voineagu¹

1. Department of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia
2. Department of Mathematics and Statistics, University of New South Wales, Sydney, Australia

Introduction

- Topological data analysis (TDA) has found applications for multiple types of biological data, from image analysis to microbial populations, gene expression analyses, and single cell RNA-sequencing [1-4].
- The most widely used version of TDA is the Mapper algorithm [5], which aims to identify a graph that characterizes the topological properties of a cloud of data-points
- TDAview is a free, open-source web tool for biologists and clinicians to investigate high dimensional data using the Mapper algorithm



Figure 1: Mapper algorithm (1 dimension, 8 bins)

Capabilities and parameters

Coloring and resizing networks based on cell-type, patient age or any other metadata you import is simple and displays the result in real-time.

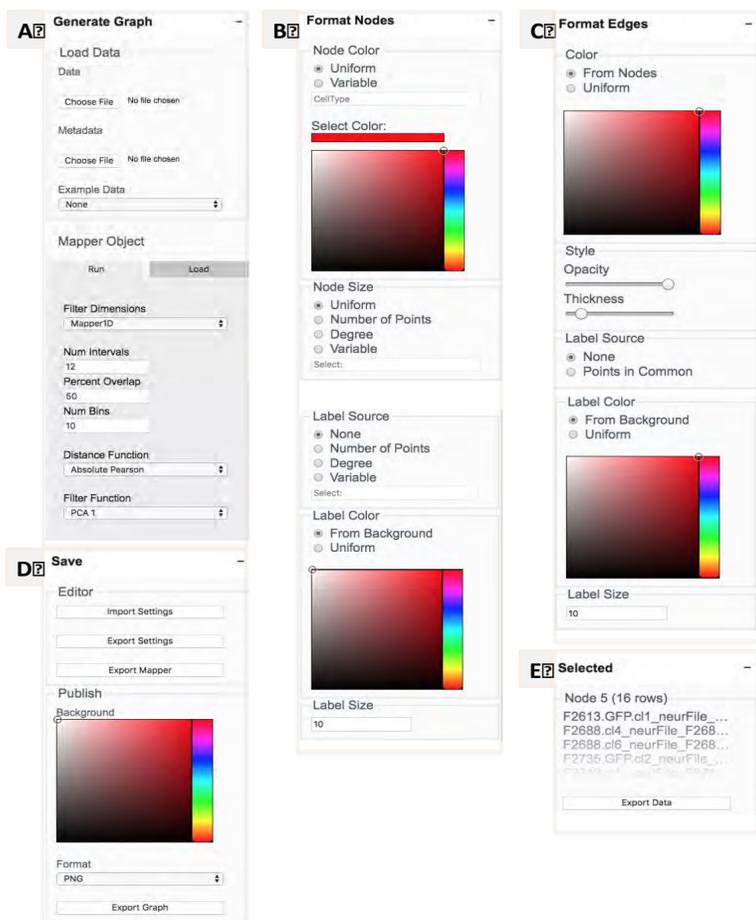


Figure 2A: Generate mapper object (Run tab) or load one in JSON format (Load tab)

Figure 2B, 2C & 2D: Modify topology using metadata and export

Figure 2E: Displays the contents of the currently selected node/cluster

Design and implementation

TDAview works entirely in browser and is implemented in JavaScript, hosted statically on GitHub.

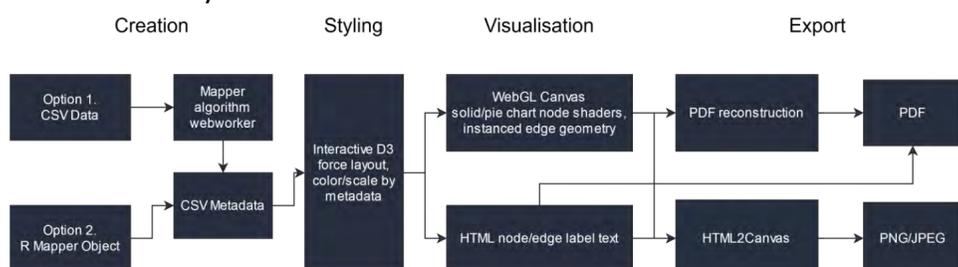


Figure 3: Technical workflow

Examples of use

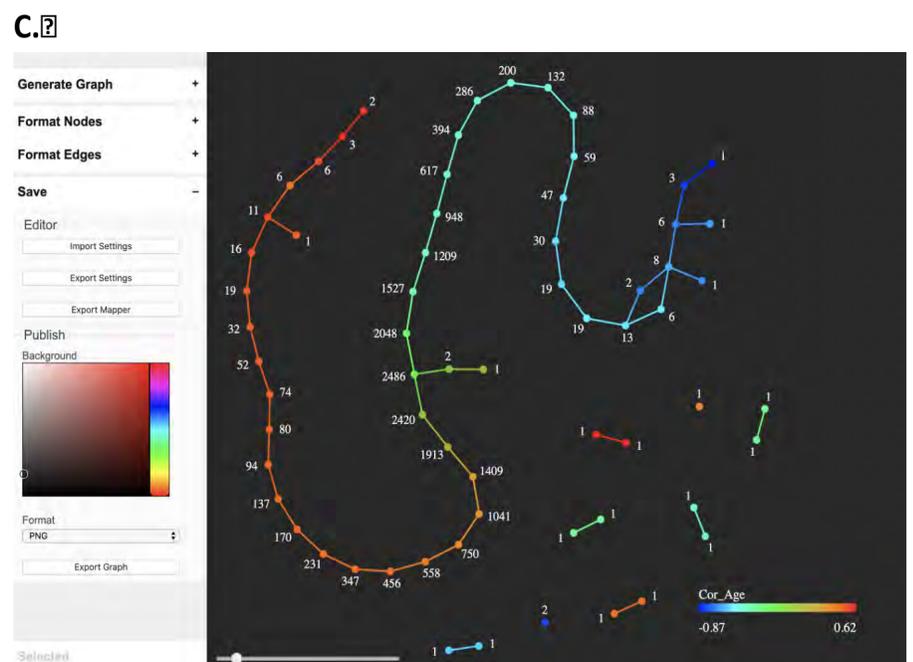
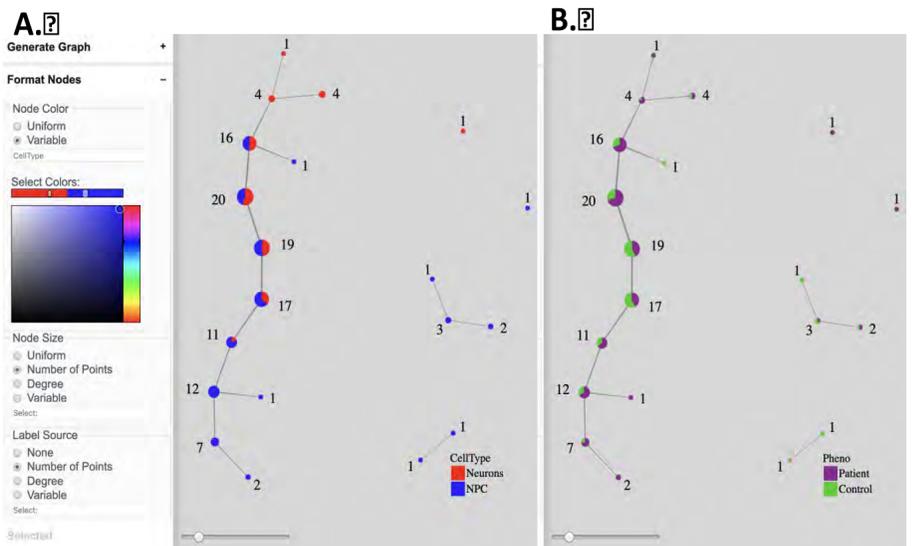


Figure 4A & 4B: Neuronal differentiation timecourse data (distance measure: $1 - |\text{Pearson correlation}|$, filter function f : first principal component (PC1), $n = 12$, $p = 50$, and $k = 10$)

Figure 4C: Gene expression data from prefrontal cortex (distance measure: Euclidean distance, filter function: PC1, $n = 12$, $p = 50$, and $k = 10$)

Limitations

TDAview has a pie chart category limit of 16 on most hardware. JavaScript PCA libraries also don't match native R/C++ speeds, which we have attempted to mitigate by offering a JSON import tool. In future this could enable other algorithms to be used.

Acknowledgements

This work was supported by an ARC Future Fellowship and a UNSW Scientia fellowship to IV.

References

1. Rizvi, A.H., et al., *Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development*. Nat Biotechnol, 2017. **35**(6): p. 551-560.
2. Garside, K., et al., *Topological data analysis of high resolution diabetic retinopathy images*. PLoS One, 2019. **14**(5): p. e0217413.
3. Liao, T., et al., *tmap: an integrative framework based on topological data analysis for population-scale microbiome stratification and association studies*. Genome Biol, 2019. **20**(1): p. 293.
4. Shnier, D., M.A. Voineagu, and I. Voineagu, *Persistent homology analysis of brain transcriptome data in autism*. J R Soc Interface, 2019. **16**(158): p. 20190531.
5. Singh, G.M., Facundo; Carlsson, Gunnar, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*. Eurographics Symposium on Point-Based Graphics, 2007. **SPBG**(SPBG07): p. 091-100.