

MDS-based Visual Survey of Biological Data Visualization Techniques

A. Kerren¹, K. Kucher¹, Y.-F. Li², and F. Schreiber^{2,3}

¹Department of Computer Science, Linnaeus University, Sweden

²Faculty of Information Technology, Monash University, Australia

³Department of Computer and Information Science, University of Konstanz, Germany

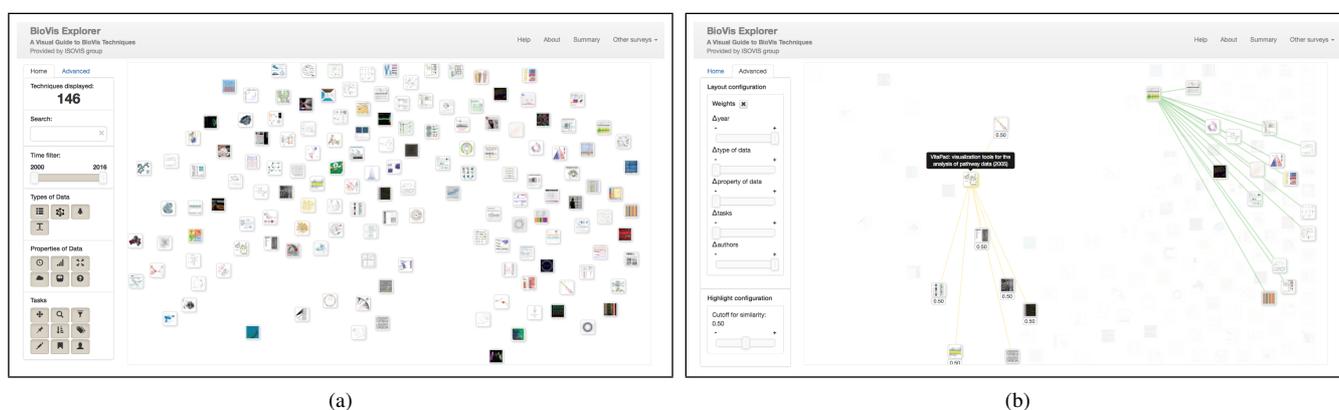


Figure 1: (a) The web-based user interface of our interactive visual guide to BioVis techniques, called BioVis Explorer. Each publication that presents a new BioVis technique is represented by a small rectangular thumbnail. The thumbnails are arranged in the 2D plane by using multidimensional scaling, i. e., similar techniques (according to a number of factors) are close to each other. By using the interaction panel on the left hand side, researchers can look for specific BioVis techniques and filter out entries with respect to a set of categories. The survey currently contains 146 categorized BioVis techniques published between 2000 and 2016. (b) Demonstration of interactions such as highlighting and selection pinning for a layout affected by modified distance weight factors. The links to similar techniques are displayed for the currently highlighted and pinned techniques in yellow and green, respectively.

Abstract

Data visualization is of increasing importance in the Biosciences. During the past 15 years, a great number of novel methods and tools for biological data visualization have been developed and published in various journals and conference proceedings. As a consequence, keeping an overview of state-of-the-art visualization research has become increasingly challenging for both biology researchers as well as visualization researchers. To address this challenge, we have reviewed visualization research for the Biosciences and created an interactive web-based visualization tool, the BioVis Explorer. BioVis Explorer allows the exploration of published visualization methods in interactive and intuitive ways, including faceted browsing and associations with related methods.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Graphical user interfaces (GUI)

1. Introduction

Data visualization is increasingly important in the Biosciences as stated by O'Donoghue et al. [OGG*10]. A large number of novel methods and tools for biological data visualization (BioVis) have been developed and published in various journals and conference

proceedings in the last 15 years. As a consequence, keeping an overview of state-of-the-art visualization research is crucial for both biology researchers (who would like to use such techniques for their individual analysis of problems) as well as visualization researchers (who would like to develop novel techniques). The challenge when learning more about biological data visualization is that

the methods available differ in a number of aspects, such as types and properties of the data to be visualized, interaction methods, and the concrete tasks which should be solved by the method. Usually only visualization experts have a good overview of these methods. To address this challenge and to provide a review of current visualization methods, we have created an interactive web-based visualization tool, called BioVis Explorer, which is freely available at

<http://biovis.lnu.se>

BioVis Explorer allows the investigation of visualization methods in intuitive ways based on filtering regarding, for example, types and properties of the data as well as associations with related visualization methods. In this work, we have focused on novel and dedicated visualization techniques for biological data sets proposed in publications from top-tier journals and conferences in the Biosciences (such as Bioinformatics and BioVis Conferences) and Information Visualization / Visual Analytics (such as IEEE TVCG and IEEE VIS) within a time range from 2000 to 2016.

2. Related Work

Interactive Visual Surveys/Guides Interactive web-based surveys address the amount and rate of emergence of research literature on visualization which are challenging for traditional survey articles. Since the introduction of the first survey websites on visualization of trees [Sch11] and time-dependent data [AMST11], many more have been launched, for example for text visualization [KK15], dynamic graph visualization [BBDW14], etc. There are also more generic approaches which may be instantiated with any scientific literature collection, such as the SurVis tool [BKW16]. Our proposed BioVis Explorer goes beyond these existing approaches by not just listing the literature in a specific order or providing standard filter features. We follow another layout strategy that spatially arranges the visual paper representations according to similarity.

Surveys on Biovisualization Techniques O'Donoghue et al. [OGG*10] greatly motivate the needs, trends and challenges of using interactive visualization techniques in the Biosciences. Kerren and Schreiber [KS14] present the state of the art in network visualization for the Life Sciences. Another example is the somewhat older article of Pavlopoulos et al. [PWS08] that also illustrates the functionality, limitations and specific strengths of network visualization tools. A similar orientation has been taken by the survey of Dinkla and Westenberg [DW12]. The minority of available surveys have chosen another core area, such as visualizing time-dependent biological data [SS14], visualizing live cell imaging [PKE16], or visualizing spatial multivariate medical data with glyphs [ROP11]. Neither do the aforementioned surveys provide a web-based interactive visual guide through the maze of various visualization techniques, nor do they encompass as many techniques and similarity measurements as proposed in this work.

3. BioVis Explorer: Design and Features

BioVis Explorer is an interactive web-based visualization tool that provides an electronic review of published visualization methods in interactive and intuitive ways, including faceted browsing and associations with related methods. Technically, BioVis Explorer is

based on a galaxy metaphor [AKS*02] and provides a coherent view of BioVis publications in order to achieve a low initial learning curve (see Fig. 1(a)). The overall design of BioVis Explorer is inspired by our previous interactive survey [KK15]. Its interface comprises a side panel with filters and statistics, and the main view with small thumbnails representing individual visualization techniques or papers, respectively. In contrast to many other interactive surveys discussed in Sect. 2, BioVis Explorer does not simply provide a thumbnail list ordered by a certain attribute, for instance, publication year. Instead, the 2D-layout of thumbnails is produced by multidimensional scaling (MDS) [MKB79] for the matrix of distances between the techniques which is based on differences in publication year, set of authors, and set of assigned categories. Thereby it is easy to get a quick overview of similar visualization methods whose thumbnails are located close to each other.

BioVis Explorer supports several user interactions including zooming and panning. By clicking on a thumbnail, the user can open a dialog box with detailed information on the publication as well as a list of similar techniques ordered by decreasing similarity. The user can also examine selected techniques and individually compare them with similar ones, as displayed in Fig. 1(b). Hovering over a technique thumbnail blurs the thumbnails for dissimilar techniques, and displays the links to similar techniques as well as the actual similarity value. Right-clicking a technique thumbnail pins it in order to provide persistent selection. In this way, the mental map of the thumbnail layout is not destroyed during the examination process.

The side panel shown on the left of Fig. 1(a) provides several filtering techniques. Filtering can be performed according to keyword search, publication year, as well as a broad taxonomy. The taxonomy describes important facets of BioVis techniques, organized into three high-level categories: *biological data types* (i.e., the structure of the data when collected or stored), *biological data properties* (i.e., specific characteristics of the single data records that are typically not exclusive), and *visualization tasks* that a user might want to perform with the visualization. These categories are themselves again divided into subcategories. An overview of all papers and their assigned categories can also be shown on demand.

The tool also provides more advanced settings for users interested in exploratory analysis. The users can change the weights of the distance factors when computing their dissimilarity, for instance, to reduce the importance of one or more factors, such as co-authorship. In consequence, the arrangement of thumbnails will be recomputed using the altered combination of factors, as displayed in Fig. 1(b), where all the weights except for year and authors have been minimized. It is also possible to adjust the value for the similarity cutoff/threshold to control which techniques will be considered similar or dissimilar during interactive exploration.

Acknowledgments: This work was partially supported by the framework grant “The Digitized Society—Past, Present, and Future” from the Swedish Research Council (Vetenskapsrådet) [grant number 2012-5659]. We also acknowledge the support of the Discovery Project grant “Visualisation of large, complex networks through small, beautiful diagrams” from Australian Research Council (ARC) [grant number DP140100077].

References

- [AKS*02] ANDREWS K., KIENREICH W., SABOL V., BECKER J., DROSCHL G., KAPPE F., GRANITZER M., AUER P., TOCHTERMANN K.: The InfoSky visual explorer: Exploiting hierarchical structure and document similarities. *Information Visualization* 1, 3-4 (2002), 166–181. doi:10.1057/palgrave.ivs.9500023. 2
- [AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*. Springer-Verlag London, 2011. URL: <http://www.informatik.uni-rostock.de/~ct/timeviz/timeviz.html>, doi:10.1007/978-0-85729-079-3. 2
- [BBDW14] BECK F., BURCH M., DIEHL S., WEISKOPF D.: The state of the art in visualizing dynamic graphs. In Borgo et al. [BMV14], pp. 83–103. URL: <http://dynamicgraphs.fbeck.com>, doi:10.2312/eurovisstar.20141174. 2
- [BKW16] BECK F., KOCH S., WEISKOPF D.: Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 180–189. doi:10.1109/TVCG.2015.2467757. 2
- [BMV14] BORGO R., MACIEJEWSKI R., VIOLA I. (Eds.): *EuroVis' 14 State-of-the-Art Reports* (2014), Eurographics. 3
- [DW12] DINKLA K., WESTENBERG M. A.: Network visualization in cell biology. *Tsinghua Science and Technology* 17, 4 (Aug. 2012), 365–382. doi:10.1109/TST.2012.6297584. 2
- [KK15] KUCHER K., KERREN A.: Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the 8th IEEE Pacific Visualization Symposium* (2015), PacificVis '15, IEEE, pp. 117–121. URL: <http://textvis.lnu.se>, doi:10.1109/PACIFICVIS.2015.7156366. 2
- [KS14] KERREN A., SCHREIBER F.: Network visualization for integrative bioinformatics. In *Approaches in Integrative Bioinformatics*, Chen M., Hofestädt R., (Eds.). Springer Berlin Heidelberg, 2014, pp. 173–202. doi:10.1007/978-3-642-41281-3_7. 2
- [MKB79] MARDIA K. V., KENT J. T., BIBBY J. M.: *Multivariate Analysis*. Academic Press, 1979. 2
- [OGG*10] O'DONOGHUE S. I., GAVIN A.-C., GEHLENBORG N., GOODSSELL D. S., HÉRICHÉ J.-K., NIELSEN C. B., NORTH C., OLSON A. J., PROCTER J. B., SHATTUCK D. W., WALTER T., WONG B.: Visualizing biological data—now and in the future. *Nature Methods* 7 (2010), S2–S4. doi:10.1038/nmeth.f.301. 1, 2
- [PKE16] PRETORIUS A. J., KHAN I. A., ERRINGTON R. J.: A survey of visualization for live cell imaging. *Computer Graphics Forum* (2016). online version published before print. doi:10.1111/cgf.12784. 2
- [PWS08] PAVLOPOULOS G. A., WEGENER A.-L., SCHNEIDER R.: A survey of visualization tools for biological network analysis. *BioData Mining* 1, 1 (2008), 12. doi:10.1186/1756-0381-1-12. 2
- [ROP11] ROPINSKI T., OELTZE S., PREIM B.: Visual computing in biology and medicine: Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers and Graphics* 35, 2 (Apr. 2011), 392–401. doi:10.1016/j.cag.2011.01.011. 2
- [Sch11] SCHULZ H.-J.: TreeVis.net: A tree visualization reference. *IEEE Computer Graphics and Applications* 31, 6 (Nov. 2011), 11–15. URL: <http://treevis.net/>, doi:10.1109/MCG.2011.103. 2
- [SS14] SECRIER M., SCHNEIDER R.: Visualizing time-related data in biology, a review. *Briefings in Bioinformatics* 15, 5 (2014), 771–782. doi:10.1093/bib/bbt021. 2