Visualization research at the Software Engineering and Information Systems Laboratory (SEIS)

Celmar Guimarães da Silva
School of Technology – University of Campinas
Limeira, São Paulo, Brazil 13484-332
Email: celmar@ft.unicamp.br
https://www.ft.unicamp.br/seis

***Abstract*—This paper summarizes twelve years of visualization research at the Software Engineering and Information Systems Laboratory (SEIS), located at the University of Campinas. This research comprises theoretical and applied research in visualization and embraces themes such as the development of matrix reordering algorithms, the creation of heuristics for visualization, and the application of visualization techniques to areas such as oil engineering and educational data mining.**

I. Introduction

The Software Engineering and Information Systems Laboratory (SEIS) is located at the School of Technology at the University of Campinas. Its research is currently focused on visualization, human-computer interaction, and software engineering areas. The laboratory is the home of the Information and Systems Engineering Group ("Grupo de Engenharia de Informação e Sistemas"), headed by Prof. Regina Moraes. Prof. Celmar G. da Silva has been supervising the visualization works of this lab since the creation of the research group in 2009. These works are focused on the reordering of matrices and graphs, educational data mining, and the application of visualizations to solve problems from areas such as environmental monitoring and oil engineering.

The following sections summarize the visualization projects of our laboratory over the last twelve years and their contributions.

II. Matrix Reordering research

The laboratory started its research on matrix reordering algorithms inspired by the use of heatmaps to show the participation of students in e-learning courses [1] and on the use of PQR-trees to solve permutation problems from computational biology. Based on these two inspirations, the laboratory proposed using PQR-trees to help permute rows and columns of matrix visualizations to reveal visual patterns. The research started with binary matrices [2] and then evolved to matrices with quantitative values [3]–[7]. Our research focused on revealing canonical data patterns [8] in matrices.

Our main contributions are summarized as follows:

- The *Matrix Reordering Analyzer tool* (MRA) [2] was created to compare the reordering algorithms according to evaluation measures applied to the output of these algorithms and also according to their execution time. This tool has been extensively used and enhanced in most of our researches related to matrix reordering.
- The *Feature Vector-based Sort* algorithm (FVS) [9] sorts the rows (columns) of an input matrix according to the mean values of each row (column). It can reveal Simplex and Equi-correlation patterns [8] on matrices.
- The *Block Reordering* algorithm [4] tries to reveal if a Block pattern [8] is present in some possible permutation of an input matrix. Its strategy is based on the equilibrium of zeroes and ones in each matrix column, which enables the choice of representative columns to guide the reordering process.
- The *Polar Sort* and *PSF* algorithms [5] [7] are focused on revealing Band and Circumplex patterns [8] on matrices. They are based on multidimensional projections of rows and columns of an input matrix on a plane, and on creating a list of the projected points according to their circular or arc format related to the underlying pattern of the input matrix. Both algorithms use these lists of rows and columns to reorder the matrix. The PSF's assymptotic time complexity $(O(n \log n))$ to reorder a $n \times n$ matrix is lower than the $O(n^3)$ time complexity of Polar Sort without reducing the quality of the resulting matrices.
- The *Hybrid Sort* method [6] combines the pattern-focused reordering algorithms Polar Sort, Block Reordering, and FVS and a matrix classifier to reveal the five canonical data patterns [8]. The method starts with the extraction of a feature vector of the input matrix. This feature vector is

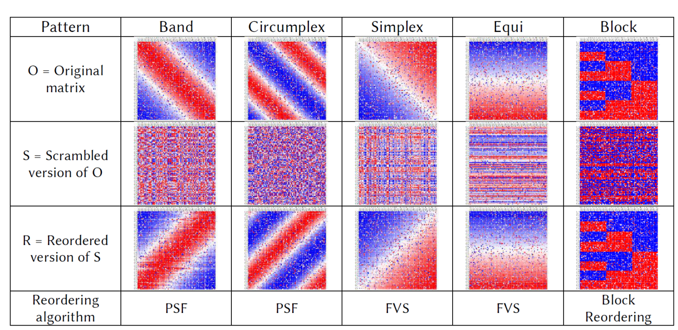

Fig. 1. Examples of the results of some pattern-focused reordering algorithms. Top: input matrices; middle: scrambled versions of these matrices; bottom: reordered versions of the scrambled matrices.

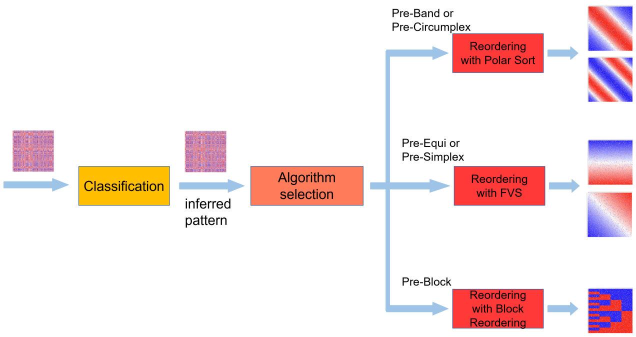

Fig. 2. Overview of the steps of the Hybrid Sort method.

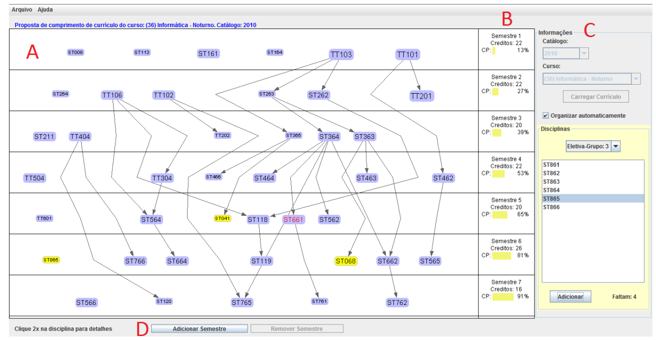

Fig. 3. CourseViewer's main user interface: (A) subjects, prerequisites and academic terms; (B) information about academic terms; (C) options for choosing courses and adding subjects; (D) options for managing academic terms. Source: [11].

processed by a decision tree that classifies it as a matrix that can be permuted to reveal one of the five mentioned patterns. According to the inferred pattern, one of the three algorithms is executed to reorder the matrix.

Fig. 1 exemplifies the results of reordering matrices with the five canonical data patterns using FVS, PSF, and Block Reordering. Red, white, and blue cells stand for high, intermediate, and low values, respectively. Fig. 2 represents the main steps of Hybrid Sort.

A current research also proposed the use of interactive heatmap matrices to present the relationship between pairs of variables. A heatmap matrix is to a heatmap just as a scatterplot matrix is to a scatterplot. It was proposed initially in a static version [10] and was extended recently to include interactive resources such as diverse ways to reorder the matrix variables, a filter based on association rules, and mechanisms to discretize quantitative variables so that they also can be represented by a heatmap matrix.

III. Educational Data Mining & Learning Analytics Research

The laboratory has been studying how to apply visualization to help users with distinct roles in the educational process: students, professors, and course coordinators. We proposed the CourseViewer tool [11], which aims to provide ways to visualize, handle and understand course curricula and student transcripts. The tool represents subjects and their prerequisites as nodes and edges of a layered directed acyclic graph represented as a node-link diagram. The horizontal layers represent the semesters, and the horizontal position of the nodes aims to minimize edge crossing. Some color mapping schemas can be selected, such as distinct colors for mandatory, elective, or extracurricular subjects; distinct colors for intervals of grades in each subject; or a color gradient for a measure of the relevance of a subject s for a given curriculum, according to the number of subjects that are reachable from s in the graph. CourseViewer provides interactive resources such as selecting, highlighting, adding, and deleting subjects; moving subjects from one semester to another one; and adding or removing semesters. Fig. 3 is an example of a course curriculum visualization.

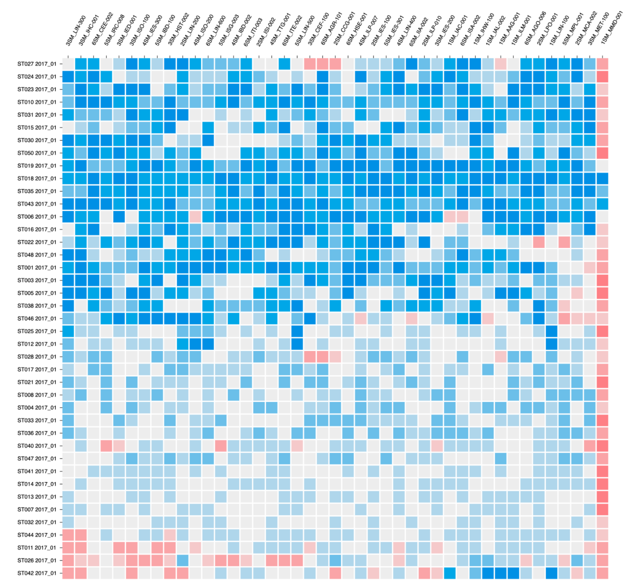

Fig. 4. A reorderable heatmap representing an overview of students grades per subject of a course. Red, white, and blue cells stand for low, intermediate, and high grades, respectively. Source: [12].

This research inspired our team to think about the potential that visualization techniques have to help to analyze not only a single transcript, but a set of transcripts that share some common features, such as the same starting year, the same base curriculum, or the same course. We developed a prototype [12] that enables the visualization of multiple transcripts as a reorderable heatmap whose rows, columns, and cells are mapped to students, subjects, and grades, respectively (Fig. 4). Current research, which is continuing this work, had conducted interviews with potential users to understand which data are relevant for professors and course coordinators to make better decisions regarding a course and its students. Visualization techniques to support the analysis of these data are under selection.

IV. Visualization Applied to Oil Engineering

In partnership with the Center for Petroleum Studies (CEPETRO/UNICAMP) and the Energy Production Innova-

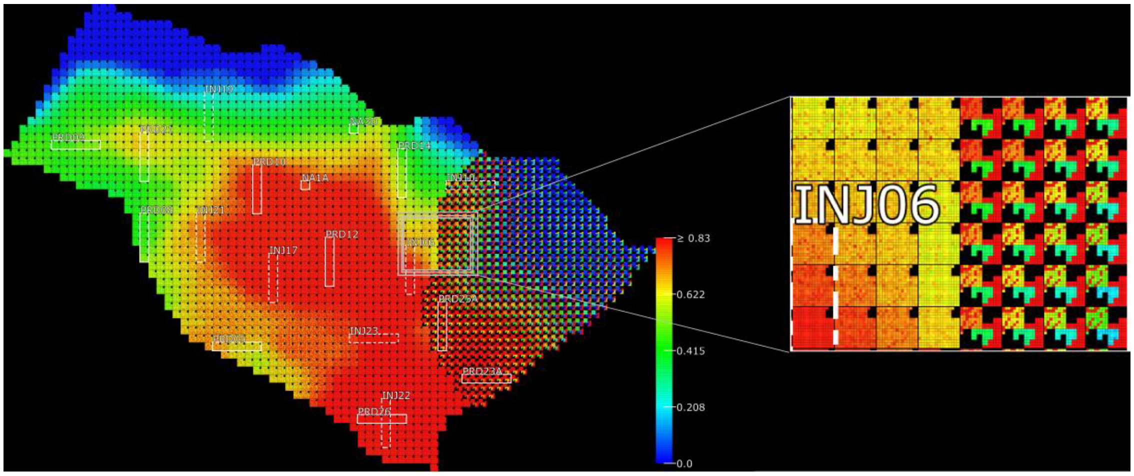

Fig. 5. Left: Pixelization-based visualization of oil saturation from 214 reservoir models (benchmark UNISIM-I-D). Right: Details of the region near to the injection well INJ06, with similar values for all the models regarding the cells at left, and groups of models with distinct values in the cells at right.

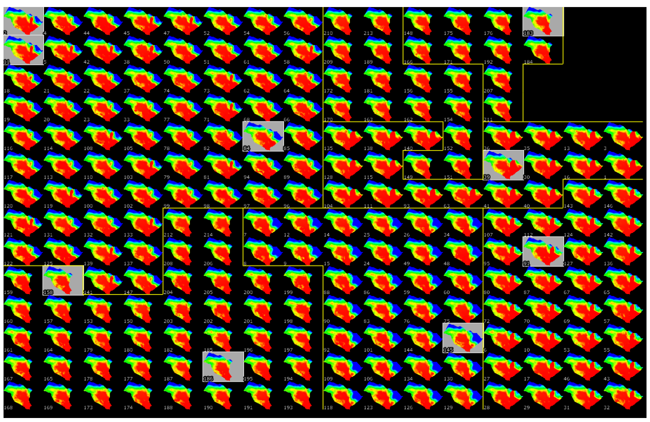

Fig. 6. Visualization of oil saturation from 214 reservoir models (benchmark UNISIM-I-D), based on Small Multiples technique, with 7 clusters of models. Representative models are highlighted.

tion Center (EPIC/UNICAMP), our lab has been applying visualization techniques to help analysts to understand data related to the oil engineering area [13] [14].

Oil reservoir development and management processes include a set of activities such as estimating the presence of oil, water, and gas inside a reservoir and defining strategies for well placement and opening. These activities deal with many intrinsic uncertainties, given that the analysts do not know exactly the geophysical properties of all the parts of a reservoir. The analysts may use sampling techniques to create a large set of reservoir models that represent those uncertainties. This set of models is usually filtered by history matching or data assimilation procedures, and then by the choice of representative models. This reduction is desired to make feasible the execution of flow simulations with these models – a computationally expensive task that is executed many times, due to the need to test and select a set of exploitation strategies.

In one of the works related to this theme [13], we propose two strategies for visualizing sets (ensembles) of reservoir models in a plane. The first approach (Fig. 5) uses Pixelization to combine inside each cell (i, j) of a reservoir heatmap the correspondent cells (i, j) from all models under analysis. The second approach (Fig. 6) uses Small Multiples to draw side-by-side 2D versions of all models, using a layout based on Peano-Hilbert curves and clustering. In both figures, red, green, and blue stands for high, intermediate, and low saturations, respectively.

V. Heuristics for visualization

Our laboratory has also been investigating the definition of heuristics for evaluating visualization software. Our starting work in this research theme [15] surveyed a set of 62 heuristics and guidelines from the InfoVis literature, and grouped these heuristics in order to create a reduced set with 15 heuristics. Later works (to be published) increased the number of heuristics to be grouped and involved visualization experts to collaborate on the task of creating this reduced set of heuristics.

VI. Other research themes

The following list summarizes other research themes developed by our lab:

- *Graph reordering* : we proposed strategies based on PQR-trees to permute elements of layers in a layered directed acyclic graph to reduce edge crossing, aiming to make these graphs more easily readable [16].
- *Visualization as support to portfolio selection activities*: we proposed visualizations based on interactive timelines that aim to help users to understand results of optimization procedures related to the allocation of large sets of tasks in an enterprise along time, with resource restrictions [17]. The visualization uses strategies such as clustering and semantic zooming for grouping tasks along time, and the use of interaction for drill down inside clusters, and for changing starting times of tasks.

- *Visualization applied to environmental monitoring*: we proposed the use of reorderable heatmaps to help to monitor an area in the ocean that received dredging material for several months. The sampling sites were monitored based on a list of sampling campaigns. Each sample was analyzed for a set of chemical and physical parameters [18], and also for a set of biotic indexes to assess benthic communities' structure and temporal variability [19]. We proposed alternative overviews of the dataset, based on reorderable heatmaps focused on a selected parameter, sample, or campaign.

VII. Conclusion

Visualization research is challenging, but at the same time rewarding. This paper summarizes some challenging and rewarding projects from almost 12 years of visualization research in the Software Engineering and Information Systems Laboratory at the University of Campinas. Its main contributions comprise theoretical results, such as creating a set of pattern-focused algorithms for matrix reordering and researching compact sets of heuristics for visualization. Besides, these contributions also embrace the application of visualization concepts and techniques to help analysts solve problems in areas such as environmental monitoring, portfolio selection, and more recently, educational data mining and oil engineering.

Acknowledgment

The author would like to thank the various FAPESP and FAEPEX/PRP/UNICAMP grants over these years, as well as all the current and former students of our research group. All this work would not be possible without them.

References

[1] C. G. da Silva, "An information visualization-based approach for exploring databases: A case study for learning management systems," in *Innovative Approaches of Data Visualization and Visual Analytics*, M. Huang and W. Huang, Eds. IGI Global, 2014, pp. 288–315. [Online]. Available: http://doi:10.4018/978-1-4666-4309-3.ch014

[2] C. G. da Silva, M. F. de Melo, F. P. Silva, and J. Meidanis, "PQR sort: using PQR trees for binary matrix reorganization," *Journal of the Brazilian Computer Society*, vol. 20, no. 1, 2014.

[3] C. G. Silva, B. F. Medina, M. R. Silva, W. H. Kawakami, and M. M. N. Rocha, "A fast feature vector approach for revealing simplex and equi-correlation data patterns in reorderable matrices," *Information Visualization*, vol. 16, no. 4, pp. 261–274, 2017.

[4] A. M. Santos, B. F. Medina, and C. G. Silva, "Block reordering: Um algoritmo para evidenciar padrão block em matrizes reordenáveis," in *Workshop of Undergraduate Works (WUW) in the 29th Conference on Graphics, Patterns and Images (SIBGRAPI'16)*, J. dos Santos and A. Maciel, Eds., São José dos Campos, SP, Brazil, 2016. [Online]. Available: http://gibis.unifesp.br/sibgrapi16

[5] C. G. Silva, "Polar Sort: combining multidimensional scaling and polar coordinates for matrix reordering," in *Proceedings of the International Conferences Interfaces and Human Computer Interaction 2019; Game and Entertainment Technologies 2019; and Computer Graphics, Visualization, Computer Vision and Image Processing 2019*, Porto, Portugal, 2019, pp. 239–243, (Best paper award).

[6] ——, "Hybrid Sort - a pattern-focused matrix reordering approach based on classification," in *Proceedings of the International Conferences on Computer Graphics, Visualization, Computer Vision and Image Processing 2020, Big Data Analytics, Data Mining and Computational Intelligence 2020, and Theory and Practice in Modern Computing 2020*, Online, 2020, pp. 35–43, (Best paper award).

[7] ——, "Using fast multidimensional projections to reveal band and circumplex patterns in reorderable matrices," in *Proceedings of the International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing 2021; Connected Smart Cities 2021; and Big Data Analytics, Data Mining and Computational Intelligence 2021*, Online, 2021, pp. 3–10, (Best paper award).

[8] L. Wilkinson, *The Grammar of Graphics*. Springer Science & Business Media, 2005.

[9] M. M. N. Rocha, P. K. Roque Jr., and C. G. da Silva, "Circumplex sort: a two-phase method for revealing circumplex data patterns in reorderable matrices," in *Proceedings of International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing 2017 and Big Data Analytics, Data Mining and Computational Intelligence 2017*, 2017, pp. 181–188.

[10] M. M. N. Rocha and C. G. da Silva, "Heatmap matrix: a multidimensional data visualization technique," in *Proceedings of the 31st Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018. [Online]. Available: http://sibgrapi.sid.inpe.br/rep/sid.inpe.br/sibgrapi/2018/10.18.02.37

[11] M. R. de Oliveira, P. J. C. de Mendonça, A. F. G. Leardine, V. P. Gonçalves, B. F. Medina, and C. G. da Silva, "CourseViewer – a prototype for visualizing undergraduate courses and student transcripts," in *Workshop of Undergraduate Works (WUW) in SIBGRAPI 2013 (XXVI Conference on Graphics, Patterns and Images)*, D. M. Maria Andreia Rodrigues, Ed., Arequipa, Peru, august 2013. [Online]. Available: http://www.ucsp.edu.pe/sibgrapi2013/eproceedings/

[12] R. T. C. Barros, T. G. Mendes, and C. G. Silva, "Use of reorderable matrices and heatmaps to support data analysis of students transcripts," in *Companion Proceedings of the 32nd SIBGRAPI — Conference on Graphics, Patterns and Images*. Porto Alegre, RS, Brasil: SBC, 2019, pp. 219–222. [Online]. Available: https://sol.sbc.org.br/index.php/sibgrapi_estendido/article/view/8334

[13] C. G. da Silva, A. A. S. Santos, and D. J. Schiozer, "Visualization of ensembles of oil reservoir models based on pixelization, small multiples and reservoir similarities," in *Offshore Technology Conference Brasil*, 2019. [Online]. Available: https://doi.org/10.4043/29788-MS

[14] C. G. Silva, L. A. A. Meira, A. A. S. Santos, and D. J. Schiozer, "Using reorderable matrices to compare risk curves of representative models in oil reservoir development and management activities," in *Proceedings of the International Conferences on Computer Graphics, Visualization, Computer Vision and Image Processing 2020, Big Data Analytics, Data Mining and Computational Intelligence 2020, and Theory and Practice in Modern Computing 2020*, 2020, pp. 19–26.

[15] M. R. de Oliveira and C. G. da Silva, "Adapting heuristic evaluation to information visualization – a method for defining a heuristic set by heuristic grouping," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) - Volume 3: IVAPP*, 2017, pp. 225–232. [Online]. Available: https://www.scitepress.org/papers/2017/61332/61332.pdf

[16] J. R. Marchete Filho and C. G. Silva, "Using PQR-trees for reducing edge crossings in layered directed acyclic graphs," in *Workshop of Works in Progress (WIP) in SIBGRAPI 2013 (XXVI Conference on Graphics, Patterns and Images)*, S. M. Alejandro C. Frery, Ed., Arequipa, Peru, august 2013. [Online]. Available: http://www.ucsp.edu.pe/sibgrapi2013/eproceedings/

[17] C. G. da Silva, J. Meidanis, A. V. Moura, M. A. Souza, P. Viadanna, M. R. de Oliveira, M. R. de Oliveira, L. H. Jardim, G. A. Costa Lima, and R. S. de Barros, "An improved visualization-based approach for project portfolio selection," *Computers in Human Behavior*, vol. 73, pp. 685–696, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563216309128

[18] C. G. da Silva and G. de Aragão Umbuzeiro, "Enhancing visual analysis of environmental monitoring data through an information visualization-based prototype," in *Anais do XXX Congresso da Sociedade Brasileira de Computação*, Belo Horizonte, 2010, pp. 575–584. [Online]. Available: https://doi.org/10.4043/29788-MS

[19] M. Flynn, C. G. Silva, P. Y. G. Sumida, A. Z. Guth, B. G. R. Alves, and G. Umbuzeiro, "Vissed software as a tool in monitoring programs for benthic data interpretation," *RevInter*, vol. 8, pp. 22–35, 2015. [Online]. Available: http://autores.revistarevinter.com.br/index.php?journal=toxicologia&page=article&op=view&path%5B%5D=216&path%5B%5D=428