

ProRank – Software for Partial Order Ranking

Stefan Pudenz, Criterion–Evaluation & Information Management, Mariannenstr. 33, D-10999
Berlin (Germany)

Abstract

This paper presents a computer program for partial order ranking. It comprises a user-friendly surface and can be executed on most of the common operating systems. Import and export filters allow the exchange with other formats. Different graphical illustrations of Hasse diagrams facilitate the visual analysis of partially ordered sets (posets) as well as the handling for presentation purposes. Beyond the partial order ranking data matrices can be analysed and filtered by means of Boolean and logical operators. Therefore the program has data mining qualities too. A component manager enables an easy navigation between the different program windows, which is in particular important when several data matrices are to be considered and investigated respectively. Posets can be analysed and characterised by different methods and parameter, for instance sensitivity analysis, order ideals, filters and stability considerations.

Introduction

Meanwhile, partial order ranking is used for solving many scientific problems as shown by the multitude of publications in different scientific journals. A rough classification could be the following:

- Partial orderings in chemistry, e.g. [1],
- Environmental Science, e.g. [2,3],
- Ecology, e.g. [4,5,6].

Little is known about applications of partial order ranking beyond the scientific field, for instance as a routine instrument for industrial applications or public authorities.

User-oriented working needs software instruments dealing with partial orders. As far as the author knows, many researchers use in-house software (e.g. RANA [7], TK-HASSE [8] or even spread-sheet formats [9]) or the more professional program WHASSE [10]. Systat Inc. offers a statistic program that includes a partial order scalogram analysis [11]. Therefore WHASSE seems to be the only program (licence for research and teaching only) having a graphical interface and many features for analysing partially ordered sets, for instance:

- Display of Hasse diagrams
- Similarity index for comparing two partial ordered sets
- Linear extensions and their analysis
- Sensitivity analysis of evaluation with respect to the attributes used
- Several properties characterising the Hasse diagram of a partially ordered set like stability, comparabilities, incomparabilities, etc.
- Classification of elements
- Aggregation of attributes
-

For details of the program and the above mentioned features the reader is referred to [10,12,13]. However, WHASSE will probably not be enhanced in the future [14]. Moreover, its features are not sufficient for large data sets where often the use of e.g. many data sheets at the same time, data exploring or data mining facilities are required.

In this paper the program ProRank[®] for partial order ranking focused on user-friendly features is presented.

Technical structure and general features

In order to provide a platform independent solution the program is coded in Java[™], i.e. it can be executed on most of the common computer systems. The only condition is the installation of the Java[™] Runtime Environment.

Specific flexibility within the program structure was taken into account by following known model-view-controller-concepts, i.e. a strict separation of the three levels

- Model
- MetaGUI
- GUI

Model: this level and package respectively, contains all classes (objects) which present the interior structure of the governed data, independent of the GUI (graphical user interface). For instance the classes responsible for the data matrix and the partial ordered set belong to the model.

MetaGUI: This level contains all classes based on the model and providing the functions for modifications of the model, e.g. calculating a sensitivity analysis or specific properties like the number of comparabilities in a Hasse diagram. These classes do not contain any data for saving. What exactly can be defined as a Meta-component is controlled by a specific interface. Further examples for Meta-components are e.g. the table that allows a view on the data matrix, or Diagram that allows a view on the Hasse diagram.

GUI: This level is responsible for real presentation and its layout on the screen. It contains the component manager which accepts meta-components and translates their content into Java-components.

This partitioning into Model, MetaGUI and GUI enables high flexibility with regard to modifications of specific features. For instance, the visual design of the program (user interface) can be changed by replacing or modifying GUI-classes in the way that no programming at the level of Model or of MetaGUI is necessary. Furthermore it enables changes in data management (e.g. caching) without the need for changes at the MetaGUI and GUI. Finally additional functionalities can be implemented by writing a new Meta-component without touching the MetaGUI or GUI.

Program Overview

Example data

To facilitate the program overview and to show features, an example file about toxicity test results from different sediments samples of the German river Rhein will be introduced.

The file consists of 29 samples which are differentiated in sediment core and surface samples. Each sample is characterised by results of toxicity investigations concerning the

microbiological inhibition of different organisms. The aim here is to evaluate/rank different sediment samples by means of their toxic characteristics.

Data exchange and formats

To enable data exchange with other programs an import/export filter for the MS Excel® format is implemented, because this is the most common table format of Windows® applications. Additionally ASCII files can be read.

Basic statistical information

Sometimes in ranking procedures it is useful to get additional information about basic statistical properties. ProRank enables display of maximum, minimum, range, mean, median, variance and standard deviation of each attribute and element respectively. Figure 1 shows the main and the table window after opening the example data. The "attribute info" contains the statistical properties of a selected attribute. Furthermore, by choosing another table view statistical properties of both attributes and elements are displayed next to the identifiers (not shown in figure 1).

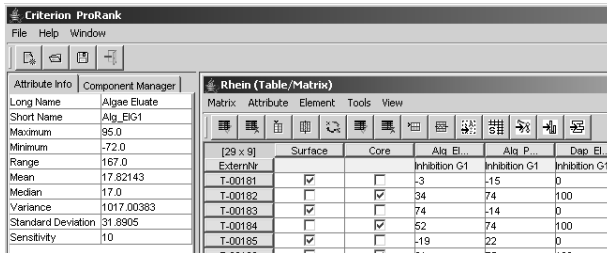


Figure 1: Main and table window of ProRank after opening the example file. Statistical properties of a selected attribute are listed in the attribute info or in an extended table view (not shown).

Context specific selection of attributes or elements

In certain cases elements may be additionally characterized by Boolean variables. Here, for instance the samples in the example file are distinguished between sediment surface and core and therefore it is interesting to compare the differences between the evaluations by means of each sample type. Then the first step is a data mining step, i.e. a context specific selection of elements which are fulfilling certain criteria – here, to be sediment's core sample. Figure 2 shows the dialog for element selection by using Boolean algebra and/or logical operators. If

the data table already contains Boolean variables, as shown in the table view in figure 1 (true and false are always indicated by checkmarks), then all of them are listed in the register by their long name, here for example, "core" and "surface" samples (not shown in figure 2). To select all sediment surface samples the negation operator is chosen twice indicating "select all elements which are not not-sediment samples". Additionally, only those elements should be selected where the microbiological inhibition of the algae test (in the eluate sample) is greater than 20%.

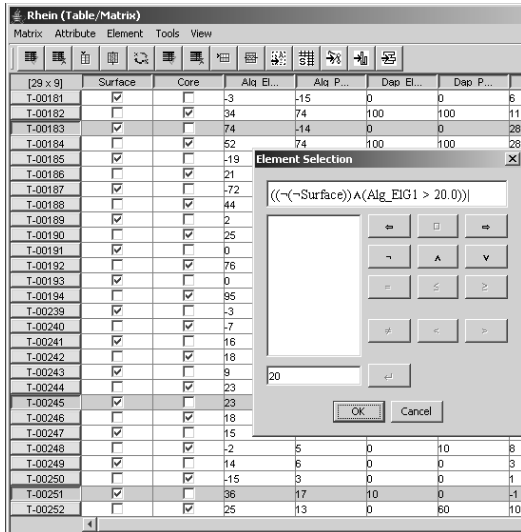


Figure 2: Context specific selection of elements.

Grouping and aggregation of attributes

In case only a part of the attribute set (information base) is of interest for evaluation or simple aggregations are to be performed then the grouping/aggregation dialog provides any features. Figure 3 shows a group called "eluate" that contains all tests of the example file performed as eluate samples and additionally the nematode test. All generated groups/aggregations are arranged in a tree. For aggregations one can choose between four methods: sum, arithmetic

mean, weighted sum and weighted mean. By selecting attributes or groups/aggregations in the left-side register, the following functions can be used:

- Creating a new table/matrix
- Calculating a Hasse diagram
- Sensitivity analysis

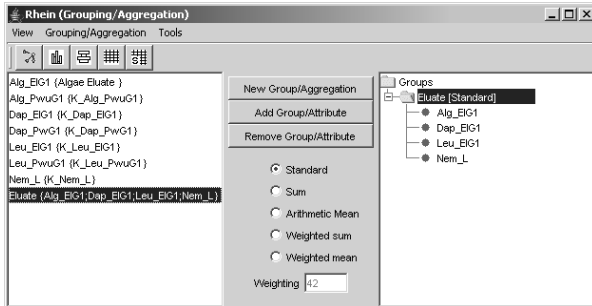


Figure 3: Grouping/aggregation dialog.

Working with Hasse diagrams

Hasse diagrams and partial orders respectively, can be visualised in three different modes (Figure 4):

1. Circles with identifiers for the elements
2. Bar diagrams for each element indicating the values of each attribute
3. Level diagram

The first one is the usual presentation of Hasse diagrams. Comparabilities between elements are indicated by lines between the circles and the elements are arranged in levels, where elements are assigned to the highest level possible. Basically elements with higher attribute values are located above those having lower values.

By visualisation of elements as bar diagrams differences between elements concerning their attribute values can be identified (see figure 4). For instance, by selecting the two incomparable elements T-00247 und T-00251 a new Hasse diagram can be calculated that simplifies a detailed visual identification of differences between the elements.

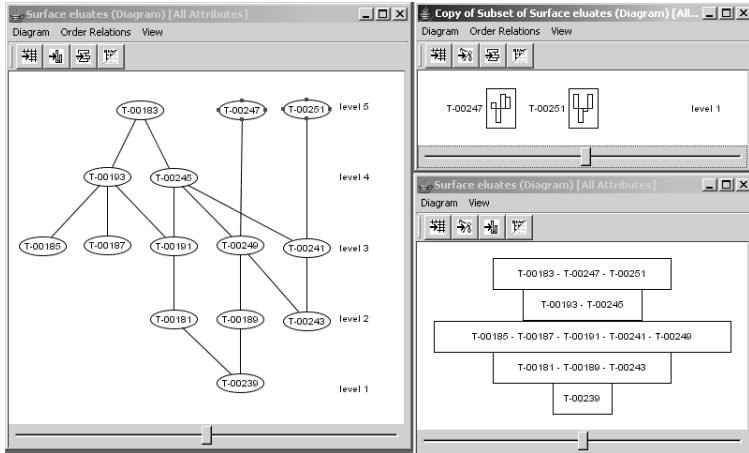


Figure 4: Visualization of Hasse diagrams and partial orders respectively.

Analyzing Hasse diagrams

The diagram window has an additional info panel that contains the following information about the Hasse diagram:

- The "level tree" gives an overview about the elements in each level.
- By the option "attribute values", the values of one or more selected elements can be shown. Additionally the selected elements can be sorted by their names or by their attribute values.
- The option "maximal chains" shows the longest sequence of lines in the Hasse diagram. For example in figure 4 the Hasse diagram has three longest chains, namely the sets $\{T-00183, T-00193, T-00191, T-00181, T-00239\}$, $\{T-00183, T-00245, T-00181, T-00239\}$ and $\{T-00245, T-00249, T-00189, T-00239\}$.
- "Properties" contains parameter, which characterize the partially ordered set (for details see also [8,10]), namely
 - Local comparabilities: all comparabilities counted for a selected element (see e.g. element T-00241 in figure 5).
 - Global comparabilities: all comparabilities of the partially ordered set.

- Internal comparabilities: all comparabilities counted within a selected group of elements of the Hasse diagram. For example, when the elements $\{T-00193, T-00187, T-00183\}$ are selected as set of interest, the internal comparabilities of this "sub-poset" are three.
- External comparabilities: comparabilities between two groups of elements, which have been selected from the Hasse diagram. For example, regarding the two sets $\{T-00185, T-00193, T-00187\}$ and $\{T-00181, T-00239\}$ the external comparabilities are counted as two, where the external incomparabilities are two.
- With respect to incomparabilities the terms local, global, internal and external have the same meaning as for comparabilities.
- Stability quantifies the influence of omitting or adding an attribute on ranking.

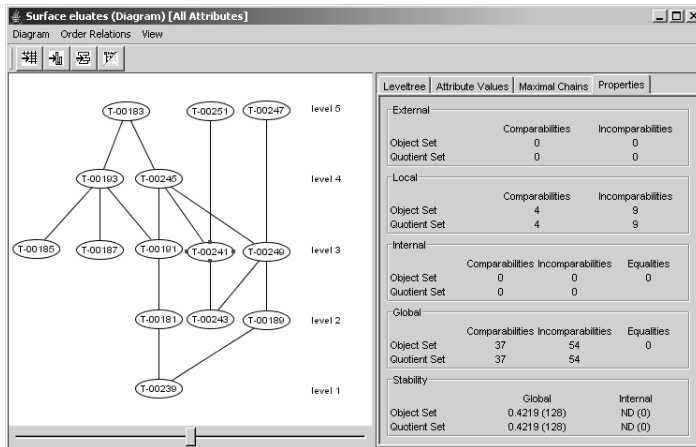
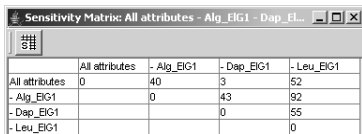


Figure 5: Properties which characterize the partial ordered set.

When those elements are of specific interest, which are located lower than a certain element (key element) the menu "order relations" provides the option "order ideal" to highlight or to create a new diagram that contains all elements that can be reached from the selected element(s) by a downward path ("successor set", see also [10]). In case all those elements are of specific interest, which are located higher than one or more selected elements, the option

"order filter" selects all elements (ten "predecessors") that can be reached from the selected element(s) by an upward path.

Sometimes it may be useful to search for important attributes. The sensitivity analysis quantifies the influence of attributes on the ranking by omitting each attribute successively and measuring the difference to the original Hasse diagram based on the whole attribute set. This calculation is based on the W-matrix or dissimilarity matrix (for details, see e.g. [10]) and the larger the matrix-entries, the greater is the difference between the Hasse diagrams. Figure 6 shows the result of the sensitivity analysis of the visualized poset in figure 5 (because the dissimilarity matrix is a symmetrical matrix, only the right side entries are shown). It is seen, that the Hasse diagram based on the attributes Alg_ELG1 and Dap_ELG1 (i.e. omitting attribute Leu_ELG1) has the greatest distance to the Hasse diagram based on all attributes. Therefore attribute Leu_ELG1 is most important for the ranking compared to the other attributes, whereas omitting attribute Dap_ELG1 results in minor changes of the Hasse diagram.



	All attributes	- Alg_ELG1	- Dap_ELG1	- Leu_ELG1
All attributes	0	40	3	52
- Alg_ELG1		0	43	92
- Dap_ELG1			0	55
- Leu_ELG1				0

Figure 6: Sensitivity matrix/analysis of the visualized poset in figure 5.

Data and project management

ProRank provides a component manager that gives an overview of all opened windows during a ranking session. It enables the navigation between all windows, i.e. Hasse diagrams, tables/matrices and dialogs. Moreover, connections between subsets of tables can be easily reproduced (e.g. subset of the data matrix "surface eluates" in figure 7).

Summary and Outlook

ProRank provides several features that allow to deal with partial order ranking and Hasse diagrams, respectively. In particular, it facilitates operations in routine work where often large data matrices are the basis and data mining features are required. Moreover, by the tree-structure of the component manager several data matrices can be investigated at the same

time, i.e. during a ranking session, without loosing the survey of possible relations between them.

An important step for the future is the implementation of tools for a preprocessing of data, e.g. classification methods like cluster analysis. A detailed description how to use fuzzy clustering methods can be found in [13,14]. Furthermore, in order to facilitate decision processes and to come up to a total order, the concept of the averaged rank and the ranking probability developed by [15] will be considered in the next version of the program.

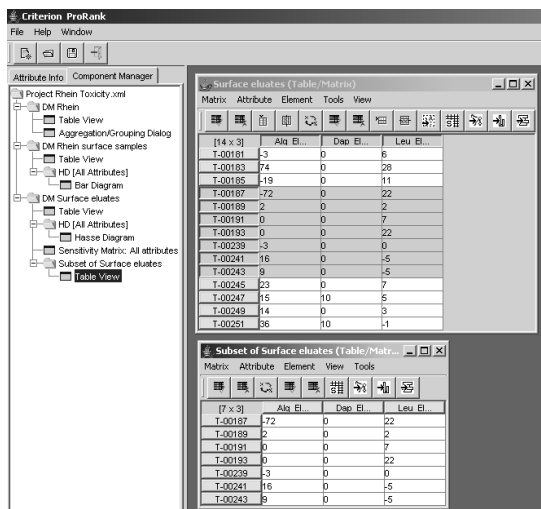


Figure 7: Component manager.

References

- [1] Klein, D.J. (2000). Prolegomenon on Partial Orderings in Chemistry: MATCH Commun. Math. Comput. Chem. 42, 7-21.
- [2] Sørensen, P.B., Brüggemann, R., Lerche, D.B., Voigt, K., Welzl, G., Simon, U., Abs, M., Erfmann, M., Carlsen, L., Gyldenkerne, S., Thomsen, M., Fauser, P., Mogensen, B.B., Pudenz, S., Kronvang, B. (2003). Order Theory in Environmental Science. Integrative Approaches. The 5th workshop held at the National Environmental Research Institute (NERI),

Roskilde, Denmark, November 2002. National Environmental Research Institute, Denmark. 161pp – NERI Technical Report no. 479, <http://technicalreports.dmu.dk>.

[3] P. B. Sørensen, R. Brüggemann, L. Carlsen, B. B. Mogensen, J. Kreuger, S. Pudenz (2003). Analysis of Monitoring Data of Pesticide Residues in Surface Waters using Partial Order Ranking Theory. *Environ. Tox. & Chem.*, 22 (3), 661-670.

[4] Ade, M., Brüggemann, R., Mess, A., Frahnert, S. (2004). Organismische Biologie als Grundlage für die Bewertung von Umweltauswirkungen durch Eingriffe in die Landschaft. - Pilotstudie mit Hilfe der Hasse-Diagramm-Technik-. *UWSF - Z. Umweltchem. Ökotox.* 16: 105-112.

[5] Brüggemann, R., Walz, N., Brauns, M., Ostendorf, W. (2004). Beitragsserie: Seeufer, ein vergessenes Ökoton; Beitrag 3: Gedanken zum Schutzziel "Artengemeinschaften". *UWSF - Z. Umweltchem. Ökotox.* 16: 48-56.

[6] Vogt, R.D., Akkanen, J., Andersen, D.O., Brüggemann, R., Chatterjee, B., Gjessing, E., Kukkonen, J.V.K., Larsen, H.E., Luster, J., Paul, A., Pflugmacher, S., Starr, M., Steinberg, C., Schmidt-Kopplin, P., Zsolnay, A. (2004). Key Site Variables Governing the Functional Characteristics of Dissolved Natural Organic Matter (DNOM) in Nordic Forested Catchments. *Aquatic Sci.* 66: 195-210

[7] Pavan, M. (2003). Total and Partial Ranking Methods in Chemical Sciences. 1-277. Doctoral Thesis. University of Milan - Bicocca, Milano Chemometrics and QSAR Research Group, Dept. of Environmental Science (www.disat.unimib.it), Italy.

[8] Behrendt, H., Altschuh, J., Sixt, S., Gasteiger, J., Höllering, R., Kostka, T. (1999). A unified Approach to Exposure Assessment by Computer Models for Degradation Reactions and Soil Accumulation: The Triazine Example. *Chemosphere* 38, 1811-1823.

[9] Sørensen, P.B. (2004). Personal Communication.

[10] Brüggemann, R., Halfon, E. (1995) Theoretical Base of the Program "Hasse". GSF-Bericht 10/95, München.

[11] SYSTAT, Systat Software Inc. (2000). California, USA, www.systat.com.

[12] Brüggemann, R., Halfon, E., Welzl, G., Voigt, K., Steinberg, C. (2001). Applying the Concept of Partially Ordered Sets on the Ranking of Near-Shore Sediments by a Battery of Tests. *J. Chem. Inf. Comput. Sci.* 41, 918-925.

[13] Brüggemann, R., Bücherl, C., Pudenz, S., Steinberg, C. (1999). Application of the Concept of Partial Order on Comparative Evaluation of Environmental Chemicals. *Acta Hydrochim. Hydrobiolog.* 27 (3), 170-178.

- [14] Brüggemann, R. (2004). Personal communication.
- [15] Pudenz, S., Brüggemann, R., Luther, B., Kaune, A., Kreimes, K. (2000). An Algebraic/Graphical Tool to Compare Ecosystems with Respect to their Pollution V: Cluster Analysis and Hasse Diagrams. *Chemosphere* 40 (12), 1373-1382.
- [16] Luther, B., Brüggemann, R., Pudenz, S. (2000). An Approach to Combine Cluster Analysis with Order Theoretical Tools in Problems of Environmental Pollution. *MATCH Commun. Math. Comput. Chem.* 42, 119-143.
- [17] Brüggemann, R., Sørensen, P.B., Lerche, D., Carlsen, L. (2004). Estimation of Averaged Ranks by a Local Partial Order Model. *J. Chem. Inf. Comput. Sci.* 44, 618-625.