Introduction to partial order theory exemplified by the Evaluation of Sampling Sites

Rainer Brüggemann[1] and Lars Carlsen[2]

[1] Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department: Ecohydrology, Germany

[2] Awareness Center, Roskilde, Denmark

Abstract

The first part of this chapter gives a detailed introduction to partial order ranking and Hasse Diagram Technique (HDT). Thus, the construction of Hasse diagrams is elucidated as is the different concepts associated with the diagrams. The analysis of Hasse diagrams is disclosed including structural analysis, dimension analysis and sensitivity analysis. Further the concept of linear extensions is introduced including ranking probability and averaged rank. The evaluation of sampling sites is, in the second part of the chapter, used as an illustrative example of the advantageous use of partial order ranking and Hasse Diagram Technique.

When a ranking of some objects (chemicals, geographical sites, river sections etc.) by a multicriteria analysis is of concern, it is often difficult to find a common scale among the criteria and therefore even the simple sorting process is performed by applying additional constraints, just to get a ranking index. However, such additional constraints, often arising from normative considerations are controversial. The theory of partially ordered sets and its graphical representation (Hasse diagrams) does not need such additional information just to sort the objects.

Here, the approach of using partially ordered sets is described by applying it to a battery of tests on sediments of the Lake Ontario. In our analysis we found: (1) the dimension analysis of partially ordered sets suggests that there is a considerable redundancy with respect to ranking. The partial ranking of the sediment sites can be visualized within a two-dimensional grid. (2) Information, obtained from the structure of the Hasse diagram: For example six classes of sediment sites have high priority, each class ex-

hibits a different pattern of results. (3) The sensitivity analysis identifies one test as most important, namely the test for Fecal Coliforms/*Escherichia coli*. This means that the ranking of samples is heavily influenced by the results of this specific test.

Introduction

Overview

In the present chapter an alternative way of analysing objects, which are characterized by several quantities is presented. Hence, instead of examining the variance (for example leading to principal component analysis) or the distances among objects (for example leading to cluster analysis), we focus on the use of partial ordering in ranking. An important aspect within the concept of partial ordering is the visualization by Hasse diagrams.
More specifically, we study

- the system of comparabilities and incomparabilities between objects which arises, if an order relation between them is defined
- how to set priorities and to detect the pattern which identify objects of high priority
- how to define logical non-contradictory sequences and
- how the selection of criteria influences the ranking of a set of objects.

Beyond this we analyse

- the role of the structure of the Hasse diagram, i.e. levels, hierarchies, articulation points
- the role of order preserving maps among partial orders and especially those order preserving maps whose results are linear orders and
- how we can derive an averaged rank, probability distributions from them and how structural properties of the Hasse diagram can be detected in probability distributions.

Further is discussed, the important concept of

- poset dimension, and
- latent variables are related to partial orders.

Partial order rankings may advantageously be visualized through Hasse diagrams. The program „WHASSE“ allows the construction of Hasse diagrams and provides several tools, which helps to analyse partially ordered sets (Brüggemann et al. 1999 a). A historical and personal view about the development of the programs around partial order is given by Halfon see p. 385. A matrix **W** is introduced, that quantify the importance of the single criteria on the eventual ranking. Additional to the contributions in this book, several textbooks and monographs and journal's publications are recommended to the reader (see reference list: "Introductory references", p. 393).

Partial order

Introductory remarks

Hasse diagrams show the relations of partially ordered sets (posets). In the following is explained why partial order is a useful concept on ranking.
Ordering is a logical way to give objects a structure: If for example chemical substances are characterized by their persistence then these substances can be sorted according to the increasing persistence, the sequence of substances corresponds to one characteristic number, namely the persistence. Often however, a single number is not sufficient to characterize objects. For example not only the persistence but also the bioaccumulation of a chemical substance may be important to explain the environmental behaviour of the substance. For further examples see contributions of this book.

Common to these examples is that each object (geographical sites, waste disposal sites, databanks, chemicals, managing options) is characterized by more than one quantity. Objects that are characterized by several quantities (we call them “attributes“ -see later for details-) often cannot be ordered, because there are conflicts between their attributes. Metaphorically we are talking of comparing apples and oranges.

An example may help to understand this. We may have five objects {A, B, C, D, E} characterized by e.g. their environmental persistence „P“ and by their ability to bioaccumulate „B“. As often is the case both attributes do not behave parallel, i.e. it is not automatically given that a persistent substance also is the most bioaccumulating.

We can arrange the five substances according to P or B (Fig. 1):

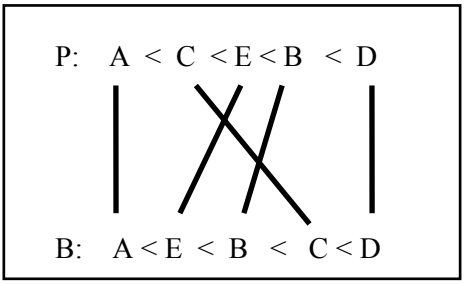


Fig. 1. "Permutation diagram": Two sequences of objects according to two different characteristics.

The type of diagram in Fig. 1 is called a permutation diagram (Urrutia 1989). It shows that there are inversions between the two sequences. Some objects will mutually exchange their positions in dependence which quantity is used to define the sequence (for example C, E). Some other objects do not change their relationships to others, if the sequence defining quantity (here: persistence or bioaccumulation) is changed (For example: A < D or E < B independent whether the persistence or the bioaccumulation is selected; other examples can also be found). Obviously some „rest of order" remains, if both quantities are considered at the same time and this fact motivate the term „partial order". Within the given example of five objects partial ordering arises because more than one quantity is used to characterize the single substances. This is often the case, where the complexity of nature prevents the use of a single ranking index (therefore many applications can be found in biology, ecology, ecotoxicology, and chemistry as disclosed through various chapters of this book. Partial order is further a typical tool within operation research, many decision support systems are based -at least implicitly - on partial order. For example in versions of ELECTRE (Roy 1972, 1990) or PROMETHEE (Brans & Vincke 1985, Brans et al. 1986, Heinrich, 2001) a partial order is at least an interim step (see also chapter by Brüggemann et al., p. 237). An access to recent literature may also be found in (Colorni et al. 2001, Lerche et al. 2002).

Obviously, the concept of partially ordered sets appears rather useful in environmental sciences. The „usual" order, namely the order in which each object can be compared with each other, can be considered as a special case of partial order, i.e., the term "linear" or "total" order is used.
Permutation diagrams become confusing if many objects are included and especially if more than two attributes characterize the objects. In such cases a corresponding number of sequences may arise and for each pair of

sequences a permutation diagram can be drawn. Instead of this troublesome procedure which leads to m·(m-1)/2 pairs of permutation diagrams (m attributes used) the technique of Hasse diagrams provides a useful tool for visualization. From the permutation diagram (Fig. 1) it can be concluded that A < C < D and A < E < B < D, respectively, whereas we cannot say anything concerning the relations between C and E and B, respectively, if both persistence and bioaccumulation are taken into account. Thus, the partially ordered set of the five objects (cf. Fig. 1) is visualized in a "Hasse diagram" (Fig. 2).

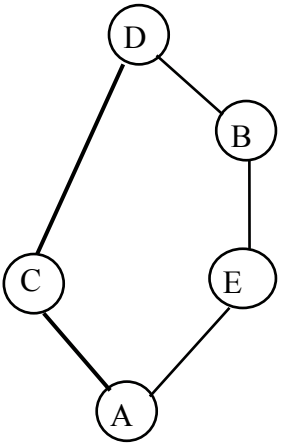


Fig. 2. The Hasse diagram as an alternative to the permutation diagram shown in Fig. 1

The name "Hasse diagram" becomes popular by the German mathematician Helmut Hasse, who worked in Marburg, Berlin and Hamburg, see also chapter by Halfon, p. 385. Often this kind of diagram is simply called "line diagram" or even only "the diagram" (Rival 1985b).

The rationale of using Hasse diagrams

The concept of partial order is described in the chapter of El-Basil, p. 3. Therefore we concentrate ourselves on the specific order relation we are using here, which is known as "Hasse Diagram Technique". In this technique we specifically consider any component of a sequence separately, as it bears its own valuable information with respect to the evaluation. Techniques, motivated by the work of Muirhead 1900, 1906 and Karamata, see

Beckenbach & Bellmann 1971, or by Young diagrams (Ruch 1975, Ruch & Gutman 1979) may be useful too. However, applying these techniques, the components of the sequence, i.e. the attributes would loose their individual meaning, which often is disadvantageous. (with respect to Young diagrams see the chapters by El-Basil, p. 3 and Seitz, p. 367.)

The basis of the Hasse Diagram Technique (HDT) is that we can perform a ranking without the use of a single ordering index (called a "ranking index"), i.e. we rank objects by maintaining all information about them. If an ordering index were used to force the object into a linear order, then information is lost. For example, an object might be ranked higher according to one criterion but lower according to another. Two objects might not be ordered unambiguously because their data are "contradictory" to each other. This ambiguity is not immediately evident when we use a ranking index, still worse: by using a ranking index the two attribute can compensate each other. That means a "bad" value in one attribute can be compensated by a "good" value in another one. Metaphorically speaking you can put one hand in boiling water and one hand in ice water. Discomforting? Yes! However, on an average basis you should feel quite comfortable! Such kinds of potential compensations or conflicts among attributes are immediately evident in a Hasse diagram.

Many problems are governed simply by comparisons, i.e. by the analysis of the order-relation. Typical examples can be found in textbooks of chemistry, when concepts like electronegativity, hardness or softness of compounds, etc is discussed. Many other problems are reducible to an order - relation. Often for example objects may be characterized by a binary bit pattern, representing whether a property is given or not. For example existence or non-existence of chemical functional groups lead to a binary bit pattern, for which a partial order can be defined (see for an example the chapter by Klein & Ivanciuc, p. 35). Partial orders help to analyse Quantitative Structure Activity Relationships (Randić 2002, Brüggemann et al. 2001), see also chapters by Carlsen, p. 163 and Pavan et al., p. 181 and references therein). Other examples are biomarker responses on certain stress factors in ecosystems (see for example, Brüggemann et al. 1995a, 1995b) and the analysis of data sources, see chapter by Voigt and Brüggemann, p. 327 and references therein.

To explain partial order and its visualization by Hasse diagrams, some useful theoretical notations are given in the following section.

Prerequisites

- *Criteria* comprise both quantitative and qualitative properties. Often it is useful to define a criteria hierarchy: Starting with a general criterion, which is hardly quantifiable, one looks for subcriteria to specify the general one, the subcriteria in turn may further specified, until a set of precise criteria is found, which may be quantified by attributes.
- *Attributes* are quantitative, measurable data. We denote these attributes as q_1, q_2,..., q_m. It is useful to define the information basis of the evaluation, *IB* to be the set of these attributes: $IB = \{q_1,q_2,...,q_m\}$. Some authors denote attributes as descriptors or possibly parameters. These terms are used synonymously.
- A *case* is a subset of selected attributes, taken from the ground set of attributes, *IB*. The attributes are specific to the problem. Each case corresponds to exactly one Hasse diagram. Thus, a given set of attributes induces a Hasse diagram. More definitions will follow in the text as the need arises.
- An *object* is the item of interest that may be characterized by attributes. Examples of objects can be chemical substances, or geographical sites (see chapter by Myers et al., p. 309), or strategies (see chapter by Simon et al., p. 221) etc. Objects are ranked graphically by Hasse diagrams (see for example Fig. 2). Generally the objects are considered to belong to a set "*E*". Therefore the objects are also often called "elements" and *E* is called a *ground set* or *object set*. The ground set corresponding to the Hasse diagram in Fig. 2 is thus $E = \{A, B, C, D, E\}$ (note: set *E* but element E). We assume that we have n elements of the set *E*.
- *Data* are the numerical values corresponding to each criterion by which a given object is characterized.
- *Equivalent objects* in Hasse diagrams: Different objects that have the same data with respect to a given set of attributes. Equality with respect to a given set of attributes defines an *equivalence relation*, "$\Re$". Objects having the same values of all their attributes form disjoint subsets of *E*, the *equivalence classes*. An equivalence class with only one object is called a *singleton* and is called trivial. The equivalence classes can be considered as elements of a set, the quotient set $E/\Re$. Usually the partial order is based on the quotient set and -if necessary- the equivalent elements are associated with that vertex, where a representative element out of the equivalence

class is drawn. Examples will be given below. Further details, see Patil & Taillie 2004 or Brüggemann & Bartel 1999.

- The *cardinality* of a (finite) set is the number of elements of the set, denoted by card *G* for a set *G*.
- *Numerical representation of objects*: Objects are considered to be elements of the object set *E*. Each object is characterized by attributes. We can create a table where the rows represent the objects and the columns the data of each object corresponding to the column-defining attribute
- Taken an element of *E*, the corresponding row consisting of the data of $q_1,...,q_m$ is often called a *tuple*, and abbreviated by q.
- *Attribute profile or pattern*: If the order of attributes is fixed then the sequence of attribute values for a given object x can be thought of as visualized by a bar diagram. This we have in mind when we are speaking of a profile or pattern.

Further - more specific - terms are explained later.

Graphs and Hasse diagrams

The construction of Hasse diagrams:

A set that has an order relation is called a *partially ordered set (poset)*. An order for a set, for example for the set *E*, is denoted by $(E, \leq)$, the set *E* often being called the ground set (of objects). As the application of partial order, presented here, is based on attributes, just *IB* influences the partial order. Therefore, we often write (*E*, *IB*). If the quotient set is used, then we write $(E/\Re, IB)$.

Partially ordered sets can be visualized through Hasse diagrams, which are quite useful if not too many objects are included. Let a and b be two elements of the object set *E*. Each object is characterized by a set of attributes. The relation '≤' between a and b is valid, if and only if this relation holds for <u>all</u> attributes of a and b. In other words: $a \leq b$, if all components of the tuple of a are smaller or equal to the corresponding component of the tuple of b. With help of the notation $q_j(i)$ with i the index for any element of *E*, and j as index for any attribute of *IB* we give a formula:

$$a, b \in E: a \leq b : \text{if and only if } q_j(a) \leq q_j(b), \text{ for all } q_j \in IB \qquad (1)$$

We call equation (1) the generality principle, because this equation defines dominance of b over a if all properties of b confirm the ≤-relation.

To illustrate the above, an example may be useful. Consider three objects a, b, c. They are characterized by two attributes, as Table 1 shows.

Table 1. Fictitious example

	q_1	q_2
a	1	1
b	1	2
c	2	1

Obviously a < b and a < c, respectively. With respect to the first attribute: a = b and with respect to the second attribute: a < b. Therefore a < b. A similar argument holds for the a-c relation. Objects for which the ≤-relation holds are *comparable* to each other. Often it is useful to have a shorthand notation for comparable objects (without specifying the orientation). Thus, if a < b, or b < a, we write a ⊥ b.

However, the relation ≤ does not hold for the objects b and c, because with respect to attribute q_1: b < c, and with respect to attribute q_2: b > c. Hence, objects that cannot be compared with each other, like b, c are called *incomparable*. A shorthand notation to describe two incomparable objects is b || c.

Cover relation:

If there is no element „x“ of *E*, for which a ≤ x ≤ b, x ≠ a, b, a ≠ b holds, then a is *covered* by b, or b *covers* a. Often the cover relation is referred to by its own symbol <•. Obviously in our example a <• b and a <• c, the corresponding graphical representation is given in Fig. 3.

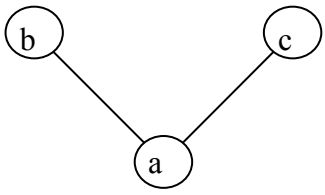


Fig. 3. Visualization of the order relation, induced by the data matrix, shown in Table 1.

Partial orders can be visualized in different ways, see also Chapters written by El-Basil, p. 3 and Seitz, p. 367. An interesting variant can be found in the chapter by Myers et al., p. 309). Other presentations are discussed in Neggers & Kim 1998. In the present chapter, the construction of Hasse diagrams is explained according to the software WHASSE (Brüg-

gemann et al. 1999 a) and is performed with the help of the cover relation as follows:

1. *E* may be represented by a configuration of circles and with an identifier for the objects within and each circle is located in the two-dimensional plane.
2. Note that the program WHASSE only displays a representative within the circle; other objects, having equal data tuples are shown in an extra field of the screen.
3. If a cover-relation holds, then a line between the corresponding object-pair is drawn. The covering pair is oriented corresponding to the ≤-relation.
4. The covered object in the ≤-related pair is located at a lower position on the page. (Alternatively we can, instead of the connecting line segment, draw an oriented arrow, beginning at the covering object and directed towards the covered object; in this case the locations in the two-dimensional plane of the Hasse diagram can be selected arbitrarily. In the practice it is more convenient to select the positions in the plane of the figure, according to the cover-relation.) By this step the lines become an *orientation*, for example "good-bad" or "high-low". See also in chapter by Helm, p. 291.
5. Finally, not all line segments for which the ≤- relation holds are to be drawn. Because of the logical rule of transitivity (which holds by definition for partial orders) lines corresponding to the pair x, z with $x \leq y$ and $y \leq z$ concluding $x \leq z$ are omitted. They do not present a cover-relation.

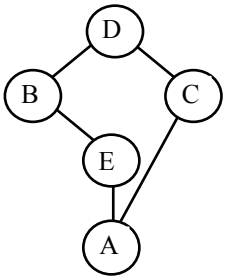


Fig. 4. The Hasse diagram of the example of Fig. 1 and Fig. 2, respectively, drawn by the program WHASSE

In order to introduce further concepts another Hasse diagram is drawn (Fig. 5):

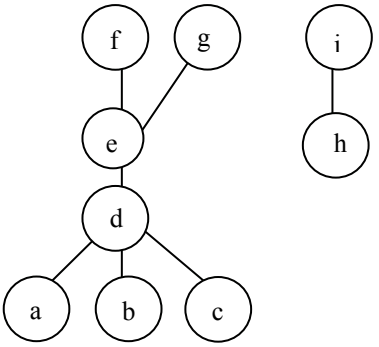


Fig. 5. Arbitrary Hasse diagram

- Elements, which are not covered by other elements, are called *maximal elements*, or -as done for example in the chapter by Carlsen and Walker, p. 153 simply as *maximals*. In Fig. 5 such elements are f, g, i.
- Elements that do not cover any other element are called *minimal element*, or simply *minimals*.
- If there is only one maximal element, then this is also called a greatest element. In Fig. 5, there is no greatest element, however in Fig. 4 element D is a greatest element.
- If there is only one minimal element, then this is also called a least element. In Fig. 5, there is no least element, however in Fig. 4 element A is a least element.
- If in a Hasse diagram there are parts that are not connected then these parts are called *hierarchies*. The suborders ({a, b, c, d, e, f, g}, ≤) and ({h, i}, ≤) are such hierarchies.

Details of the construction of Hasse diagrams „by hand“ are explained by Halfon et al. 1989. There is a useful "four-point-program" how step-by-step Hasse diagrams may be constructed (nevertheless quite tedious, if done by hand). See for a detailed description, (Voigt and Brüggemann, p. 327). There are still many ways to draw a Hasse diagram and some mathematicians are thinking about that point as art, Rival 1989. For example the program WHASSE would draw the Hasse diagram of Fig. 2 as depicted in Fig. 4. In the specific case that a poset can be considered as lattice, i.e. fulfils the axioms of lattices, then Freese 2004 gives an advice how to draw automatically lattices.

According to the scientific background the actual diagram may be constructed such that the results are presented as clear as possible. If there is no such specific background, the Hasse diagram is drawn as symmetric as

possible. Incomparable objects are, conservatively located at the same height and as high as possible on the page. For example the object C in Fig. 4 could be located everywhere between objects D and A without hurting the order relations. Because of the above-mentioned convention, incomparable objects are arranged in *levels*. Sometimes a compromise between the symmetry demand and the general clearness of the diagram is to be accepted. The concept of levels is further discussed below.

The concept of order preserving maps plays an important role in applications of Hasse Diagram Technique (HDT). For an introduction, this concept will be exemplified by the so-called level construction (see Fig.6).

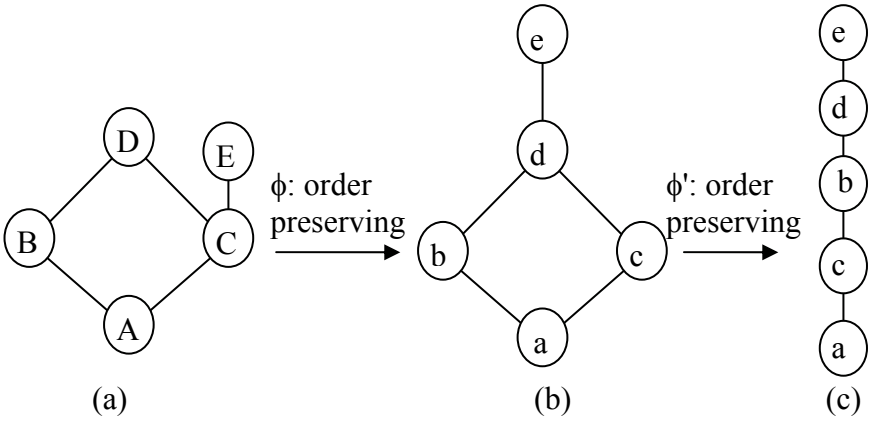


Fig. 6. The Hasse diagram (left side (a)) is mapped onto the Hasse diagram (b). All order relations of the domain set, and order relations (left side) are preserved in the range of the mapping ϕ (right side). Finally an order-preserving map ϕ': $F \rightarrow F$ is applied to obtain a linear order (diagram (c))**.**

Let *E* be a set of objects and *F* another set. Let x_1, x_2, ... be the objects of *E* and y_1, y_2, ... the objects of *F*. An assignment $f(x_i) = y_i$ is order preserving, if any order relation $x_i \leq x_j$ is maintained, i.e. $f(x_i) \leq f(x_j)$ or $y_i \leq y_j$. Thus, if a set {A, B, C, D, E} (Fig. 6a) is assigned to the set {a, b, c, d, e} as follows: f(A) = a, f(B) = b, f(C) = c, f(D) = d, f(E) = e then in order to obtain an order preserving map one has to demand: a < b, b < d, a < c, c < e, c < d as e.g. in Fig. 6b and c. It should be noted that the order C < E is maintained. Thus, f(C) < f(E) or c < e. This is not affected by the creation of a new order d < e. Indeed: Very often an order-preserving map is associated with an enrichment of comparabilities.

Assignments as ϕ are often called mappings, the mapping relate one set (the domain) to another one (the range of a map). Often it is very useful that the order of the image is a linear one. Especially in QSAR applications

as shown in chapter by El-Basil, p. 3 the quantity of interest, for example a toxicity of substances, induces a linear order, whereas information on chemicals (say: topological indices or other codes of the chemical structure) leads to a partial order (for example visualized by Young diagrams). Then the art is, to find such topological indices that the partially ordered set can be related to the linear order by an order-preserving map.

There are several possibilities to construct linear orders. Theoretically very important is the concept of linear extensions, which is explained later (vide infra).

Another concept is that of the "levels". Linear orders by a level construction encompasses in HDT the following steps:

1. Set i = 1
2. Consider for the first steps of construction the quotient set $E/\mathfrak{R}$ (not the set of objects E). The set of the maximal elements, MAX, is thus the subset of $E/\mathfrak{R}$.
3. Identify the maximal elements (in $E/\mathfrak{R}$) and label the set MAX_1
4. Reduce the set $E/\mathfrak{R}$ by the maximal elements MAX_1, $E/\mathfrak{R}_{new} = E/\mathfrak{R}_{old} - MAX_1$
5. Draw the elements of MAX_1 in top-position in the drawing plane. All elements of MAX_1 get the same vertical position.
6. Add 1 to i. I.e. $i_{new} = i_{old} + 1$.
7. Identify the new maximal elements of ($E/\mathfrak{R} - MAX_{i-1}$, IB) . Label the new set MAX by i.
8. Reduce the set $E/\mathfrak{R}$ by the maximal elements MAX_i, $E/\mathfrak{R}_{new} = E/\mathfrak{R}_{old} - MAX_i$
9. Draw the maximal elements MAX_i in the same vertical position. Elements of MAX_{i-1} will located below those of MAX_i .
10. Repeat the steps 6-9 till $E/\mathfrak{R}$ is exhausted. The corresponding i is C_{max}, the number of elements in the maximal chain of ($E/\mathfrak{R}$, IB).
11. Corresponding to the intended application: a) give the top elements the level no C_{max} and the lower levels C_{max} -1 , C_{max}-2, ... ,1 or b) keep the i-labelling as level-label. In that case the bottom elements get the level number C_{max} and the top elements 1.
12. If wanted, the order relations can be added as edges.

This construction is order preserving.

A detailed example may be helpful:

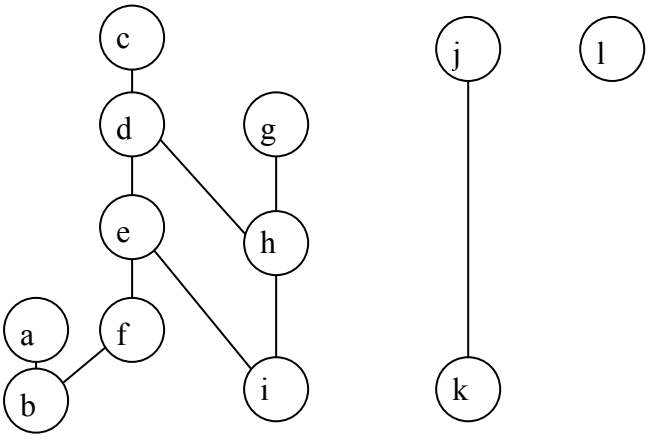


Fig. 7. Hasse diagram of 12 elements. E = {a, b, c, d, e, f, g, h, i, j, k, l}. Note that the Hasse diagram is not drawn following the convention of the program WHASSE in order to clarify the construction

- Step 1: i = 1
- Step 2: E can be identified with $E/\Re$, because there are only trivial equivalence class (i.e. singletons).
- Step 3: MAX_1 = {a, c, g, j, l}
- Step 4: $E/\Re_{new}$ = {a, b, c, d, e, f, g, h, i, j, k, l} - {a, c, g, j, l} = {b, d, e, f, h, i, k}
- Fig. 8 shows the resulting Hasse diagram:
- Step 5: (see Fig. 9)
- Step 6: i = 2
- Step 7: MAX_2 = {d, k} (see Fig. 8)
- Step 8: $E/\Re_{new}$ = {b, d, e, f, h, i, k} - {d, k} = {b, e, f, h, i}
- Step 9: (see Fig. 9)
- Step 6: i = 3 (iteration)
- Step 7: MAX_3 = {e, h} (see Fig. 8).
- Step 8: $E/\Re_{new}$ = {b, e, f, h, i} - {e, h} = {b, f, i}
- Step 9: (see Fig. 9)
- Step 6: i = 4 (iteration)
- Step 7: MAX_4 = {f, i}
- Step 8: $E/\Re_{new}$ = {b}
- Step 9: (see Fig. 9)
- Step 6: i=5

- Step 7: MAX_5= {b}
- Step 8: $E/\Re_{new} = \phi$
- Step 12: C_{max} = 5. We follow the labelling of a). See Fig. 9 left side for the level structure and right side for the diagram, supplied with the order relations:

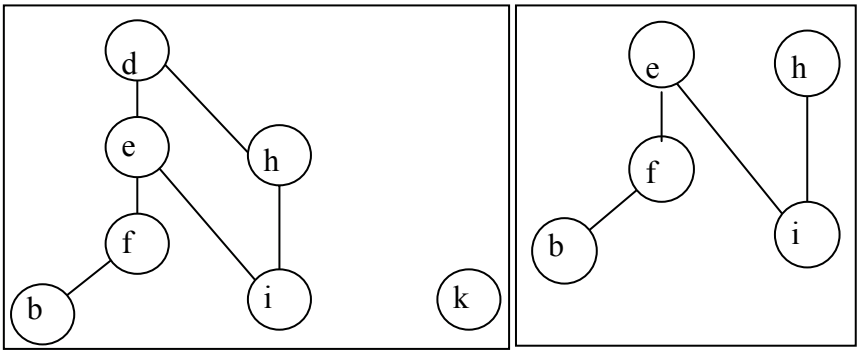


Fig. 8. The resulting poset and its visualization after subtracting the maximal elements of ($E/\Re$, *IB*) after the start and the first iteration

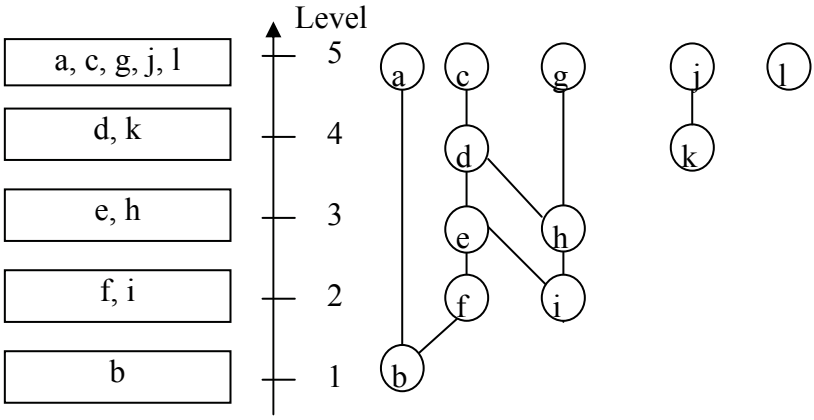


Fig. 9. Example to determine the level structure (Left side: Assignment to levels Right side: the Hasse diagram redrawn)

These steps sound difficult, however they are easily understandable, just by doing! Here some examples (Fig. 10)

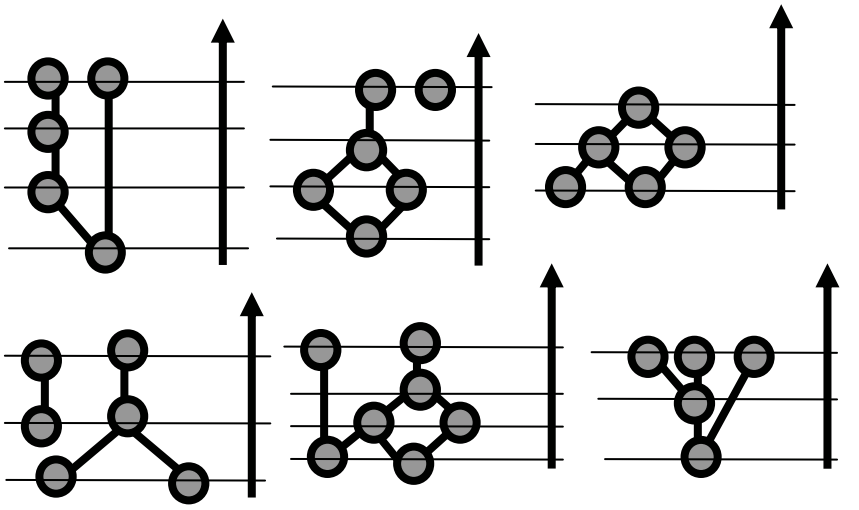


Fig. 10. Example, how to assign the levels. If one vertex contains several equivalent objects, than these objects belong all to the same level. The vertical arrow symbolizes the order induced by the vertical arrangement of the vertices

The levels may be considered as a first very crude evaluation: If a high level is associated with a high hazard, then the sequence of increasing levels coincides with increasing hazard.

In the above advices 1-6, the rule 4 needs additional explanations. In order to do this, we introduce first the concept of graduation and of the rank-function, respectively. If there is a rank function r, then for any element of the ground set the levels are uniquely found. Hence, a poset is graded or possesses a rank function if:

a) $x > y$ implies $r(x) > r(y)$ (order preserving!) and
b) for x covering y a unique function r can be found, such that $r(x) = r(y)+1$.

In the case, shown in Fig. 11 (a) such a rank function exists, whereas in Fig. 11 (b) one cannot find a function r. Obviously, for the Hasse diagram in Fig. 11 (a) all five objects are located at specific levels, whereas the hatched object in the diagram in Fig. 11 (b) may be located either at the level of x or the level of y, respectively. However, corresponding to the level construction the element u belongs to MAX_2. The elements u and z have therefore the same vertical position and are below the top element, which belongs to MAX_1.

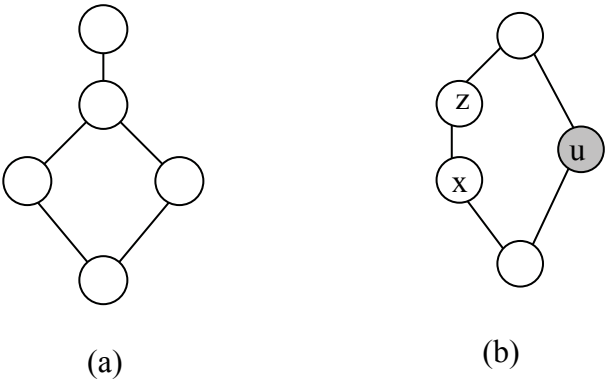


Fig. 11. Graded (a) and non-graded (b) posets (visualized by Hasse diagrams) Grey circle, x and z: (see text)

Posets, which do not have a rank function, give the user of Hasse diagrams the additional freedom, for example to introduce further information. Locating an element as high as possible obviously is a conservative approach. Thus, in, e.g., risk assessment, high values of attributes are associated with high risk. Locating an element of a poset as high as possible has thus a warning function.

A useful theorem to find out whether a rank function r exists, is the so-called Jordan-Dedekind Chain Condition (JDCC) (see also Birkhoff 1984), stating that all maximal chains between the same endpoints have the same finite length. Thus, if a poset satisfies JDCC, a rank function can be found. In Fig. 11 (a) there can be found two maximal chains. Both have the same length. In Fig. 11 (b), once again two maximal chains can be found. However, they differ in their length. Hence the JDCC is hurted in case of the poset, visualized in Fig. 11 (b). A generalization of rank functions for lattices is given in Freese (2004). However, as most empirical posets do not satisfy the axioms of lattices, we will not deepen this concept here.

Hasse diagrams as digraphs

Hasse diagrams can be interpreted as mathematical graphs, i.e. they are called d*igraphs* (directed graph), because of the orientation of the lines. Following the definitions of order the digraphs are acyclic. Interpreted as ordinary graphs, Hasse diagrams are *triangle-free*: Due to the rule of transitivity, line segments corresponding to a < c can be omitted if a < b and at the same time b < c. A digraph consists of a set E (or $E/\Re$ if the quotient set is to be partially ordered) of vertices (circles in Hasse diagrams) and a

set of oriented edges each connecting two vertices. If the vertices are drawn in the diagram according to the above rules (defining the level-construction) then the arrows can be simply be represented by lines, because then the element x will be arranged below y, if $x < y$. Therefore the orientation of the line is replaced by the vertical location in the drawing plane. The circles are the objects of E, or elements of the set $E/\mathfrak{R}$ to be ranked.

The basic essence is that by the order relation a data matrix is represented by a mathematical graph with objects as vertices and that the structure of this graph tells us somewhat about the data structure. As the data matrix arises from external studies (experimental work, modelling, empirical data) the resulting graph is called an "empirical graph", which may have (hitherto hidden) regularities. A main task in performing partial order as an exploring tool is just to detect (by abstraction, by simplification) regularities or structures in the graph. Helpful, however still not yet fully developed, is that one can establish an algebra among a set of posets, which reveals different kinds of sums, products and exponentiation, see for example Jonsson 1982.

The concepts "hierarchy", "articulation points", "chains" and "anti-chains" are very basic and simple ones, which direct into the structural analysis of digraphs. These concepts will be explained in the next section.

Simple elements of interpreting a Hasse diagram

Overview

The basics to consider Hasse diagrams are to check

1. the system of comparabilities and incomparabilities
2. the priority elements
3. pattern of attributes and
4. identifying data structures.

Almost all these kinds of analyses of Hasse diagrams can be found in the different chapters of this book.

Example

A simple example is given in the following (Table 2 and Fig. 12).

There are three hierarchies. One of them is a trivial hierarchy as it consists of one element only, i.e., element f that is not comparable to any other elements. Such elements are also called *isolated elements*. If only few isolated elements are found, whereas almost all other are comparable, then the

isolated elements should be examined carefully as very often specific data structures are the reason for their isolation.

Table 2. A more extended example, demonstration of isolated hierarchies

objects\attributes	q_1	q_2	q_3
a	2	2	3
b	1	2	2
c	2	1	2
d	4	3	1
e	4	2	1
f	0	0	5

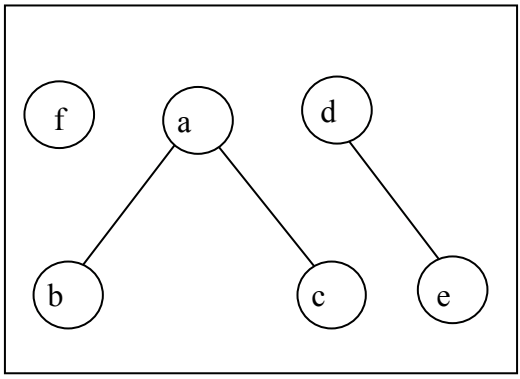


Fig. 12. The partially ordered set of objects of Table 2 has a Hasse diagram with three isolated hierarchies, namely ({a, b, c}, {q_1, q_2, q_3}) , ({d, e},{q_1, q_2, q_3}) and ({f}, {q_1, q_2, q_3})

If, on the other hand, all elements of E (or $E/\Re$) are isolated then the attributes should be checked for the degree of anti-correlation (Spearman rank correlation). It depends on the scientific question, whether such a trade-off among attributes (a decreasing sequence of values of one attribute is always accompanied by an increasing sequence of another attribute) should be maintained in the study. There are methods to deal with such cases, see the chapter by Simon et al., p. 221 and by Sørensen et al., p. 259. However, this shall not be further discussed here. The subsets {d, e} as well as {a, b, c} form *nontrivial hierarchies*. Hence, we have three order relations: $b \le a$, $c \le a$, and $e \le d$. The fact that the set E can thus be partitioned into three disjoint subsets is always of great interest with respect to the data structures. Further structural elements, which are of interest in the analysis of Hasse diagrams, are subsequently discussed:

Chain: Subset of the ground set, where all elements are mutually comparable. An example is the chain ({d, e},{q_1, q_2, q_3}) another: ({b, a},{q_1, q_2, q_3}) (Fig. 12). Often it is sufficient, simply to write {d, e} is a chain. Any other element of the ground set added would led to at least one incomparability and thus hurts the definition. Therefore the chains {b, a}, {c, a}, {e, d} are maximal. The identification of chains is of high interest with respect to exploring data structures, because the generality principle demands that for all attributes of objects of a chain it is valid. Thus, if $x < y$, x, y being elements of a chain, then $q_i(x) < q_i(y)$ implies $q_j(x) \le q_j(y)$ for all $j \ne i$. Fol-

lowing the elements of a chain in one direction (from top to bottom or (exclusively) from bottom to top) the attributes are increasing in a weak monotonous manner.

Anti-chain: Subset of the ground set, where all elements are mutually incomparable. An example is the anti-chain ({f, a, d}, {q_1, q_2, q_3}). Any other element of the ground set added to the set {f, a, d} would introduce a comparability. Therefore {f, a, d } is a maximal anti-chain. Attribute profiles being results of monotonous variations as seen in chains are not considered as essentially different. Contrary, attribute profiles through anti-chains are essentially different. Hence the width, Wd(*E*), of the poset is considered as a measure of diversity.

Maximal elements (often also called simply "maximals"): Elements of the ground set $E/\mathfrak{R}$, x_i, for which no $y_i \in E/\mathfrak{R}$ can be found with $x_i \leq y_i$. Maximal elements in the Hasse diagram, shown in Fig. 12 are: f, a, d.

Minimal elements (often also called simply "minimals"): Elements of the ground set $E/\mathfrak{R}$, x_i, for which no $y_i \in E/\mathfrak{R}$ can be found with $x_i \geq y_i$. Minimal elements in the Hasse diagram, shown in Fig. 12 are: f, b, c, e.

Isolated elements: Elements that are both: Minimal and Maximal elements. Maximal/Minimal elements which are not isolated, are often called proper maximal/minimal elements. An isolated element in the Hasse diagram, shown in Fig. 12 is: f.

Hierarchy: Let $E'/\mathfrak{R}$ and $E''/\mathfrak{R}$ be two subsets of $E/\mathfrak{R}$. If for all $x \in E'/\mathfrak{R}$, and all $y \in E''/\mathfrak{R}$: x||y then ($E'/\mathfrak{R}$, *IB*) and ($E''/\mathfrak{R}$,) are hierarchies. In a Hasse diagram they can often be recognized as non connected parts.

Articulation point: If the elimination of one element of E/$\mathfrak{R}$ enhances the number of hierarchies in the residual poset, then this element is called an articulation point. In the Hasse diagram, Fig. 12 the element a is an articulation point.

Long chains, hierarchies and articulation points indicate specific data structures. The role of hierarchies will be explained by a two dimensional scheme (Fig. 13): Several objects may be located as points within the two rectangles H_1 and H_2. Comparing one object of H_1 with one of H_2 will lead to q_1 (of $x \in H_1$) > q_1 (of $y \in H_2$), whereas q_2 (of $x \in H_1$) < q_2 (of $y \in H_2$). Hence no object of H_1 is comparable with that of H_2. In Neggers & Kim 1998 a rather nice wording is found for the objects belonging to the field *F* and *P*: These are the future objects relative to the objects in the field *P*, which are called the objects in the past.

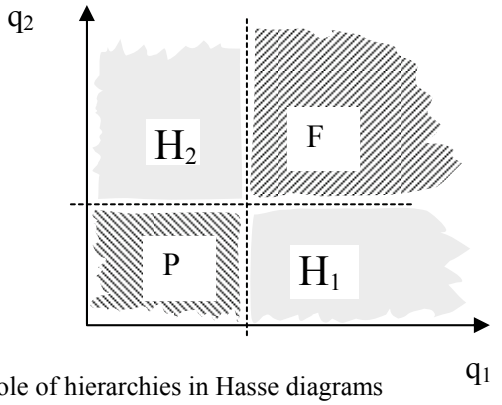


Fig. 13. Role of hierarchies in Hasse diagrams

Note that by construction of levels any level is to be considered as anti-chain. However, this anti-chain may not necessarily be a maximal one. The evaluation of sampling sites for sediment samples of the Lake Ontario is used as a further illustrative example (cf. pp. 94). More details can be found in Brüggemann & Halfon (1997) and Brüggemann et al. (2001 b).

Characterizing a Hasse diagram as a whole

Characteristic Numbers of Posets

In the present section a series of simple characterising numbers is introduced. They are useful to give a general overview and impression of the poset and the corresponding visualizing graph, the Hasse diagram. It is recommended to read the careful discussion by Pavan & Todeschini 2004 and in this book, chapter by Pavan, p. 181. The Hasse diagram of Lake Ontario will exemplify all numbers.

- NECA: Number of equivalence classes with more than one object, i.e., the number of nontrivial equivalence classes.
- Wd(E): The width of a Hasse diagram. It is the maximum number of elements of $E/\mathfrak{R}$, which are found in an anti-chain. In the context of Young diagrams (see Seitz, p. 373) also called a “breadth”.
- L(E): The length of a Hasse diagram: The number of line segments in the chain with a maximum number of elements of $E/\mathfrak{R}$.

- H(E): The height of a Hasse diagram = C_{max}. H(E)=L(E)+1. H(E) is the number of objects (of $E/\mathfrak{R}$) in the maximum chain.
- NL, the number of levels = H(E).
- NEL, the number of elements (of $E/\mathfrak{R}$) in the level, which contains the most elements of $E/\mathfrak{R}$; note that this number is not necessarily the same as Wd(E).
- NMAX: The number of maximal elements (called: number of maximal equivalent classes because this information is related to $E/\mathfrak{R}$).
- NMIN: The number of minimal elements (notation as for the maximal elements).
- Z: Number of all equivalence classes, including singletons, i.e., Z = card $E/\mathfrak{R}$. Note that Z and NECA differ. If NA is the number of elements of E, which are contained in nontrivial equivalence classes (NECA) then the following equation holds
  card E = NA + Z - NECA (2)
  Some other numbers are also interrelated, for example the relation
  NL = L(E)+1 (3)
- P(IB): stability of ranking. This quantity is a measure for the effect of extending or reducing the set of attributes on the structure of the Hasse diagram. It is calculated as the quotient of all incomparabilities, U_{total} and Z·(Z-1)/ 2:

$$P(IB) = \frac{2 \cdot U_{total}}{Z \cdot (Z-1)} \quad (4)$$

  If P(IB) is near 1 or 0, respectively, then extending or reducing, respectively, the set attributes should have a minor effect.

Linear Extensions

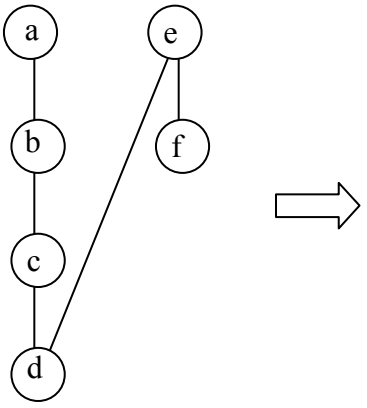
The linear extensions are the basis of the dimension theory of posets. Besides the dimension of posets other characterizations may be derived from linear extensions (Carlsen et al. 2002, Lerche et al. 2003, Lerche & Sørensen 2003).

Extensions may be explained by the following: Given a poset (E, ≤) then we can assign another poset ($EX(E)$, ≤) which

1. supplies some ||-relations of (E, ≤) by < or > -relations
2. maintains all comparabilities of E in the correct orientation

Extensions are order-preserving maps from the ground set E into the ground set E; see Davey & Priestley 1990. Linear extensions ($LEX(E)$, $\leq$) are order-preserving maps from E to E, which assign to (E, $\leq$) a linear order.
In Fig. 6 (b) an extension is shown (identify A with a, B with b, etc), but not a linear one. An additional preserving map leads to a linear order (Fig. 6 (c)). The diagram in Fig. 6 (c) is a linear extension of that in Fig. 6 (a). Given a poset (E, $\leq$) then several linear extensions ($LEX(E)$,$\leq$) are possible. A systematic procedure is described by Atkinson (1989), especially for trees a closed formula can be derived Atkinson (1990). A useful formula to calculate the number of linear extensions is also given by Stanley (1986).

Each relation $x \leq y$, $x, y \in E$ is reproduced in $LEX(E)$. However, the reverse statement is not true. All in all, any linear extension is an image of an order-preserving map. The diagram (Fig. 14) visualizes the concept. All comparabilities $x < y$, $x, y \in E$, of (E,$\leq$) are reproduced in the first fourteen lines of the table, whereas the last sequence (16^{th} row) in the table illustrates a non order preserving map. The relation $d \leq e$ of (E, $\leq$) is reversed. This sequence therefore is no linear extension of the poset (shown in the left side of Fig. 14). If the sequences (1) to (14) are considered as partially ordered sets, then they have comparabilities, which are not found in the original poset. For example the elements b and e are comparable in the 14 sequences of Fig. 14, but are incomparable in the original poset. The incomparability of b, e is expressed in the linear extensions by the fact that there are some, where $b > e$, and some where the opposite is true.



1	a	b	c	e	f	d		2
2	a	b	c	e	d	f		2
3	a	e	f	b	c	d		2
4	a	e	b	f	c	d		4
5	a	e	b	c	f	d		4
6	a	e	b	c	d	f		3
7	e	f	a	b	c	d		1
8	e	a	f	b	c	d		3
9	e	a	b	f	c	d		3
10	e	a	b	c	f	d		3
11	e	a	b	c	d	f		2
12	a	b	e	f	c	d		2
13	a	b	e	c	f	d		4
14	a	b	e	c	d	f		3
	a	b	c	d	e	f		

Fig. 14. Poset (E, $\leq$) (left side) and its 14 linear extensions ($LEX(E, \leq)$) ((1) to (14).

In Fig. 14 the first column labels the linear extensions, which are represented as sequences in rows 1 - 14. A sequence a b c ... is to be read as a > b > c. Furthermore there is a sequence (last row in the table) which is **not** a linear extension of (E, $\leq$). Vertical bold lines indicate jumps (see below). The last column indicates the number of jumps of each single linear extension. Consecutive elements in linear extensions ($LEX(E)$, $\leq$), which have no correspondence in (E, $\leq$) are called "*jumps*" (see Fig. 14, the vertical bold lines indicating jumps). The jump number, jump ($LEX_i(E, \leq)$), obviously depends on the actual selected linear extension. The jump number of a poset (E, $\leq$), jump (E, $\leq$), is just min(jump(LEX_i (E, $\leq$))), whereby the minimum is to be found by checking all linear extensions. Beside the jump - number there is also a bump - number. Once again the bump number is to be referenced to a specific linear extension. A bump is a consecutive pair of elements in a linear extension, which are comparable in the underlying poset. The bump number of a poset is the maximum about all bump numbers found for the linear extensions. If a linear extension of n elements is formed then n-1 consecutive relations are found in a linear extension. Therefore

$$\text{jump}\,(LEX_i(E, <)) + \text{bump}\,(LEX_i(E, <)) = n - 1 \qquad (5)$$

Linear extensions of a minimal jump number of specific interest: These linear extensions (also called "greedy linear extensions") preserve as much as possible the chain-structure of a Hasse diagram (Rival 1983, Rival & Zaguia 1986). As one can see in Fig. 14 that linear extension with jump number = 1 preserves both chains a > b > c > d and e > f. Consequently, the jump number of a poset may be considered as indicator for "chainyness": Thus, a low jump number indicates that the poset contains subposets, which are long chains. In operation research or queuing plans a jump implies often some cost-intensive rearrangements. Therefore linear extensions with a small number of jumps are preferred in organisation of work. Contrary to that, Patil and Taillie 2004 are discussing in their paper that the jump number may also serve to weight linear extension, where the linear extension with the largest number of jumps gets the highest weight.

If a specific element, say $x \in E$ is selected then its *spectrum* is of interest (Atkinson 1990). It should be noted that other authors (for example Trotter 1991, Schröder 2003) also call the spectrum a projection. However, we favour "spectrum" as the more suitable name for the following construction. Thus, let LT be the number of linear extensions of a poset, then we can find the rank of an element x in the i^{th} linear extension: rank(i, x). Note that this construction should not be confused with the rank function, we discussed above. Conventionally, the bottom element of a linear extension is given the rank 1, thus the top element has the rank n (card E = n). However, if appropriate the top element may be assigned the first priority, such that bottom elements will get numbers > 1 (see for example chapter by Carlsen, p. 163). We call $\lambda_k(x)$ the frequency, how often $x \in E$ gets the rank k. The spectrum spec(x) is a tuple containing n components ($\lambda_1(x)$, $\lambda_2(x)$, ... ,$\lambda_n(x)$). Thus for example the spectrum of element b in Fig. 14 as follows: spec(b) = (0, 0, 3, 6, 5, 0). (i) There is no linear extension, where the rank of b is 1, 2 or 6. (ii) There are 3 linear extensions, where the rank of b is 3. (iii) There are 6 linear extensions, where the rank of b is 4. (iv) There are 5 linear extensions, where the rank of b is 5. Obviously:

$$LT = \sum \lambda_k(..) \qquad k = 1,...,n \tag{6}$$

The set of linear extensions is the basis for probability considerations: Dividing $\lambda_k(x)$ by LT the quantity prob (rk(x) = k) = $\lambda_k(x)/LT$ can be interpreted as (ordinal) probability to get the rank k, sometimes also called “absolute rank”. Hence, an averaged rank, Rkav can be derived by

$$Rkav(x) = \sum k \cdot \lambda_k(x)/LT \tag{7}$$

and the elements x ∈ *E* can be ordered by their Rkav-values. Therefore a total order, however, often including equivalence classes can be derived from a poset, without the numerical combination of attributes to one ranking index. This concept is widely used based on the following arguing: If the attributes are combined, say by weighted sums or any other positively monotonous function, then the result must be (besides ties) one of the linear extensions, as the set of all linear extensions encompasses all results of order preserving maps. However, there are still many open problems due to computational difficulties in handling large object sets, advices can be found in Lerche et al. 2003 or in Patil & Taillie 2004. If n objects are mutually incomparable, then n! linear orders are possible, corresponding to n! permutations. (See also chapter of Sørensen, Lerche, Thomson, p. 259, for a discussion of entropy, related to the number of linear extensions). Hence a crude upper estimation of the number of linear extensions is n! A ground set containing for example 17 elements may have at most ca. $3.5\ 10^{14}$ linear extensions.

Recently an alternative was discussed, in order to use a local model of the partial order, which describes the environment in the directed graph around the element of interest. For further discussions two recent publications should be consulted (Brüggemann et al. 2004, 2005).

A rather good approximation for an element of interest, x, may be obtained, if the successors (all elements "below" x) and predecessors (all elements "above" x), respectively, are organized into a so-called “S-x-P” chain, all remaining elements, i.e. those incomparable to x being considered as isolated. From a combinatorial study follows that the averaged rank of an element x can be expressed as

$$\mathrm{Rkav} = \frac{\sum_{k=0}^{k=U} (S+1+k) \cdot \binom{U}{k} \cdot (S+1)^{k} \cdot (P+1)^{U-k}}{\sum_{k=0}^{k=U} \binom{U}{k} \cdot (S+1)^{k} \cdot (P+1)^{U-k}} \tag{8}$$

which can be transformed (Brüggemann et al. 2004) into

$$\mathrm{Rkav} = (S+1)\cdot(S+P+U+2)/(S+P+2) \tag{9}$$

Since N = S+P+U+1 the averaged rank of an element x may be expressed by the following simple relation [1]

$$\mathrm{Rkav}(x) = \frac{(S(x)+1)\cdot(N+1)}{N+1-U(x)} \tag{10}$$

with:

- S(x): = |{y ∈ E : y < x}| is the number of successors of x
- N the total number of elements and
- U(x): = |{y ∈ E : y || x}| is the number elements incomparable to x.
- P(x): = |{y ∈ E : y > x}| is the number of predecessors.

The principle is illustrative demonstrated by determining the averaged rank of element b in the Hasse diagram depicted in Fig. 15.

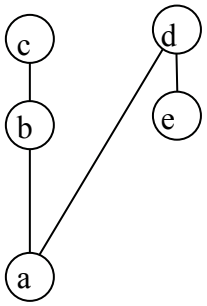


Fig. 15. Example for application of equation 10. The averaged rank of element b is to be estimated

It is immediately seen that N = 5, S(b) = 1 (element c), P(b) = 1 (element a), and U(b) = 2 (elements d and e). Hence, according to equation (10) the averaged rank of element b is estimated to be Rkav(b) (estimated) = (1+1)·(5+1)/(5+1-2) = 3, the exact value - calculated after equation 7 - being Rkav(b) (exact) = 2.889.

Dimension of a poset

The dimension of a poset is based on the set of linear extensions. A linear extension can be considered as a set of ordered pairs. For example the linear extension no 1 in Fig. 14 (right side):

[1] Counting from bottom to top.

{(a, a), (b, a), (c, a), (e, a), (f, a), (d, a), (b, b), (c, b), (e, b), (f, b), (d, b), (c, c), (e, c), (f, c), (d, c), (e, e), (f, e), (d, e), (f, f), (d, f), (d, d)}. Each pair denotes a ≤ -relation. For example (e, b) means that in the linear extension no 1 e ≤ b.
A similar set could be found for any other linear extension, for example no 2:
{(a, a), (b, a), (c, a), (e, a), (d, a), (f, a), (b, b), (c, b), (e, b), (d, b), (f, b), (c, c), (e, c), (d, c), (f, c), (e, e), (d, e), (f, e), (d, d), (f, d), (f, f)}
The intersection of these two sets of pairs leads to:
{(a, a), (b, a), (c, a), (e, a), (d, a), (f, a), (b, b), (c, b), (e, b), (d, b), (f, b), (c, c), (e, c), (d, c), (f, c), (e, e), (d, e), (f, e), (d, d), (f, f)}.
This intersection does not coincide with the set of ordered pairs of the poset itself (Fig. 14 (left side)):
{(a, a), (b, a), (c, a), (d, a), (b, b), (c, b), (d, b), (c, c), (d, c), (e, e), (f, e), (d, e), (f, f), (d, d)}
Thus this kind of troublesome check has to be repeated until the intersection of the set of ordered pairs of the linear extensions coincide with that of the poset. The lowest number of linear extensions -written as ordered pairs as shown above- whose intersection is the actual poset (together with its transitive relations), is its dimension. Following the explanation above one would have to check 14·13/2 intersections, just to verify that the dimension equals 2. If such pair of ordered sets, derived from any two linear extensions is found, one has found a "*realizer*" of the poset (Trotter 1991).

Note, it is not a good policy to derive the dimension by finding explicitly the realizers. Here five useful theorems are taken from the literature (Trotter 1991):

- $\dim (E,\leq) \leq \mathrm{Wd}(E)$ (for further on Wd(E), see p. 81) (11)
- Let $(E,\leq)$ a poset and $(C,\leq)$ a chain, $C \subset E$. Then $\dim(E,\leq) \leq 2 + \dim(E\text{-}C,\leq)$ (12)
- Let $(E,\leq)$ a poset, and n:=card $E \geq 4$, then: $\dim (E,\leq) \leq n/2$ (13)
- Let $E_A \subset E$ an anti-chain of a poset $(E, \leq)$, then $\dim (E, \leq) \leq \max(2, \mathrm{card}(E\text{-}E_A))$ (14)
- If the Hasse diagram, supplied (if necessary) by a greatest and least element can be drawn in the plane without crossing of lines, then the dimension of the poset is 2, (15)
- Let $(E, \leq)$ be a poset and $(E', \leq)$ be a subset of E, then $\dim (E, \leq) \geq \dim (E', \leq)$ (16)

We apply equation 12 to determine the dimension of the poset shown in Fig. 16:

- Step 1: As the poset is not a linear order we conclude: dim $(E, \leq) > 1$
- Step 2: We select a chain: $C = \{c, b, a\}$
- Step 3: The ground set is now $E\text{-}C = \{d, e\}$. The poset $(\{d, e\}, \leq)$ is a chain.
- Step 4: dim $(\{d, e\}, \leq) = 1$
- Step 5: $1 <$ dim $(E, \leq) \leq 2 + 1$. Thus the dimension of $(E, \leq)$ is either 2 or 3.

Equation 11 would be more useful: As Wd(E) of the poset, shown in Fig. 15 is 2, the dimension must be 2. Generally, for the purposes intended in this chapter equations (15) and (16) are the most interesting theorems.

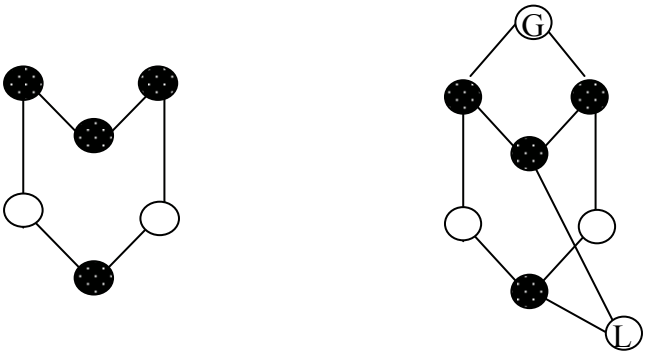
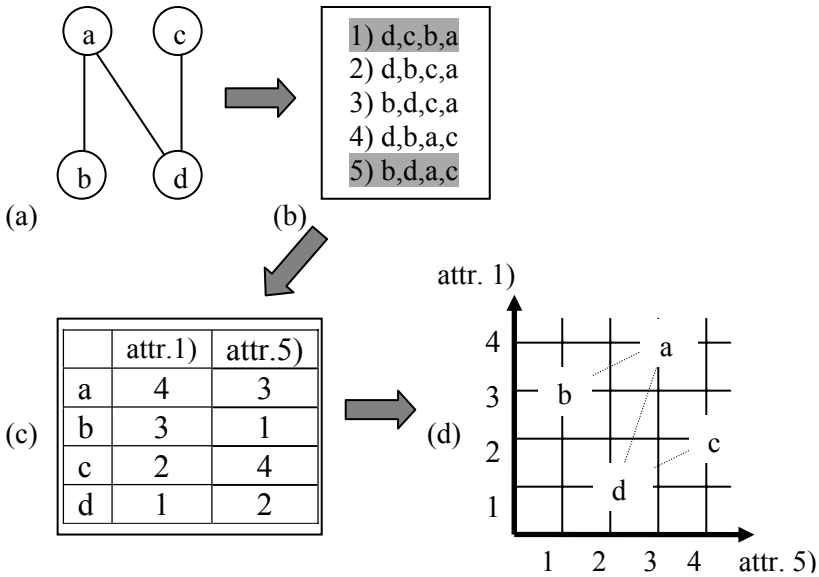


Fig. 16. Hasse diagram of a poset with dimension 3. The Hasse diagram on the left side follows not the convention explained earlier! The Hasse diagram on the right side is supplied by a greatest "G" and least element "L".

The poset, whose Hasse diagram is shown in Fig. 16 (left side) has the dimension 3. A priori, as obviously there is no crossing of lines, the dimension would be expected to be 2. However, this poset must be extended by a greatest, G, and a least element, L. Then a crossing of lines within a plane is not avoidable. Thus posets having such substructure have at least dimension 3. For other examples, compare Trotter 1991. Why is the dimension of posets so interesting? Let us assume we got a Hasse diagram by using 5 attributes. If now, the dimension of the partial order would be 2 then we knew in advance that two linear extensions are sufficient to reproduce the partial order. As each single linear extension can be considered as the linear order induced by an unknown attribute, two attributes are sufficient to obtain the same partial order as by the original five ones. Usually these two attributes cannot be found as a subset of the information base.

They are called latent variables. The original five attributes may be (in a complex manner) mapped onto 2 latent variables.



	attr.1)	attr.5)
a	4	3
b	3	1
c	2	4
d	1	2

Fig. 17. The partial order of four elements and the concept of dimension

In Fig. 17 a Hasse diagram (a) and its 5 linear extensions (ordered for increasing values) (b) are shown. Two realizers (grey hatched) are identified and are considered as new attributes ("attr. 1"; "attr.5") (c). The objects are located in a rectangular grid (d) due to values of attr. 1 and attr. 5. A rotation of the coordinate system around ca 45^{o} would reproduce the original Hasse diagram. If the dimension of posets is 2 or 3 then it may be useful, to embed the poset into a two- or three-dimensional grid (see Brüggemann 2001 b). For an example of embedding a poset into a two-dimensional coordinate system, see also the chapter of El-Basil.

On the other hand, any two-dimensional scatter plots can be interpreted as a partial order, if the generality principle is applied to the both coordinates of any point.

Sensitivity study

Mathematical Notation and Background

Preferably a maximal element should be chosen as a starting point for the analysis. This choice, however, is not mandatory. Thus, other elements of E or $E/\Re$ could be chosen too. This selected element is called "key element". We may simultaneously select more than one key element even all elements (no restrictions apply here). For the sake of convenience all key elements are supposed to form a set K ($\subseteq E$).

The analysis of a key element implies a search of all elements located lower than that of the key element, i.e. all elements that can be reached from the key element by a path, a sequence of connecting edges. (Therefore the selection of maximal elements rather than other elements is more meaningful). These elements together with elements equivalent but not identical to the key element are called successors. The set of all successors of the key element "k" is denoted as $G(k,A)$, $A \subseteq IB$. We include the information about the actual set of attributes (i.e. the case) by A. Note the similar concept of "down-sets" in Davey & Priestley (1990): The order ideal (or down set), generated by the key element will be denoted by $O(k,A)$. Then it is valid:

$$G(k,A) = O(k,A)-\{k\}$$

The operation "-" is the set theoretical subtraction. For example:
{a, b, c, d} - {a, e} = {b, c, d}

Those elements of the first set, which also are in the second set, are eliminated. By definition $G(k)$ does not include the key element itself. The successor sets and their cardinalities are the heart of the sensitivity analysis shown here. The successor sets found for two Hasse-diagrams resulting from two attribute-subsets of IB are used to quantify certain differences. The cardinality of successor sets (denoted: card $G(k)$) and of their set theoretical combinations play an important role here.

Residual sets

To assess the influence of each attribute on ranking, we compare Hasse diagrams that arise from subsets *B, C* of *IB*. A straightforward method to perform this task is to choose a key element and quantify the effect of each attribute set on its successor set. For this purpose the residual set, R, i.e. is now introduced.

$$R(\mathrm{k}, B, C) := (G(\mathrm{k}, B) \setminus G(\mathrm{k}, C)) \tag{17}$$

By Venn-Euler diagrams residual sets can easily be understood (see Fig. 18):

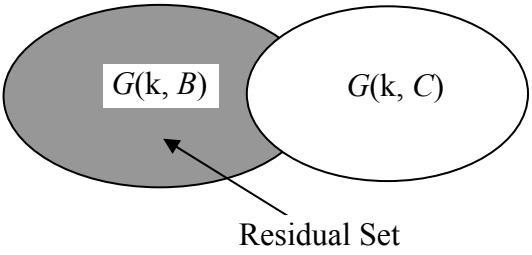


Fig. 18. Venn-Euler diagram of the residual set $R(\mathrm{k}, B, C) = G(\mathrm{k}, B) - G(\mathrm{k}, C)$

In general $R(\mathrm{k}, B, Ç) \neq R(\mathrm{k}, C, B)$. Therefore the symmetric difference set "$W(\mathrm{k}, B, C)$" of the sets $G(\mathrm{k}, B)$ and $G(\mathrm{k}, C)$ is introduced:

$$W(\mathrm{k}, B, C) := R(\mathrm{k}, B, C) \cup R(\mathrm{k}, C, B) = [G(\mathrm{k}, B) - G(\mathrm{k}, C)] \cup [G(\mathrm{k}, C) - G(\mathrm{k}, B)] \tag{18}$$

If the cardinality of $W(\mathrm{k}, B, C)$ is small, i.e.

$$\mathrm{W(k, B, C)} << \min [G(\mathrm{k}, B), G(\mathrm{k}, C)])$$

then subsets *B* and *C* lead to not very different Hasse diagrams. If the difference is large then the two corresponding Hasse diagrams are dissimilar to each other. Those attributes, by which *B* and *C* differ, play a key role in ranking. This finding motivates the introduction of the matrix **W**.

Definition of the matrix W

Calculating the matrix **W**

The matrix **W**(k) assesses the difference of Hasse diagrams induced by the two subsets of attributes with respect to a key element k. This matrix, which is at the heart of the analysis, is called the "dissimilarity-matrix", because the larger the matrix-entries are, the greater is the difference between the successor sets for the element k and hence between the Hasse diagrams (see for more details, below). We define the entry W(k, *B*, *C*) of matrix **W** to be:

$$W(k, B, C) := \text{card}\,[R(k, B, C) \cup R(k, C, B)] \tag{19}$$

For any key-element k the residual sets *R*(k, *B*, *C*) and *R*(k, *C*, *B*) are determined, their elements being counted and summed. The entries of the matrix **W** are subsequently calculated by adding the cardinalities of the *R*-sets. To simplify notation, we now write W(k, i, j) for W(k, *B*, *C*).

Search for the important attributes

Several W(k, i, j)'s, $k \in K$ (*K* is any set of key elements) can be compared to see how a change in attributes affects the partial order with respect to the set of several key elements:

$$W(K, i, j) := \sum_{k \in K \subset E} W(k, i, j) \tag{20}$$

W(*K*) is a symmetrical matrix. **W**(*E*) is the total dissimilarity matrix of the set of *E*. Let be n:=card *E*. Mainly the **W**(k) and the **W**(*E*) matrices are useful. The final steps towards a sensitivity are:

1. If we are interested in comparisons of the full attribute set *IB* with all subsets $A_i \subset IB$, A_i only one row of the matrix **W** is of interest. We can choose the first one without loss of generalization, thus we are left with W(k, 0, 1), W(k, 0, 2), ... , W(k, 0, p), where the index 0 denotes the full attribute set *IB* (i.e. $A_0 \equiv IB$) and $p=2^m -1$.
2. To see the influence of single attributes on a Hasse diagram we compare the Hasse diagrams induced by *IB* with those induced by those attribute sets $A_j \subset IB$ with only m-1 attributes ($A_i = IB - \{q_i\}$). Therefore the effect of dropping exactly one attribute is given by the remaining m entries: W(k, 0, 1), W(k, 0, 2), ..., W(k, 0, m).

3. The m entries W(k, 0, 1), W(k, 0, 2), ..,W(k, 0, m) are put together to form a "sensitivity tuple" of the key element k, s(k) being $[s_1, ..., s_m]$.
4. The larger s_i the larger is the symmetrized difference between G(k, IB) and G(k, A_i) and correspondingly the larger the influence of attribute q_i on the position of key element k within the Hasse diagram under IB compared with that under A_i .
5. The matrix **W**(k) depends on the selection of the key element k. If however, more objects are to be analyzed we generalize according to equation (20).
6. **W**(E) will be used as a measure of sensitivity. Accordingly we quantify the sensitivity by:
7. $\sigma(i) := W(E, IB, A_i)\ 1 \leq i \leq m$ (21)
with the enumeration scheme of step 3).
8. It can be shown that $\sigma(i)$ has values between 0 and $n\cdot(n-1)$. Hence a measure of attribute's sensitivity, independent of the number of objects is:
9. $\sigma_{norm}(i) = W(E, IB, A_i)/[n\cdot(n-1)]$. $0 \leq \sigma_{norm}(i) \leq 1$

Evaluation of Sampling Sites

Sediment samples of Lake Ontario as object set and the tests of the battery as information base

A battery of tests developed by Dutka et al. 1986 to test the sediments of near-shore sites of Lake Ontario (Canadian part) is used to exemplify the definitions and some results of HDT. In Lake Ontario 55 sediment samples were tested, thus, the set E contains 55 objects. Dutka et al. classified their results and used discrete scores instead of the measured (raw) data. For our analysis we have adopted their classification. Thus, s_i denotes the score of the i-th test of the battery. Five specific tests form the actual battery: (1) Fecal Coliforms „FC“, as an indicator designed to control the health state of the sediments, (2) Coprostanol „CP“ and (3) Cholesterol „CH“ both being indicators of loadings by fecals, (4) Microtox tests „MT“ and (5) Genotoxicity tests „GT“ disclosing some kind of acute toxicity and the potential for carcinogenicity, respectively (see Table 3).

Table 3. Scores of the 5 test battery results for representatives of the equivalence classes of $E/\Re$

identifier	FC	CP	CH	MT	GT	identifier	FC	CP	CH	MT	GT
1	2	0	0	4	0	17	3	0	0	6	0
2	1	0	0	2	0	18	1	0	0	2	4
3	2	0	0	2	0	23	1	0	0	0	4
4	3	0	0	0	0	25	4	0	0	0	0
5	3	3	2	0	0	27	5	0	0	0	0
7	2	0	0	8	0	31	4	5	4	0	0
9	1	0	0	6	2	32	3	0	0	8	0
11	1	0	0	0	0	91	2	0	0	0	0
12	3	0	0	2	0	92	3	0	0	4	0
14	1	0	0	8	0	95	3	5	2	6	0

By scoring the data many equivalence classes (in fact 20) arise (vide infra). It is convenient to refer only to these classes by specifying a representative for each class Thus, besides the sensitivity study we apply the concept of quotient sets. With the equivalence relation $\Re$ meaning equality in all five scores s_{FC}, s_{CP}, s_{CH}, s_{MT} and s_{GT}, the following sediment samples appeared as equivalent, (Table 4) the quotient set being denoted as $E/\Re$.

Table 4. Nontrivial equivalence classes and their battery of tests pattern. No. of sites in bold letters are later used as representatives for the whole equivalence class ec_i

Equivalence Class (ec)	card (ec)	FC	CP	CH	MT	GT
ec_1={**2**,8}	2	1	0	0	2	0
ec_2={**4**,6,10,13,19,21,22,29,30,48,94}	11	3	0	0	0	0
ec_3={**11**,16,40,41,42,43,44,45}	8	1	0	0	0	0
ec_4={15,**92**}	2	3	0	0	4	0
ec_5={**17**,35}	2	3	0	0	6	0
ec_6={20,24,26,28,34,37,39,49,50,51,**91**,93}	12	2	0	0	0	0
ec_7={**23**,60}	2	1	0	0	0	4
ec_8={**27**,33,46,47}	4	5	0	0	0	0

The sites, referred to as site numbers in bold letters are later used as representatives for the whole equivalence class. The site numbers are used as object identifiers.

The quotient set $E/\Re$ consists of the 8 equivalent classes {ec1, ec2, ec3, ec4, ec5, ec6, ec7, ec8} together with remaining 12 singletons {1}, {3}, {5}, {7}, {9}, {12}, {14}, {18}, {25}, {31}, {32}, {95}.

Now we apply all the characteristic numbers of Hasse diagrams, introduced earlier in this chapter.

n = card E = 55, Z = card $E/\Re$ = 20, NECA = 8, NA = 43

Clearly: card E = NA + SG, SG the number of singletons (here: SG = 12) and SG = Z - NECA.

The information base of the battery of tests is: IB: = $\{s_{FC}, s_{CP}, s_{CH}, s_{MT}, s_{GT}\}$. The partial ordering of the samples arises as explained in sections 2 and 3. The visualization of the partial order by HDT is depicted in Fig. 19.

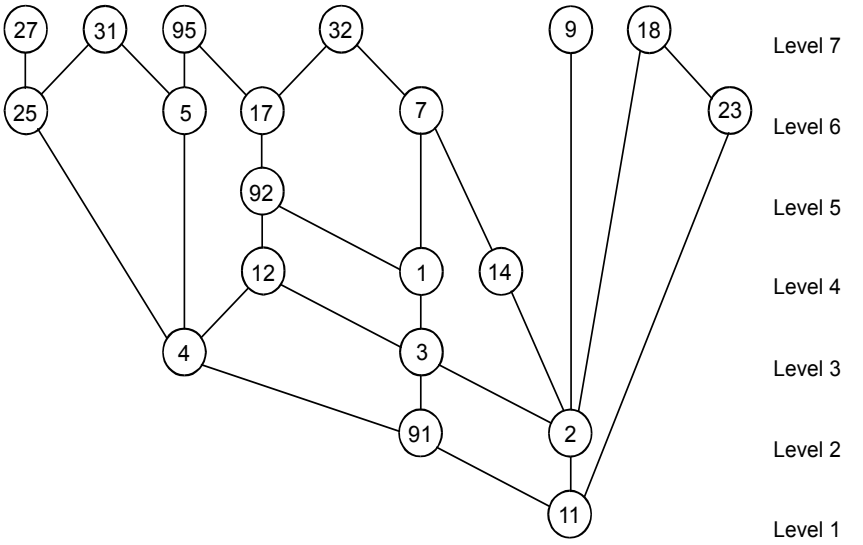


Fig. 19. The comparative evaluation of samples of the Lake Ontario, as generated by the WHASSE software. Hasse diagram of the poset $(E/\Re, \leq)$.

We check now the items discussed in former sections by some illustrative examples. Note, that in the following sections the term sites is used for the single objects/elements covered by the ranking exercise, reflecting the actual nature of the data material.

Comparability:

Taking site 31 as an example it is immediately seen that due to the transitivity (see El-Basil, p. 3) this site is comparable to (and worse than) site 4. Thus, 31 ≥ 4 as we have the sequence 31 ≥ 5 and 5 ≥ 4 from which 31 ≥ 4 follows logically. Likewise, through a longer chain 32 ≥ 17 ≥ 92 ≥ 1 ≥ 3 ≥ 91 ≥ 11 it follows that 32 ≥ 11. We say that 32 are connected to, or comparable

to 11, because there is a path, which can be followed without changing the orientation. On the other hand, site 17 is not connected/comparable to site 14, because there is no path, which can be followed from 17 to 14 without changing the orientation: $17 \geq 92 \geq 1 \geq 3 \geq 2$, however: $2 \leq 14$. The relation between the sites 31 and 5, displaying $31 \geq 5$, is a cover relation, whereas the relation between the sites 31 and 4, although $31 \geq 4$ is not a cover relation, as there is an in-between element, i.e., site 5 located between site 31 and site 4. What does comparabilities or chains tell us? By identifying chains we know that the upper object (e.g. site 32) is in all aspects worse than the lower object (e.g. site 3). All attributes increase simultaneously when the path from the lower element, i.e., site 3 to the site 32 is followed. In mathematical terms this can be described as a weak positive monotonous function, i.e. equal or increasing values of all attributes simultaneously following a chain.

Incomparability and Anti-chain:

Site 32 is, e.g., incomparable to site 9 as well as to many others. There is no path (in the digraph) by which we can start from site 32 and stop at site 9 without changing the orientation. It should be remembered that in an ordinary graph there is a path: $32 \rightarrow 7 \rightarrow 14 \rightarrow 2 \leftarrow 9$. However, the arrows recall that in the digraph we have an orientation, whereas in the ordinary graph we only have a line. The set {25, 5, 17, 7, 23} is an example of an anti-chain (cf. Fig. 19). However, this anti-chain is not of maximum length as site 9 could be added without violating the definition of an anti-chain. Large anti-chains indicate a high diversity of attribute profiles. Incomparabilities arise if at least one pair of attributes is antagonistic: i.e. a "walk" from an object x to an object y is accompanied with increasing of at least one attribute and decreasing of at least one other. For an illustration, take the incomparable sites 95 and 32. As the incomparability arises from the fact that CP, CH increase, FC and GT do not change, whereas MT decreases if the path from site 32 to site 95 is followed (cf. Table 3 and Fig. 19).

Priority elements:

As the sampling sites with high responses of the test-battery are of most interest, the maximal elements are taken as priority elements, i.e. the equivalence classes {27, 33, 46, 47}, {31}, {95}, {32}, {9}, {18}. From this we conclude that a) the sites 27, 33, 46, 47, 31, 95, 32, 9, 18 are of specific importance, and b) the set of sites {27, 33, 46, 47} has the same profile of scores, thus, they may be remedied by the same methods, whereas the attribute profiles differ among all other priority objects.

Characterizing numbers:

With the Hasse diagram of Fig. 19 at hand it is easy to derive the remaining characterizing numbers discussed in former sections. Hence, we find L(E) = 6, H(E) = 7, and NL = H(E) =7. These numbers give an impression in which detail the steps from a minimal element to a maximal element may be disclosed. This informs us here about the maximum possible differentiation in the degree of hazards.

In the present case (cf. Fig. 19) a partitioning of E (or $E/\mathfrak{R}$) into levels of increasing hazard prevails. Thus, {ec3} < {ec6, ec1} < {ec2, {3}} < {{12},{1},{14}} < {ec4} < {{25}, {5}, ec5, {7}, ec7} < {ec8, {31}, {95}, {32}, {9}, {18}}, the "<" sign reflecting that the sets are ordered corresponding to their level number.

We further find that NEL = 6, which in the present case coincides with NMAX = 6. The number of minimal elements NMIN = 1.

Finally the stability is to be calculated: P(IB) = 0.574

This means than on one hand the Hasse diagram will change remarkably, if an attribute is omitted or if an additional attribute is included, leading to new P(IB) values of 0.247 and 0.832, respectively. Hence, omitting an attribute changes the Hasse diagram towards a chain, whereas adding a new attribute causes the appearance of several hierarchies, eventually leading to an anti-chain.

See also for another example in chapter by Helm, p. 298.

Linear extensions:

As 20 objects (elements) of E/$\mathfrak{R}$ (= Z) are a rather high number, we would have to expect up to $2 \cdot 10^{18}$ linear extensions we restrict our study to the order ideal O(95). Its Hasse diagram is shown in Fig. 20.

For the poset, shown in Fig. 20 a total of 66 different linear extensions are possible. In the present context it makes no sense to list them all. For illustration a random selection of 5 linear extensions is listed below:

L1: 11 < 2 < 91 < 3 < 1 < 4 < 5 < 12 < 92 < 17 < 95
L2: 11 < 91 < 2 < 3 < 4 < 1 < 12 < 5 < 92 < 17 < 95
L3: 11 < 2 < 91 < 4 < 3 < 1 < 12 < 5 < 92 < 17 < 95
L4: 11 < 91 < 4 < 2 < 3 < 1 < 12 < 92 < 5 < 17 < 95
L5: 11 < 91 < 2 < 4 < 3 < 1 < 5 < 12 < 92 < 17 < 95

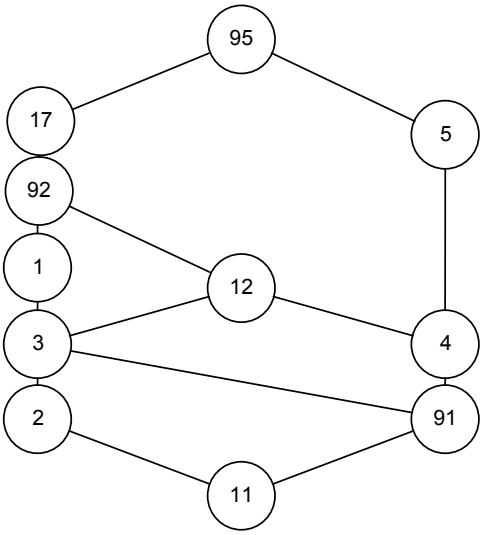


Fig. 20. *O*(95) , the order ideal generated by object (sampling site) 95. All considerations here are based on $E/\Re$

The jump-numbers are 3, 6, 4, 3, 5 for L1 to L5, respectively, the jumps in, e.g., L1 being found between $2 < 91$, $1 < 4$ and $5 < 12$.

What kind of information can be derived from this? If we represent a poset by a set of linear extensions, then those of major interest are those preserving the chains of the poset as far as possible. It is obvious that L2 is a correct representation of the partial order. However, the chains that can be identified (cf. Fig. 20) are separated by many elements, which originally did not belong to chains.

A further use of linear extension is the probability scheme (ranking probabilities) that they provide. Probability plots are depicted for the three sites 1, 17 and 91 (Fig. 21a) and for site 5 (Fig. 21b), respectively. (See also the contributions, chapters by Voigt and Brüggemann, p. 327; Brüggemann et al., p. 237; Carlsen, p. 163.

Remarkable differences can be noted. Thus, in the case of the three sites 1, 17, and 91 rather sharp maxima are developed, indicating that they can safely be assigned to a rank near the maximum of their probability plot. However, the sites differ in their individual ranking position. Thus, site 91 takes a lower rank site 1 a medium rank and site 17 a rather high rank. Therefore a mutual ranking sequence of the sites 1, 17, and 91, i.e., $91 < 1 < 17$, can be given since the minimum rank of the one site apparently does not overlap significantly with the maximum rank of a lower positioned site.

The site 5, on the other hand, differs from the above discussed sites as a rather smeared out probability plot is disclosed. Thus, the eventual assignment of a rank for site 5 is uncertain. This can also be seen directly from the visualization in the Hasse diagram (Fig's. 20 and 19). The site 5 is not as strongly connected as the other three elements. In more detail the consequences are discussed in Brüggemann et al. (2001b). We can calculate the local quantity U(x), i.e. the number of incomparabilities of an element x. The larger the values of U(x) the more uncertain the rank of x is. In the case of site 5 it turns out that U(5) = 6 , whereas the corresponding values for the site 91, 1, 17 are U(91) = 1 , U(1) = 2 , and U(17) = 1, respectively. Therefore the measure of uncertainty about the ranks is U(91) = U(17) < U(1) << U(5).

As stated earlier, it is possible to calculate averaged ranks; the full list of information is given in Table 5, where the minimum, maximum rank and the local incomparabilities are displayed.

Table 5. Summary of the analysis by linear extensions

Identifier	Min	Rkav	Max	U(x)
1	5	6.67	8	3
2	2	2.85	5	3
3	4	4.70	6	2
4	3	4.33	6	3
5	4	7.67	10	6
11	1	1	1	0
12	6	6.94	8	2
17	9	9.82	10	1
91	2	2.39	3	1
92	8	8.63	9	1
95	11	11	11	0

The analysis by linear extensions is very attractive as it helps to derive a linear ranking, without any subjective preferences. The data lead to a poset, the poset may be analyzed with respect to its structure, this is a combinatorial problem, and finally a ranking probability can be derived. Crucially in this procedure is that very different attribute profiles may lead to the same Hasse diagram and thus to the same set of linear extensions and therefore finally to the same probability characteristics: Thus, the attribute profiles a) (0,0), (1,0), (0,1), (1,1) and b) (0,0), (1,0), (0,5), (4,7) lead to identical Hasse diagrams.

A priori this is fine as the first attribute definitely should compensate the second one. However, the sites, which belong to (1,0),(0,1) on the one side and (1,0), (0,5) on the other side will get the same averaged rank! Thus, the analysis by linear extensions alone should be carried out with appropriate care. We continue the analysis of the poset and discuss the attribute profiles.

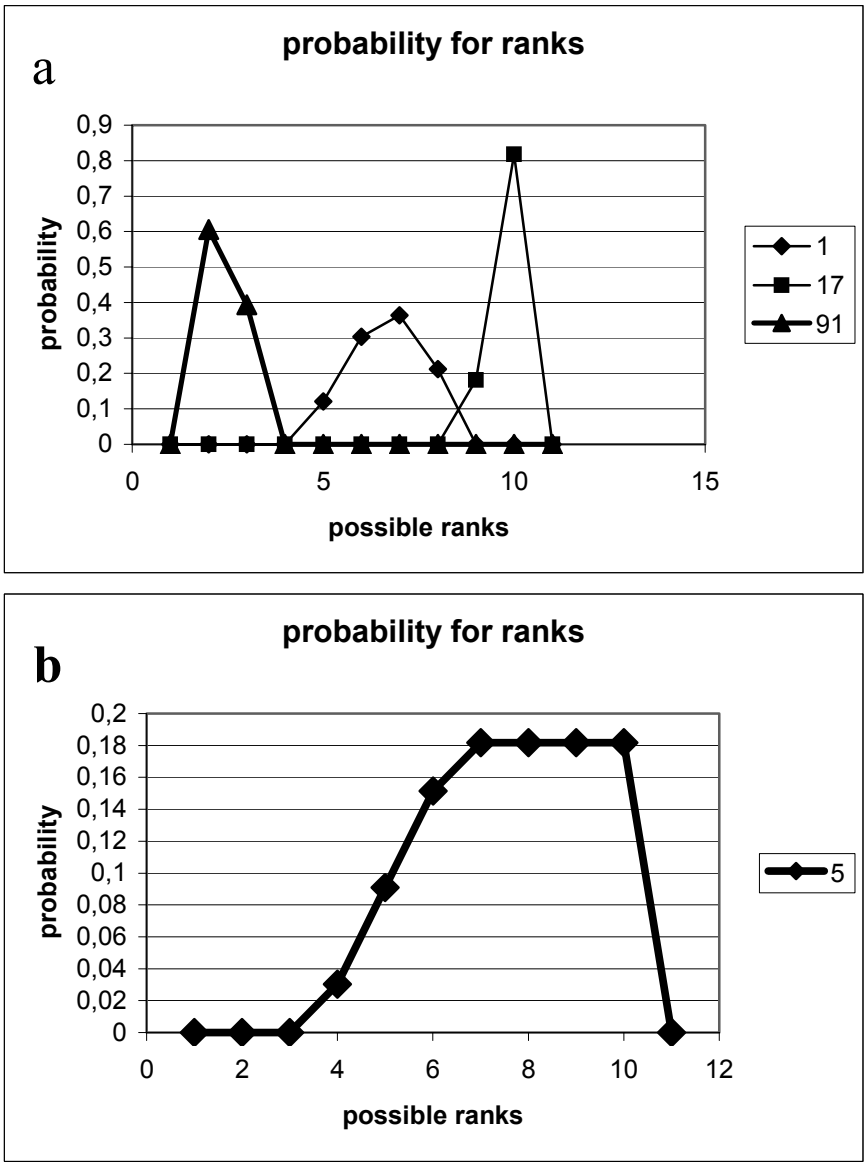


Fig. 21. Probability plots for 4 elements of the poset, shown in Fig. 20

Up to now, we have a quite good overview about the ranking of sites in Lake Ontario. However, does the test battery comprise redundancies? The subsequent dimension analysis will disclose this.

Dimension analysis

We find, applying equation 15 that the poset shown in Figure 19 has dimension 2. Once the dimension d of a poset is found with d < card *IB*, then corresponding many new latent ordering variables l_1, l_2, ... l_d may be used to form the same Hasse diagrams as found by the original attributes. Hence, the same ranking must be possible by a lower number of latent ordering variables and a redundancy within the battery appears possible. However, the numerical relation between the original attributes and the latent ordering variables may be rather difficult to derive and, if even then hard to interpret as it is often the case, e.g., in principal component analysis.

Corresponding to the dimension d = 2, the poset shown in Fig. 19 can alternatively be visualized by a two-dimensional grid as is shown in Fig. 22. Both visualizations have their advantages. Structures within a Hasse diagram, e.g., successor sets, or sets of objects separated from others by incomparabilities, can be more easily disclosed by a representation like that of Fig. 19. In multivariate statistics reduction of data is typically performed by principal components analysis or by multidimensional scaling. These methods minimize the variance or preserve the distance between objects optimally. When order relations are the essential aspect to be preserved in the data analysis, the optimal result is a visualization of the sediment sites within a two-dimensional grid.

Some scores of the test battery are additionally shown. From them the values of the scores of other objects can be estimated or exactly calculated. For example, for site 17, FC must have the value 3, because the lower object 92 and the higher object 95 have $s_{FC} = 3$. The value of CP must be 0 because $s_{CP}(32) = 0$, which is the lowest value. Similarly $s_{CH}(17) = 0$ and $s_{GT}(17)=0$, whereas for $s_{MT}(17)$ only the interval $4 \leq s_{MT}(17) \leq 8$ can be predicted from the knowledge of the neighbours in the Hasse diagram.

The grid (Fig. 22) can be thought of as being a coordinate system, with one axis of a latent order variable l_1 and another by l_2, according to d = 2. By these two latent ordering variables, each element $\in E/\Re$ can be characterized by a pair, which represents correctly the order relations (Compare Figure 17) that are important for ranking but which is clearly not unique with respect to a numerical representation. The interpretation of the latent variables l_1 and l_2 is supported by checking the configurations within the two-dimensional grid in terms of its a priori content (variables FC, CP, CH, MT, GT). A clear correlation can be detected between FC and the latent variable l_1 and also between GT and the latent variable l_2.

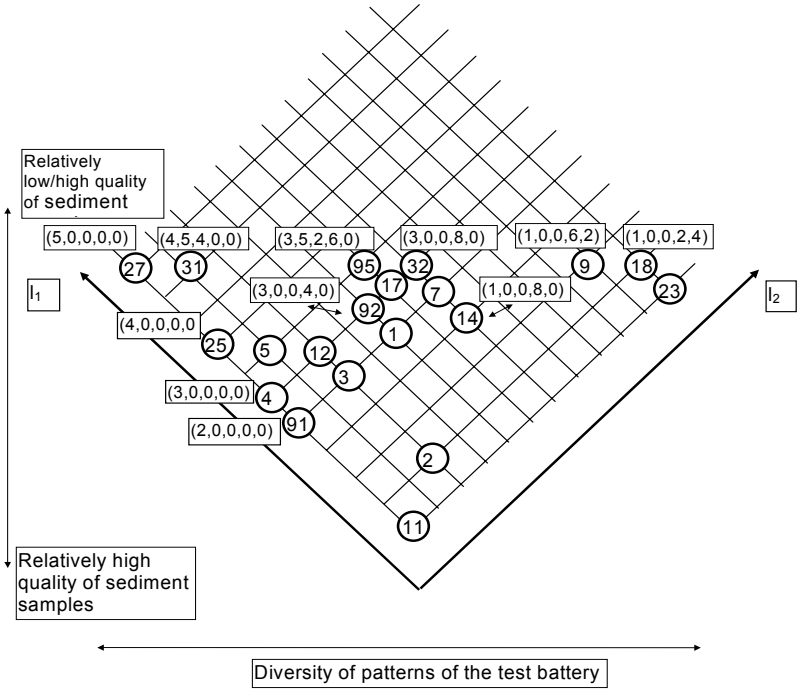


Fig. 22. Visualization of the ranking result of the sediment samples of Lake Ontario after dimension analysis

Sometimes these variables FC and GT with primary meaning are called polar items Shye 1985, Borg & Shye 1995. For further elucidation see also the multivariate technique posac (partial order scalogram analysis with coordinates), which is explained in Brüggemann et al. 2003, Voigt & Welzl, 2002 and for which a tool is provided in the software package Systat (R) 2000.

The other variables accentuate the possibility of discrimination in a nonlinear manner. Therefore, in a qualitative sense, the ranking of the sediment sites of the Lake Ontario seems to be determined by a hygienic and an ecotoxicological component. Some objects could be embedded into the grid on alternative ways. However, the order theoretical information, namely the comparabilities and incomparabilities are maintained. This can be easily proved by verifying that the Hasse diagram induced by five attributes (Fig. 19) is isomorphic to that, induced by the two latent variables (Fig. 22). If the ranking is in mind, then obviously the five tests apparently contain some redundancies, because the decision for "good" or "bad" could also be given on the basis of two coordinate values.

Sensitivity analysis of the ranking

For our example the matrix **W** has the following values (Table 6)

Table 6. Values of the matrix **W** for different combinations of attribute

W	case 0 FC,CP,CH, MT,GT	case 1 CP,CH,M T,GT	case 2 FC,CH,M T,GT	case 3 FC,CP, MT,GT	case 4 FC,CP,C H,GT	case 5 FC,CP,C H,MT
case 0	0	795	0	0	360	124
case 1	-	0	795	795	1155	919
case 2	-	-	0	0	360	124
case 3	-	-	-	0	360	124
case 4	-	-	-	-	0	484
case 5	-	-	-	-	-	0

It is seen that cases 1 to 5 excludes one after another FC, CP, CH, MT and GT, respectively. Thus, comparing these cases to case 0, including all 5 attributes, will disclose the relative importance of the 5 tests comprising the battery. Thus, from this matrix the sensitivities are $\sigma(FC) = 795$, $\sigma(CP) = \sigma(CH) = 0$, $\sigma(MT) = 360$ and $\sigma(GT) = 124$, respectively, unambiguously disclosing the test "FC" as the most important within the attribute set containing the five tests. The tests CP and CH apparently do not have any influence at all on the order theoretical structure of the set of samples, i.e. they do not influence the prioritization of the sites. Their low sensitivities are also found by Dutka et al. 1986, who established a regression model between the two quantities. It is emphasized that this conclusion refers to the classified values of the battery of tests. Hence, the result with respect to FC should be carefully examined as the high sensitivity may be induced by the scoring process.

Fig. 23 shows the Hasse diagram (generated by the computational software, WHASSE (Brüggemann et al. 1999 a) therefore drawn in its standard format: circles, and each object as high as possible in the drawing plane):

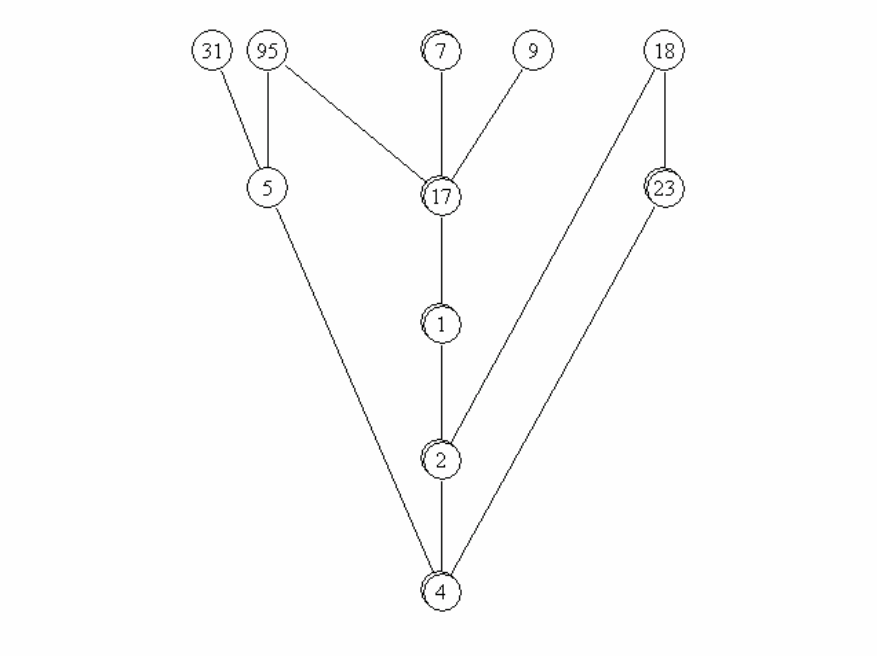


Fig. 23. 55 samples evaluated with the test battery of Dutka, excluding the FC test. Note that many samples are members of non-trivial equivalence classes

The dramatic changes compared to the original Hasse diagram (Fig. 19) are immediately seen.

Discussion and Conclusion

The battery of tests approach helps to evaluate sites using different criteria simultaneously: The decision of which sites are "good" or "bad", i.e. the sorting process is more difficult the larger the number of samples and especially the larger the number of tests, since there is more information that can be used to differentiate among the tested objects. This, in turn, leads to difficulties for ranking, because the complexity of a well-designed battery is being lost, if in order to compare the tested objects, a ranking index like

$$\Gamma = \sum g_i \cdot q_i \qquad (22)$$

is constructed. The presentation by a Hasse diagram avoids the arbitrariness in constructing a ranking index. Applying concepts of partially ordered sets must not be performed in isolation. All results depend on the data representa-

tion used. The present study aimed at demonstrating the HDT using, and extending the results of Dutka et al. 1986. Therefore we did not need statistical analyses. However, generally, the appropriate data representation is of much concern, Brüggemann & Welzl 2002. The use of cluster analysis and principal component analysis may be helpful to obtain a statistical relevant data representation and to avoid insignificant numerical differences of the attributes, which in turn would lead to insignificant comparabilities and incomparabilities and thus to very complex Hasse diagrams.

A combination of Hasse Diagram Techniques and explorative statistical methods could be a very promising approach to future tasks in environmental sciences. Approaches in this respect were followed on the pollution of regions in Germany with heavy metals, cf. Brüggemann et al. 1999b) and on the contents of environmental databases, cf. Voigt et al. 2004.

The analysis of empirical datasets may lead to empirical partial orders, which do not necessarily fulfill the axioms of lattices. The school around Wille (Wille 1987 and Ganter & Wille 1996) has shown how it is possible nevertheless to construct a lattice. The resulting lattices and the analysis based on them is called "Formal concept analysis". As lattices fulfill more axioms than posets generally, one gets a richer theory of them. Especially it is possible to generate a set of implications. See chapter by Kerber, p. 355 for introductory examples.

The main advantage of a ranking by HDT is that it can be performed without any normative constraints. HDT simply sorts the objects without any additional information. Beyond sorting, many conclusions may be drawn from the Hasse diagrams as they represent a well-defined mathematical structure. Summarizing the following recommendations can be given:

- If the battery of tests is used to test many objects, perform a cluster analysis to get rather numerically robust results. Instead of the measured results for each object use some characteristic values of the cluster (mean values or some other quantities, describing a cluster center).
- Apply HDT to look for priority objects, to identify objects or subsets with characteristic patterns (in mathematical terminology: find "order ideals") or to select sequences (in order theoretical terminology: "chains") of objects.
- Perform a dimension analysis to estimate the redundancy of the test system and a sensitivity study to identify important or less important attributes. The rational for the importance of each attribute cannot be drawn from the HDT; here the scientific background is needed: What are the characteristics for all the tested objects are there any internal correlations among the attributes?

- If an aggregation is done, as, e.g., by eqn. 22 then note that the weights may have an important influence on the ranking results via Γ if objects have an high degree of incomparability, i.e. have a large value for U(x).

References

Atkinson MD (1989) The complexity of Orders. In: Rival, I. (Ed.) Algorithms and Order NATO ASI series. Series C Mathematical and Physical Sciences Vol. 255 Kluwer Academic Publishers, Dordrecht, pp 195-230

Atkinson MD (1990) On the computing the Number of Linear Extensions of a Tree. Order 7:23-25

Beckenbach EF, Bellman R (1971) An Inequality of Karamata. In: Inequalities, 3 edn. Springer-Verlag, Berlin, pp 1-107

Birkhoff G. 1984. Lattice theory. Providence, Rhode Island: American Mathematical Socienty, Vol XXV

Borg I, Shye S (1995) Facet Theory - Form and Content. Sage Publications, Thousand Oaks, California

Brans JP, Vincke PH (1985) A Preference Ranking Organisation Method (The PROMETHEE Method for Multiple Criteria Decision - Making). Management Science 31:647-656

Brans JP, Vincke PH, Mareschal B (1986) How to select and how to rank projects: The PROMETHEE method. European Journ Oper Research 24:228-238

Brüggemann R, Bartel HG (1999) A Theoretical Concept to Rank Environmentally Significant Chemicals. J Chem Inf Comput Sci 39:211-217

Brüggemann R, Bücherl C, Pudenz S, Steinberg C (1999a) Application of the concept of Partial Order on Comparative Evaluation of Environmental Chemicals. Acta hydrochim hydrobiol 27:170-178

Brüggemann R, Halfon E (1997) Comparative Analysis of Nearshore Contaminated Sites in Lake Ontario: Ranking for Environmental Hazard. J Environ Sci Health A32(1):277-292

Brüggemann R, Halfon E, Welzl G, Voigt K, Steinberg C (2001b) Applying the Concept of Partially Ordered Sets on the Ranking of Near-Shore Sediments by a Battery of Tests. J Chem Inf Comp Sc 41:918-925

Brüggemann R, Pudenz S, Carlsen L, Sørensen PB, Thomsen M, Mishra RK (2001a) The Use of Hasse Diagrams as a Potential Approach for Inverse QSAR. SAR and QSAR Environ Res 11:473-487

Brüggemann R, Pudenz S, Voigt K, Kaune A, Kreimes K (1999b) An algebraic/graphical tool to compare ecosystems with respect to their pollution. IV: Comparative regional analysis by Boolean arithmetics. Chemosphere 38:2263-2279

Brüggemann R, Schwaiger J, Negele RD (1995a) Applying Hasse diagram technique for the evaluation of toxicological fish tests. Chemosphere 30(9):1767-1780

Brüggemann R, Simon U, Mey S (2005) Estimation of averaged ranks by extended local partial order models. Math Commun Math Comput Chem 54(3):489-518

Brüggemann R, Sørensen PB, Lerche D, Carlsen L (2004) Estimation of Averaged Ranks by a Local Partial Order Model. J Chem Inf Comp Sc 44:618-625

Brüggemann R, Welzl G (2002) Order Theory Meets Statistics -Hassediagram technique-. In: Voigt K and Welzl G (eds) Order Theoretical Tools in Environmental Sciences - Order Theory (Hasse diagram technique) Meets Multivariate Statistics-. Shaker-Verlag, Aachen, pp 9-39

Brüggemann R, Zelles L, Bai QY, Hartmann A (1995b) Use of Hasse Diagram Technique for Evaluation of Phospholipid Fatty Acids Distribution in Selected Soils. Chemosphere 30(7):1209-1228

Brüggemann R, Welzl G, and Voigt K (2003) Order Theoretical Tools for the Evaluation of Complex Regional Pollution Patterns. J Chem Inf Comp Sc 43:1771-1779

Carlsen L, Lerche DB, Sørensen PB (2002) Improving the Predicting Power of Partial Order Based QSARs through Linear Extensions. J Chem Inf Comp Sc 42:806-811

Carlsen L, Sørensen PB, Thomsen M (2001) Partial Order Ranking - based QSAR's: estimation of solubilities and octanol-water partitioning. Chemosphere 43:295-302

Carlsen L, Sørensen PB, Thomsen M, Brüggemann R (2002) QSAR's based on Partial Order Ranking. SAR and QSAR in Environmental Research 13:153-165

Colorni A, Paruccini M, Roy B (2001) A-MCD-A, Aide Multi Critere a la Decision, Multiple Criteria Decision Aiding. JRC European Commission, Ispra

Davey BA, Priestley HA (1990) Introduction to Lattices and Order. Cambridge University Press, Cambridge

Dutka BJ, Walsh K, Kwan KK, El Shaarawi A, Liu DL, Thompson K (1986) Priority site selection from degraded areas based on microbial and toxicant screening tests. Water Poll Res J Canada 21(2):267-282

Freese R. 2004. Automated Lattice Drawing. In: Eklund P, Ed. ICFCA 2004, Berlin: Springer-Verlag, p 112-127

Ganter B and Wille R (1996) Formale Begriffsanalyse Mathematische Grundlagen, Springer-Verlag, Berlin, pp 1-286

Halfon E, Hodson JA and Miles K (1989) An algorithm to plot Hasse diagrams on microcomputer and Calcomp plotters. Ecol Model 47:189-197

Heinrich R (2001) Leitfaden Wasser - Nachhaltige Wasserwirtschaft; Ein Weg zur Entscheidungsfindung. Wasserforschung e.V., Berlin

Jonsson B (1982) Arithmetic of Ordered Sets. In: Rival I (Ed) Ordered Sets. D. Reidel Publishing Company, Dordrecht, pp 3-41

Lerche D, Brüggemann R, Sørensen PB, Carlsen L, Nielsen OJ (2002) A Comparison of Partial Order Technique with three Methods of Multicriteria Analysis for Ranking of Chemical Substances. J Chem Inf Comp Sc 42:1086-1098

Lerche D, Sørensen PB (2003) Evaluation of the ranking probabilities for partial orders based on random linear extensions. Chemosphere 53:981-992

Lerche D, Sørensen PB, Brüggemann R (2003) Improved Estimation of the Ranking Probabilities in Partial Orders Using Random Linear Extensions by Approximation of the Mutual Probability. J Chem Inf Comp Sc 53:1471-1480

Muirhead RF (1900) Inequalities relating to some Algebraic Means. Proc Edingburgh Math Soc 19:36-45

Muirhead RF (1906) Proofs of an Inequality. Proc Edingburgh Math Soc 24:45-50

Neggers J, Kim HS (1998) Basic Posets. Singapore: World Scientific Publishing Co

Patil GP, Taillie C (2004) Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. Environmental and Ecological Statistics 11:199-228

Pavan,M. and R.Todeschini, 2004. New indices for analysing partial ranking diagrams. Analytica Chimica Acta 515: 167-181.

Randić M (2002) On Use of Partial Ordering in Chemical Applications. In: Voigt K and Welzl G (Ed.) Order Theoretical Tools in Environmental Sciences - Order Theory (Hasse Diagram Technique) Meets Multivariate Statistics. Shaker-Verlag, Aachen, pp 55-64

Rival I (1981) (Ed.) Ordered Sets, NATO ASI Series, Series C. Mathematical and Physical Sciences Vol. 83, Kluwer Academic Publishers, Dordrecht

Rival I (1983) Optimal Linear Extensions by Interchanging Chains. Proc AMS 89:387-394

Rival I (1985a) (Ed.), Graphs and Order, NATO ASI Series, Series C. Mathematical and Physical Sciences Vol. 147, Kluwer Academic Publishers, Dordrecht

Rival I (1985b) The diagram. In: Rival I (1995 a) pp 103-133

Rival I (1989) (Ed.) Algorithms and Order, NATO ASI Series, Series C. Mathematical and Physical Sciences Vol. 255, Kluwer Academic Publishers, Dordrecht

Rival I, Zaguia N (1986) Constructing Greedy Linear Extensions by Interchanging Chains. Order 3:107-121

Roy B (1972) Electre III: Un Algorithme de Classements fonde sur une representation floue des Preferences En Presence de Criteres Multiples. Cahiers du Centre d'Etudes de Recherche Operationelle 20:32-43

Roy B (1990) The outranking approach and the foundations of the ELECTRE methods. In: Bana e Costa (Ed.) Readings in Multiple Criteria Decision Aid. Springer-Verlag, Berlin, pp 155-183

Ruch E (1975) The diagram lattice as Structural Principle. Theor Chim Acta 38:167-183

Ruch E, Gutman I (1979) The Branching Extent of Graphs. J of Combinatorics-Inform System Sciences 4:285-295

Schröder BSW (2003) Ordered Sets An Introduction. Birkhäuser, Boston

Shye S (1985) Multiple Scaling - The theory and application of partial order scalogram analysis. North-Holland, Amsterdam

Stanley P (1986) Enumerative Combinatorics. Volume I, Wadsworth & Brooks/ Cole, Monterey, California

Systat Program http://www.spss.com/software/science/SYSTAT/, SPSS Inc., 2000

Trotter WT (1991) Combinatorics and Partially Ordered Sets Dimension Theory; John Hopkins Series in the Mathematical Science. The J Hopkins University Press: Baltimore

Urrutia J (1989) Partial Orders and Euclidian Geometry, in Rival, I, Ed. 1989, pp 387-434

Voigt K, Welzl G, Brüggemann R (2004) Data analysis of environmental air pollutant monitoring systems in Europe. Environmetrics 15:577-596

Wille R (1987) Bedeutungen von Begriffsverbänden: Beiträge zur Begriffsanalyse (Hrsg.: Ganter B, Wille R, Wolff KE). BI Wissenschaftsverlag Mannheim pp 161-211