MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **66** (2011) 231-251

ISSN 0340 - 6253

A New Tool to Analyze Partially Ordered Sets
Application: ranking of polychlorinated biphenyls and alkanes/alkenes in river Main, Germany

Rainer Bruggemann[1], Kristina Voigt[2]

[1] *Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department Ecohydrology, Berlin, Müggelseedamm 310, 12587 Berlin, Germany, E-mail:* brg_home@web.de

[2] *Helmholtz Center Munich, Institute of Biomathematics and Biometry, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany, E-mail:* kvoigt@helmholtz-muenchen.de

(Received November 26, 2010)

Abstract

Partial order renders a helpful tool in many ranking problems. Especially when multi-indicator systems come into play, the product order as a special order relation is an adequate mathematical structure. A typical outcome of partial order analysis are the chains, where the values of the indicators are weak monotonously increasing. However, one also finds in partial order analysis incomparable elements. In this paper we suggest the construction of a special type of a binary relation, a tripartite graph. By means of the tripartite graph the role of indicators causing incomparabilities can be clarified. Furthermore, quantities derived from tripartite graphs help to decide whether or not the sets of incomparable elements are topologically connected. Hence, tripartite graphs enlighten the interpretation of complex data sets.

1. Introduction

Chemicals can be harmful to humans and the environment. How do we find out whether they are hazardous? There are many time-consuming and expensive investigations necessary to perform a risk assessment. After experimental determination of chemicals' properties the exposure and its toxicological consequences for humans and the environment is determined. One of the first attempts to develop mathematical models was the software package E4CHEM, which was developed 1984-86 [10], now the software package EUSES is widely

used (see e.g. [41,42]). Even with the simplified version of EUSES, described by [41], there is still much effort to perform the risk assessment. Hence the question is: with which chemicals to begin at first? Thus ranking is needed to give the more involved investigations a reasonable operating sequence ([29] or in the context of EUSES: [25]). Once accepted that a ranking is needed, we discover that there is no intrinsic property of a chemical through which it can be ranked. Hence, several aspects of a chemical need to be simultaneously considered and the central question arises: how to rank chemicals characterized by several attributes?

Partial order theory as a discipline of discrete mathematics can be very helpful in ranking studies, see for example [7,9,11,19,21,26,31-34,36-38]. Here we introduce at a new tool supporting the analysis of partial order especially with respect to incomparabilities and introduce the "tripartite graph".

The paper is organized as follows:

First we start with the basics of partial order, and then we introduce the example, a set of chemicals to be ranked. After discussing the example we explain the concept of a tripartite graph, derive some properties and apply this tool on our example. A discussion concludes the paper.

2. Basics of the theory of partially ordered sets

2.1 Product order

Although the application of partial order on data matrices is explained in detail in many references, we introduce some concepts at the sake of convenience for the reader. For more theoretical background, see e.g. [23,28,35]. See also a review of application of partial order in the field of chemistry, [27].

Let us suppose that an "object set" X (in technical terms also called a "ground set") is of interest. Suppose that X is a finite set (and we do not mention it further). We wish to compare objects $\in$ X. Therefore we use the symbol $\leq$ as a binary relation among the objects. In our case - an n x m data matrix is at hand, the special order relation called product- or componentwise order must be applied: Let x, y be two different objects of the object set X. Let Q be the space of measurements may be of different scaling levels. Let q(x) be the data row for x and q(y) that for y, i.e q(x), q(y) $\in$ Q: $x \leq y$ if and only if $q(x) \leq q(y)$, $q(x) \leq q(y)$ if and only if $q_j(x) \leq q_j(y)$ for all j, j = 1, 2, …,m. We call q_j the attributes or indicators of the study.

An analysis of the data matrix by partial order is meaningless without the information about the aim of ranking. When the aim of ranking is known, the indicators are to be checked for their common orientation with respect to the aim. If necessary, the columns of the data matrix must be appropriately transformed (for example by multiplication by -1).

If x, y are different objects but $q(x) = q(y)$, i.e. $q_j(x) = q_j(y)$ for all j, then the objects x and y are called equivalent under the equivalence relation 'equality'. Equivalence is denoted as: $x \cong y$. If we want to exclude equivalence, then we write $x < y$. Consequently: $x < y$ if and only if $q(x) \leq q(y)$, $q(x) \leq q(y)$ with at least one q_{j*}, for which $q_{j*}(x) < q_{j*}(y)$ is valid. Consider representatives of each equivalence class and add the information about the other equivalent elements, whenever needed.

Component: when in a directed graph any two vertices are reachable without taking care of the orientation of the edges, then we speak of a weak connection and the maximal subset of vertices which is weakly connected is called a component of the graph.

Further notations: (i) Sometimes it is necessary to specify the $\leq$ or $<$ relation. In that case $\leq$ or $<$ get an appropriate subscript, eg. $\leq_{\{qi,qj\}}$ indicates that a partial order based on the indicators q_i and q_j, is considered. (ii) Posets based on a data matrix are indicated by (X, IB), whereby $IB=\{q_1,..,q_m\}$ is the set of attributes (indicators), also called the information base [8]. (iii) We speak of 'elements' if their membership to a set is considered and of objects in a more general sense. (iv) If neither $x \leq y$ nor $x \geq y$ then x and y are incomparable, notated as: $x \parallel y$. (v) If we want to indicate comparability for objects x and y without specifying the orientation, we write: $x \perp y$. (vi) If a partial order is to be denoted without reference to an attribute set, we use the notation $(X, \leq)$. (vii) When A is a finite set, we denote by $|A|$ the number of its elements.

2.2 Hasse diagrams

We first introduce the 'cover-relation': x is covered by y if there is no element $z \in X$ for which $x < z$ and $z < y$. We write this as $x \leq: y$. With the cover-relation at hand, we can get a diagrammatic representation of the partially ordered set: Let us consider x and y, and assume that $x \leq: y$. Then we draw x in a vertical plane below y and connect both with a straight line. This is repeated for every ordered pair, i.e. for all pairs of two objects for which $\leq:$ - relation holds. The resulting diagram is known as Hasse diagram. *Example:* X= {a, b, c, d, e, f} (Figure 1):

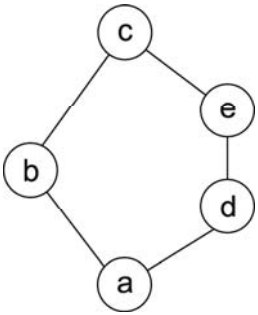

Figure 1: Hasse diagram representing a poset (X, ≤)

2.3 Some additional concepts:

1. The fact that a ≤ e can be easily deduced from the Hasse diagram: because of transitivity, no line appears for a ≤ e.
2. If there is no incomparability, then we speak of a complete, total or linear order. In the case of a complete order, the objects x ∈ X can be arranged in a sequence $x_1 < x_2 < \ldots < x_n$, i.e. a ranking is found.
3. Chain: If a subset X' ⊂ X can be found such that for all (x, y) ∈ X' × X', a complete order can be found, then this subset, together with the partial order relation, is a chain. Let C be a chain, if there is no outside element, whose inclusion into a C is possible maintaining the comparabilities, then C is a maximal chain. In Figure 1 ({d, a, b}, ≤) is a chain, but is not maximal, because object e can be included.
4. Bifurcation of chains: Two chains C_1 and C_2 have a subset of objects in common. For example: ({a, d, e}, ≤) and ({a, b}, ≤).
5. Weak order: Representative elements of equivalence classes are in a chain, but there are nontrivial equivalence classes.
6. Antichain: If a subset X' ⊂ X can be found such that, for no (x, y) ∈ X' × X', x ⊥ y, i.e. x ≤ y or y ≤ x holds, then this subset, equipped with the partial order relation, is called an antichain.

Hasse diagrams provide answers to "where is an object and why it is, where it is", because chains, bifurcation of chains and subsets of objects with common pattern of indicator values can be identified (Figure 1). If the Hasse diagram is too messy to get chains by inspection, software tools (e.g. PyHasse, [15,39,40]) help to find chains.

2.4 Formal Concept Analysis

Formal concept analysis (FCA) is a well known method which also belongs to the theory of partially ordered set. There are many publications available, see [1,3,22,24]. For the sake of the convenience of the reader we explain FCA following the lines of [2]: Formal concept Analysis is based on a triple (X,Q,I) , where X is the set of objects, Q the set of properties (in the case of multivalued attributes a preprocessing step, called scaling, has to be performed, [24], and I is a relation between X and Q. We write (x I q) if $x \in X$ has the property $q \in Q$ or if $q \in Q$ is realized by $x \in X$. Having all possible relations, I, a set of concepts can be derived. A concept is a pair of object subset X' and property subset Q'' such that $\forall\ x \in X'$ and for all $q \in Q''$ it is valid: (x I q). The concepts can be partially ordered on the basis of the inclusion relation applied to the object subsets of each context. In [43], it has shown that the resulting poset has the properties of a lattice.

The main point of FCA which is of interest here is the "symmetric" view on objects and properties, whereas a Hasse diagram, as shown in Figure 1 only shows the order relations focusing on the objects.

3. Example

The basis of this example is a study, in which chemicals in the German river Main were monitored in order to support management plans for that river. It was found that 19 organic compounds are of special interest. By applying a simple simulation model EXWAT, see [5,6,10] which is part of the evaluation package E4CHEM indicators were calculated which describe the fate of these chemicals on basis of their substance properties and the environmental properties of the river Main (for details, see [12]).

EXWAT couples as far as possible deterministically (that means on the basis of known processes and first principles like mass conservation, thermodynamics and kinetic laws) chemical properties like log K_{OW}, which is a measure for accumulating in solids (soils, sediments, suspended particles), and like Henry Law Coefficient, which is a measure about the thermodynamical tendency to accumulate in the air.

These properties are purely related to the chemical substances and are therefore depending on the chemical structure. The application of the model relates these chemical properties to actual tendencies to volatize, or to accumulate in the sediments, or just to stay in the water body. These are the three environmental fate descriptors used in this example.

The indicators obtained from the model EXWAT are neither describing only the chemical nor the river itself but the effect of transport and distribution mechanism due to the hydrological conditions on the chemicals. The indicators were transformed into scores, see Table 1.

Table 1: Scores of three indicators describing the fate of polychlorinated biphenyls (pcb) and some chloroalkanes,- alkenes (alk) in the river Main (Vl: Volatilization flow, Se: Sedimentation flow, Ad: Flow downstreams). pcb: Polychlorinated biphenyls, alk: Alkanes/Alkenes *)

	name	Volatilization (Vl)	Sedimimentation (Se)	Advection (Ad)
Pcb 12	PCB 28	3	3	2
Pcb 13	PCB 52	2	3	2
Pcb 14	PCB 101	2	4	1
Pcb 15	PCB 138	2	4	1
Pcb 16	PCB 153	1	4	1
Pcb 17	PCB 180	1	4	1
alk ch	chloroform	4	1	2
alk tt	tetrachloromethane	4	1	3
alk tn	trichloroethane	4	1	3
alk tr	trichloroethene	4	2	2
alk pe	tetrachloroethene	3	2	3

*) Some of these data were used as an example concerning Environmental Impact Assessment [20]

From the scores the following Hasse diagram was obtained (Figure 2). As labels we do not include the chemical classes ,i.e. the additional information like " pcb" or "alk".

When an evaluation, decision or even a ranking is wanted then this Hasse diagram is a poor result, as the length of the maximal chains (number of elements in a chain) is 2, which is compared to the possible length of a linear order a very small degree of separation: 2/8 = 0.25. The methodology of partial order offers several tools to obtain nevertheless a linear or weak order. In [17], some methods are discussed, as well as in [32], the crucial role of weights is discussed in case one is considering of deriving a weighted sum from the indicators.

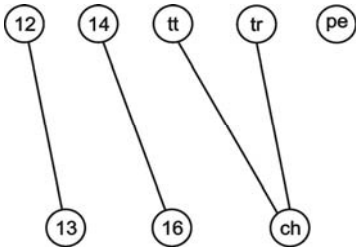

Figure 2: Fate of 8 organic compounds in the river Main, due to the simulation model EXWAT. The abbreviations are given in Table 1. Equivalent Objects: {tt, tn} {14, 15} {16, 17}. At the top are the relatively hazarduous chemicals.

Here, however we want to focus on the nonlinearity of partial order, which is expressed by the most striking pictorial effect, namely the bifurcation of chains and the appearance of

components. We see that ({14, 16}, ≤) is one component of the directed graph (Hasse diagrams are acyclic directed graphs whose underlying usual graph does not contain triangles), other component are ({13, 12}, ≤), ({pe}, ≤) (an isolated element) and {tr, ch, tt}, ≤). Furthermore we see that there is no connection between the chemical (contextual) defined subset of {tr, tt, ch, pe} and the polychlorinated biphenyls, which themselves are not in one component.

We discuss two tools, one is the separability, which is already published [16] and the other: the tripartite graph, which is the new tool we want to explain in this paper in some detail.

4 Separability and tripartite graph as tools in partial order analysis

4.1 Separability

Let us identify two disjoint subsets of $X/_{\cong}$: X_1 and X_2 . Two subsets X_1 and X_2 are separated, if $x \in X_1$ and $y \in X_2$ imply $x||y$ $\forall x \in X_1$ and $\forall\ y \in X_2$. We also want to speak of a degree of separation of subsets. Therefore we introduce the 'separability' Sep(X_1, X_2, IB): The possible number of relations (i.e. of < or > or ||-relations) N(X_1, X_2) between X_1 and X_2 is: N(X_1, X_2) = $|X_1|*|X_2|$. We define: U(X_1, X_2, IB) = {(x,y): $x \|_{IB} y$, $x \in X_1$, $y \in X_2$, $X_1 \cap X_2 = \varnothing$}. Then the separability between X_1 and X_2 is calculated as follows: Sep(X_1, X_2, IB) = |U(X_1, X_2, IB)| /N(X_1, X_2). We note: Sep(X_1, X_2, IB) = Sep(X_2, X_1, IB). If Sep(X_1, X_2, IB) = 1 then X_1 and X_2 are separated.

We call IB' ⊆ IB the set of antagonistic attributes/indicators and abbreviate it by AIB(X_1, X_2) if $\forall$ (x,y) ∈ X_1*X_2 with Sep(X_1, X_2, IB) = 1 we find: $x \|_{IB'} y$. For more details, see [16].

When |AIB| = 2 then we can visualize the order relations in a 2D scatter plot. Several situations may appear, Figure 3 shows some examples:

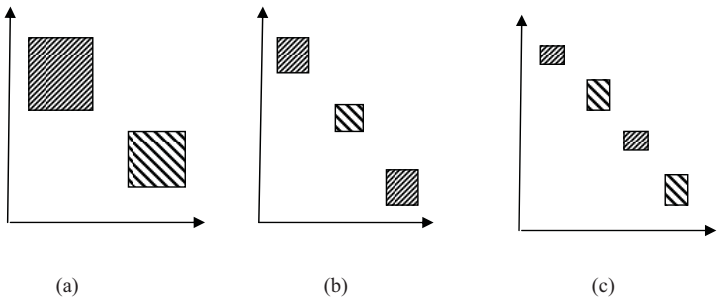

Figure 3: The two coordinates are the attributes of AIB, with |AIB| = 2. The different patterns show to which area the objects belonging to the two separated subsets belong. (a): example of topologically connected, (b) and (c): examples of topologically non-connected.

Observation 1: In a space, spanned by the attributes of AIB the objects of X_1 and X_2, with $Sep(X_1, X_2, AIB) = 1$ do not necessarily belong to topological connected subsets of IR^2. (two dimensional space of real numbers).

In case of |AIB| >2 one can find separated subsets and a minimum set of coordinates. However a) the graphical display is difficult and b) the possibility of non-topological connected subsets of $IR^{|AIB|}$ will also be hard to be identified.

4.2 Application of FCA

Basically the problem can be formulated as follows: Which values of which attribute lead to certain positions of the objects within a Hasse diagram. Formal concept analysis with its "symmetric view" on objects and attributes could be applied. The problem is that especially in multivalued contexts in general we obtain a large number of concepts, so that it is wise to look for alternatives. Here however with a pretty small set of objects we can consider the lattice of concepts, (see Figure 4).

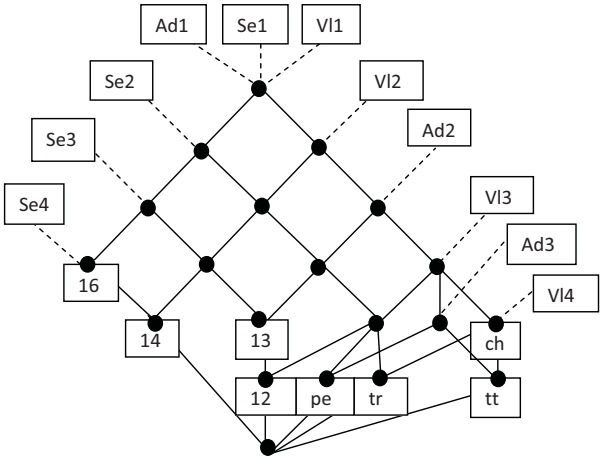

Figure 4: Lattice of the eight chemicals and three attributes, transformed by an ordinal scaling. Explanation of abbreviations are found in Table 1.

Why for example are {16,14} and {12, 13} separated? By inspection of Figure 5 we see that common for both sets are Se3, Vl1, and Ad1 (Ad<i> means for example: advection with the value i). Furthermore one can see that Se4 is exclusively a property common for {16, 14} and the maximal value with respect to Advection Ad is Ad1. In contrast: {12, 13} have the property Se3 (are not in that extent underlying a sedimentation process), but Ad2. Hence the

separation of the two subsets is explained by Se4, Ad1 of {16, 14} but Se3, Ad2 of {12, 13}. We further see that two attributes are sufficient to explain the separation of the two sets. We also see that we can find in IR^2 spanned by advection and sedimentation two open sets, covering {16, 14} on the one side and covering {12, 13} on the other side. {tr, pe} and {tt} is another example of separated subsets. Se2 and Vl3 are common for {tr, pe} but for {tt} we as antagonistic indicator at first hand only Se1. Ad3 is common for pe and tt, cannot separate the two sets, vl4 is common for tr and tt. Only if we accept beside attribute Se also the attributes Vl and Ad we can characterize the separation: Se2, Vl3: {16, 14}, {tt} is characterized by Se1, Vl4 and Ad3. Without Ad or without Vl the separation cannot be explained.

4.3 Multi-coordinate representation

How the attributes and their values explain the separation of {tr, pe} and {tt} can also be visualized by interpretation of the product order as an order of intervals, displayed by directed lines, representing each attribute (Figure 5):

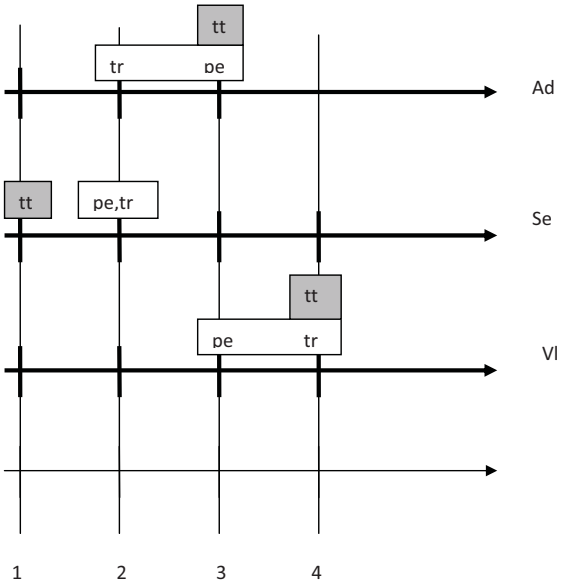

Figure 5: Display of separated subsets as intervals in a multi-coordinate representation. Grey hatched the position of tt, to indicate that tt belongs to another separated subset (see Table 1).

By Ad and Se tt and tr become incomparable, by V_1 and Se tt and pe become incomparable. The determining fact is that we observe the elements of the one set once to the left and once to the right side of the other set along the coordinates.

Observation 2: To identify separated subsets in a multi-coordinate representation we do not need the single numerical values of the objects, but only whether the ranges of attributes representing subsets of the objects are in that way mutually oriented that one family of subsets constitutes the one, and another family of subsets constitutes the other separated subset.

4.4 Tripartite graph

4.4.1 Introduction

Motivated by observation 2 we define a tripartite graph as follows:

$V_2 = \{(x_1,x_2): \ x_1 \in X_1 \ , x_2 \in X_2, \ Sep(X_1,X_2,IB)=1\}$

and two copies of vertices formed by the attribute's names, namely:

$V_1 = \{q_i \in IB \}$

$V_3 = \{q_i \in IB \}$

In the drawing plane we arrange V_2 in the middle, V_1 at the left and V_3 at the right side.

We connect vertices of V_1 with a vertex $\in V_2$ if and only if $x_1 >_{qi} x_2$, $(x_1, x_2) \in V_2$

and connect V_3 with a vertex of V_2 if and only if : $x_1 <_{qi} x_2$, $(x_1,x_2) \in V_2$. Figure 6 shows the tripartite graph in case of the separated subsets {12,13},{14, 16}:

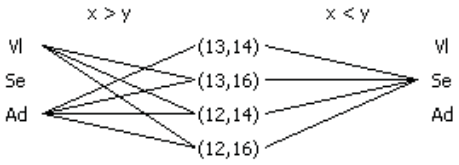

Figure 6: Example of a tripartite graph, constructed from the separated subsets {12,13} and {14, 16} (PyHasse software). Vl: Volatilization, Se: Sedimentation, Ad: Advection.

For convenience in the PyHasse outcomes the lines indicating a >-relation are coloured red, whereas those of the <-relation blue software: PyHasse).

We see for example the pair of incident edges (Ad, (13,14)), ((13,14),Se) : It describes that 13 $>_{Ad}$ 14 but 13 $<_{Se}$ 14, as it must be, as 13||14.

An obvious observation: (*observation 3*) Incident edges ($q_i \in V_1$, (x, y)), ((x, y), $q_i \in V_3$) cannot appear, because such appearance would imply: $x >_{qi} y$ and at the same time $x <_{qi} y$, which is impossible.

As Ad $\in V_1$ one the one side and Se $\in V_3$ on the other side are connected with all pairs $\in X_1^* X_2$, we conclude that two coordinates namely Ad and Se are sufficient to explain the separation of the subsets {12, 13} and {14, 16}. Following observation 3 Se $\in V_1$ cannot have any connection to one of the pairs (x, y) $\in X_1^* X_2$, because Se $\in V_3$ is connected with all pairs. Similarly Ad $\in V_3$ cannot be connected with pairs $\in X_1^* X_2$ because Ad $\in V_1$ is with all pairs connected.

Observation 4: Reconstruction of the partial order: Let X_1 , X_2 be two disjoint object subsets, which are not necessarily separated. Then there may be pairs (x, y) $\in X_1^* X_2$, with x < y. In the tripartite graph such a pair would have exclusively incident edges either with V_1 (x > y) or with V_3 (x < y). Hence it is possible, also to analyze subsets X_1, X_2, whose Sep(X_1, X_2, IB) <1.

In the following we first show, how the separated subsets {tr, pe} and {tt} can be analyzed within the framework of tripartite graphs, then we will extend the analysis to "almost-separated subsets".

In Figure 7 the example {tr, pe}, {tt} is revisited:

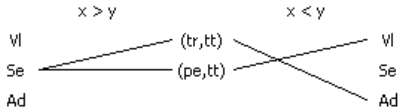

Figure 7: Separated subsets {tr, pe} and {tt} and their tripartite graph (PyHasse software).

From Figure 7 we learn:

1. With respect to Se all pairs obey the relation (x>y).

2. Two coordinates are not sufficient. For a separation Sep(X1,X2,IB') =1 we need {See on the one side and both: Vl and Ad on the other side.

In the last example we demonstrate how the tripartite graph looks when we select the disjoint but not separated subsets {pe, tr} and {tt, ch} (Figure 8).

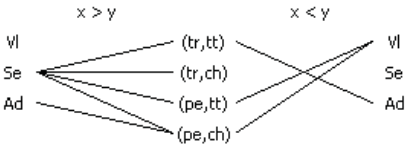

Figure 8: Tripartite graph of the disjoint subsets {pe, tr}, {tt,ch} (PyHasse software).

We see that (tr,ch) has no connection to the vertices of V_3. Hence we have tr > $_{Se}$ ch. As there are no further incident edges, tr and ch must have equal values with respect to Vl and Ad., which is true as an inspection of Table 1 verifies.

Observation 5: A Hasse diagram shows in a compact manner comparabilities. In the case of incomparabilities, however, the Hasse diagram is of only limited use: One cannot see in which way the attributes are leading to incomparabilities. A tripartite graph renders this information: Looking for example to (tr, tt): We see that (tr, tt) has edges incident to V_1 and to V_3, therefore it is valid: tr || tt. We see furthermore that tr > tt with respect to Se, but tt > tr with respect to Ad.

In the last example we study the separated subsets {12, 13, 14, 16} and {tt, tr, ch, pe} which also obey the chemical distinction between Alkanes/Alkenes and the Polychlorinated biphenyls (Figure 9).

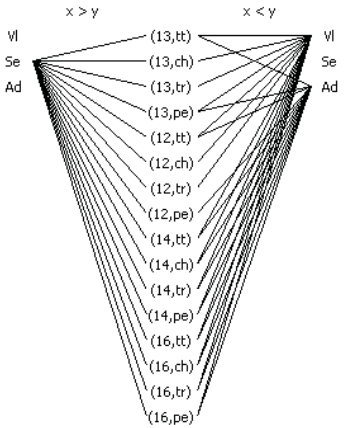

Figure 9: Tripartite graph of {12, 13, 14, 16} and {tt, tr, ch, pe} (PyHasse software).

One can see that (i) Se is causing that the PCBs are all greater than the alkanes/alkenes, therefore Se ∈ V_3 is not connected to any pair ∈ X_1* X_2 and (ii) not all pairs are at the same time connected with two vertices of V_3, hence AIB must contain all three attributes, i.e. AIB = IB.

4.4.2 Individual analyses

Two practical questions appear:

1) Considering the graph in Figure 10 it is hard to decide, which descriptor V_1 or Ad is more contributing to the fact that Alkanes/Alkenes have larger scores than the PCBs.

2) Although the graph in Figure 10 is simple one may have more complex situations, therefore we need tools to “zoom in”.
We answer the second question first:
a) We can select one pair (x, y) (here (13, pe) and see their relations to V_1 and V_3 , b) we can select one attribute q_i (here Vl, volatilization) and show the subgraph induced by the vertex $q_i \in V_1$ and $q_i \in V_3$,

c) we can select either one element x from the pair $(x,y) \in V_2$ to see all its relations (x,.) to V_1 and V_3 or d) y in order to see all relations (.,y) to V_1, V_3. In Figure 10 all four variants are shown:

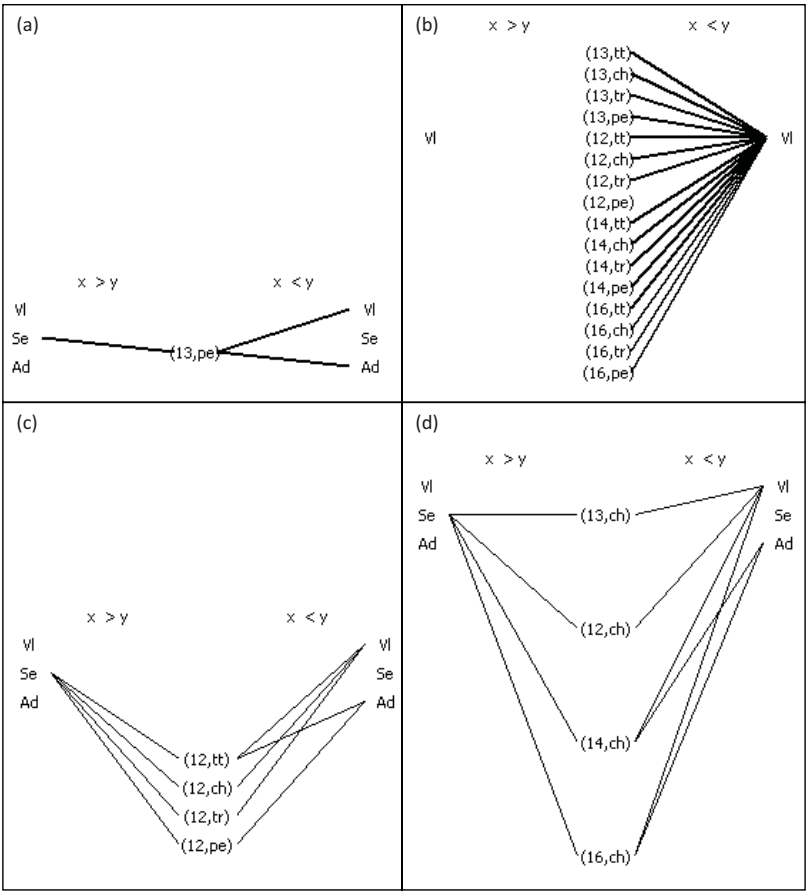

Figure 10: Individual analysis of the tripartite graph (PyHasse software)

By means of Figure 10 b) we can count how often each attribute is contributing to >, =, or < -relation. However, and this answer for the first question we can do this job simultaneously for all attributes (Figure 11).

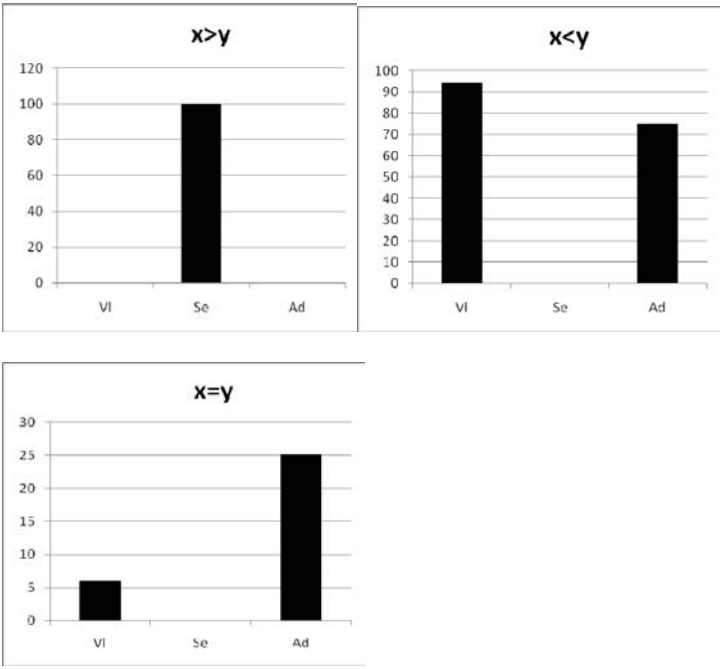

Figure 11: Contribution of attributes to the >, < and =-relation between the objects of the two separated subsets {12, 13, 14, 16} and {tt, tr, ch, pe} (PyHasse software).

4.4.3 Mixing analysis

We started with the problem as to how far separated subsets are forming topological connected sets in IR^m. Inspecting Figure 4 we see that topological non-connected subsets should have the property that one attribute is once responsible for a >- relation and once for a <-relation. We formalize this observation as follows: Let X_{1i} subsets of X_1 and X_{2i} subsets of X_2 with $Sep(X_1, X_2, IB)= 1$. Furthermore we write $q_i(X_{1i}) > q_i((X_{2j})$ to indicate that for all $x \in X_{1i}$ and for y X_{2j} $q_i(x) > q_i(y)$. A situation like that in Figure 3 (b) appears if we find:

$q_i(X_{rj1}) > q_i(X_{si}) > q_i(X_{rj2})$ r=1 or 2 then s=2 or 1.

The attribute q_i must therefore appear once as a connected vertex of V_1 and once as a connected vertex of V_3. We calculate the mixing degree as follows:

Let $val(q_i(V_k))$ (k=1 and 3) the valence of a selected vertex $q_i \in V_k$. We count how often q_i of V_1 and at the same time q_i of V_3 has a connection with $(x, y) \in X_1 * X_2$. Hence mixing-degree, mix, is proportional to $val(q_i(V_1)) * val(q_j(V_3))$, $j \neq i$. The product is unequal zero if both factors at the same time have values > 0. We normalize by: $floor(m^2/4)$ which is the maximum value accessible for any $val(q_i(V_1)) * val(q_j(V_3))$. In the example case there is no mixing degree $\neq 0$. So we can imagine that for the alkanes/alkenes vs PCBs a representation must be possible with X_1 and X_2 as topological connected subsets. To show how a situation where topological non-connected subsets appear (unfortunately the chemical examples does not show this feature), we examine a fictitious data matrix (Table 2).

Table 2: Fictitious data matrix for demonstration of topological non-connected separated subsets.

	q_1	q_2	q_3
a	1.0	1.0	1.0
b	3.0	1.0	1.0
c	4.0	4.0	4.0
d	2.0	3.0	2.0
e	2.0	4.0	1.0
f	4.0	5.0	6.0
g	6.0	5.0	4.0
h	1.0	6.0	2.0

In Figure 12 the Hasse diagram and the tripartite graph (separated subsets: X_1= {b, d} and X_2={h, e}) are shown:

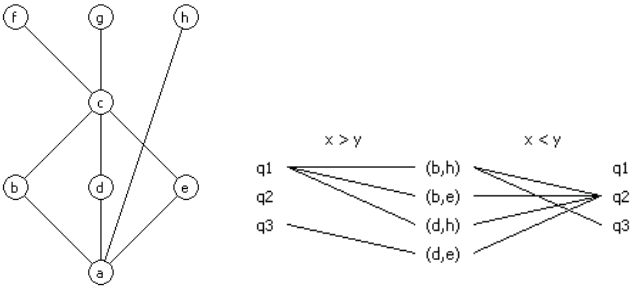

Figure 12: Hasse diagram of the data of Table 2 and the tripartite graph (see text)

In Figure 13, we see that q_3 is as a connected vertex in V_1 as well as in V_3. Therefore the mixing degree will be $\neq 0$ for q_3 (Table 3)

Table 3: Mixing degree for the fictitious example (PyHasse software):

	q_1	q_2	q_3
mix%	0	0	0.5

Indeed for a subset of X_1 , namely {b} we find that it is less a subsets of X_2, namely {h, but another subset of X_1, namely {d} is larger than {e} $\subset X_2$. Schematically written: b $<_{q3}$ {h, e} $<_{q3}$ d, i.e. set X1 is not topologically connected with respect to coordinate q3 or "q_3 has some mixing degree $\neq$0. character".

4.4.4 Reduction of the tripartite graph

We construct an adjacency matrix, T, of the tripartite graph as follows:

There are m*(m-1) rows of all pairs (q_i, q_j) i $\neq$ j. the set A, and $|X_1|*|X_2|$ columns, corresponding to the pairs (x, y) $\in X_1*X_2$. An entry T_{rs} at the crossing of the r_{th} row (q_i, q_j) and s_{th} column (x, y) is unequal zero if it is valid:

$x >_{qi} y$ and $x <_{qj} y$.

As in Formal Concept Analysis a mapping g is defined: (q_i, q_j) $\in$ A $\rightarrow$ (x,y) if the corresponding entry of the matrix T has a value unequal 0. Obviously those attribute pairs are sufficient to explain a separation of X_1, X_2 for which is valid: $\cup$ g(q_i, q_j) $\supseteq X_1*X_2$. If T is considered as the usual adjacency matrix the entries are 0 or 1.

A simple meta-algorithm for a reduction is now:

Start: set B = ø , the set of pairs to be found

(1) Find the row with the maximum of the row sums.

(2) Mark the corresponding row with a "*"

(3) Set B $\leftarrow$ g(q_i, q_j) of row *.

(4) Set all entries in rows other than the *-row = 0, where the entry in the row * unequal zero.

(4) Repeat the process by starting with (1) until B = X_1*X_2.

The crucial point in this meta-algorithm is that with an adjacency matrix with 0, and 1 as entries we can find in step (1) several rows, having the same maximal value of their row sum.

Therefore the entry in T is not 0 and 1 but 0 and $|q_i(x)-q_j(y)|$. The maximum row sum corresponds then to that pair of q_i and q_j with the largest discrimination between x and y. By this modification the chance to get the same row sum is somewhat reduced. When there are still several rows with the same maximal row sum, then the first row is taken.

By this procedure we obtain a reduced tripartite graph by appropriate translation of the remaining pairs of A together with the pairs of X_1*X_2 into a tripartite graph as explained in 4.3.1. For the separation of the alkanes/alkenes versus PCBs we obtain the reduced tripartite graph, shown in Figure 13:

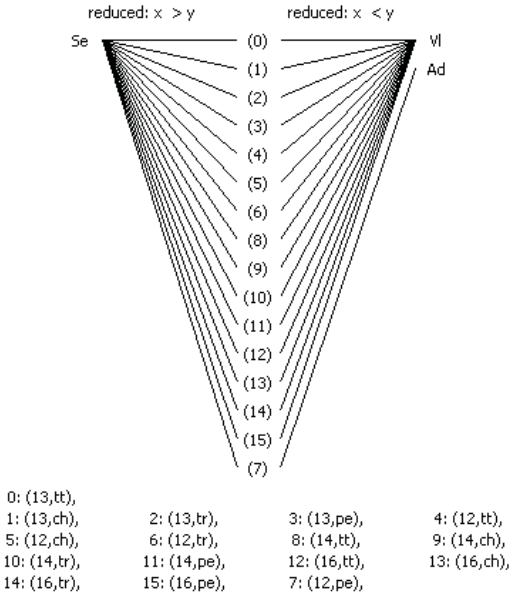

Figure 13: Reduced tripartite graph.

Fifteen pairs (out of 16) $\in X_1*X_2$ have the same constellation: namely Se as attribute responsible for $x \in X_1 > y \in X_2$. and V_1 for $x \in X_1 < y \in X_2$. Only because of one pair: pair no 7 we need the third property Ad to obtain the needed separation.

5. Discussion and Conclusion

We started from two questions (problems):

1) If we find separated subsets, what can be said about the attribute values, causing the separation and

2) Can we give a characterization of the two subsets in terms of their topological connectedness?

By means of the tripartite graph and eventually with the possible individual analyses we obtain an insight about the reasons why two object subsets are separated as an outcome of the partial order analysis.

With the mixing degree we can quantify how attributes are causing a "in-between" situation of subsets of the separated subsets. Therefore we consider tripartite graphs as a useful supplement of partial order analysis.

It should be clear that the simple fact of $|AIB|= 2$ does not necessarily mean that X_1 and X_2 have a simple, namely topological connected structure.

There are still open points:

a) Is the mixing degree sufficient enough to deduce statements about the topological character of the two separated subsets?

b) Instead of a tripartite graph the more common bipartite graph with the set A (see section before) together with set $X_1 * X_2$ would also do the job of allowing an insight how attributes are causing the separation of the subsets X_1 and X_2. However in the case of a bipartite graph the user has to be aware which role the first component and which role the second component in the pair (q_i,q_j) plays.

c) The reduction algorithm is not unique. Even if the entries are numbers unequal zero and deviating from 1 one may find several maximal rows and to select just the first row makes the procedure dependent on the enumeration of the attributes by the user.

d) Often separated subsets are found from the context as in the example studied here. Can we find directly from any posets separated subsets (without an a priori contextual background)? In the software PyHasse simple search routines are available. A systematic approach seems to

be possible by applying Formal Concept Analysis [4]. However, other algorithms without applying Formal Concept Analysis (with its need of scaling) are necessary.

e) Even by reduction, the tripartite graph may be large and complex. It is planned in a further upgrade of PyHasse to define equivalence classes among the pairs $\in X_1 * X_2$, where the equivalence relation is called: incidences with the same vertex sets. We indicated this kind of reasoning in 3.4.4.

Summarizing: We showed that it is feasible to interpret the vertical component of Hasse diagrams, which are more or less the chains. Now, with the tripartite graph also a way is found to interpret the horizontal dimension of a Hasse diagram, namely the set of antichains.

References

[1] P. Annoni, R. Bruggemann, The dualistic approach of FCA: A further insight into Ontario Lake sediments, *Chemosphere* **70** (2008) 2025–2031.

[2] P. Annoni, R. Bruggemann, Exploring partial order of European countries, *Soc. Indicators Res.* **92** (2009) 471–487.

[3] H. G. Bartel, Formalbegriffsanalytische Untersuchung ausgewählter Aromatizitätskriterien, *MATCH Commun. Math. Comput. Chem* **30** (1994) 9–35.

[4] H. G. Bartel, J. Mucha, Personal communication, publication to be submitted to Match in preparation, 2010.

[5] R. Bruggemann, S. Trapp, Release and fate modelling of highly volatile solvents in the river Main, *Chemosphere* **17** (1988) 2029–2041.

[6] R. Bruggemann, S. Trapp, M. Matthies, Behavior assessment for a volatile chemical in the middle and lower German part of the Rhine river, *Envir. Tox. Chem.* **10** (1991) 1097–1103.

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

[8] R. Bruggemann, J. Schwaiger, R. D. Negele, Applying Hasse diagram technique for the evaluation of toxicological fish tests, *Chemosphere* **30** (1995) 1767–1780.

[9] R. Bruggemann, E. Halfon, G. Welzl, K. Voigt, C. Steinberg, Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests*, J. Chem. Inf. Comp. Sci.* **41** (2001) 918–925.

[10] R. Bruggemann, U. Drescher-Kaden, *Einführung in die modellgestützte Bewertung von Umweltchemikalien - Datenabschätzung, Ausbreitung, Verhalten, Wirkung und Bewertung*, Springer–Verlag, Berlin, 2003.

[11] R. Bruggemann, U. Simon, S. Mey, Estimation of averaged ranks by extended local partial order models, *MATCH Commun. Math. Comput. Chem.* **54** (2005) 489–518.

[12] R. Bruggemann, G. Restrepo, K. Voigt, Structure – fate relationships of organic chemicals derived from the software packages E4CHEM and WHASSE, *J. Chem. Inf. Model.* **46** (2006) 894–902.

[13] R. Bruggemann, K. Voigt, Basic principles of Hasse diagram technique in chemistry, *Comb. Chem. & High Throughput Screen.* **11** (2008) 756–769.

[14] R. Bruggemann, K. Voigt, G. Restrepo, U. Simon, The concept of stability fields and hot spots in ranking of environmental chemicals, *J. Environ. Model. & Soft.* **23** (2008) 1000–1012.

[15] R. Bruggemann, K. Voigt, Analysis of partial orders in environmental systems applying the new software PyHasse, in: J. Wittmann, M. Flechsig (Eds), *Simulation in Umwelt- und Geowissenschaften- Workshop Potsdam 2009*, Shaker–Verlag, Aachen, 2009, pp. 43–55.

[16] R. Bruggemann, G. P. Patil, Multicriteria prioritization and partial order in environmental sciences, *Environ. Ecol. Stat.*, in press.

[17] R. Bruggemann, L. Carlsen, An improved estimation of averaged ranks of partial orders, *MATCH Commun. Math. Comput. Chem.* **65** (2011) 383–414.

[18] L. Carlsen, B. N. Kenessov, S. B. Batyrbekova, A QSAR/QSTR study on the human health impact of the rocket fuel 1,1-dimethylhydrazine and its transformation products. Multicriteria hazard ranking based on partial order methodologies, *Environ. Tox. Pharm.* **27** (2009) 415–423.

[19] L. Carlsen, Partial order ranking of organophosphates with special emphasis on nerve agents, *MATCH Commun. Math. Comput. Chem.* **54** (2005) 519–534.

[20] L. Carlsen, R. Bruggemann, Partial order ranking as a tool in environmental impact assessment. PAH and PCB pollution of the river Main as an illustrative example, in: G. T. Halley, Y. T. Fridian (Eds.), *Environmental Impact Assessment,* Nova Science Publishers, 2009, pp. 335–354.

[21] L. Carlsen, The interplay between QSAR/QSPR studies and partial order ranking and formal concept analyses, *Int. J. Mol. Sci.* **10** (2009) 1628–1657.

[22] C. Carpineto, G. Romano, *Concept Data Analysis*, Wiley, Chichester, 2004.

[23] B. A. Davey, H. A. Priestley, *Introduction to Lattices and Order,* Cambridge Univ. Press, Cambridge, 1990.

[24] B. Ganter, R. Wille, *Formale Begriffsanalyse Mathematische Grundlagen*. Springer–Verlag, Berlin, 1996.

[25] B. G. Hansen, A. G. Van Haelst, K. Van Leeuwen, P. Van der Zandt, Priority setting for existing chemicals: European union risk ranking method, *Environ. Tox. Chem.* **18** (1999) 772–779.

[26] S. Kardaetz, T. Strube, R. Bruggemann, G. Nützmann, Ecological scenarios analyzed and evaluated by a shallow lake model, *J. Environ. Manag.* **88** (2008) 120–135.

[27] D. J. Klein, J. Brickmann, Partial orderings in chemistry, *MATCH Commun. Math. Comput. Chem.* **42** (2000) 1–290.

[28] J. Neggers, H. S. Kim, *Basic Posets*, World Sci. Publish., Singapore, 1998.

[29] A. Newman, Ranking pesticides by environmental impact, *Environ. Sci. Technol.* **29** (1995) 324A–326A.

[30] G. P. Patil, C. Taillie, Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization, *Environ. Ecol. Stat.* **11** (2004) 199–228.

[31] S. Pudenz, ProRank – software for partial order ranking, *MATCH Commun. Math. Comput. Chem.* **54** (2005) 611–622.

[32] G. Restrepo, R. Bruggemann, M. Weckert, S. Gerstmann, H. Frank, Ranking patterns, an application to refrigerants, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 555–584.

[33] U. Simon, R. Bruggemann, S. Mey, S. Pudenz, METEOR – application of a decision support tool based on discrete mathematics, *MATCH Commun. Math. Comput. Chem.* **54** (2005) 623–642.

[34] P. B. Sørensen, R. Bruggemann, L. Carlsen, B. B. Mogensen, J. Kreuger, S. Pudenz, Analysis of monitoring data of pesticide residues in surface waters using partial order ranking theory, *Envir. Tox. Chem.* **22** (2003) 661–670.

[35] W. T. Trotter, *Combinatorics and Partially Ordered Sets Dimension Theory*, The Johns Hopkins Univ. Press, Baltimore, 1992.

[36] S. Tsakovski, V. Simeonov, Hasse diagrams as explanatory tool in environmental data mining: A case study, in: J. Owsinski, R. Bruggemann (Eds.) *Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues,* Sys. Res. Inst. Polish Acad. Sci., Warsaw, 2007, pp. 50–68.

[37] K. Voigt, G. Welzl, R. Bruggemann, Data analysis of environmental air pollutant monitoring systems in Europe, *Environmetrics* **15** (2004) 577–596.

[38] K. Voigt, R. Bruggemann, Ranking of pharmaceuticals detected in the environment: Aggregation and weighting procedures, *Combin. Chem. High Through. Screen.* **11** (2008) 770–782.

[39] K. Voigt, R. Bruggemann, M. Kirchner, K.-W. Schramm, Influence of altitude concerning the contamination of humus soils in the German Alps: a data evaluation approach using PyHasse, *Environ. Sci. Pollut. Res.* **17** (2010) 429–440.

[40] K. Voigt, R. Bruggemann, H. Scherb, H. Shen, K. H. Schramm, Evaluating the relationship between chemical exposure and cryptorchidism by discrete mathematical method using PyHasse software, *J. Environ. Model. Soft.* **25** (2010) 1801–1812.

[41] F. A. M Verdonck, G. Boeije, V. Vandenberghe, M. Comber, W. de Wolf, T. C. J. Feijtel, M. Holt, V. Koch, A. Lecloux, A. Siebel-Sauer, P. A. Vanrolleghem, A rule-based screening environmental risk assessment tool derived from EUSES, *Chemosphere* **58** (2005) 1169–1176.

[42] T. G Vermeire, P. T. J. van der Zandt, H. Roelfzema, C. J. van Leeuwen, Uniform system for the evaluation of substances I principles and structure, *Chemosphere* **29** (1994) 23–37.

[43] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, in: I. Rival (Ed.), *Ordered Sets*, D. Reidel Publishing, Dordrecht, 1982, pp. 445–470.