Chapter 2
Multivariate Datasets for Inference of Order: Some Considerations and Explorations

Ganapati P. Patil, Wayne L. Myers, and Rainer Brüggemann

Abstract Ideal formulation of a multi-indicator system (MIS) would be to define, design, and acquire the entire construct with complete consensus among all concerned. However, such would be an extreme rarity in actuality. Experts have differing views. Factors may not express monotonically, as when either extreme is unfavorable. The entirety cannot be assessed and must be sampled. Empirical experience to validate expectations is inadequate. Consequently, exploratory examination of any available datasets collected for collateral purposes can augment insights relative to suitable surrogates for ideal indicators, with particular attention to ordering relations for subsets of quantifiers and ensembles of entities (objects, cases, instances, etc.).

Multivariate datasets are comprised of several quantifiers (variates or variables) as columns recorded for multiple entities as rows. The data matrix thus realized is not necessarily directly useful nor fully informative for analytically inferring order among entities. In this chapter, some approaches are discussed which may be helpful in extracting insights on ordering properties that are embodied in multivariate

Expanded from the keynote lecture by, Patil (2012), at the International Workshop on Partial Order Theory and Modeling held in Berlin, Germany during September 27–29, 2012. This research and outreach effort have been partially supported by US NSF Project 0307010 for Digital Governance and Hotspot Geo-Informatics of Detection and Prioritization.

G.P. Patil (✉)
Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA
e-mail: gpp@stat.psu.edu

W.L. Myers
Penn State Institutes of Energy and Environment, The Pennsylvania State University, University Park, PA 16802, USA

R. Brüggemann
Department of Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, Berlin, Germany

R. Brüggemann et al. (eds.), *Multi-indicator Systems and Modelling in Partial Order*,
DOI 10.1007/978-1-4614-8223-9_2, © Springer Science+Business Media New York 2014

datasets and applicable in configuring suites of indicators. These procedures may be particularly helpful in finding suitable surrogates and applying partial order theory when expediency is essential. We consider orientation, crispness of data, and culling of candidates according to importance in respect of some desirable criteria.

2.1 Introduction

The ideal is often not attainable in formulating multiple indicator systems (MIS), and ad hoc adoption of multivariate datasets as suites of surrogate indicators is not directly defensible when contested. The proponents of an MIS have an obligation to analyze the comparative capacity for objective ordering contained in the constituents of a data matrix before advocating adoption as a suite of surrogates for addressing more abstract aspects. Exploratory examination of any available datasets collected for collateral purposes can augment insights relative to suitable surrogates for ideal indicators, with particular attention to ordering relations for subsets of quantifiers and ensembles of entities.

A multivariate data matrix consists of several quantifiers (variates or variables) as columns recorded on multiple entities (objects, cases, instances, etc.) as rows. The given form of data matrix is not necessarily directly useful nor fully informative for analytically inferring order among entities. In this chapter, we consider some approaches which may be helpful in extracting insights on ordering properties that are embodied in multivariate datasets and applicable in configuring suites of indicators.

Commonality of orientation is crucial, and we begin by exploring implications of alternate orientations. Incomparability is adverse with regard to ordering and provides an initial basis for such assessment. We seek succinctness and consider culling candidates regarding redundancy. When suppressing a candidate who has little impact on comparability, it can be considered relatively redundant. When relatively redundant candidates have been suppressed, interest shifts to the incomparability induced by each of the remaining indicators. An indicator having substantial specificity with regard to comparability is speaking to possible distinctive ensembles of entities (objects) that can be further investigated otherwise, such as through clustering. An indicator that induces considerable general incomparability and isolation suggests a distinctive separate dimension of comparison.

Scaling and crispness of data are concerns, with some aspects being considered and/or illustrated. Compositing and weighting of indicators are also discussed. Empirical versus inferential aspects and approximation are of interest as constrained by computational complexity and practicality of graphical depiction. The Hasse Diagram Technique (HDT) has constraints of practicality due to rapid loss of interpretability with increasing number of object entities. Some inferential aspects can be introduced by repeated sampling and compiling frequencies of incomparability by sample size along with constructing representative Hasse diagrams for different

sample sizes. Approximations and diagrammatic representations can be obtained from local partial order models and posetic features of up-sets and down-sets. When the entities to be ordered are samples from a larger universe, questions arise as to the global meaning of orderings within samples—particularly so for a single sample or time series of samples.

In overview, the task of determining a sequence, i.e., a linear or weak order or meta-sequence (sequence of sequences, as for example a poset) usually begins with definition of the entities (objects) of interest, i.e., in defining the **object set**. Then the purpose and nature of the intended sequencing is to be defined. Often the purpose has no counterpart in a measurable quantity, so surrogates are formulated to provide a **suite of indicators** as a proxy. If this is done in a knowledge-based manner, we call it here the (preferred) a priori approach. Each of several such indicators should reflect as least a difference of nuance so that redundancy is avoided. The alternate a posteriori approach of primary focus here is to exploit existing multivariate dataset(s) collected for some parallel purpose such as monitoring, diagnosis, or investigation. Questions then arise as to which (column) variables in the dataset can be considered as indicative for the current purpose and in what manner.

After having defined the object set and selected a suite of candidate indicators, additional aspects must be addressed before sequencing is useful:

- Is the data representation appropriate, i.e., how does one handle numerical differences? Is there any slight numerical difference indicative of order within a pair of objects?
- How should one consider candidate indicators of different scaling levels? Whereas in natural sciences metric information is often available; in other sciences the information is ordinal as numbers are assigned to linguistic descriptors. Should all indicators then be expressed in ranks as the "lowest common denominator" of scaling strength?
- Can indicators be culled (suppressed, set aside, eliminated) on the basis of relative redundancy as reflected in marginal impact on comparability? Do certain candidates speak primarily to small ensembles of specific objects, thus being optional for "fine tuning?"

These questions are among those considered in this chapter with the intent of suggesting methodological approaches when the context of sequencing does not give thorough guidance.

The chapter is organized as follows:

1. Matters of orientation and reducing redundancy with the theory of partially ordered sets
2. Formulation of matrices for retrospective indicator systems
3. Analytical aspects of structure and structuring (since many chapters in this book are concerned with these aspects, this is only in overview)
4. Conflict analysis (this is also in overview as a subject of several chapters)

2.2 Orientation and Partitioning

2.2.1 Theoretical Framework of Partially Ordered Sets (Posets)

Much of what follows is framed in the theory of partially ordered sets. There are many references available, as for example, Brüggemann and Patil (2011) or Brüggemann and Voigt (2008).

2.2.1.1 Axioms of Partial Orders

Suppose that objects of interest comprise an "object set" X and further that X is a finite set. For five objects a, b, c, d, and e:

$$X = \{a,b,c,d,e\}.$$

Objects of the set X are to be compared and arranged (meta)-sequentially. The symbol $\leq$ is used as a binary relation among the objects. The role of this relation is specified by axioms:

Axiom 1 :	Reflexivity :	$x \in X : x \leq x$	(2.1)
Axiom 2 :	Anti-symmetry :	$x \leq y, y \leq x$ implies $y = x$	(2.2)
Axiom 3 :	Trannsitivity :	$x \leq y$ *and* $y \leq z$ implies $x \leq z$	(2.3)

Reflexivity: An object can be compared with itself.

Antisymmetry: If both comparisons are valid, i.e., $y \leq x$ and at the same time, $x \leq y$, then this axiom demands that x is identical with y.

Transitivity: Transitivity is present if the objects are characterized by properties which are at least ordinal scaled.

These axioms give the set X an algebraic structure, namely that of a partially ordered set (poset). For comparison purposes, the selection of $\leq$ must be considered as a narrowing. A more general theory of comparisons is the tournament theory (see for instance Bartel and Mucha 2014).

2.2.1.2 Quotient and Object Sets

Several objects may have the same numerical values as different individuals (ties). In such case the objects are considered as equivalent, expressing that they have identical rows in the data matrix, but are nevertheless distinct. These objects form an equivalence class, and one object from the equivalence class is often selected to represent all the others.

To make a clear distinction:

- The set of equivalence classes under an equivalence relation R is called quotient set, denoted by *X*/R.
- From any equivalence class one object may be selected as representative of the others for purposes of comparison.
- The object set retains identical objects as individual elements.
- It is common practice to consider the data matrix, consisting of n rows/objects and m columns/indicators.

2.2.1.3 Comparative Structure Induced by Hasse-Diagram Technique

Let x, y be two different objects of the object set X, with $q(x)$ being the data row for x and $q(y)$ for y. Then

$$\begin{aligned} &x \le y \text{ if and only if } q(x) \le q(y), \\ &q(x) \le q(y) \text{ if and only if } q(x) \le q(y) \text{ for all i.} \end{aligned} \tag{2.4}$$

If x, y are different objects but $q(x) = q(y)$, i.e., $q_i(x) = q_i(y)$ for all i, then the objects x and y are equivalent, denoted as: $x \cong y$

When some $q_i(x) < q_i(y)$ but some others $q_i(x) > q_i(y)$, then x and y are incomparable, denoted as: $x \parallel y$.

When mutual incomparability appears for all objects of a subset $X' \subseteq X$, then X' is an "antichain."

When for the objects x, y, it is valid that $q(x) \le q(y)$ or $q(x) \ge q(y)$, then x and y are comparable, denoted as: $x \perp y$.

$$An\ object\ x\ generates\ a\ down-set\ as\ x \in X : O(x) := \{y \in Y : y \le x\} \tag{2.5}$$

$$An\ object\ x\ generates\ an\ up-set\ as\ x \in X : F(x) := \{y \in Y : y \ge x\} \tag{2.6}$$

$$\begin{aligned} An\ object\ x\ may\ generate\ &a\ set\ of\ x-incomparables\ as : x \in X : U(x) : \\ &= \{y \in Y : y \parallel x\} \end{aligned} \tag{2.7}$$

We call $q_i(.)$ an attribute or indicator interchangeably.

2.2.1.4 Difficulties in Dealing Directly with Hasse Diagrams and Underlying Issues

Equation (2.4) can be applied to every dataset insofar as there are no data gaps, giving rise to an associated Hasse diagram representing the order relations among objects. There are some issues of interpretability for Hasse diagrams, however, and as to how far application of (2.4) is practically informative.

The difficulty that most often becomes apparent is that the graphical representation of a poset by a Hasse diagram increases very rapidly in its perceptual complexity with increasing number of objects, becoming a blur of intersecting lines that is not visually informative. The more complex the comparative structure, the more rapid the degradation of interpretability for a Hasse diagram. The diagram also does not reveal comparative substructure that may reside in column-wise subsets of the indicative data matrix, which is an underlying issue for the Hasse Diagram Technique (HDT).

Isolation of an object is a foundational feature of HDT that stymies further direct interpretation of that object relative to other objects since it lacks comparability to any other (see for a heuristic Al-Sharrah 2014). The extreme degenerate case is with every object as an isolate whereby there is complete lack of both comparability and sequencing. Since Eq. (2.4) encompasses the entire set of prospective indicators, it then offers no insights regarding comparative substructure that may reside within the candidate data. The difficulties with Hasse diagrams arise from the presumption that such diagrams directly provide in every case a foundational representation of the comparative content of an informational context. There is need for supplemental processing and/or alternative depictions that effectively allow generalized views of comparative content and substructure for the candidate data.

Partial orders can be visualized in coordinate systems such as scatter plots, with the basic idea being to extract from the posets coordinate values for each object containing comparative information of interest. Such information can often be obtained from frequencies of features in the Hasse diagram. For example the difference between number of objects in a down-set and number of objects in the up-set may be one coordinate, whereas the other coordinate counts $U(x)$. One may even differentiate among the elements of $U(x)$, according to whether they are isolated elements or not. For the concept of posetic coordinates, see also Myers and Patil (2013a).

Formal concept lattices with their symmetric view on indicators and objects constitute a theoretical approach for exploration beyond HDT that is relevant but not presently pursued (for details see for instance, Annoni and Brüggemann (2008) or Brüggemann and Patil (2010, 2011)].

2.2.2 Reorientation of Candidate Indicators

Unless it is apparent from the context, attention should be given to orientation of column variables as candidate indicators. Statistical evidence of orientation issues in the data come as substantial negative rank correlations and affect HDT by increasing incomparables $U(x)$. The statistical signal can be observed by inspection of a rank correlation matrix, but the effect on $U(x)$ may not be readily detectable by simple inspection.

2.2.2.1 Benchmarking

If several experts are involved who can reach consensus regarding appropriate order for a subset of objects, an expedient approach is to compare original and reversed orientation for each column separately to see which best reproduces the order of the experts for those particular objects. If this can be satisfactorily accomplished, then the more laborious combinatorial approach below can be avoided.

2.2.2.2 Comparative Combinatorial Reorientation

Exploratory analysis can be conducted with regard to amount of incomparability in relation to changing orientation. The "reorientation" is to be exemplified by first considering two fictitious datasets and then a sample of actual data.

Dataset 1:
As one can easily see, quantifier q_2 is just the reverse of quantifier q_1. If one of the first two columns is reversible to accord better with the third, then the first two columns become redundant. The third column differs from the first by pair-wise reversals for object *a* with *b* and object *e* with *f*, along with a tie for objects *c* and *d*. The third column differs more relative to the second, since reversed pairs are reversed. If reorientation is not admissible, then this dataset offers no context of crisp comparability.

An indicator function of incomparability (2.8) as a variant of $U := |U(x)|$ serves here as the criterion for choosing among combinations of reversals:

$$Ugr(X) := \left| \left\{ (x,y) : x, y \in X, x \parallel y \right\} \right| \tag{2.8}$$

with (2.8) being a function on the set of all data matrices resulting from the original matrix by reversing a subset of indicative quantifiers. This set of data matrices has 2^m members including the original data matrix. Bit patterns are used to designate combinations of columns, with a 1 in the string of binary digits (bits) designating that the column is reversed. Examples of bit patterns are as follows:

[0, 0, 0] no data column is reversed
[1, 0, 0] first data column is reversed
[1, 1, 0] first and second data columns are reversed

There are some redundancies among the bit-patterns for Table 2.1 with respect to comparability of objects since reversing an entire table does not change comparability; for instance [1, 0, 0] represents the same set of order relations as [0, 1, 1] in an upside-down manner. The indicator function Ugr([0, 1, 1]) is the same as Ugr([1, 0, 0]), except that the Hasse diagrams of [1, 0, 0] and of [0, 1, 1] are dual (see Brüggemann and Patil 2011). Note also that the pattern [1, 0, 0] is the complement of [0, 1, 1] in Boolean algebra.

Applying (2.8) for all possible eight data matrices of Table 2.1 leads to the following graph (Fig. 2.1).

Table 2.1 First fictitious dataset

Objects	Quantifiers		
	q_1	q_2	q_3
a	1	6	2
b	2	5	1
c	3	4	4
d	4	3	4
e	5	2	6
f	6	1	5

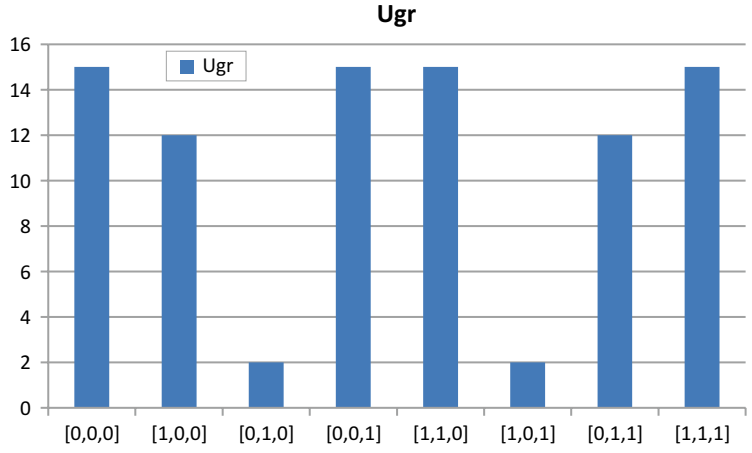


Fig. 2.1 Ugr [Eq. (2.8)] as a function of the eight different reorientation patterns

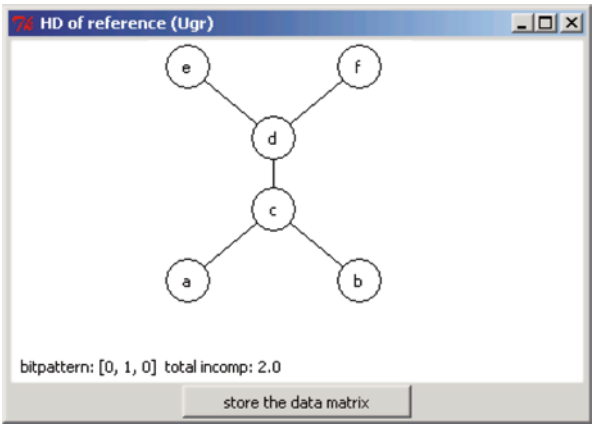


Fig. 2.2 Hasse diagram from data matrix having the second column reversed (PyHasse module: orientation1.py)

One can see that reversing the second column [0, 1, 0] would lead to the lowest Ugr and to a Hasse diagram richest with respect to the order relations $\perp$ (Fig. 2.2). For the reason given above, the complement of this pattern [1, 0, 1] also carries the same comparability but with the sense of the entire table being reversed. Likewise,

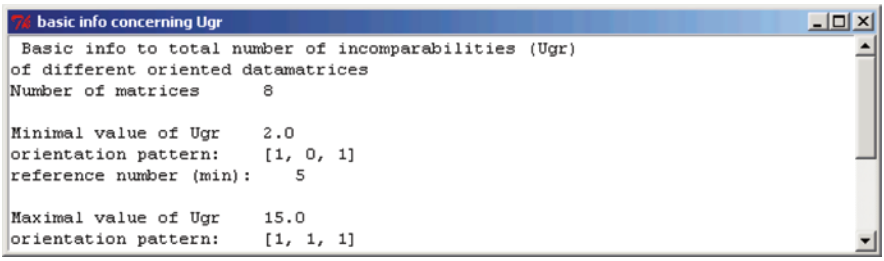


Fig. 2.3 PyHasse interface for basic information about the set of data matrices with all combinations of attributes reversed

Table 2.2 Information about U-scanning through all possible (Table 2.1) column orientations

Basic info on total number of incomparabilities (Ugr) of differently oriented data matrices	
Number of matrices	8
Minimal value of Ugr	2.0
Orientation pattern:	[1, 0, 1]
Reference number (min):	5
Maximal value of Ugr	15.0
Orientation pattern:	[1, 1, 1]
Reference number (max):	7
Mean value of Ugr	11.0
Reference numbers of minimal values of Ugr:	2, 5

reversing all columns is as bad as not reversing any. Reversing the first column is not the same as reversing the second column, since the third column is more like (higher rank correlation) the first than the second. Reversing both, second and third column, compounds conflict with the third column. It should also be noted with reference to Fig. 2.2 that presence of a completely redundant column does not increase the incomparability of objects, since one conflict is sufficient to make objects incomparable under the product-order relation used in the definitions above.

However, in general, it is not a tractable proposition to draw figures such as Fig. 2.1 since combinatorial expansion would entail a diagram of 1,024 bars for as few as ten quantifiers. Tabular condensation is needed as shown in Fig. 2.3 where the minimal and maximal values of Ugr are reported together with their bit-patterns (extracted from full information provided by the PyHasse module "orientation1.py" as given in Table 2.2). In order to refer to a bit-pattern, a reference number is introduced as follows.

Let $[a_m,a_{m-1},\ldots,a_1]$ be components of a bit-pattern with a_k referring to contribution of 2^{k-1}, then

$$Reference\,number = \sum a_i * 2^{i-1}\,(i = 1,\ldots,m)\;and\;a_i \in \{0,1\}. \tag{2.9}$$

Table 2.2 includes a reference number for maximal value of Ugr which was truncated from Fig. 2.3.

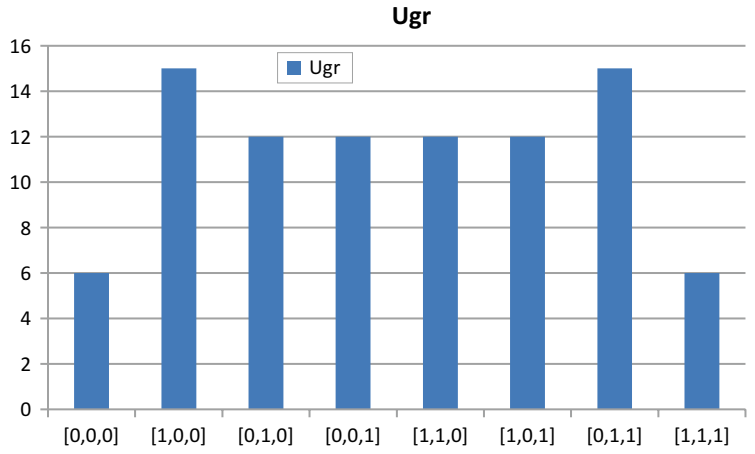


Fig. 2.4 Variation of the Ugr-function [Eq. (2.8)] in the case of data matrix of Table 2.3

Table 2.3 Second fictitious dataset

	q_1	q_2	q_3
a	1	3	1
b	2	2	2
c	3	1	3
d	4	5	6
e	5	6	5
f	6	7	4

Table 2.4 Basic results of orientation1.py applied on data of Table 2.3 (second fictitious dataset)

Basic info to total number of incomparabilities (Ugr) of different oriented data matrices	
Number of matrices	8
Minimal value of Ugr	6.0
Orientation pattern:	[1, 1, 1]
Reference number (min):	7
Maximal value of Ugr	15.0
Orientation pattern:	[1, 0, 0]
Reference number (max):	4
Mean value of Ugr	11.25
Reference numbers of minimal values of Ugr:	0, 7

Dataset 2:
Here there are six objects and three quantifiers without any completely countercurrent columns. The first column is the same as for dataset 1. The value 4 does not appear in column q_2.

The result of scanning column reversals for dataset 2 is given in Table 2.4. In this dataset, column reversals change Ugr by a factor of 2.5 (= Ugr_{max})/Ugr_{min}). However, the minimum is for the original orientation with any change increasing the number of incomparabilities—so none should be done. Figure 2.4 shows the bar diagram of the Ugr values for different combinations of column orientations.

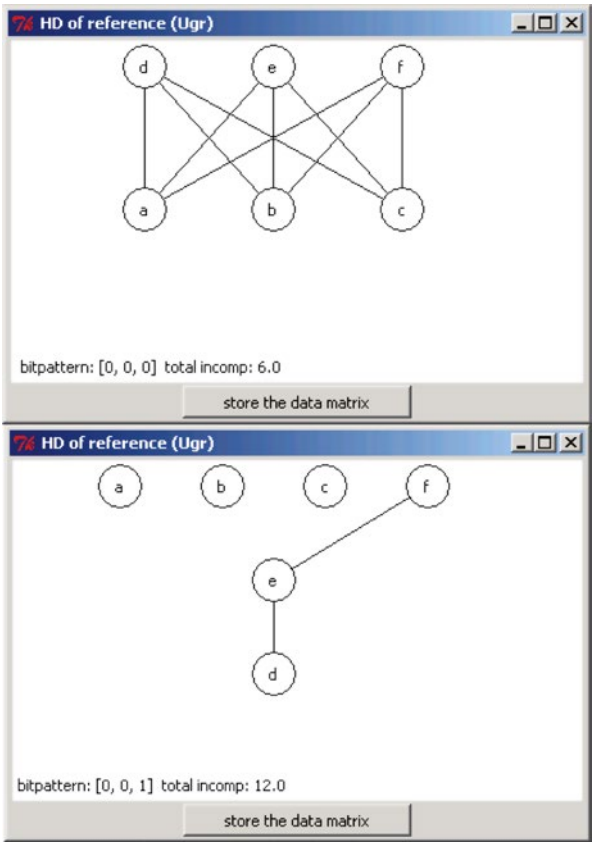


Fig. 2.5 Two examples of Hasse diagrams based on Table 2.3

In Fig. 2.5 two examples of Hasse diagrams are shown: one for the original configuration [0, 0, 0] and one for the configuration [0, 0, 1]. Note that reversal of the third indicator has made half of the objects isolates with an ostensible complete lack of comparability. Here, the possible analysis of antichains and other variants (Carlsen and Brüggemann 2014) are not shown.

Dataset 3:
It may be of interest to check a sample of actual data. For this, a data matrix is selected which is a part of a larger matrix discussed by Myers and Patil (2013a). There are ten indicative attributes as percentages of different kinds of land cover. For regions defined in terms of biology and geomorphology, different kinds of land cover have intrinsic implications as indicators of landscape ecology in a particular region. Indicator Pct40 is related to the fact that a given region has forest as a natural land cover if ecological succession is allowed to proceed without intervention. Natural disturbance dynamics have transitional cover (Indicator Pct33) where disturbed patches are again reverting to forest. Humans are a major (non-natural) disturbance agent in this region. The most extreme and enduring scars of

humanization in these landscapes are quarries, strip mines, and gravel pits (Pct32). Development (Pct20) along with urban/recreational grasses (Pct85) constitutes a prevalent aspect of humanization as long-term disturbance. Agriculture comprises a second prominent aspect of humanization comprised of crops (Pct82) and hay/pasture (Pct81) with crops involving annual exposure of the soil surface to erosion in this relatively steep terrain. Loss of natural wetlands to humanization has led to restrictions on further conversion of woody wetlands (Pct91) and emergent herbaceous wetlands (Pct92) which together currently comprise a small fraction of total landscape area. Water (Pct11) naturally occurs primarily as rivers and streams, but a large impoundment called Raystown Lake is also present. The objects are geographical units called OCTIVs (OCTagonal Integrating Vicinities). The number of objects is restricted to a sample of 12 for a local context with the matrix of reduced rows being shown in Table 2.5. This data matrix exemplifies the degenerate case wherein all objects are isolates with none being comparable to any others. Table 2.6 shows the result of reorientation, whereby some comparability can be induced.

When we take the algebraic complement of [1, 1, 1, 0, 1, 0, 0, 1, 1, 1] namely [0, 0, 0, 1, 0, 1, 1, 0, 0, 0], we see that three attributes reversed would give an enriched poset, where Ugr is reduced by a factor of 1.2 (66/55). There seems to be no single attribute which is completely countercurrent to the others. The number Ugr = 66 corresponds to the maximal possible U. The Hasse diagram, corresponding to the bit pattern [1, 1, 1, 0, 1, 0, 0, 1, 1, 1] is shown in Fig. 2.6, having only three isolates.

The attributes for which reversal is indicated are:

PCT33: Transitional

PCT 81: Hay/pastures

PCT 82: Row crops

These three attributes belong to those land cover characteristics which are considered as neither pro nor con (see Myers and Patil 2013a). Hence the reorientation study suggests reversing three attributes in order to get a more enriched Hasse diagram (which would be the dual poset of that shown in Fig. 2.6).

Additional computational and structural considerations

For practical purposes, exploration of orientation also entails consideration of computation. The Ugr scan is comprehensively combinatorial and therefore fully informative. However, it is computationally most intensive. On way to ease such computational burden is to make the approach somewhat sequential rather than completely combinatorial. Accordingly, an initial sweep can invert each column individually and retain the one inversion that has most impact. The remaining columns in this modified matrix can then each be examined individually for additional improvement, with the best of these being retained in a twice modified matrix. This progression of successive modification continues until no improvement is obtained. Still another strategy would be to reduce computation by investigating random samples from datasets having large numbers of objects (rows).

If the burden of combinatorial computation is reduced in the foregoing manner, then it also becomes feasible to incorporate disjoint features of the Hasse diagrams in the progressive scans. The most obvious of these features are isolates, since HDT structure that is disjoint will preclude any sort of inter-subset inference.

Table 2.5 Twelve geographical units called OCTIVs (OCTagonal Integrating Vicinities) by Myers and Patil (2013a, b) with ten IVI (Integrative Vicinity Indicator) attributes as percentages of different kinds of land cover

ID	Pct11	Pct20	Pct32	Pct33	Pct40	Pct81	Pct82	Pct85	Pct91	Pct92
25	0.007961	0.007961	0.000000	0.732425	32.20285	59.159302	7.889499	0.000000	0.000000	0.000000
26	0.000000	0.007959	0.000000	0.127358	86.93783	10.745840	1.265621	0.000000	0.875587	0.039799
36	42.629608	0.326511	0.000000	0.238910	56.31918	0.015927	0.103527	0.000000	0.000000	0.366329
37	1.751871	0.055741	0.000000	0.541487	95.54069	1.600573	0.445930	0.000000	0.000000	0.063704
38	0.047778	0.000000	0.000000	0.000000	90.38063	8.862876	0.668896	0.000000	0.000000	0.039815
39	0.055719	0.047759	0.000000	0.692509	89.74767	7.689246	1.767093	0.000000	0.000000	0.000000
48	0.063704	0.509635	0.000000	0.597228	54.32394	36.765408	7.684344	0.000000	0.031852	0.023889
49	4.872223	0.063689	0.000000	0.000000	93.25691	1.249900	0.326407	0.000000	0.055728	0.175145
50	4.139468	0.103486	0.000000	1.010985	92.38975	1.807037	0.167170	0.000000	0.151249	0.230854
51	1.480655	1.369208	1.098551	0.175131	94.77790	0.573157	0.310460	0.000000	0.000000	0.214933
59	0.055736	2.508161	0.000000	0.127398	62.49701	26.690023	4.912811	2.508161	0.270722	0.429970
60	3.495222	2.866242	0.000000	0.987261	87.67516	2.300955	1.815286	0.000000	0.597133	0.262738

Table 2.6 Reorientation of the data matrix of Table 2.5

Basic info on total number of incomparabilities (Ugr) of differently oriented data matrices	
Number of matrices	1,024
Minimal value of Ugr	55.0
Orientation pattern:	[1, 1, 1, 0, 1, 0, 0, 1, 1, 1]
Reference number (min):	935
Maximal value of Ugr	66.0
Orientation pattern:	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Reference number (max):	1023
Mean value of Ugr	65.45
Reference numbers of minimal values of Ugr:	88, 935

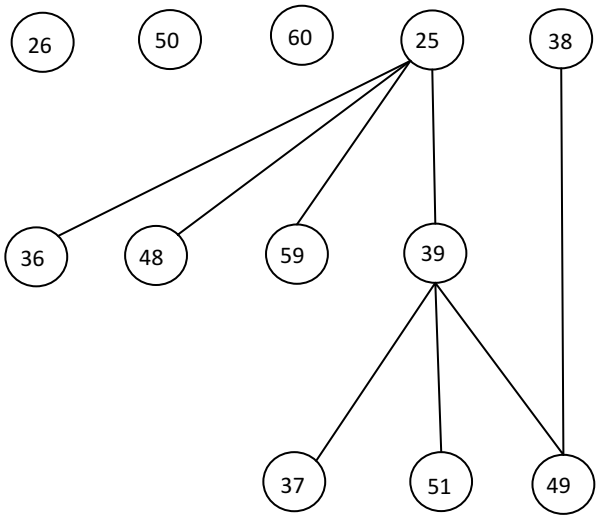


Fig. 2.6 Twelve regions in Pennsylvania under ten attributes (partially reoriented data matrix)

2.2.3 Elimination of Individual Candidate Indicators

One may ask as to how far single indicative quantifiers are individually crucial with respect to order relations. Therefore another kind of exploratory analysis can be performed whereby eliminating the effect of one candidate indicator is considered. The number of different matrix outcomes is far less for such single eliminations than for combinatorial reorientation. The procedure is similar to the sensitivity analysis with the central concept of the matrix W (see Brüggemann and Patil 2011; Brüggemann et al. 2001), however, the indicative quantities are not the entries of the matrix W but $|U(x)|$.

Since the scope of scanning is different, another variant of $U := |U(x)|$ is denoted by Uli as the analytical indicator (or sensitivity) function. The set of matrices for which Uli is calculated contains $m+1$ data matrices with these being the original and m data matrices where one single column as a candidate indicator is eliminated. Uli has a value corresponding to elimination of each individual column indicator

Table 2.7 Basic information on (Uli) incomparability for single elimination in Table 2.1

Indicator	1	Not considered: Uli=	12.0
Indicator	2	Not considered: Uli=	2.0
Indicator	3	Not considered: Uli=	15.0

Table 2.8 Basic information on (Uli) incomparability for single elimination in Table 2.2

Indicator	1	Not considered: Uli=	6.0
Indicator	2	Not considered: Uli=	3.0
Indicator	3	Not considered: Uli=	3.0

which is the number of incomparabilities remaining after elimination of the respective column. This is exemplified by applying the PyHasse module on the data matrices of Tables 2.1 and 2.2.

2.2.3.1 Dataset 1 (Table 2.1)

An investigator would normally explore elimination after doing any indicated reorientation. For comparison with the results of the earlier reorientation study, however, we work directly on the data of Table 2.1. The result for the dataset of Table 2.1 is shown in Table 2.7 and is confirmatory to and consistent with prospective reorientation. Eliminating the second indicator reduces the incomparabilities to 2. This is the same number obtained by reversing the second indicator to be completely redundant with the first indicator. This verifies that a completely redundant indicator has no effect on the number of incomparabilities. Hence the finding of reorientation and appropriateness of elimination are both confirmed.

2.2.3.2 Dataset 2 (Table 2.3)

The result for column (candidate indicator) elimination for dataset 2 (Table 2.3) is shown in Table 2.8. For this dataset, the exploration of orientation was not suggestive of any change. Here the order structure is most sensitive to elimination of the first column. It is less sensitive, but equally so, to elimination of either column 2 or 3. This illustration sets the stage for what follows.

2.2.3.3 Sample of Actual Data (Table 2.5)

For the sample of actual data in Table 2.5 (without reorientation) the result of single elimination is not informative. No elimination of a single attribute from the original indicator matrix reduces Uli. The complete isolation of objects for the full dataset remains so for all single-column eliminations.

Table 2.9 Impact values from PyHasse module (sensitivity18_3.py) for reoriented land cover matrix

Property	Sensitivity	Property	Sensitivity	Property	Sensitivity
Pct32	0	Pct92	0	Pct40	3
Pct81	0	Pct11	2	Pct33	7
Pct85	0	Pct20	2		
Pct82	0	Pct91	2		

2.2.4 Culling of Candidate Indicators

Analysis of the impact of each single attribute on the structure of a poset (see Sect. 2.2.1) has attracted considerable interest (see Brüggemann et al. 2001). On that basis, attention is focused here on application of the cumulative ambiguity graph concept (Brüggemann and Patil 2011, 2010). This entails a sensitivity analysis using incomparability as a sensitivity criterion. The attributes are ordered from high to low impact on the poset and successively incorporated beginning with the one that is most influential. Therefore, m data matrices are to be checked. Again $U := |U(x)|$ provides an indicator function for ambiguity and a plot is obtained with U (scaled as fraction of maximum ambiguity) on the ordinate and number of included attributes, "natt," on the abscissa. The indicator function can attain values "near" the maximum either when natt approaches m or even at values of natt = m^* << m. In the latter case, a segregation of the indicative attributes is motivated. The first m^* attributes are considered as important for retention, whereas the remaining $m-m^*$ attributes are considered as relatively redundant. Those attributes having relative redundancy are either eliminated or set aside as optional "fine tuning" (nuance) attributes for later inclusion if needed.

This method of sensitivity analysis is to be applied considering that the suggested column reversals provide an oriented matrix (odm) of land cover data. Table 2.9 shows results obtained from PyHasse module (sensitivity18_3.py) for sensitivity in terms of HDT mismatches induced by deleting the respective indicator. The cumulative ambiguity maximum (CAM) graph is shown in Fig. 2.7.

There are some ties with respect to the attribute sensitivity data, with ties being ordered arbitrarily:

Pct33 > Pct40 > Pct91 = Pct20 = Pct11 > Pct92 = Pct82 = Pct85 = Pct81 = Pct32.

This graph and Table 2.9 is suggestive of the first five (ordered) attributes being important for retention and the rest being essentially disposable. The user interface of PyHasse module (sensitivity19_1.py) is shown in Fig. 2.8. A sensitivity analysis for this subset of five (oriented) attributes is given graphically in Fig. 2.9.

From Fig. 2.9:

"Emergent Herbaceous Wetlands" (Pct 92) > "reoriented Hay Pasture" (Pct81) > "reoriented Row Crops" (Pct82) > "Quarries etc." (Pct32) > "Urban/Recreational Grasses" (Pct32).

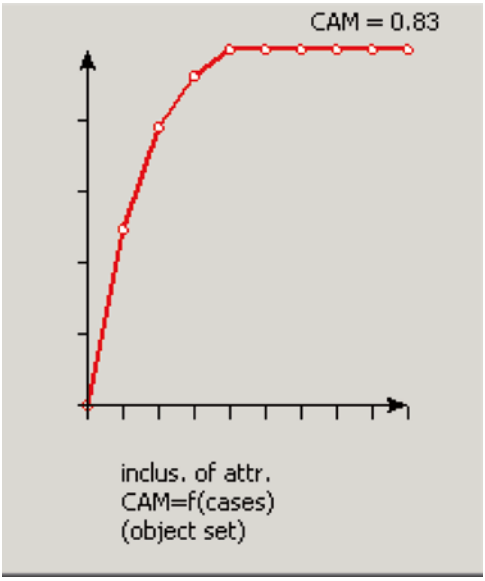


Fig. 2.7 CAM (cumulative ambiguity maximum) graph of the partially reoriented land cover data with ordinate as fraction of maximum ambiguity (incomparability)

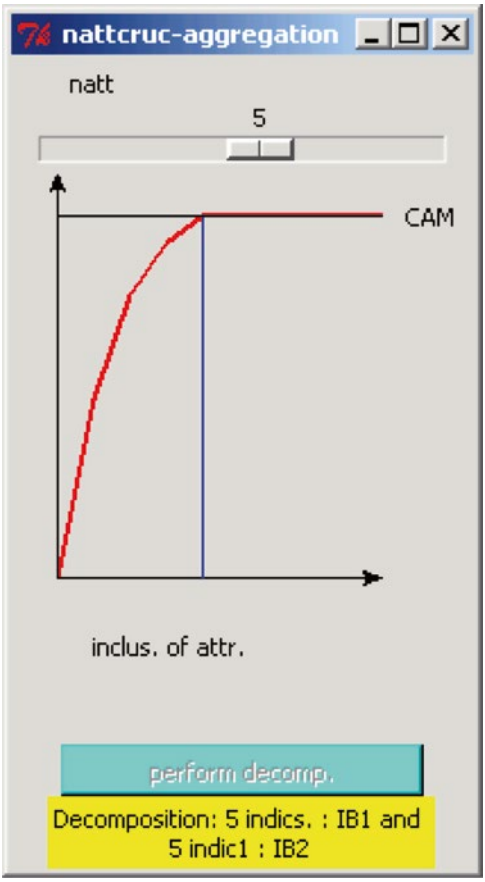


Fig. 2.8 Interactively selecting the most important indicative attributes on the basis of a CAM-graph

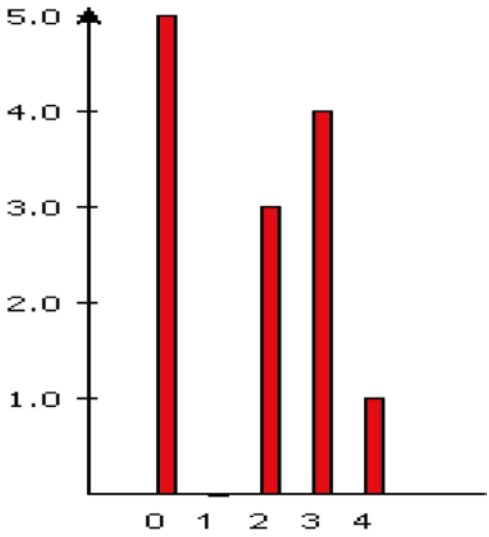


Fig. 2.9 Sensitivity (HDT impact) of five indicators: Pct92 (0), Pct85 (1), Pct82 (2), Pct81 (3), Pct32 (4)

Taking these five most important (reoriented) attributes, a poset results with $U = 23$, whereas U of the reoriented (odm) complete matrix is 55. Seven HDT levels are obtained (Fig. 2.10), indicating substantial gain of order relations. Object 25 is the single maximal element (i.e., a greatest element), with objects 36, 51, and 59 being minimal elements. There are now no isolated elements. We call this data matrix with the five most important (reoriented) indicators as a redacted or "refined" data matrix (rdm).

Explorations of culling attributes need not be limited to the foregoing dichotomy of "major" and "minor" indicators. Two indicators that order the objects in the same manner are order-redundant, and with only a few perturbations there is still a high degree of such redundancy. Given that one such indicator is in the suite, another will only marginally alter the perspective on order relations. The statistical signal of nearly redundant indicators is a strong rank correlation. Conversely, indicators with weak or negative rank correlation will order objects differently and thus introduce appreciable ambiguity. This is essentially the same perspective as above, but framed somewhat differently. This framing would suggest rearranging the indicators so that higher degrees of redundancy are reflected in a gradational block diagonal pattern for the rank correlation matrix. If such a pattern emerges, it can motivate culling of the indicators by selecting one that is central to each block. This reduced suite of indicators can then be compared in terms of effect to the major ("important") indicators emerging from the foregoing method. Carrying this line of exploration further, each block of indicators can be segregated and separately analyzed to obtain its characteristic structural signals in HDT. If distinctive structural patterns emerge in this manner, then the implication is that there are distinct domains of induced order that merit separate characterization followed by study of how those domains interact.

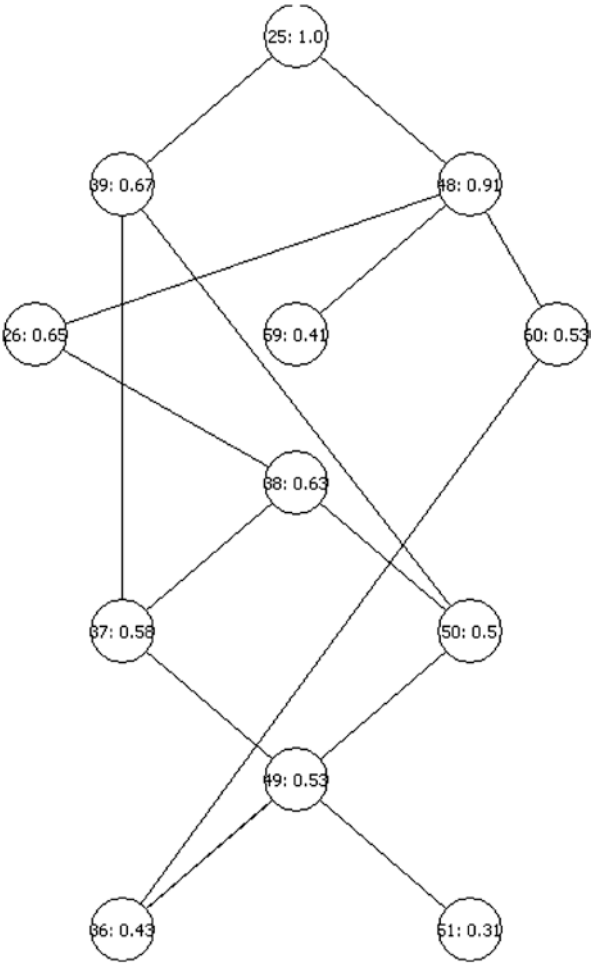


Fig. 2.10 Hasse diagram for land cover data using five major indicators after reorientation and culling of the indicator set. Note: second values inscribed in *circles* are referenced subsequently

2.3 Quantification and Generalization of Indication

2.3.1 Shared Sense of Sequence and Diminutive Differences

In the foregoing section the focus was on reorientation, elimination, and partition of attributes. However, the result of a partial order study also depends heavily on the data representation. One might raise a question regarding which numerical differences of metric attributes are meaningful for sequencing (in terms of ranks). Implicit in such a question, however, is the assumption that magnitudes of differences are meaningful. This assumption holds only for data that are scaled at a level of interval strength, whereas sequencing as ranks need only entail ordinal quantification. Once again, this can be explored with the land cover data of Table 2.5, which is especially suited to this purpose because it addresses compositional components necessarily

Table 2.10 Attribute value differences according to their distribution in 0.1 intervals, based on [0, 1]-normalized data matrix

	Pct11	Pct20	Pct32	Pct33	Pct40	Pct81	Pct82	Pct85	Pct91	Pct92
0.0…0.1:	43	25	55	9	24	19	20	55	28	16
0.1…0.2:	12	12	0	11	7	18	13	0	9	5
0.2…0.3:	0	0	0	5	0	1	4	0	2	3
0.3…0.4:	0	2	0	5	3	3	2	0	8	8
0.4…0.5:	0	8	0	10	7	9	3	0	0	11
0.5…0.6:	0	1	0	8	11	4	4	0	1	9
0.6…0.7:	0	0	0	4	6	3	2	0	9	2
0.7…0.8:	0	2	0	4	0	0	4	0	0	3
0.8…0.9:	1	9	0	6	2	3	2	0	1	4
0.9…1.0:	4	4	0	2	5	5	11	0	2	3

ranging from 0 to 100 %. For compositional data (before any reorientation) the column indicators are additive across a row with upper limit on the total of 100 %. Table 2.10 shows the distribution of differences in values for the compositional indicators in fractional terms; or, more generally, on the basis of a (0, 1)-normalized data matrix. It can be seen that Pct11 has data differences mainly concentrated on the lower values and some few near the maximum possible difference. Pct32 is still more extremely distributed over the ten intervals, since there are only differences in the lowest interval (0.0…0.1). Pct82 has differences spread out over all ten intervals, and the lowest interval is only slightly more frequently represented.

A client might tend to question, for example, whether the slight difference for Pct11 between objects 25 (0.007961 %) and 26 (0.000000 %) in the non-normalized original land cover data matrix should really be taken into account for a ranking, when at the same time differences up to around 42 % are possible. A seemingly apparent approach, namely to declare two objects as equivalent when the numerical difference of an attribute is less than ε (ε arbitrarily selected), is not even possible because of the transitivity of the equivalence relation. Consider three objects a, b, and c with data values 0.1, 0.2, and 0.3 and select $\varepsilon = 0.15$, then certainly a and b are equivalent. In like manner b, c are equivalent; hence a, b, and c must be in the same equivalence class, although the difference between a and c is larger than ε! In Brüggemann and Patil (2011) some possibilities for overcoming this problem are described in detail. Three such methods are discretization, cluster analysis, and application of fuzzy concepts. Cluster analysis and fuzzy concepts lead to additional interesting aspects of collective interaction among objects as seen through a lens of multiple indicators.

2.3.2 Discretization

The seemingly simplest variant would be to select a range of data for each of the attributes and to select the number of equidistant intervals covering the data range. Instead of $q_i(x)$, object x gets a score, $s_i(x)$, indicating to which interval $q_i(x)$ belongs.

In statistics this is generalization in class intervals, and in numerical analysis it is the strategy of generalization by "binning."

It can be argued that such discretization carries drawbacks of arbitrariness in selecting data range and number of classes and that data values having only slight numerical difference can be assigned to different classes (see Brüggemann and Welzl 2002). Although these cautions are well founded, they have not deterred application in either statistics or numerical analysis for achieving generalization. Likewise, it is ill-advised to dismiss the approach because of these cautions in connection with HDT. A major issue for HDT as raised earlier is difficulty in filtering fine structure of a poset so that coarse structure becomes evident. Some occasional perturbations at cutoffs will affect fine structure rather than coarse structure, and generalization should be tunable with regard to its severity. Doing HDT on quotient sets with different bin widths should therefore serve to help elucidate coarse structure as is needed. The variant that is actually simplest is generalization by orders of magnitude through rounding or truncation of successive digits in multi-digit values.

2.3.3 Clustering, Collectivity, and Posetic Features

Cluster analysis is one of the several modes of data analysis in the statistical genre that merit greater usage in conjunction with posetic information (Myers and Patil 2012, 2013c). As is often the case, some of the more common configurations for general purposes are not particularly suitable in this regard as with the version called Kmeans which requires specific knowledge about the number of clusters (Bock 1974, 1979; Bock et al. 2003; Diday 1979; Brüggemann et al. 2013a). However, others among the many clustering modalities (Luther et al. 2000) invite imaginative adaptation (see Myers and Patil 2013b).

Since clustering most often addresses objects through (dis)similarity of their attributes, coupling these analytical domains is most readily done through posetic features of individual objects such as down-set for $x \in X : O(x) := \{y \in Y : y \leq x\}$, up-set for $x \in X : F(x) := \{y \in Y : y \geq x\}$, and the x-incomparables $x \in X : U(x) := \{y \in Y : y \| x\}$. This is stressed by Myers in several papers (Myers et al. 2006; Myers and Patil 2008, 2010, 2013a), because in areas like data mining the Hasse diagram loses its attractiveness quickly . This extends to more flexible graphic depictions than Hasse diagram itself as is well illustrated by the FOU-plot. The FOU-plot uses as one coordinate the difference of the contents of the down-set ($|O(x)|$) and the up-set ($|F(x)|$) and as second coordinate the content of the incomparable elements with $x : |U(x)|$ (see also Brüggemann et al. 2013c). A diagram of this type which corresponds to the Hasse diagram of Fig. 2.10 is shown in Fig. 2.11.

Leftward in a FOU-plot is less favorable and rightward is more favorable, with higher positions carrying greater incomparability. Thus, the most definitely favorable circumstance is for object 25 at the lower right marked in blue. Similarly, the most definitely unfavorable circumstance is for object 36 marked in blue nearest the lower left.

With regard to clustering, the hierarchical agglomerative mode is most adaptable to the present purpose. It gives rise to a dendrogram by which to view evolution of

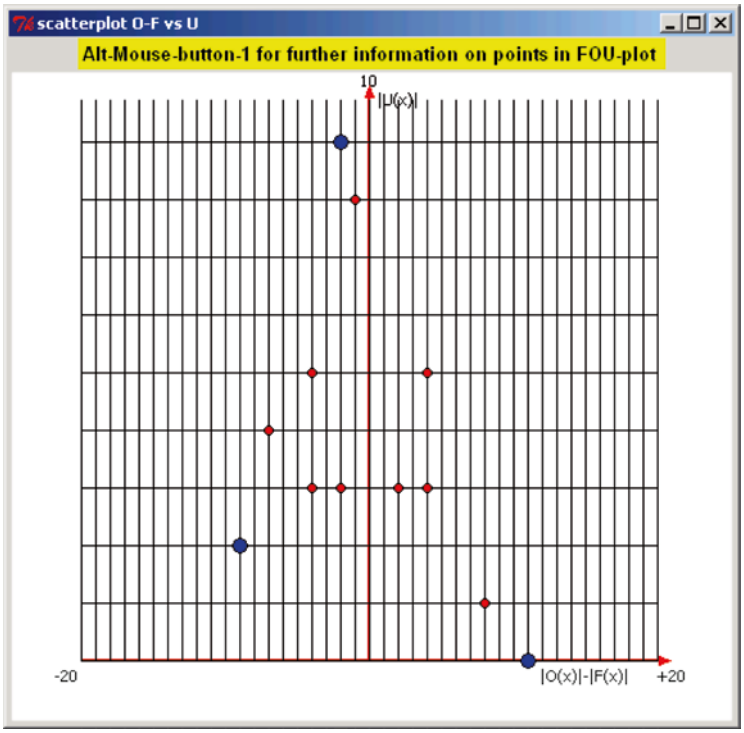


Fig. 2.11 FOU-plot corresponding to the Hasse diagram of Fig. 2.10. *Upper blue point* marks object 59, *left-most blue point* marks object 36, and *right-most blue point* marks object 25

grouping and either graphical or tabular summarization of the objects at any interesting level of grouping can be rendered. The posetic features as illustrated above as well as posetic rankings can be attached to member objects of particular groups with other groups being suppressed for purposes of simplified pattern perception. In this way, posetic characteristics can be carried into the zoom-in and zoom-out generalization of a dendrogram (see Myers and Patil 2013b).

Depicting posetic and attribute characteristics of clusters in lattices of pair-wise plots can be particularly effective for heuristic examination of degrees of joint variation (or what may be called "collectivity") of the indicators. This in turn can suggest exploratory grouping of indicators into different domains of propensity for ordering.

2.3.4 *Fuzzy Concept*

Fuzzy concepts are described in Brüggemann and Patil (2011) and Brüggemann et al. (2011) and are basically derived from a fuzzy subsethood, where the <-relation among objects is transformed into a fuzzy subsethood (see also Van de Walle et al. 1995).

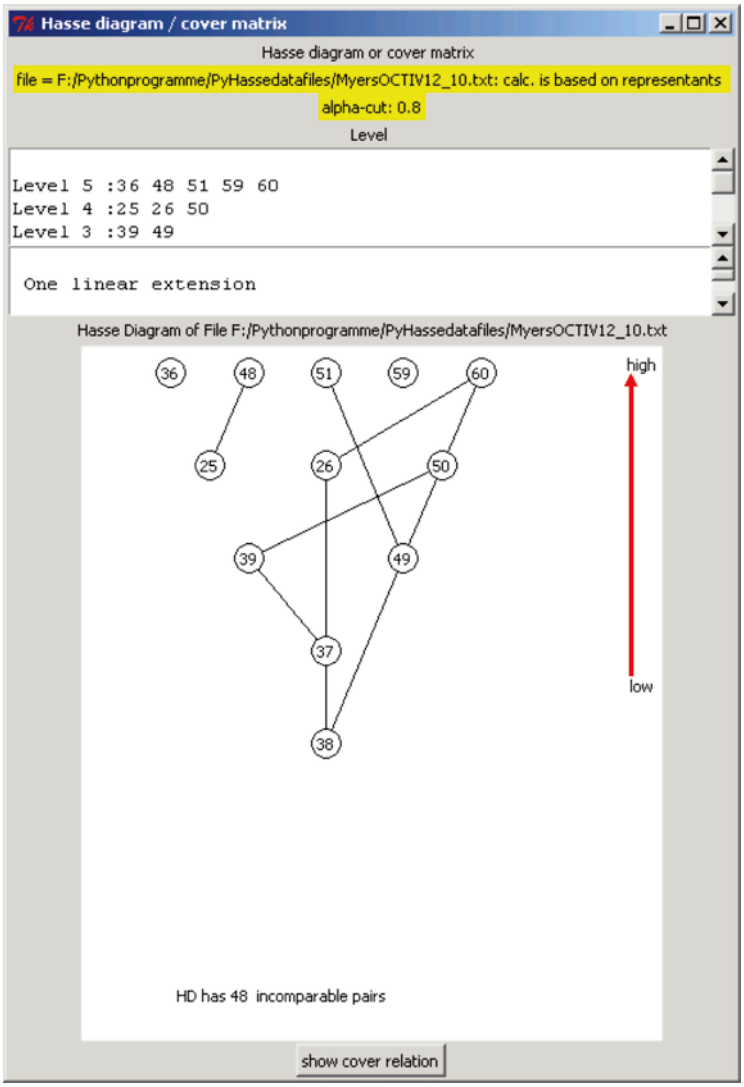


Fig. 2.12 Fuzzy results for initial (unmodified) land cover data in Table 2.5

The Fuzzy concept is illustrated here to show possibilities. Exemplification can be done in terms of the land cover data, first using the unmodified data in Table 2.5 with the fuzzyHD13.py module of PyHasse software and then performing defuzzification to obtain the Hasse diagram shown in Fig. 2.12.

Figure 2.12 may be compared to Fig. 2.6 which shows the Hasse diagram obtained after modifying the initial data matrix by orienting columns, while keeping in mind that the data matrix of Table 2.5 gave a complete antichain (all objects as

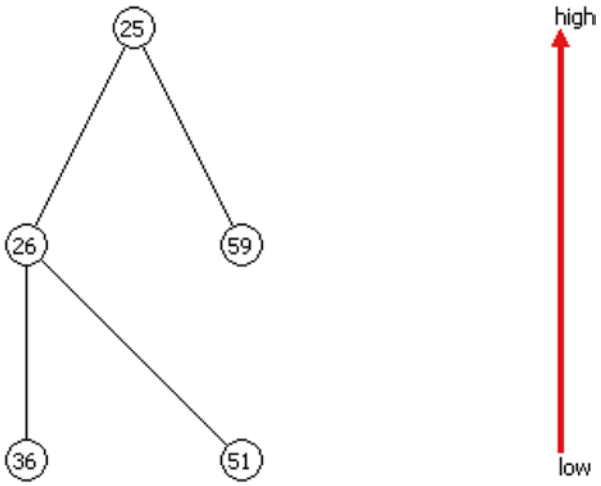


Fig. 2.13 Fuzzy poset for five-column-refined land cover data (compare to Fig. 2.10)

Table 2.11 Equivalence classes of fuzzy poset for five-column-refined land cover data

object 25:	48
object 26:	37 38 39 49 50 60

isolates) before orientation. The fuzzy approach has clearly found structure that was not evident in the crisp approach.

It is of further interest to continue with the fuzzy versus crisp approaches on the five-column-refined data matrix and compare the result to the Hasse diagram of Fig. 2.10. The results of doing so are shown in Fig. 2.13 and Table 2.11.

It is of interest that the indicator function U is now 4, signaling that (a) many conflicts are due to slight numerical differences and (b) the reorientation eliminated many conflicts. Table 2.12 shows that a cost of the reconciliation comes as ties. Comparative results are shown in Table 2.12.

A careful mathematical description based on an example of refrigerants can be found in Brüggemann et al. (2011). Along with questions of combining attribute values in the Kosko-formula (see Van de Walle et al. 1995), the main challenge facing a researcher becomes how to defuzzify the result.

2.4 Weak Order Provided by Partial Order Approaches

In partial order theory, the linear extensions play a large role. A linear extension is a linear order which preserves the order relation of a poset (for details, see Trotter 1992). Scanning all possible linear extensions, objects will be located at different

Table 2.12 Comparative results for fuzzy and crisp approaches

Data matrix	U	Ties	Remarks
Initial data (Table 2.5)	Gets its maximum possible value as the poset is a complete antichain	No ties	Many incomparabilities may be caused by small numerical differences as can be deduced from Table 2.10
Oriented (Fig. 2.6)	A reduction from $U=66$ (initial) to $U=55$ (oriented)	No ties	The poset shows some structure There are three isolated vertices and a vertex (no 38) loosely connected with the set {25, 36, 37, 39, 48, 49, 51, 59}
Five-column refined (Fig. 2.10)	$U=23$	No ties	Seven levels, hence at least one chain with seven elements, i.e., a subset of X which can be ranked without any additional information
Initial data fuzzy (Fig. 2.12)	$U=48$	No ties	Five levels, several graph–theoretical components, which indicate peculiarities in the data
Five-column-refined fuzzy (Fig. 2.13)	$U=4$	Many ties	A weak order can be deduced from X, however, the number of ties needs additional steps to break them

heights. Therefore the average height (hav) is a measure for the rank and is called the "average rank" (Rkav). The exact calculation is computationally difficult (NP-hard), however, certain approximations are derived (see Brüggemann and Carlsen 2011) and the most simple one, the "Local Partial Order Model 0" (LPOM0) is just based on the number of objects and the two coordinates also used in the FOU-plot.

Using the rather rough expedient of treating $U(x)$ as a single super-object gives the approximation:

$$hav(x) = \left[O(x) + U(x)\right]\left[O(x) / \left\{O(x) + F(x)\right\}\right] + O(x)\left[F(x) / \left\{O(x) + F(x)\right\}\right]. \tag{2.10}$$

Therefore we use this extremely simple method to get an approximate measure regarding the weak order of the 12 objects as land-use regions, with results contained in Table 2.13. See also the chapter of Rocco and Tarantola (2014) for an application of the LPOM-concept. Interestingly, Eq. (2.10) is also obtainable as the mean value of a Binomial statistical model with parameters $n-1$ and $[O(x)/\{O(x)+\mathrm{F(x)}\}]$.

We find

$36 < 51 < 59 \cong 50 < 49 < 37 \cong 60 < 38 < 26 < 39 < 48 < 25$

The result of the exact method, i.e., counting the heights of any object in linear extensions, whereby the linear extensions are obtained as paths in a lattice of

Table 2.13 LPOM0 of F:/Pythonprogramme/PyHasse/pdt-files/Myersdecomp2_29.10.2012

Object	Rkav	$O(x)$	$U(x)$
25	12.0	12	0
26	9.1	7	3
36	1.182	1	2
37	5.2	4	3
38	7.8	6	3
39	9.75	6	5
48	10.833	10	1
49	3.9	3	3
50	3.25	2	5
51	1.444	1	4
59	3.25	1	9
60	5.2	2	8

Table 2.14 Average ranks (Rkav) according to method of Oliver Wienand

25:	12.0
26:	9.212
36:	1.538
37:	5.693
38:	7.63
39:	9.024
48:	10.793
49:	4.086
50:	4.419
51:	2.043
59:	5.397
60:	6.166

down-sets of the original poset ("lattice theoretical method," K. De Loof et al. 2006, program code: O. Wienand) is displayed in Table 2.14.

This sequence is

$36 < 51 < 49 < 50 < 59 < 37 < 60 < 38 < 39 < 26 < 48 < 25$

With the exact Wienand method the ties within LPOM0 disappear, and there are some rank inversions. From a methodological point of view this shows that the focus on $|O(x)|$, $|F(x)|$, and $|U(x)|$ is not sufficient. However, it is not claimed that sequence obtained from purely order theoretical aspects should replace all rankings obtained from other conventional multi-criteria decision methods (MCDM),since additional knowledge such as weights of each indicator come into play. Nevertheless, information extracted from partial order theory extended to incorporation of indicator weights may be useful for discussion of MCDM-rankings (see Patil and Joshi 2014), where a concept of reconciliation plays a major role. The theoretical framework is explained in that chapter.

2.5 Weighted Composite Indicators

2.5.1 Weighting for Indicators and Linear Extensions

As indicated above, a generalization of the concept of linear extensions by including weights for any single linear extensions (see Patil and Taillie 2004) opens a wide variety of analysis tools. Tools based on this generalization are subsumed under the concept of comparative knowledge discovery (CKD). "Reconciliation of weights of indicators" as one such tool also belongs to CKD, where it is of interest how far empirical weights of indicators can be adjusted to be coincident with subjective indicator weights of stake holders (see Patil and Joshi 2014).

Here, however, discussion will be limited to less sophisticated tools; namely two relatively simple methods to compare partial order findings with results of formulating composite indicators determined as weighted combinations of basic indicators.

2.5.2 Weight Intervals

The basis of this method is discussed by Brüggemann et al. (2013b) and follows the idea that exact weights may not be known but can be treated as intervals. Instead of stating a certain weight value for an indicator such as Pct11, we argue that the weight may be in an interval [$\text{weight}_{\text{min}}$, $\text{weight}_{\text{max}}$]. After giving all five indicators a weight interval, a Monte Carlo simulation can be performed by picking a certain weight value out of each interval (renormalizing if necessary so that the sum is one) and calculating a composite indicator (CI).

Clearly there are many CIs possible; however, some may vary in their numerical values without yielding different rankings. The central idea is that the set of CIs derived from the weight intervals constitutes a new multi-indicator system (MIS), which may now also take into account (through the weights) knowledge of stakeholders, decision makers, etc. If the weights are exactly known then clearly only one CI is obtained and a linear or weak order results. In general, however, several candidate CIs are obtained which render different ranks and the question is as to how far common order relations can be found. Thus the new MIS, consisting of the set of CIs, corresponding to the weight intervals is used to get a new and in general enriched poset. The refined data matrix (rdm) of five important reoriented land cover indicators is used to illustrate the procedure by applying weight intervals.

2.5.2.1 Weight Values Being Known

The PyHasse module (conflict7.py) is first applied to compare the average ranks, Rkav, (of the LPOM0-approximation) with the CI for equal weights. From the result given in Fig. 2.14 it is seen that CI and Rkav are not in conflict.

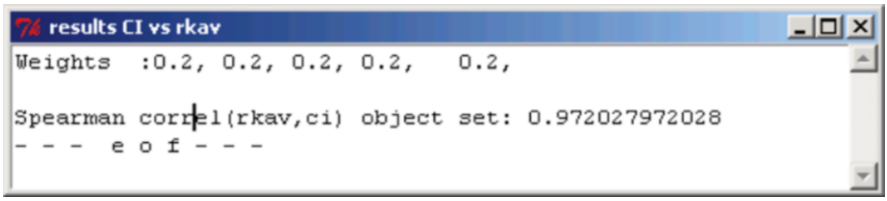


Fig. 2.14 Result of conflict-analysis based on equal weights of the CI and the LPOM0-approximation

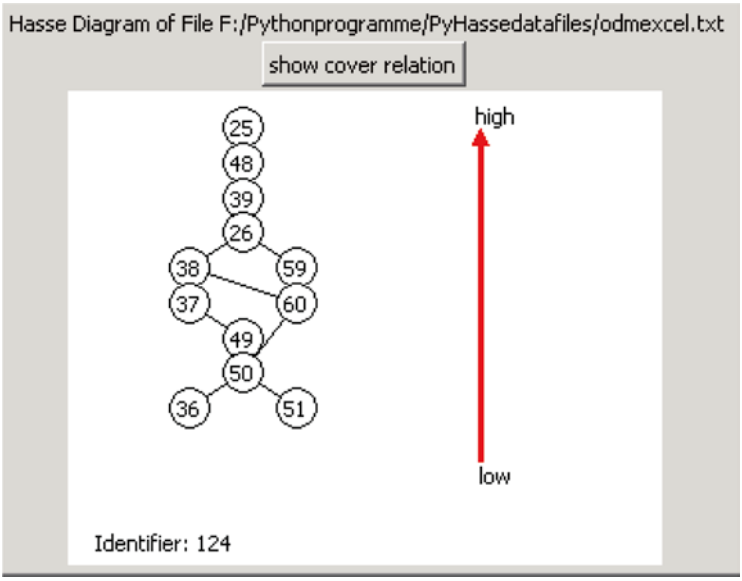


Fig. 2.15 Hasse diagram of CI's based on weight intervals [0.1, 0.3] for each of the five indicators

Returning to Fig. 2.10, the circles of that Hasse diagram contain the corresponding values of the CI(x) along with the object identifiers. Note that region 50 is covering 36, and there is an artificial overlap of the covering line with the circle representing object 49.

2.5.2.2 Weight Values as Intervals

Next, knowledge about weights is relaxed and an interval is assumed for each weight of [0.1, 0.3].

As the set of all CIs consistent with the weight intervals, the result shown in Fig. 2.15 is obtained whereby it is seen that even this narrow range of weight intervals leads to conflicts.

If the weight interval is further narrowed to [0.15, 0.25] for all five weights, a similar Hasse diagram is obtained with both objects 59 and 60 being quite "resistant" to change with regard to their incomparability.

2.5.3 Weight Evolution

Weighted composite indicators as well as a set of multiple indicators can be built up in a stepwise manner as follows:

Let $\Gamma(j)$ be the composite indicator obtained after calculating the weighted sum for indicators 1,…,j.

$$\Gamma_j(x) = \sum_{k=1}^{j} g_k \cdot q_k(x). \tag{2.11}$$

Let $|O(x,j)|$ be the number of elements in the down-set of x, taking the indicators $q_1,\ldots,q_j$ as multi-indicator system into account. Using a criterion of entry order such as impact on orderings, the sequence of indicators is fixed. Thus as the first indicator in (2.11) that indicator is selected having the largest, whereas as q_j the indicator with the relative least impact on the partial orders chosen. Then the evolution of the ordering is tracked in terms of $O(x)$ and/or $F(x)$ during the progression ($j=1,\ldots,m$). When this is done with the refined subset of land cover indicators, the following evolutionary behavior based on

$$\chi_j(x) = \frac{\Gamma_j(x)}{|O(x,j)|} \tag{2.12}$$

is observed, whereby we are differentiating pretty qualitatively between "smooth" behavior and "crucial" changes in χ:

Region-objects 25, 26, 36, 48 evolve smoothly.

Region-objects 37, 39 have a crucial step of marked change at entry of indicator Pct81.

Region-object 60 has a crucial step of marked change at entry of Pct32.

Region-objects 49, 50, 51 have crucial steps of marked change at entry of Pct81 and Pct82.

Region-object 59 has crucial steps of marked change for three indicators—Pct81, Pct82 and Pct32.

Pronounced shifts in patterns of progression are signposts of interesting interactions between objects and indicators. Critical steps usually indicate that a particular indicator has pronounced effects on certain objects. This is notably true in this excerpt from a larger dataset, since the Pct32 indicator is non-zero only for object 51. Thus, critical steps give clues to special subsets among the objects as seen through the lens of particular indicators.

2.6 Insights, Implications, and Inference

A multivariate dataset that is completely lacking in context should not be of much interest for any serious seeking of sequences. Formal statement of context for a dataset comes as metadata, and the metadata should be fully exploited in assessing prospects as a multi-indicator system (MIS). However, it is seldom that a multivariate dataset collected for some collateral or parallel purpose such as monitoring will serve well for direct use as an MIS. It will generally be necessary to engage in exploratory extraction of latent information on joint order that may reside in the dataset and that has been the topic of current consideration. There is a parallel with the subject of image enhancement wherein the purpose is to strengthen and bring to the foreground the imaging information that resides in digital image data, so it is hereby proposed to consider this undertaking as Joint Order Enhancement (JOE).

It has been repeatedly shown here that incomparability under partial order theory constitutes a useful criterion for pursuit of joint order enhancement, starting with joint orientation. A different scoping for the tally of incomparability enables assessment of joint effect attributable to each candidate indicator, as being either relatively minor or more major. Selecting the subset of (oriented) major ones will give a refined matrix of candidates for further exploration. That exploration can include probing groups of indicators taking into account the signals provided by the rank correlation matrix. It is only reasonable to couple statistical signals with protocols of partial order theory throughout the JOE endeavor.

It is an expedient aspect of exploration to have the capacity for examining order relations in overview, which the conventional Hasse diagram technique does not provide. Posetic features of "down-set" and "up-set" can be coupled with incomparability to provide an "FOU" space for exploring broader aspects of ordering, beginning with an FOU-plot. Objects positioned high on the ordinate exhibit considerable incomparability. Objects at the lower right are definitely prominent prospects in the order endeavor, and objects at the lower left are definitely subordinate. Thus, "marker" objects can be identified in the respective positions to serve as comparators for other objects that are less distinctive. The identifier of an object might be prefaced by one of the symbols ⌝ ⌜ ⌞ ⌟ to signify its incomparability and have as a trailing (subscript) one of the symbols →↔← to signify whether or not it has favorability for the balance of $O(x)$ and $F(x)$. With reference to the highlighted objects in Fig. 2.11, the symbology would be: ⌝59$_{\leftrightarrow}$, ⌟25$_{\rightarrow}$, ⌟36$_{\leftarrow}$. The coordinates of an FOU-plot also provide opportunity for progressive hierarchical clustering to suggest posetic proximity groups of objects.

Generalization of ordering can be induced by directly generalizing the scales of individual attributes, but considerable caution is warranted in this regard, and it should be done with full cognizance of any metadata that is available. The simplest version for multi-digit values (other than ranks) is rounding or truncation to effect generalization by an order of magnitude. This is particularly appropriate in

circumstances where intermediate computer processing has induced additional digits. Binning is intentional introduction of equivalence classes that may occasionally perturb transitivity for objects in terms of the raw data and is essentially declaration that the original scaling is inappropriate. Rank data can be binned, but it is then incumbent to rank the bins as the values to put forward. Commensurate generalization for different attributes is problematic for this approach, so each generalization of an individual attribute should be evaluated before proceeding to generalize another attribute.

Fuzzy concepts provide a different genre of generalization that entails greater sophistication and complexity. An experiment on actual data shows herein that a fuzzy approach can be informative with regard to joint order enhancement in a manner that is not readily accomplished otherwise.

Formulating and evaluating composite indicators obtained as weighted summations of original attributes provide a flexible and fertile avenue of joint order enhancement. The flexibility of weighting also allows exogenous information to be incorporated in the evaluation of order relations. Venturing further along such directions leads into the domain of comparative knowledge discovery.

It is all too easy when delving into order relations to treat a dataset as fully encompassing a context. If this were generally the situation, then there would have been no incentive to intellectualize regarding inferential statistics as an extension of descriptive statistics. Particularly when working with opportunistically acquired data, it is not usually a tenable assumption that a context is fully embedded within a dataset. This is patently not true for many settings of environmental sampling such as water, air, and soil. Replicate sampling quickly makes evident the almost instantaneous variability in time and space. To treat an order extracted from a set of sample data as being immutable truth is problematic at best and more often fundamentally deceptive. If replicate observations are not available by which to explore order variation in terms of sampling variation, then some recourse to distributional modeling of perturbations in the data is needed. Trying to cling tenaciously to every distinction of order that arises from nth digits is simply misguided.

As a minimum with regard to exploration of joint order enhancement, we would recommend:

(a) Investigating evidence of conflicting orientation.
(b) Investigating evidence of relative redundancy among indicators for consideration of culling.
(c) Using local partial order models as approximations for computing average height.
(d) Depict general features of comparability using posetic graphics such as FOU plots.

More sophisticated aspects can then be pursued as deemed appropriate, such as fuzzy analysis and weighted composite indicators.

References

Al-Sharrah G (2014) Ranking hazardous chemicals with a heuristic approach to reduce isolated objects in Hasse diagrams. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 10. Springer, New York, NY

Annoni P, Brüggemann R (2008) The dualistic approach of FCA: a further insight into Ontario Lake sediments. Chemosphere 70:2025–2031

Bartel HG, Mucha J (2014) Measures of incomparability and of inequality and their applications. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 3. Springer, New York, NY

Bock HH (1974) Automatische Klassifikation. Vandenhoeck & Ruprecht, Göttingen

Bock HH (1979) Clusteranalyse mit unscharfen Partitionen. In: Bock HH (ed) Studien zur Klassifikation, Bd 6 Klassifikation und Erkenntnis III - Numerische Klassifikation. Gesellschaft für Klassifikation, Frankfurt, pp 137–163

Bock HH, Gaul W, Schader M (2003) Studies in classification, data analysis, and knowledge organization. Springer, Berlin

Brüggemann R, Carlsen L (2011) An improved estimation of averaged ranks of partially orders. Match Commun Math Comput Chem 65:383–414

Brüggemann R, Patil GP (2010) Multicriteria prioritization and partial order in environmental sciences. Environ Ecol Stat 17:383–410

Brüggemann R, Patil GP (2011) Ranking and prioritization for multi-indicator systems – introduction to partial order applications. Springer, New York, NY

Brüggemann R, Voigt K (2008) Basic principles of Hasse diagram technique in chemistry. Comb Chem High Throughput Screen 11:756–769

Brüggemann R, Welzl G (2002) Order Theory Meets Statistics -Hassediagram technique-. In: Voigt K, 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, Halfon E, Welzl G, Voigt K, Steinberg C (2001) Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests. J Chem Inf Comp Sci 41:918–925

Brüggemann R, Kerber A, Restrepo G (2011) Ranking objects using fuzzy orders, with an application to refrigerants. Match Commun Math Comput Chem 66:581–603

Brüggemann R, Mucha HJ, Bartel HG (2013a) Ranking of polluted regions in South West Germany based on a multi-indicator system. Match Commun Math Comput Chem 69(2):433–462

Brüggemann R, Restrepo G, Voigt K, Annoni P (2013b) Weighting intervals and ranking, exemplified by leaching potential of pesticides. Match Commun Math Comput Chem 69(2):413–432

Brüggemann R, Carlsen L, Voigt K, Wieland R (2013c) PyHasse software for partial order analysis: scientific background and description of some modules. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 19. Springer, New York, NY

Carlsen L, Brüggemann R (2014) Indicator analyses, what is important – and for what? In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 18. Springer, New York, NY

De Loof K, De Meyer H, De Baets B (2006) Exploiting the lattice of ideals representation of a poset. Fundam Inform 71:309–321

Diday E (1979) Problems of clustering and recent advances. In: Bock HH (ed) Klassifikation und Erkenntnis III. Gesellschaft für Klassifikation, Frankfurt, pp 3–16

Luther B, Brüggemann R, Pudenz S (2000) An approach to combine cluster analysis with order theoretical tools in problems of environmental pollution. Match 42:119–143

Myers WL, Patil GP (2008) Semi-subordination sequences in multi-measure prioritization problems. In: Todeschini R, Pavan M (eds) Data handling in science and technology, vol 27. Elsevier, New York, NY, pp 161–170

Myers WL, Patil GP (2010) Preliminary prioritization based on partial order theory and R software for compositional complexes in landscape ecology, with applications to restoration, remediation, and enhancement. Environ Ecol Stat 17:411–436

Myers WL, Patil GP (2012) Multivariate methods of representing relations in R for prioritization purposes: selective scaling, comparative clustering, collective criteria and sequenced sets. Springer, New York, NY

Myers WL, Patil GP (2013a) Coordination of contrariety and ambiguity in comparative compositional contexts: balance of normalized definitive status in multi-indicator systems. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 8. Springer, New York, NY

Myers WL, Patil GP (2013b) Higher-order indicator with rank-related clustering in multi-indicator systems. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 16. Springer, New York, NY

Myers WL, Patil GP (2013c) Statistical geoinformatics for human environment interface. CRC/Taylor & Francis, Boca Raton, FL

Myers WL, Patil GP, Cai Y (2006) Exploring patterns of habitat diversity across landscapes using partial ordering. In: Brüggemann R, Carlsen L (eds) Partial order in environmental sciences and chemistry. Springer, Berlin, pp 309–325

Patil GP (2012) Keynote lecture on partial orders and composite indicators for multivariate ranking in multivariate nonparametric statistics at the international workshop on partial order theory and modeling held in Berlin, Germany

Patil GP, Joshi S (2014) Comparative knowledge discovery with partial order and composite indicator. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 6. Springer, New York, NY

Patil GP, Taillie C (2004) Multiple indicators, partially ordered sets, and linear extensions: multicriterion ranking and prioritization. Environ Ecol Stat 11:199–228

Rocco C, Tarantola S (2014) Evaluating ranking robustness in multi-indicator uncertain matrices: an application based on simulation and global sensitivity analysis. In: Brüggemann R, Carlsen L, Wittmann J (eds) Multi-indicator systems and modelling in partial order, Chap 13. Springer, New York, NY

Trotter WT (1992) Combinatorics and partially ordered sets dimension theory. The Johns Hopkins University Press, Baltimore, MD

Van de Walle B, De Baets B, Kersebaum KC (1995) Fuzzy multi-criteria analysis of cutting techniques in a nuclear dismantling project. Fuzzy Set Syst 74:115–126



http://www.springer.com/978-1-4614-8222-2

Multi-indicator Systems and Modelling in Partial Order
Bruggemann, R.; Carlsen, L.; Wittmann, J. (Eds.)
2014, XIX, 437 p. 139 illus., 60 illus. in color., Hardcover
ISBN: 978-1-4614-8222-2