

# Exploring Partial Order of European Countries

Paola Annoni · Rainer Brüggemann

Accepted: 20 July 2008 / Published online: 27 August 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Partial Order Theory has been recently more and more employed in applied science to overcome the intrinsic disadvantage hidden in aggregation, if a multiple attribute system is available. Despite its numerous positive features, there are many practical cases where the interpretation of the partial order can be rather troublesome. In these cases the analysis of underlying dimensions could be useful to uncover particular data structures. The paper shows a way of addressing the problem with the help of an actual case study, which deals with European opinions on services of general interest. In particular, a partial order of countries is firstly provided and then a method to detect dimensions is discussed and applied. The analysis stems directly from the Partially Order Set (poset) and Lattice theory with particular references to dimension theory and Formal Concept Analysis. The study is eventually able to pinpoint role and relevance of different attributes characterizing EU countries which are used to define the partial order.

**Keywords** Hasse diagrams · Dimension theory in posets · Planar lattices · Formal Concept Analysis · Association rules · Public services

## 1 Introduction

The study of underlying dimensions, components or factors is by all means a milestone in multivariate statistics. The usual way to tackle the issue consists in some kind of data reduction to summarize only relevant information. This way of doing often implies the setting-up of synthetic scores, one for each reduced dimension, which should simplify data interpretation for example by means of linear and non-linear multivariate analysis (Gifi 1990).

---

P. Annoni (✉)

Department of Economics, Business and Statistics, University of Milan, via Conservatorio 7, 20122  
Milano, Italy  
e-mail: paola.annoni@unimi.it

R. Brüggemann

Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin,  
Germany  
e-mail: brg\_home@web.de

The approach is here completely the opposite. Starting from a case study, which involves European countries and a suite of items (attributes) which describes citizens' opinion on major services of general interest, a multiple attribute technique is adopted to compare countries keeping attributes separated. Multiple comparison may be considered as the counterpart of linear ranking. The standard approach is to aggregate attributes, for setting-up indicators from which a linear ranking may be obtained, or to perform an outranking (Brans and Vincke 1985; Lansdowne 1997; Vincke 1999). In spite of its simple output, linear ranking has a crucial disadvantage, recently pointed out in many scientific contributions: synthetic indicators are indeed frequently based on arbitrary choices which are rarely justified (Myers et al. 2005; Patil and Taille 2004; Voigt et al. 2004; Brüggemann et al. 2001, 2004). The focus of this paper is not to provide indicators but to introduce a rather novel methodology which might be of practical use when dealing with comparison and evaluation in social sciences. Our approach is based on simple elements of partial order theory which keep attributes separated and allows for detecting eventually conflicting attributes.

For the case under study, countries are the objects to be ordered due to their attribute values; although there is not necessarily a linear order we speak of "ranking" in a broader sense. Once multiple prioritization of countries is gained, underlying dimensions are detected without any data reduction and avoiding information loss that is due to setting-up of a single highly aggregated numerical value. Instead useful results are obtained due to the general theory of dimensions in partially ordered sets, posets, and lattices, which are a particular form of posets (Dushnik and Miller 1941; Trotter 1975, 1992).

In addition to this type of poset analysis, which assigns an asymmetric role to objects, namely objects are ordered due to their attributes ("objects order"), a similar still different approach is adopted: Formal Concept Analysis, FCA (Ganter and Wille 1999; Carpineto and Romano 2004). In FCA each element of the pair {objects; attributes} plays a symmetric role, namely an object and attribute order. Therefore FCA represents a formalism for exploring hierarchies for correlations, similarities, anomalies and also inconsistencies from the point of view of both objects and attributes.

FCA has been recently adopted to support various kind of tasks using different and heterogeneous types of data, in the field of information retrieval, text mining, rule mining or environmental data analysis (Annoni and Brüggemann 2008; Pudenz et al. 2002; Bartel and Brüggemann 1998; Bartel and Nofz 1997; Brüggemann et al. 1997; Brüggemann and Voigt 1995). For the case under study, FCA helps in the classification of countries and attributes being guided by their mutual relationships.

In comparison to classical approaches, which involve scaling and dimensionality reduction, the methodologies adopted here have two main features: they are metric-free and parametric free.

First:

- (a) Since we start from ordinal variables we are exploiting only ordinal properties of variables without imposing a ratio or metric scale on them. Then partial order methods are of special value when quantitative variables are not available, as it often occurs in social sciences.
- (b) Many statistical methods may be used that provide optimal quantifications of categories for qualitative variables, such for example Homogeneity analysis for nominal and Non-linear PCA for ordinal variables. On the one side they allow for numerical evaluation, on the other their outcome is not always statistically robust. Most of them are based on the minimization of a loss function solved by ALS (Alternating Least Square) algorithms which can be sensitive to low marginal frequencies, as recently showed by

some empirical studies (Annoni 2007; Linting et al. 2007). Partial order methods do not present this kind of pitfall since they are not based on scaling.

Second:

The methodology does not need any distribution assumptions making it suitable for addressing practical problems which lack distributional knowledge.

The reminder of the paper is laid out as follows. Section 2 is aimed at the description of the case study and the original data-set. The knowledge already gained about the data structure, arising from previous studies, is introduced and serves here as a new starting point. In Sect. 3 a sketch of partially ordered sets and lattices is outlined in order to better understand the results obtained for the case study. Similarly, Sect. 4 describes both FCA basic theory and its application to the present setting. Finally, discussion and conclusion are provided in Sect. 5.

## 2 Case Study and Prior Knowledge About Data

The case study stems from an European official survey which regards social and political aspects of citizens' every-day life. In particular, the data-set is the Eurobarometer EB58 survey carried out between September and October 2002, on behalf of the European Opinion Research Group (Eurobarometer 58.0 2002), which was already object of investigation by one of the authors.

The sample is composed by about sixteen thousands people from EU15 member states before 2004 enlargement and it covers the population aged 15 years and over. Sampling design and representativeness of respondents are under control of *Health and Consumer Protection General Directorate* which regularly commissions surveys within EU members.

Among various aspects of every-day life, the survey deals with consumers' opinions about eight major services of general interest (in parentheses abbreviations used in the paper): (1) mobile telephone services (mob phone) (2) fixed telephone services (fix phone) (3) electricity supply services (electricity) (4) gas supply services (gas) (5) water supply services (water) (6) postal services (postal) (7) transport services within towns/cities (transport) (8) rail services between towns/cities (rail). Five criteria are used to analyze each service: (a) access easiness to service (access) (b) price of the service (price) (c) quality of the service (quality) (d) clarity of the information aimed at consumers by service customer care (info) (e) fairness of terms and conditions of service contracts (contract).

Since a previous investigation was performed on the same data-set (Annoni 2007), prior knowledge is available. In particular two outcomes are here considered as the starting point of the study and simplify the analysis: one is related to services and the other is related to criteria. From the point of view of services, mobile phone and gas supply showed a particular behavior and are both discarded from further analysis. Variables related to mobile phone turned out to highly influence country multiple ranking.<sup>1</sup> This could be due to the fact that the performance of this kind of utility depend on providers, which in this field are highly globalized. The service features could hence reflect the particular provider the respondent has chosen with no dependence on the country. Variables related to this service are then likely to reflect other factors which do not allow for a clear country classification. With regard to gas supply, it is unevenly spread within different countries thus leading to comparability problems across countries which may affect the analysis in

---

<sup>1</sup> The influence was quantified in terms of  $W$  matrix of Hasse diagrams (Annoni 2007; Brüggemann et al. 2001).

**Table 1** Denomination of investigated variables and their categories

Criterion	Variable name	Allowed categories
Price	Price: Fix phone	Fair; neutral; unfair; excessive
	Price: Electricity	
	Price: Water	
	Price: Postal	
	Price: Transport	
	Price: Rail	
Quality	Quality: Fix phone	Very good; fairly good; neutral; fairly bad; very bad
	Quality: Electricity	
	Quality: Water	
	Quality: Postal	
	Quality: Transport	
	Quality: Rail	
Contract	Contract: Fix phone	Fair; neutral; unfair
	Contract: Electricity	
	Contract: Water	
	Contract: Postal	
	Contract: Transport	
	Contract: Rail	

various way. Moreover, from the point of view of criteria, simple descriptive analysis showed that some criteria are non-informative, with almost null variance of distribution of scores (Annoni 2007). Criteria with relative highest variability are price, quality and clearness of contract and are then more deeply analyzed in the study.

In summary, prior analysis suggested to keep a reduced data-set composed by three criteria and six services, for a total of eighteen items. These items play the role of variables of the analysis. All variables can assume a specified number of categories which ranges from two to four categories, of the type 'fair', 'unfair', 'excessive', etc. Two more categories are provided in addition: 'don't know' and 'no access/not applicable'. The jointly presence of these two supplementary categories suggested to undertake a recoding: category 'don't know' has been considered as a neutral opinion, also according to literature indications (Johnson and Albert 1999), whilst 'no access/not applicable' has been considered as a missing value, since it represents a situation where the respondent cannot express his opinion. The recoding procedure comes out with transformed variables following the Likert scale (Smith and Smith 2004), with scale intervals ranging from three to five points, all respecting the same polarity with regard to each criterion: the higher the score the lower the level of satisfaction declared by the interviewed.

Selected variables are shown in Table 1. Their denomination is intended to be explicit: the first part refers to the criterion, the last part refers to the service. So as for example the variable 'price: fix phone' is the one which records satisfaction opinions of citizens regarding the perceived fairness of price of fixed telephone.

### 3 Prioritization and Dimension Analysis by Partial Order Theory

#### 3.1 Hasse Diagrams

The Hasse diagram technique HDT has been recently used to perform multi-criteria assessment in a wide range of fields of applications (Brüggemann and Carlsen 2006). The technique can be applied whenever a set of objects is described by a set of items, attributes

or properties. The set of items is called the ‘information basis’ (IB) of the comparative evaluation. Given the IB, each object is assigned a string of values according to the items used.

Hasse diagrams are used to visualize a partially ordered set or poset  $\mathbf{P}$  which is defined as a pair  $(X, P)$  where  $X$  is a set, called the ground set, and  $P$  is a reflexive, transitive and antisymmetric binary relation on  $X$  ( $\leq$ ), which is called partial order on  $X$ .

The general idea behind HDT is to avoid numerical aggregation of attributes; as a consequence comparability and incomparability among objects appear and are indeed of primary interest. Correspondingly a short list of definitions is needed in order to introduce HDT in more technical forms. The notation  $x < y$  means  $x \leq y$  and  $x \neq y$ , with  $x, y \in X$ . If  $x$  and  $y$  are two elements of  $X$ , they are said comparable ( $x \perp y$ ) when either  $x < y$  or  $y < x$ . In the case ‘ $x < y$ ’ no attribute value observed on  $y$  is less than that observed on  $x$ , if relation ‘ $<$ ’ is assumed to describe increasing attribute values. On the contrary  $x$  and  $y$  (with  $x$  unequal  $y$ ) are said incomparable ( $x \parallel y$ ) if neither  $x < y$  nor  $y < x$ . The incomparability relation  $x \parallel y$  means that at least one attribute value observed on  $y$  is greater than that on  $x$  and, at the same time, that there is another attribute where the value on  $y$  is less than that on  $x$ . Moreover, the element  $x$  is said to be covered by  $y$  ( $x \prec y$ ) when  $x < y$  and there is no point  $z \in X$  for which  $x < z$  and  $z < y$ . An element  $x \in X$  is called maximal (minimal) element if there is no element  $y \in X$  with  $x < y$  ( $x > y$ ) (Trotter 1992).

A very simple example of poset  $\mathbf{P}$  is a family of sets partially ordered by inclusion, a type of poset that will be shortly useful in the definition of poset dimensions (Sect. 3.3).

Hasse diagrams are graphical representation of posets. In an Hasse diagram the elements of  $X$  are drawn as small circles which are connected by straight line segments if and only if a cover relation exists between the two end points elements. Only cover relations are represented since the binary relation  $P$  is transitive. In the two-dimensional graph upper points are conventionally larger than lower points, in terms of  $P$ . In contrast to FCA, explained below (Sect. 4), the HDT represents an asymmetric approach stressing relations only among the objects.

Since many former publications (Brüggemann et al. 2003, 2004; Myers et al. 2005; Patil and Taille 2004; Voigt et al. 2004; Sørensen et al. 2003; Carlsen and Walker 2003, 2006) extensively illustrate HDT theory and applications, no further information will be provided about the technique.

As already stated, our goal is here to prioritize countries on the basis of European citizens opinion regarding services of general interest. Each country is assigned a score to synthetically describe citizens opinion about each criterion describing each service under study.

The original data-set is a matrix  $16.000 \times 18$  with citizens opinion on the rows and variables describing service performance on columns. We then need to summarize data assigning scores to each country that properly describe the average citizens opinion for each aspect of investigated services. To that purpose, the conditioned median of each variable in the original data-set is computed, with country as conditioning variable. The median has the feature of simply synthesizing the distribution coherently with the ordinal measurement scale of the original variables. The collection of all medians, one for each criterion for each service, is to be intended as the IB of the study. The starting matrix for HDT analysis is then a  $15 \times 18$  item matrix with European countries on rows and items on columns. Table 2 reports such scores, which are the medians of the distributions of satisfaction opinion, expressed as literal values like ‘FG’, ‘VG’, etc., separately for each variable.

**Table 2** Medians of distribution of satisfaction scores for each service and each criterion (Codes of observed categories: F = ‘fair’; U = ‘unfair’; N = ‘neutral’; FG = ‘fairly good’; VG = ‘very good’)

	B	DK	GR	I	E	F	IRL	L	NL	P	FIN	S	A	D	UK
Price: fix phone	F	F	U	U	U	N	F	F	F	U	F	F	F	F	F
Price: electricity	F	F	U	U	N	F	F	F	F	U	F	F	F	F	F
Price: water	F	F	F	N	F	U	F	F	F	F	F	F	F	N	F
Price: postal	F	F	F	F	F	F	F	F	F	F	F	F	F	N	F
Price: transport	F	F	F	N	F	N	F	F	U	N	N	F	N	U	F
Price: rail	F	F	F	U	N	N	F	F	U	N	N	N	U	U	N
Quality: fix phone	FG	VG	FG	FG	FG	FG	VG	VG	FG	FG	FG	VG	FG	FG	FG
Quality: electricity	FG	VG	FG	FG	FG	FG	VG	VG	FG	FG	FG	VG	VG	FG	VG
Quality: water	FG	VG	FG	FG	FG	FG	VG	VG	FG	FG	FG	VG	VG	FG	FG
Quality: postal	FG	VG	FG	FG	FG	FG	VG	FG	FG	FG	FG	FG	FG	FG	FG
Quality: transport	FG	FG	FG	FG	FG	FG	FG	FG	FG	FG	FG	FG	FG	FG	FG
Quality: rail	FG	FG	FG	N	FG	FG	FG	FG	N	FG	FG	FG	FG	FG	FG
Contract: fix phone	F	F	F	U	N	F	F	F	F	F	F	F	F	F	F
Contract: electricity	F	F	F	N	F	F	F	F	F	F	F	F	F	F	F
Contract: water	F	F	F	N	F	F	F	F	F	F	F	F	F	F	F
Contract: postal	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Contract: transport	F	F	F	N	F	F	F	F	F	F	F	F	F	F	F
Contract: rail	F	F	F	N	F	F	F	F	F	F	F	F	F	F	F

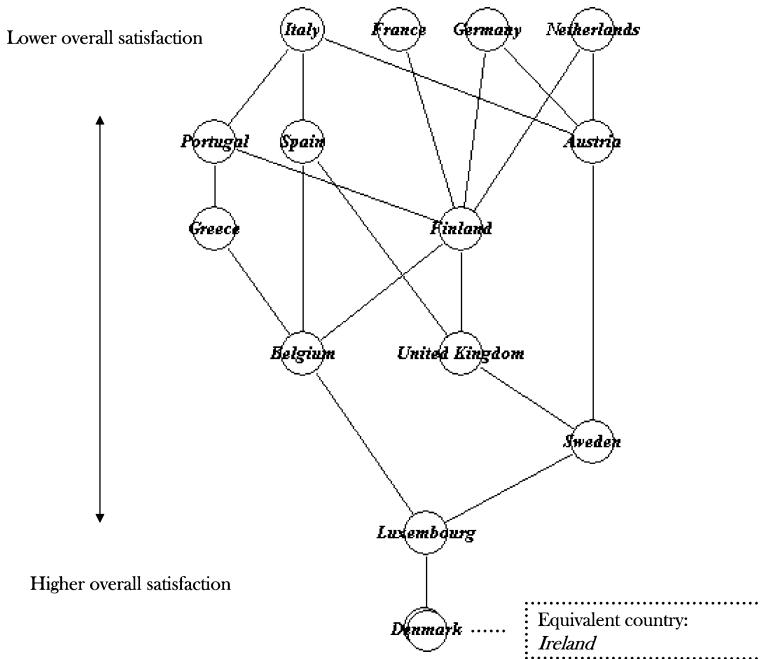
To summarize, the case study is composed by the following elements: (1) the set of country profiles  $X$ , that is the rows of the item matrix, and (2) the partial order  $P$  on  $X$ , that is the binary relation which stems naturally from the ordinal scale of items.

Hasse diagram associated to the  $15 \times 18$  item matrix is shown in Fig. 1. Sequences of comparable countries (chains) and of incomparable countries (antichains) are recognizable. For example Italy forms a chain with E, GR, B, L and DK, thus indicating that, for all variables, scores for Italy are not lower than those for Spain (with a strict relation for at least one variable); all scores for Spain, forming its profile, are not lower than those for Greece; and so on. Thus the top element of a chain (maximal element) is a country where, on average, citizens are less satisfied than citizens of countries located downward along the chain itself. An antichain includes instead countries which are not comparable with each other, thus meaning that for these countries some scores are less while some score are higher than the ones of another country of the antichain. As an instance, Italy forms an antichain with NL, F and D, which are all maximal countries. At the other end of the picture, DK and IRL are two minimal equivalent countries. Note that we consider a poset based on a set of representatives, taken from any equivalence class under the equivalence relation ‘equality’.

There are several ways in HDT to allow for a better interpretation of the data structure, here in this paper a dimension analysis is at this point performed.

### 3.2 Hierarchical Partial Ordering

A strategy similar to hierarchical partial ordering (Carlsen 2007) is adopted to get an in-depth understanding of data structure. Due to the intrinsic nature of the IB, items can be naturally ‘clustered’ by services or by criteria. The second approach is adopted since the



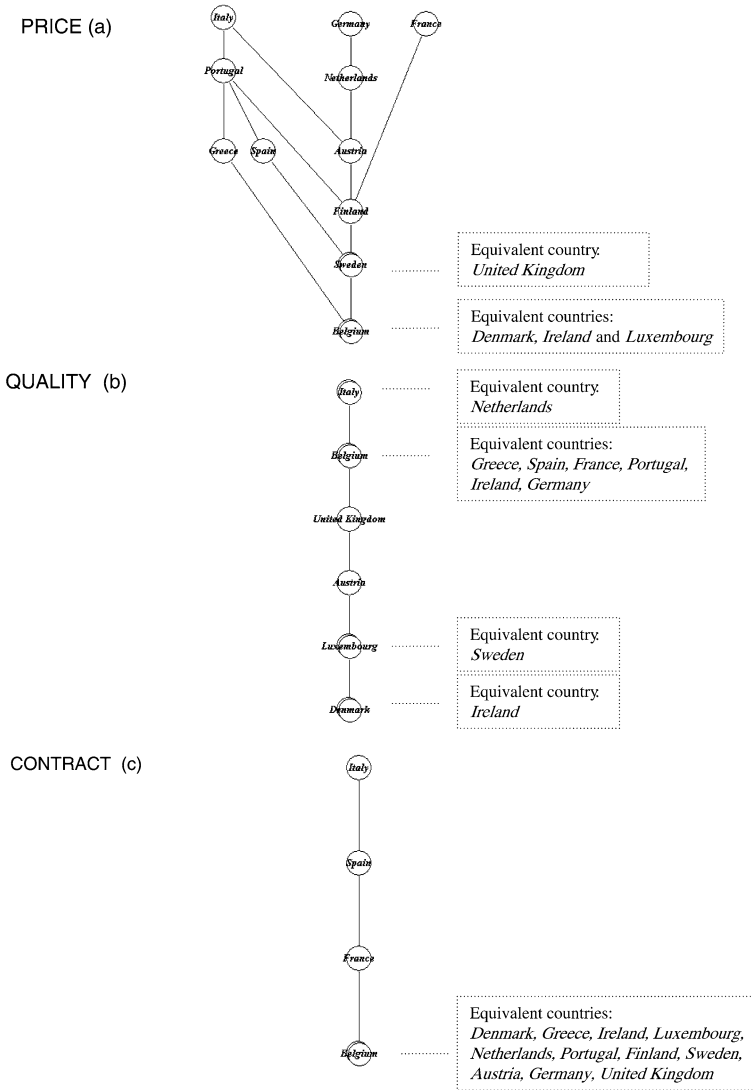
**Fig. 1** Hasse diagram of 15 European countries ranked with 18 criteria

goal is to get a country ranking on the basis of the performance of all investigated services and not separately for each service.

The original  $15 \times 18$  item matrix is then split into three sub-matrices each of dimension  $15 \times 6$ , featuring only one criterion at a time. The resulting three separate Hasse diagrams are visualized in Fig. 2 for price, quality and fairness of contract.

It is noticeable that *Quality* and *Contract* criteria yield linear rankings. However the two linear rankings are not coincident, meaning that same countries are not always at the same position in both rankings. The intersection of the two graphs due to *Quality* and *Contract* is composed by Italy on one hand, with the worst satisfaction, and Denmark and Ireland on the other hand, with the best satisfaction for both quality and contract. It can be noted a wide clustering of countries on high levels of satisfaction on service contract Fig. 2c and a more stretched situation for quality criterion Fig. 2b. This indicates a more articulated situation for quality which is likely (and reasonably) perceived as predominant criterion with respect to contract, causing a more heterogeneous fan of opinions. Finally, the presence of linear orders for quality or contract means that, for each country, a certain rank for one service is not in contradiction with the rank of any other service with respect to quality or contract. For example, France is located between Belgium and Spain in the contract Hasse diagram independently which service is considered under the contract criterion. With other words, a higher or lower satisfaction within a country does not depend on the particular service but only on the criterion under analysis.<sup>2</sup>

<sup>2</sup> On the contrary if you simultaneously consider quality and contract then conflicts would arise only due to these different criteria jointly considered. In this particular case, only Spain and Netherlands come out incomparable since Spain is worse than Netherlands for fix phone contract while it is better than Netherlands for rail quality.



**Fig. 2** Three separate Hasse diagrams of 15 European countries ranked separately by (a) price, (b) quality and (c) contract

Separate discussion is due to *Price* criterion (Fig. 2a), which shows a more varied situation and an effect of the type of service on citizens opinion. Many considerations can be drawn about this diagram directly from the lattice theory. This mathematical branch is widely described in the literature. A recent description could be found, for example, in Trotter (1992). In the following, a brief summary of dimension theory for lattices is provided together with those definitions and results which helped us in deepening the analysis of the poset for price.



### 3.3 Poset Dimensions and Planar Lattices

The concept of dimension in posets was formerly introduced by Dushnik and Miller back in 1941 (Dushnik and Miller 1941) and a more recent overview is given by Trotter (Trotter 1992). In our context the very idea is to find out another and smaller set of (latent) attributes to reconstruct the original poset. Hence the interpretation may be simplified.

In technical terms the dimension of a poset basically refers to the existence of minimal sets of linear extensions for a poset. Linear extensions are formally defined as follows. Let  $P$  and  $Q$  be two partial orders on the same set  $X$ , then  $Q$  is called an *extension* of  $P$  if, for all  $x, y \in X$ ,  $x \leq y$  in  $P$  implies  $x \leq y$  in  $Q$ . The pair  $\mathbf{P} = \{(\text{all extensions of } P); (\text{inclusion})\}$  is a poset itself where maximal elements are linear orders on  $X$  called linear extensions  $L$  of  $P$ . The set of all linear extensions of  $P$  is denoted by  $\Xi(P)$ . In less formalized terms, linear extensions of  $P$  can be thought as extensions of  $P$  which contain all the elements of  $X$  which can be represented by linear Hasse diagrams.

Now we have every element to introduce the dimension of a poset  $\mathbf{P} = (X, P)$ ,  $\dim(X, P)$ , that is defined as the least positive integer  $t$  for which there exists a family  $\{L_1, L_2, \dots, L_t\}$  of linear extensions of  $P$  so that  $P = \bigcap_{i=1}^t L_i$  (Trotter 1992). Clearly, a poset has dimension one if and only if its Hasse diagram is linear, i.e. if it is a chain.

Finding out the dimension of a poset is useful for the identification of different attributes which determine the structure of its Hasse diagram. Our goal is here to determine the dimension of the poset obtained by price criterion for fifteen countries shown in Fig. 2a, which from now on will be called poset  $\mathbf{P}_{price}$ .

At this purpose let us introduce a special case of poset, the lattice. A lattice is a poset with rich mathematical properties, therefore many more knowledge is available which may help in interpretation. Examples refer to dimension extraction, efficient visualization and elements of artificial intelligence (Brüggenmann and Voigt 1995; Ganter and Wille 1999; Trotter 1992).

Introducing lattices technically, let  $\mathbf{P} = (X, P)$  be a poset and  $S \subseteq X$ , then an element  $y \in X$  is called an upper (lower) bound for  $S$  if  $s \leq y$  ( $s \geq y$ ) for every  $s \in S$ . The difference between an upper (lower) bound and a maximal (minimal) element is that the former are defined on a subset of  $X$  whilst the latter is defined on the whole set  $X$ . A poset  $\mathbf{P}$  is called a lattice if every nonempty subsets  $S \subseteq X$  has both a least upper bound and a greatest lower bound. The least element for  $X$  is generally called zero and the greatest element one. Dimension theory about lattices has been very well developed and it is used here to find out the dimension of poset  $\mathbf{P}_{price}$ , without exploiting directly the definition of dimension which is generally a very tedious procedure.

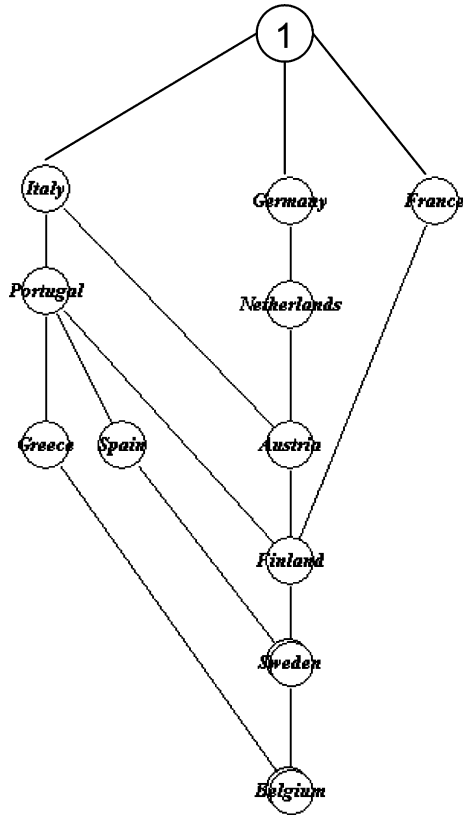
By analyzing the Hasse diagram of  $\mathbf{P}_{price}$  two remarks can be noted. Firstly there is a natural way to associate  $\mathbf{P}_{price}$  to a lattice containing  $\mathbf{P}_{price}$ : since  $\mathbf{P}_{price}$  does not have a greatest element, albeit it has the lowest element, it is possible to add a new point to serve as greatest element. Let's denote this modified poset with  $\mathbf{P}'_{price}$ . It is now easy to show that this is a lattice (see Fig. 3).

The lattice  $\mathbf{P}'_{price}$  has the further feature of being planar, that is it can be drawn without edge crossing. This is an fundamental property since many results have been achieved for planar lattices.

From the following theorems, reported by Trotter, can be deduced that the poset  $\mathbf{P}_{price}$  is two-dimensional:

**Theorem 1** (Trotter 1992) *Let  $\mathbf{P} = (X, P)$  be a lattice. Then  $\mathbf{P}$  is planar if and only if  $\dim(\mathbf{P}) \leq 2$ .*

**Fig. 3** Modified poset  $\mathbf{P}_{price}$  which becomes a lattice



**Theorem 2** (Trotter 1975): For any posets  $\mathbf{P} = (X, P)$  and any point  $x \in X$ , the following inequality holds<sup>3</sup>:

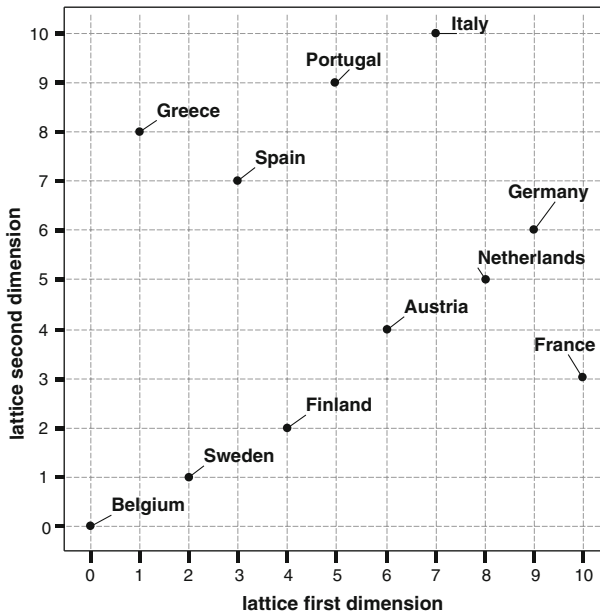
$$\dim(X - x) \leq \dim X.$$

From Theorem 1 it can be claimed that every planar lattice has dimension not higher than two. Since mono-dimensional posets are only those which are perfectly linear (chain), the lattice of Fig. 3 has dimension two. Furthermore, if we define  $X$  as the ground set of  $\mathbf{P}'_{price}$  and if  $x \in X$  is the greatest element added to get a lattice, then the ground set of poset  $\mathbf{P}_{price}$  is  $(X - x)$  and from Theorem 2 it follows:

$$2 \leq \dim(X - x) \leq \dim(X) = 2 \Rightarrow \dim(X - x) = 2$$

If  $\mathbf{P}_{price}$  is two-dimensional, it is possible to draw its perfect representation in the Euclidean space (Fig. 4). The spatial location of points of Fig. 4 has been built following the principles of Partially Ordered Scalogram Analysis with Coordinates, POSAC (Shye et al. 1994; Borg and Shye 1995). For each point in planar POSAC representation, points located in the lower left-hand quadrant with respect to that point are lower than the point itself in terms of the partial order. On the other hand, points in the upper right-hand region

<sup>3</sup> The statement of the theorem is wider, here only some results, useful to our specific case, are reported.



**Fig. 4** Perfect representation of two-dimensional lattice  $P_{price}$

are higher than the point in the same sense. Points in the remaining quadrants (upper left and lower right) are not comparable to the reference point.

Apart from the group of best countries (B, S and FIN together with equivalent countries: DK, IRL, L and UK), where citizens do not complain about price, a deeper look at the scores of remaining countries allows for the identification of two groups: one is composed by more traditionally Latin and Mediterranean countries (GR, E, P, I) where people discontentment is for price of fixed phone and electricity, the other is composed by more continental countries (A, F, NL and D), where the discontentment is focused on water, transport and rail services.<sup>4</sup> It then seems that two latent dimensions govern the scatter plot: one is mainly composed by fixed phone and electricity and the other is mainly composed by transportation services.

We have seen that when a poset is a lattice data interpretation is generally simplified. To make it plain, a lattice is indeed a structured poset in the sense that its Hasse representation has a graph-theoretical representation, where from every point there exists a path which leads either to the upper most or to the lower most point, respectively one and zero. Unlikely, the simple mathematical definition of HDT may often lead to graphs with several disaggregated and disconnected parts (Brüggemann et al. 1997). In these cases data interpretation is obviously rather troublesome. A possible way to tackle this issue can be found in Formal Concept Analysis, a method belonging to family of partial order theory methods, which for its nature always leads to a lattice. In the following the description of FCA results for the price criterion is provided.

<sup>4</sup> Price of postal service is a very low discriminating item.

## 4 Formal Concept Analysis

The ‘price’ criterion shows the highest variability. For a further insight into the order structure induced by ‘price’ criterion, another instrument is applied. It stems directly from partial order theory, namely Formal Concept Analysis, FCA (Ganter and Wille 1999; Carpineto and Romano 2004).

FCA is not based on objects and attributes as in HDT but on properly defined pairs of object and attribute subsets, called concepts. In contrast to HDT the order relation of FCA lattices is the set inclusion among the concepts.

From the mathematical point of view the starting point of FCA is the triplet: context, concept and concept lattice.

A context is a triple  $(O, A, I)$  where  $O$  and  $A$  are two sets, which contain respectively objects and attributes as elements, and  $I$  is a relation between  $O$  and  $A$ . If an object  $o \in O$  has the attribute  $a \in A$  it is written  $(o, a) \in I$ . The set of attributes associated to an object can be thought of as a binary vector with zeros indicating an ‘off’ status and ones indicating an ‘on’ status for corresponding attributes. A context is generally arranged in a table called context table which, from the statistical point of view, is an indicator matrix.

For the case study under investigation, the set  $O$  includes the fifteen countries, while the set of attributes  $A$  is specifically defined as follows. Taken each service separately, the distribution of relative frequencies of category ‘fair’ for all the fifteen countries is computed and countries with frequency below the first quartile are flagged as unsatisfied with regard to the corresponding service. The procedure gives rise to the context table shown in Fig. 5, where a cross in cell  $(i, j)$  indicates that country  $i$  belongs to the set of least satisfied countries for service  $j$ . This scaling fixes the view we wish to look at our data and, in particular, it focuses on every type of relations among countries and attributes. As stated by Wolff (1993), albeit “at first glance this scaling procedure might look as a rough description of original data, it is indeed a very powerful way to represent the data with respect to a particular view the analyst is interested in”.

Once a context table is defined, FCA looks for every possible concept implied in the table itself. A (formal) concept refers to an abstract entity dually specified by the suite of attributes needed to define it and the set of objects to which the concept applies. The formal concept is a key entity in FCA and it gives its name to the whole procedure. A

**Fig. 5** Context of the ‘fair’ category for the price criterion

country	fix phone	electricity	water	postal	transport	rail
Belgium						
Denmark						
Greece	X	X				
Italy	X	X	X	X	X	X
Spain	X	X				
France			X			
Ireland						
Luxembourg						
Netherlands					X	X
Portugal	X	X	X			
Finland						
Sweden				X		
Austria				X	X	X
Germany				X	X	X
United Kingdom						

concept can be formally defined as a pair  $C \equiv (Oo, Aa)$ ,  $Oo \subseteq O$  and  $Aa \subseteq A$ , where  $Oo$  contains all the objects that belong to the concept and  $Aa$  contains all the attributes common to objects in  $Oo$ . In FCA language  $Oo$  is called concept extent and  $Aa$  is called concept intent. Note that to be a concept the relation between extent and intent has to be bijective: an extent is related to an intent and viceversa an intent is related only to that extent. For this reason not all possible subsets of objects (or attributes) can be the extent (intent) of some concept.

For instance, in the case of context table under analysis (Fig. 5), the pair  $(\{GR, I, E, P\}; \{\text{fix phone, electricity}\})$  is a concept since all four countries share discontentment regarding fix phone and electricity and, vice versa, fix phone and electricity have a low performance only in those four countries. On the contrary,  $(\{I, S\}; \{\text{postal}\})$  is not a concept since the attribute 'postal' is shared also by Austria and Germany.

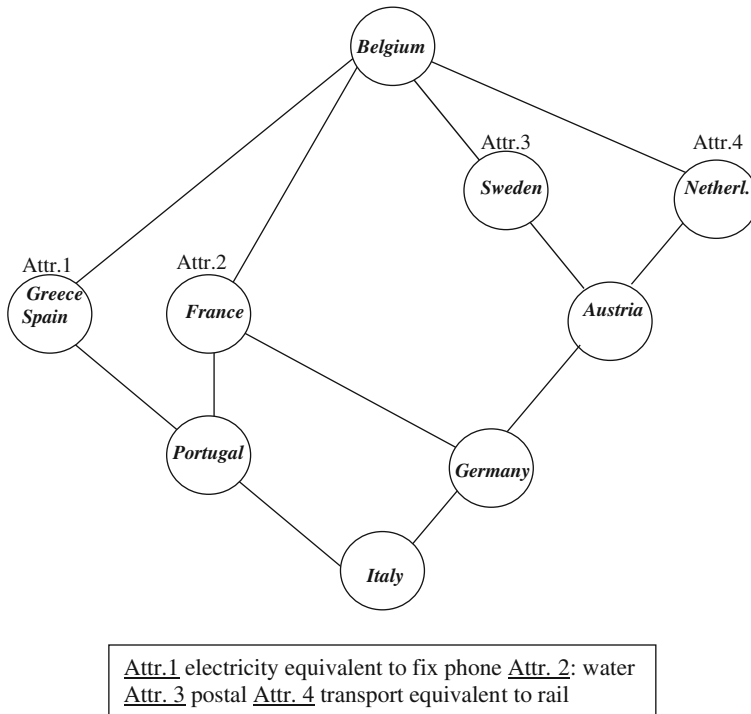
The set of all the concepts of the context  $(O, A, I)$  is usually denoted by  $\mathcal{C}(O, A, I)$  and can be assigned an order relation, namely the inclusion relation: if  $C_1 \in \mathcal{C}(O, A, I)$  and  $C_2 \in \mathcal{C}(O, A, I)$  are two concepts,  $C_1$  is said a subconcept of  $C_2$  if its extent is included in the extent of  $C_2$  or, equivalently, if its intent comprises the intent of  $C_2$ . In this case it is written  $C_1 \leq C_2$ . The ordered set  $\{\mathcal{C}(O, A, I); \leq\}$  is a lattice (as was shown by Ganter and Wille 1999) and is called the 'concept lattice'. The concept lattice is a lattice which represents order relations of all concepts of the context. In a concept lattice, concepts are represented as circles and the following reading rule applies: the extent and the intent of each concept are read by collecting all objects below and, respectively, all attributes above the concept circle. This practical rule stems from a fundamental mathematical theorem on concept lattices (Wille 1982).

Figure 6 shows the concept lattice of the price criterion associated to the context table given in Fig. 5. The group of 'price-satisfied' countries is the one with Belgium as representative country and it comprises: Belgium, Denmark, Ireland, Luxembourg, Finland, United Kingdom. Belgium is in fact located as the top element thus indicating that it does not share any attribute. It is the group of those countries without mark in the context table of Fig. 5. Apart from United Kingdom, which in the poset analysis is considered as equivalent to Sweden and Finland, the group of satisfied countries coincides with the one previously detected (Fig. 2a).

On the opposite side, Italy is located as bottom object thus sharing all the attributes (alas!). This means that people are unsatisfied about the price of all the services or, in other words, Italians are the least satisfied among 15 countries. Again this results is in agreement with previous analysis and could be easily readable from the context table of Fig. 5: the Italy row is fully crossed.

More specifically, taken Italy apart, fix phone and equivalently electricity are matter of concern for Greece, Portugal and Spain (the last country being in the equivalence class of Greece). On the contrary, transport and rail seem to be too expensive for Netherlands, Austria and Germany. In a sense two major groups of countries, geographically well outlined, come into light: Latin versus continental countries.

Because of the aforementioned symmetry between attributes and objects, either objects or attributes can serve as a starting point for a deeper analysis. For example: (a) what is in common for Portugal and France? The only common attribute is 'water', which means that the population of both countries are complaining the price of water service; (b) what is in common for the attributes electricity and transport? Now this question is directed toward the objects, i.e. the countries. It turns out that only for Italy the two attributes are jointly of concern. Note that (a) and (b) are examples of what was mentioned as 'symmetric' analysis. This symmetric analysis is the outcome of the definition of concepts and can be



**Fig. 6** Concept lattice for the context table in Fig. 5. Equivalence classes of objects: {Belgium, Denmark, Ireland, Luxembourg, Finland, United Kingdom}: Belgium as representative. {Greece, Spain}: Greece as representative

further formalized by the algebraic tool of Galois-connections (Ganter and Wille 1999; Carpineto and Romano 2004).

Another aspect results automatically from the symmetric nature of the formal concept lattice with respect to objects and attributes: the association rules, or as originally used by Ganter and Wille (1999) the ‘implications’. The subset property of objects is related to the subset of attributes defining them. Hence an inclusion of subsets of objects can be interpreted as an implication among attributes.

Relevant implications or associations rules can automatically be obtained by a software by Yevtushenko (2000). The list of association rules obtained in our case is shown in Table 3, where the typical notation of Association Rules (AR) is here adopted, as firstly introduced by Agrawal et al. (1993). In Table 3 only rules with confidence level of 100% are visualized, with confidence level being, as usual in AR theory, the proportion of countries with the specified antecedents for which the consequent is also true. The last column of Table 3 indicates in brackets the number of countries which satisfy the rule. As an example, the third rule says that if countries are unsatisfied about the price of water and postal services, then they are also unsatisfied about the transportation system. There are two countries which satisfy this rule and Fig. 6 shows that they are Germany and Italy.

Even if the application is simple in this particular case, it gives a clear example of possible outcomes which can be reachable by the lattice approach.

**Table 3** Set of association rules extracted from FCA lattice of Fig. 6

Rule #	Antecedent	Consequent	Confidence (%)	# of countries
1	Electricity, postal	Water, transport	100	{1}
2	Electricity, transport	Water, postal	100	{1}
3	Water, postal	Transport	100	{2}
4	Water, transport	Postal	100	{2}

## 5 Overall Discussion and Conclusion

This paper provides an insight on Europeans opinion based on Eurobarometer surveys, with particular focus on perceived performance of major services of general interest.

Some main points should be mentioned: (1) we identified the crucial role of criterion price versus criteria quality and contract; (2) we showed the existence of two groups of unsatisfied countries; (3) due to the symmetric feature of the FCA, simple association rules for price discontentment were derived. For a quick view at major results, Table 4 summarizes main conclusions obtained from the adopted methodologies.

We identified two crucial points: firstly dimension theory as a tool to simplify the partial order and secondly FCA as a tool to visualize relations among attributes and objects.

On one hand, dimension theory is an instrument to find appropriate scatter plots in lower dimensions and to identify latent dimensions which are responsible for points configuration. On the other hand, FCA starts with ideas which seem to be completely different from many statistical approaches used so far in environmental and social sciences: the central point of FCA is to model human thinking and how to make broad use of the natural attempt to classify entities and explore their mutual relations. In contrast to that, many statistical approaches start with another point of view: the error minimization in data as a vehicle to derive specific procedures, such for example Factor Analysis, Generalized Linear Models and Nonlinear Multivariate Analysis.

What are errors in our study and how can they be taken into account? The first source of error is uncertainty of the given score and the second source of error is the bias due to the fact that survey data are in this case of subjective type. As recently studied for example by Bertrand and Mullainathan (2001), subjective data can be rather troublesome because they are affected by cognitive problems, due to questionnaire formulation (questions ordering or

**Table 4** Summarizing major results

Adopted method	Main conclusions
Hierarchical partial order	From quality and contract criteria, extraction of two linear orders $\Rightarrow$ homogeneity of citizens' opinion across services
Dimension analysis of posets	Two latent dimensions perfectly describe the price criterion with most complaining countries well separated into two groups: {E, GR, I, P} {A, D, F, NL}
Formal concept analysis	Two groups of countries unsatisfied about price: {E, GR, P} {A, D, NL} Italy unsatisfied about every service As far as price is concerned: {fix phone; electricity} {urban transportation; rail} groups of equivalent services Extraction of relevant Association Rules for price dissatisfaction among different services

wording, scale choice, etc.), and also by respondent's attitude (reluctancy, personality, social and cultural background, etc.). The error of the first type is in some sense overcome by the quantile approach here adopted, whereas the second type of error is beyond the goal of the present job even if it still remains for possible future work.

To summarize finally, when evaluation comes into play FCA seems to be an appropriate tool to deal with complexity. In environmental sciences the deterministic analysis in terms of process based modeling has only a restricted value because on the one side the processes are still unknown and on the other side the data are still not complete for such deterministic approach. FCA symmetric view on objects and attributes provides an analysis of the interplay between groups of objects and attributes making it an ideal tool for environmental science.

Partial order and its specific variant named FCA show to provide suitable instruments to relate objects and properties based on discrete mathematics. The methods discussed here help us to perform a comparison and an evaluation without being forced to aggregate the attributes to an indicator.

Our future work is intended to further explore these approaches specifically in order to make them feasible for applied environmental as well as social sciences.

## References

- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 12th ACM SIGMOD Conference on Management of data* (pp. 207–216) Washington, DC.
- Annoni, P. (2007). Different ranking methods: Potentialities and pitfalls for the case of European opinion polls. *Environmental and Ecological Statistics*, *14*, 453–471.
- Annoni, P., & Brüggemann, R. (2008). The dualistic approach of FCA: A further insight into Ontario Lake sediments. *Chemosphere*, *70*(11), 2025–2031.
- Bartel, H. G., & Brüggemann, R. (1998). Application of formal concept analysis to structure-activity relationships. *Fresenius' Journal of Analytical Chemistry*, *361*, 23–28.
- Bartel, H. G., & Nofz, M. (1997). Exploration of NMR data of glasses by means of formal concept analysis. *Chemometrics and Intelligent Laboratory Systems*, *36*, 53–63.
- Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for Subjective Survey Data. *Economics and Social Behaviour*, *91*, 67–72.
- Borg, I., & Shye, S. (1995). *Facet theory*. Thousand Oaks, CA: SAGE Publications.
- Brans, J. P., & Vincke, P. H. (1985). A Preference Ranking Organization Method (The PROMETHEE Method for Multiple Criteria Decision - Making). *Management Science*, *31*, 647–656.
- Brüggemann, R., & Carlsen, L. (2006). *Partial order in environmental sciences and chemistry*. Berlin: Springer.
- Brüggemann, R., Halfon, E., Welzl, G., Voigt, K., & Steinberg, C. E. W. (2001). Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests. *Journal of Chemical Information and Computer Sciences*, *41*, 918–925.
- Brüggemann, R., Sørensen, P. B., Lerche, D., & Carlsen, L. (2004). Estimation of averaged ranks by a local partial order model. *Journal of Chemical Information and Computer Sciences*, *44*, 618–625.
- Brüggemann, R., & Voigt, K. (1995). An evaluation of online databases by methods of lattice theory. *Chemosphere*, *31*, 3585–3594.
- Brüggemann, R., Voigt, K., & Steinberg, C. E. W. (1997). Application of formal concept analysis to evaluate environmental databases. *Chemosphere*, *35*, 479–486.
- Brüggemann, R., Welzl, G., & Voigt, K. (2003). Order Theoretical Tools for the Evaluation of Complex Regional Pollution Patterns. *Journal of Chemical Information and Computer Sciences*, *43*, 1771–1779.
- Carlsen, L. (2007). Hierarchical partial order ranking. *Environ. Pollut.* (in press). Available online 4 Jan 2008. doi:10.1016/j.envpol.2007.11.023.
- Carlsen, L., & Walker, J. D. (2006). Prioritizing PBT substances. In R. Brüggemann & L. Carlsen (Eds.), *Partial order in environmental sciences and chemistry* (pp. 153–160). Berlin: Springer.
- Carlsen, L., & Walker, J. D. (2003). QSARs for prioritizing PBT substances to promote pollution prevention. *QSAR & Combinatorial Science*, *22*, 49–57.



- Carpineto C., & Romano, G. (2004). *Concept data analysis theory and applications*. England: Wiley.
- Dushnik, B. & Miller, E. W. (1941). Partially ordered sets. *American Journal of Mathematics*, 63, 600–610.
- Eurobarometer 58.0. (2002). September–October 2002. EORG Codebook, European Opinion Research Group.
- Ganter, B., & Wille, R. (1999). *Formal concept analysis—mathematical foundations*. Berlin: Springer.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.
- Johnson, V. E., & Albert, J. H. (1999). *Ordinal data modeling*. New York: Springer.
- Lansdowne, Z. F. (1997). Outranking methods for multicriterion decision making: Arrow's and Raynaud's conjecture. *Social Choice and Welfare*, 14, 125–128.
- Linting, M., Meulman, J. J., Groenen, P. J. F., & Vanderkooij, A. Y. (2007). Stability of principal component analysis: An empirical study using the balanced bootstrap. *Psychol Methods*, 12(3), 359–379.
- Myers, W. L., Patil, G. P., & Cai, Y. (2005). Exploring patterns of habitat diversity across landscapes using partial ordering. Technical Report Number 2005-0701 Technical Reports and Reprints Series, The Pennsylvania State University, PA 16802, 1–34.
- Patil, G. P., & Taille, C. (2004). Multiple indicators, partially ordered sets and linear extensions: Multi-criterion ranking and prioritization. *Environmental and Ecological Statistics*, 11, 199–228.
- Pudenz, S., Brüggemann, R., & Bartel, H. G. (2002). QSAR of ecotoxicological data on the basis of data-driven if-then-rules. *Ecotoxicology*, 11, 337–342.
- Shye, S., Elizur, D., & Hoffman, M. (1994). *Introduction to facet theory. Content design and intrinsic data analysis in behavioral research*, Applied Social Research Methods Series, 35. Thousand Oaks London New Delhi: SAGE Publications.
- Smith, E. V., Jr., & Smith, R. M. (2004). *Introduction to Rasch measurement: Theory, models and applications*. Maple Grove, Minnesota: JAM Press.
- Sørensen, P. B., Brüggemann, R., Carlsen, L., Mogensen, B. B., Kreuger, J., & Pudenz, S. (2003). Analysis of monitoring data of pesticide residues in surface waters using partial order ranking theory—Data interpretation and model development. *Environmental Toxicology and Chemistry*, 22, 661–670.
- Trotter, W. T. (1975). Inequalities in dimension theory for posets. *Proceedings of the American Mathematical Society*, 47, 311–316.
- Trotter, W. T. (1992). *Combinatorics and partially ordered sets: Dimension theory*. Baltimore: The Johns Hopkins University Press.
- Vincke, P. H. (1999). Robust and neutral methods for aggregating preferences into an outranking relation. *European Journal of Operational Research*, 112, 405–412.
- Voigt, K., Welzl, G., & Brüggemann, R. (2004). Data analysis of environmental air pollution monitoring systems in Europe. *Environmetrics*, 15, 577–596.
- Wille, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered sets* (Series C, Vol. 83, pp. 445–470). Dordrecht: D. Reidel Publishing Company.
- Wolff, K. E. (1993). A first course in formal concept analysis—How to understand diagrams. In F. Faulbaum (Ed.), *SoftStat'93, Advances in Statistical Software*, 4 (pp. 429–438).
- Yevtushenko, S. A. (2000). System of data analysis “Concept explorer”. In *Proceedings of the 7th National Conference on Artificial Intelligence* (pp. 127–134). Russia, K II-2000.