

Chapter 5

Quantifying Complexity of Partially Ordered Sets

Guillermo Restrepo

Abstract We discuss two complexity indicators reported in the literature for partially ordered sets (posets), the first one based on linear extensions and the second one on incomparabilities. Later, we introduce a novel indicator that combines comparabilities and incomparabilities with a Shannon's entropy approach. The possible values the novel complexity indicator can take are related to the partitions of the number of order relationships through Young diagrams. Upper and lower bounds of the novel indicator are determined and analysed to yield a normalised complexity indicator. As an example of application, the complexity is calculated for the ordering of countries based on their performance in chemical research. Finally, another complexity indicator is outlined, which is based on comparabilities, incomparabilities, and equivalences.

5.1 Introduction

Order relationships are often used in chemistry and environmental sciences, as can be recognised in statements as “less reactive than”, “less polluting than”, etc. If objects are included in the above claims, then they turn into “ x is less reactive than y ”, “ x is less polluting than y ”, etc. Now, if “is less polluting (reactive) than” is symbolised by \preceq ,¹ then we obtain $x \preceq y$. The binary relation \preceq satisfies three properties (Trotter 1992), i.e. reflexivity, antisymmetry and transitivity, which means that $x \preceq x$, that if $x \preceq y$ and $y \preceq x$, then $x = y$, and that if $x \preceq y$ and $y \preceq z$, then $x \preceq z$, respectively, with x, y, z being objects to order. If the objects are grouped into a set X , then the

¹ We use \preceq instead of $<$ to allow the relation between one object and itself.

G. Restrepo (✉)

Interdisciplinary Research Institute, Universidad de Pamplona, Bogotá, Colombia

Laboratorio de Química Teórica, Universidad de Pamplona, Pamplona, Colombia

e-mail: grestrepo@unipamplona.edu.do; guillermorestrepo@gmail.com

couple (X, \leq) is called a partially ordered set, abbreviated as *poset* (Trotter 1992). It is called “partially” because it might happen that not every couple of objects in X satisfies the relation \leq . In general, for any couple $x, y \in X$, one of these possibilities holds: (1) $x \leq y, y \leq x$; (2) $x \not\leq y, y \not\leq x$; in the first case we say that x and y are *comparable* and we write, in general, $x \perp y$ and in the second case we state that x and y are *incomparable* ($x \parallel y$) (Brüggemann and Bartel 1999). For the particular case where $x \leq y$ and $y \leq x$, we say that x and y are *equivalent* ($x \sim y$). This kind of relation is only explored in Sect. 5.2.3. If every pair of objects of X is comparable, the poset is called a *chain* and \leq a *linear order* (or *total order*). In contrast, an *antichain* is made of incomparable objects (Brüggemann and Bartel 1999). A poset may contain several chains in it. An incomparable object is a chain as well as an antichain; if the chains of a poset are collected, the chain containing more objects than all the other chains is called a *maximum chain*. Likewise, a *maximum antichain* is the antichain of a given poset having more objects than the other antichains (Trotter 1992).

The derived idea from the above generalisation is that it is not always possible to end up with a total order. A particular popular kind of total order is a ranking, where there is only one first, one second, one third, and so on. Hence, given a set of objects to order, it might occur (according to the ordering constraints) that some couples are equivalent and incomparable and others comparable. Brüggemann, Klein, and their co-workers have shown how these situations abound in daily life and in chemistry and environmental sciences (Brüggemann and Patil 2011; Restrepo et al. 2011).

Once (X, \leq) is given, a graphical representation can be generated, namely the Hasse diagram (Halfon 2006; Neggers and Kim 1998). This diagram is a directed graph $\mathcal{H} = (X, E)$ with $(x, y) \in E$ iff $x, y \in X, x \leq y$, and there exists no $z \in X$ with $x \leq z \leq y$ (symbolised as $x \leq : y$ and called a *cover relation* of y over x) (Trotter 1992); E is called the set of directed edges of \mathcal{H} . By convention, \mathcal{H} is drawn in the Euclidean plane whose horizontal/vertical coordinate system requires that the vertical coordinate of $y \in X$ be larger than the one of $x \in X$ iff $x \leq : y$ (Restrepo and Brüggemann 2008).

If $X = \{a, b, c, d, e\}$ and the following order relations hold: $a \leq : b, c \leq : b, c \leq : d, b \leq : e$, and $d \leq : e$, then the corresponding Hasse diagram is shown in Fig. 5.1a. For the ensuing discussion it is important to define linear extensions and intersection of posets. A *linear extension* of the poset (X, \leq) is a total order obtained from the poset such that it preserves the order relations contained in (X, \leq) (Trotter 1992). The corresponding linear extensions of the Hasse diagram in Fig. 5.1a are shown in Fig. 5.1b.

In spite of the simplicity gained when drawing cover relations rather than comparabilities, Hasse diagrams may become rather entangled as the number of elements and cover relations grows. For practical reasons, mainly related to the readability of a Hasse diagram, it is important to develop indicators of such readability, which may be associated with the complexity of the diagram.² Here we

²The fact of having entangled diagrams does imply that the order information of the poset is hidden. In fact, there are different approaches to extract information from posets, as discussed by Brüggemann and Patil (see Brüggemann and Patil 2011, pp. 36–38).

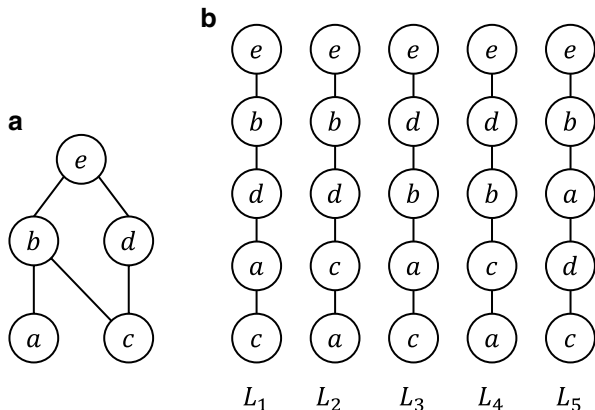


Fig. 5.1 (a) Hasse diagram and (b) linear extensions of the poset $(\{a,b,c,d,e\},\{a \leq b, c \leq b, c \leq d, b \leq e, d \leq e\})$

discuss some approaches to treating posetic complexity and moot a novel complexity indicator, which is further generalised at the end of the chapter.

5.2 Complexity of Posets

Trotter and Bogart, in 1976, related the complexity of a poset to its dimension (Trotter and Bogart 1976). The dimension of a poset was defined by Dushnik and Miller in 1941 as the minimum number of linear extensions of the poset whose intersection³ yields the poset. Hence, for the poset (X, \leq) shown in Fig. 5.1a, its dimension is 2 since $(X, \leq) = L_4 \cap L_5$. Trotter and Bogart gave an interesting justification of dimension as a posetic complexity indicator: “Suppose each of the finite number of observers expresses his individual opinion on the relative merits of a finite set of options by ranking the options in a linear order. A partial ordering on the options is obtained by ranking option x higher than option y when all observers have agreed that x is preferred to y . Conversely, the dimension of a partial order indicates the minimum number of observers necessary to produce the given partial order as a statement of those preferences on which the observers agree unanimously” (Trotter and Bogart 1976). Trotter and Bogart defined two additional complexity indicators, namely interval dimension and semi-order dimension, which are based on the original dimension definition and require the introduction of functions from the poset to the real numbers (Trotter and Bogart 1976). Another dimension complexity indicator

³Given the posets (X, \leq') and (X, \leq'') , their intersection yields the poset (X, \leq) where $\leq = \{(x,y) | x \leq' y \wedge x \leq'' y; x, y \in X\}$.

is the greedy dimension (Kierstead and Trotter 1985), which follows the same dimension principle explained above but with restrictions on the kind of linear extensions to use, called greedy linear extensions. As explained by Trotter, greedy dimension arose as a solution to the scheduling problem known as the jump number problem: “An ordered set P^4 represents a set of tasks to be performed on a single processor. If $x < y$ in P , then x must be performed before y . An admissible schedule is then a linear extension of P . Suppose that a set-up cost is paid for each pair $x, y \in P$ with x incomparable to y in P and x and y occurring consecutively in the linear extension. Find a linear extension L of P which minimizes the number of consecutive pairs of L which are incomparable in P ” (Kierstead and Trotter 1985). The minimum number of those greedy extensions able to reproduce the poset by their intersection is called the greedy dimension of the poset. Some other results on dimension theory are found in Kelly and Trotter (1982) and West (1985).

Although complexity associated with posetic dimension has generated a wealth of studies, its calculation is a difficult task, to the extent that it is an NP-hard problem⁵ (Yannakakis 1982; Yáñez and Montero 1999).

Another posetic complexity indicator was developed in 2000 by Luther et al., who developed a heuristic indicator, defined as follows:

$$J = -D \times C / X$$

where $T(s)$ is given by

$$T(s) := \begin{cases} 0 & 0 \leq s \leq \beta, \\ \frac{s - \beta}{\alpha - \beta} & \beta \leq s \leq \alpha, \\ \frac{1 - s}{1 - \alpha} & \alpha \leq s, \end{cases}$$

with

$$s := \frac{\text{number of incomparabilities}}{\text{number of objects} \times (\text{number of objects} - 1) / 2}$$

where α and β are steering parameters, which are set by the authors as $\alpha = 0.8$ and $\beta = 0.3$. To illustrate this method, let us consider the unlabelled posets of a set of three objects,⁶ the corresponding values, along with the posets, are depicted in Fig. 5.2.

⁴ Here, $P = (X, \leq)$.

⁵ Yannakakis found that for certain posets it is NP-hard to decide if their dimension is equal or lower than a particular natural number (see Yannakakis 1982).

⁶ We take unlabelled posets since what is important in the kind of complexity we are considering is the connectivity among the objects on the poset rather than their identity.

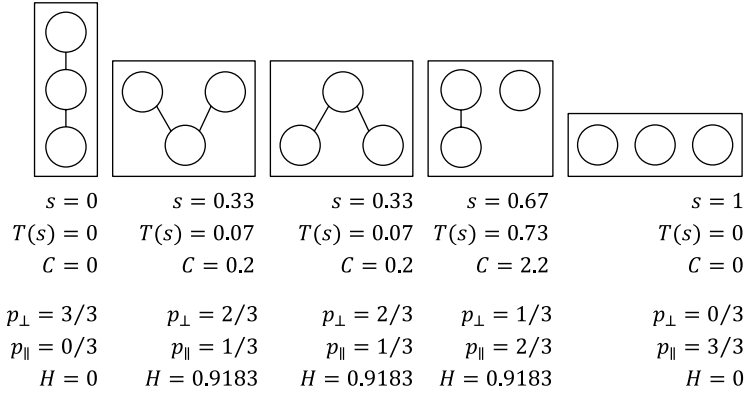


Fig. 5.2 Unlabelled posets of three objects and their respective values s and $T(s)$ needed to calculate their complexity C according to Luther et al.’s method. Values p_{\perp} , p_{\parallel} , and H are explained in Sect. 5.2.1

The complexity indicator C captures the intuitive idea that a maximum chain and a maximum antichain are not complex posets, while posets holding a mixture of comparabilities and incomparabilities are more complex. However, the method has a subjectivity, namely the setting of α and β .

In the current chapter, we develop a novel posetic complexity indicator following a Shannon’s entropy approach based on the number and distribution of order relationships in the poset.

5.2.1 An Entropic Posetic Complexity Indicator

If the system whose complexity is going to be calculated can be described as a discrete random variable with possible values $\{A_1, A_2, \dots, A_n\}$ and associated probabilities $\{p_1, p_2, \dots, p_n\}$, then Shannon’s entropy (Shannon 1948) (complexity) of the system is $H = -\sum p_i \log_2 p_i$.

The random variable we define on a poset is given by the relationships among the elements of the poset and its values are the number of comparabilities and of incomparabilities.

Definition 1 Let (X, \leq) be a poset and let $R \subset X \times X$ be the possible relationships between couples of objects to order. Let $\perp \subseteq R$ and $\parallel \subseteq R$ be the comparabilities and incomparabilities of X , respectively, such that $R = \perp \cup \parallel$. As $|R| = N(N-1)/2$ is the cardinality of R , with $|X| = N$, we define $p_{\perp} = |\perp|/|R|$ and $p_{\parallel} = |\parallel|/|R|$ as the probabilities of having a comparability and an incomparability in (X, \leq) , respectively. We say that $H(X, \leq) = -\sum p_i \log_2 p_i$, with $i \in \{\perp, \parallel\}$, is the *complexity of the poset* (X, \leq) .

For the posets of Fig. 5.2, the respective probabilities and complexities are shown at the bottom of the same figure, where the posets with maximum chain and

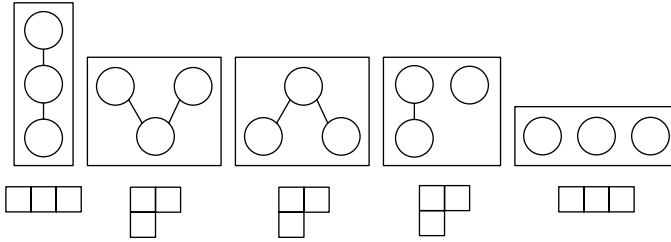


Fig. 5.3 Unlabelled posets of three elements, considering two relationships, and their associated Young diagrams for $|R|$ partitioned into one or two parts

maximum antichain have lowest complexity while all the other posets have maximum complexity. This is so since the distribution of relationships for those with maximum complexity is the same, i.e. 2 and 1, regardless of if they are comparabilities or incomparabilities. Note that for the maximum chain the number of comparabilities is 3 even if the Hasse diagram depicts only 2; this occurs because the diagram shows cover relationships rather than comparabilities.

If both complexity indicators, i.e. C and H , are normalised, then they yield similar results for the posets in Fig. 5.2, except for the one with one comparability and two incomparabilities. In Luther et al.’s indicator, this poset is the most complex one, while in our methodology it has the same complexity as the others that are neither maximum chain nor maximum antichain. This occurs because Luther et al.’s method gives more importance to incomparabilities than to comparabilities, which, as mentioned, is a subjectivity of the method. In our case, both comparabilities and incomparabilities are treated with the same importance.

Some properties of H can be explored on the basis of its relationship with integer partitions and therefore with Young diagrams (Andrews and Eriksson 2004). A Young diagram is a graphical representation of a partition; it is made up of a two-dimensional arrangement of boxes where the k th row has the same number of boxes as the k th term in the partition. If the partition of the integer n is $a + b + \dots + c$, for a list a, b, \dots, c , of r positive integers in such a way that $a \geq b \geq \dots \geq c$, then the diagram is the arrangement of n boxes in r rows (Andrews and Eriksson 2004).

As R is always an integer and \perp and \parallel partition it, then $|R| = |\perp| + |\parallel|$. Hence, the associated Young diagrams for the posets shown in Fig. 5.2 are depicted in Fig. 5.3. The Young diagrams here described always have two rows, one referring to the number of comparabilities and the other to the incomparabilities. The top row always corresponds to the relation with more cases. Hence, if there are more comparabilities than incomparabilities, then the top row refers to comparabilities. If the incomparabilities are highest, the top row corresponds to incomparabilities. In the case of having the same number of comparabilities and incomparabilities, then there is no distinction as to which relation is at the top or at the bottom.

Since our complexity indicator always requires R to be partitioned into comparabilities and incomparabilities and as extreme cases one has either $|R|$ comparabilities or $|R|$ incomparabilities, then the number of possible Young diagrams for unlabelled

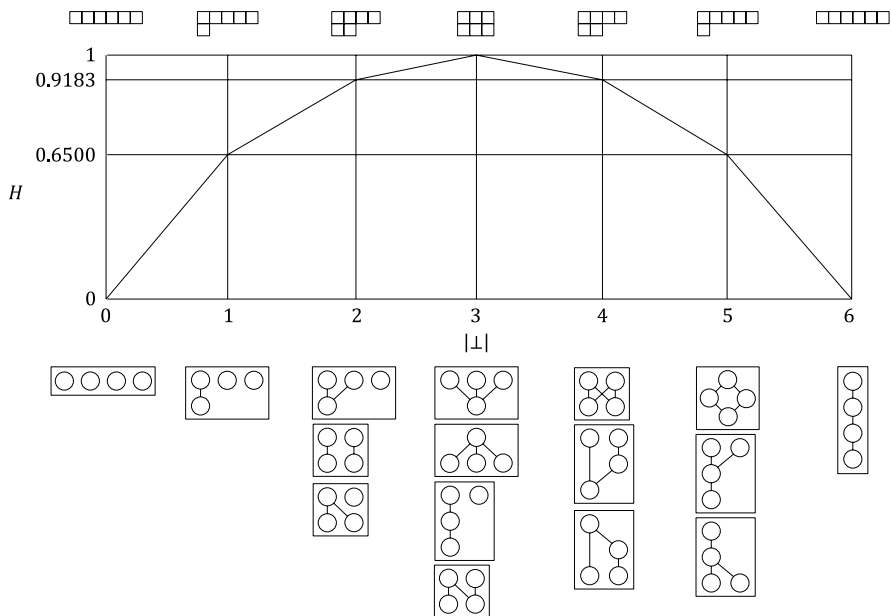


Fig. 5.4 Entropy values H of unlabelled posets of four objects, which are ordered according to their number of comparabilities $|L|$. At the top of the plot, the corresponding Young diagram for each set of equivalent posets is depicted

posets of $|R|$ relations is $\lfloor (|R|/2) + 1 \rfloor$,⁷ which is the number of partitions of $|R|$ into one or two parts (Hardy 1920). This implies that there are always $(|R|/2) + 1$ different values of H for a given $|R|$, which is the same as there are always $(N(N - 1)/4) + 1$ different values of H for the set of all unlabelled posets on a set X of cardinality N .

Proposition 1 For an unlabelled poset (X, \leq) , with $|X|=N$, there are $(N(N - 1)/4) + 1$ different values of H (Definition 1).

Proof It is already stated in the previous paragraph \square

For all 16 posets on a set of four objects, their respective complexities and Young diagrams are shown in Fig. 5.4.

Proposition 1 is important as it states that just by knowing the number of objects to order, one knows the possible complexity values the different unlabelled posets can take.

Now the question that arises is: having calculated the complexity of a poset is it possible to state that the poset is complex or not? The flat answer is no, for the entropic complexity calculated is not an absolute indicator. The problem can be overcome if one knows the total number of unlabelled posets for the N in question, i.e. the number of posets for N objects. It turns out that knowing the total number of

⁷Note that $\lfloor x \rfloor$ is the floor function that maps a real number to the largest previous following integer.

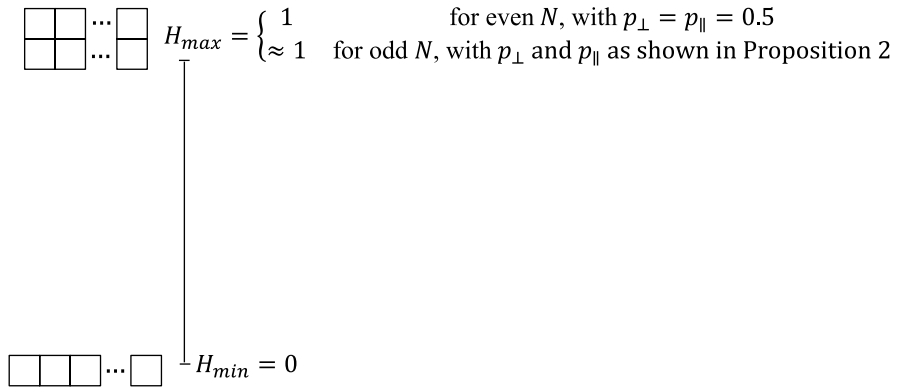


Fig. 5.5 Upper H_{max} and lower H_{min} bounds of H , with their respective Young diagrams. The line corresponds to the real scale bounded by H_{max} and H_{min}

unlabelled posets of a given N is an open question in mathematics⁸ (McKay and Brinkmann 2002). Fortunately there is a way to overcome the problem of counting posets, namely through knowing the upper and lower bounds of H . Hence, if we know the maximum and minimum values H can take, then the values of H make more sense, as a bounded scale is given for H and then one can know whether the poset and its complexity are close to the H upper bound, in which case one can claim that the poset is complex to a great extent. Or one can state that the poset is not that complex if its complexity is close to the lower bound.

Let us first consider the lower bound of H . According to the Young diagram’s representation of the partition of N induced by \perp and \parallel , the minimum value H can occur when there is either only comparabilities or only incomparabilities, i.e. $p_{\perp}=1$ and $p_{\parallel}=0$, or $p_{\perp}=0$ and $p_{\parallel}=1$, respectively. In that case, the Young diagram is made of a single row containing N boxes. The complexity associated with such a case is $H_{min}=0$, which is the lower bound for H .

The upper bound of H occurs when the number of comparabilities and incomparabilities is equal or almost equal, i.e. when $p_{\perp}=p_{\parallel}\cong 0.5$, which implies a Young diagram with two rows evenly or close to evenly populated.

In general, we have H values ranging between the real scale depicted in Fig. 5.5.

The above findings can be formalised as follows:

Proposition 2 For an unlabelled poset (X, \leq) , with $|X|=N$, the upper (H_{max}) and lower (H_{min}) bounds of H (Definition 1) are given by $H_{min}=0$ and ⁹

⁸ Counting the number of unlabelled posets for a given N is a matter of current research in order theory. McKay and Brinkmann, in 2002, developed an algorithm to count this number up to $N=16$ (see McKay and Brinkmann 2002). Some of the results they found are that for $N=4$, there are 16 posets; 16,999 for $N=8$; 1,104,891,746 for $N=12$; and 4,483,130,665,195,080 for $N=16$.

⁹ $\lceil x \rceil$ is the ceiling function that maps a real number to the smallest following integer.

$$H_{\max} = \begin{cases} 1 & \text{for even } N, \text{ with } p_{\perp} = p_{\parallel} = 0.5 \\ -\sum p_i \log_2 p_i & \text{for odd } N, \text{ with } p_{\perp} \text{ and } p_{\parallel} \text{ as follows} \end{cases}$$

$$p_{\perp} = \frac{\left\lceil \frac{N(N-1)}{4} \right\rceil}{\frac{N(N-1)}{2}} \text{ and } p_{\parallel} = \frac{\left\lfloor \frac{N(N-1)}{4} \right\rfloor}{\frac{N(N-1)}{2}}$$

or

$$p_{\perp} = \frac{\left\lfloor \frac{N(N-1)}{4} \right\rfloor}{\frac{N(N-1)}{2}} \text{ and } p_{\parallel} = \frac{\left\lceil \frac{N(N-1)}{4} \right\rceil}{\frac{N(N-1)}{2}}$$

Proof In the previous paragraph proof was given that $H_{\min}=0$. The upper bound of H is attained when N is partitioned into two parts p_{\perp} and p_{\parallel} of the same or almost the same cardinality. It is the same when N is even, in which case the total number of possible relationships $(N(N-1))/2$ is evenly divided into two parts of cardinality $(N(N-1))/4$. Hence, $p_{\perp} = p_{\parallel} = \frac{(N(N-1))/4}{(N(N-1))/2} = 0.5$; therefore $H=1$. If N is an odd integer, then $(N(N-1))/4$ is not an integer number and the most homogeneous distribution of comparabilities and incomparabilities is given when the absolute value of $(|\perp| - |\parallel|) = 1$. In that case the respective cardinalities are given by

$$\left\lceil \frac{N(N-1)}{4} \right\rceil \text{ and } \left\lfloor \frac{N(N-1)}{4} \right\rfloor$$

or

$$\left\lfloor \frac{N(N-1)}{4} \right\rfloor \text{ and } \left\lceil \frac{N(N-1)}{4} \right\rceil.$$

Hence,

$$P_{\perp} = \frac{\left\lceil \frac{N(N-1)}{4} \right\rceil}{\frac{N(N-1)}{2}} \text{ and } p_{\parallel} = \frac{\left\lfloor \frac{N(N-1)}{4} \right\rfloor}{\frac{N(N-1)}{2}}$$

or

$$p_{\perp} = \frac{\left\lceil \frac{N(N-1)}{4} \right\rceil}{\frac{N(N-1)}{2}} \text{ and } p_{\parallel} = \frac{\left\lfloor \frac{N(N-1)}{4} \right\rfloor}{\frac{N(N-1)}{2}}$$

The complexity H is then calculated as in Definition 1 \square

Having found the upper and lower bounds of H for posets of N objects, it is now possible to frame each complexity indicator for a particular poset of N objects into the range defined by the upper and lower bounds. Hence, the closer the complexity to its upper bound, the more complex it is; likewise, the closer to its lower bound, the less complex it is.

5.2.2 *An Application to the Ordering of Countries by Performance in Chemistry*

A widespread source of posets in chemistry and environmental sciences is the ordering of objects that are characterised by different properties, something that has been further explored in the so-called Hasse diagram technique (Brüggemann and Bartel 1999; Restrepo and Brüggemann 2008; Restrepo et al. 2008a). If objects $x, y \in X$ are characterised by properties $q_1(x), q_2(x), \dots, q_i(x)$ and $q_1(y), q_2(y), \dots, q_i(y)$, respectively, x is ordered lower than y ($x \leq y$) if all its properties are lower in magnitude than those of y , or if at least one property is lower for x while all others are equal. This gives place to comparabilities. If all properties of x and y are equal, both objects are equivalent. Note that in the entropic complexity approach introduced in this chapter, equivalences are not considered. If at least one property q_j satisfies $q_j(x) < q_j(y)$ while the others are opposite ($q_i(x) \geq q_i(y)$), x and y are incomparable.

Let us consider the ordering of countries according to the circulation of their scientific production in chemistry from 1996 to 2007, i.e. a bibliometric ordering of countries. We took the data from SCImago (2007). The example takes 195 countries characterised by three properties: NDoc, number of citable documents (articles, reviews, and conference papers); ACit, average citations of documents published during 1996–2007¹⁰; and Hind, the H-index (Hirsch 2005), which takes the value h if the country's number of documents has at least h citations. To make a fair comparison of countries, we took into account the 2011 population (Pop) of the countries listed in the database of the World Bank (2013), which corresponds to 183 countries. Thus, we

¹⁰ SCImago takes a citation window of 4 years less than the observation window. That is why even if the query was performed in 2012, the information corresponds to the period 1996–2007.

Table 5.1 29 countries characterised by NDoc-cap (number of citable articles, reviews, and conference papers, per capita), ACit-cap (average citations of documents published during 1996–2007, per capita), and Hind (H-index, per capita) of chemical documents published between 1996 and 2011

Country	Label	NDoc-cap	ACit-cap	Hind-cap
Austria	AUS	0.00144578	0.02508424	0.17783062
Australia	AUS*	0.00119272	0.02150471	0.19560578
Belgium	BELG	0.00173828	0.03061113	0.23466797
Canada	CAN	0.00133530	0.02683961	0.29109632
Czech Republic	CZE	0.00152683	0.01908544	0.16184449
Denmark	DEN	0.00188572	0.04201383	0.26965788
Estonia	EST	0.00110075	0.01594981	0.06164179
Finland	FIN	0.00170355	0.02878992	0.18398292
France	FRA	0.00143855	0.02438349	0.35819989
Germany	GER	0.00160919	0.03017239	0.49241341
Hong Kong	HKO	0.00125856	0.02582556	0.15480231
Hungary	HUN	0.00121212	0.01410902	0.10909036
Ireland	IRE	0.00127368	0.02455654	0.12864163
Israel	ISR	0.00149800	0.03148791	0.21271566
Italy	ITA	0.00101272	0.01722637	0.20355674
Japan	JAP	0.00125943	0.01948339	0.33878685
Monaco	MONA	0.00285093	0.04843735	0.05416773
Netherlands	NET	0.00147347	0.03455279	0.28585254
Norway	NOR	0.00111611	0.01880653	0.10379867
New Zealand	NZE	0.00115069	0.01883672	0.10816444
Portugal	POR	0.00113227	0.01710866	0.11096287
Singapore	SIN	0.00182592	0.03805209	0.21545807
South Korea	SKO	0.00100585	0.01393096	0.16898210
Slovenia	SLO*	0.00225097	0.02831726	0.15531725
Spain	SPA	0.00146586	0.02543266	0.28291083
Sweden	SWE	0.00226870	0.04827789	0.39021602
Switzerland	SWI	0.00339661	0.08121296	0.69970178
United Kingdom	UK	0.00149308	0.02938381	0.42104845
United States	USA	0.00100059	0.02415426	0.50930071

obtained three properties: NDoc-cap, number of citable documents per capita [NDoc/Pop]; ACit-cap, average citations of documents published during 1996–2007 per capita [ACit \times NDoc-cap]; and Hind-cap, the H-index per capita [H-index \times (NDoc/Pop)]. For the sake of simplicity, we selected those countries with NDoc-cap ≥ 0.001 , ACit-cap ≥ 0.01 , and Hind-cap ≥ 0.05 . Hence, we ended up with 29 countries, whose information is shown in Table 5.1. The complete information for all countries is shown in Table SP1 of the supplementary material uploaded in the following link: <https://docs.google.com/file/d/0B4tX8gtPEjlbNjJFUWsxOTk1c0k/edit>.

To assess the effect of each property upon the complexity of the poset, we considered four posets: one considering the three properties and three others where only two out of the three properties are regarded. These posets are depicted in Figs. 5.6, 5.7, 5.8 and 5.9.

Fig. 5.6 Hasse diagram of 29 countries ordered by three bibliometric properties, with complexity of 0.9831. Country's labels are found in Table 5.1.

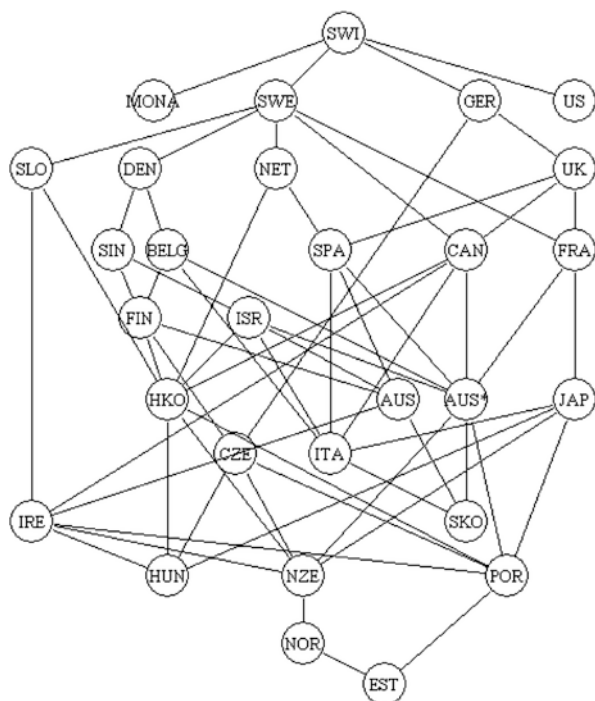


Fig. 5.7 Hasse diagram of 29 countries ordered by two bibliometric properties: ACit-cap (average citations of documents published during 1996–2007, per capita) and Hind-cap (H-index per capita) of chemical documents published between 1996 and 2011, with complexity of 0.9073. Country's labels are found in Table 5.1

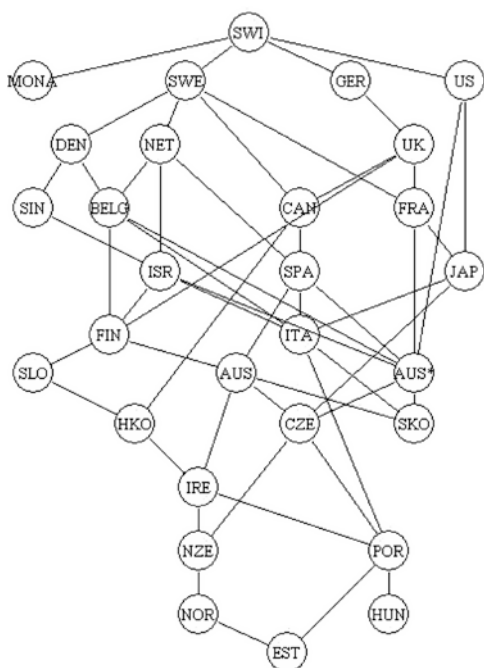
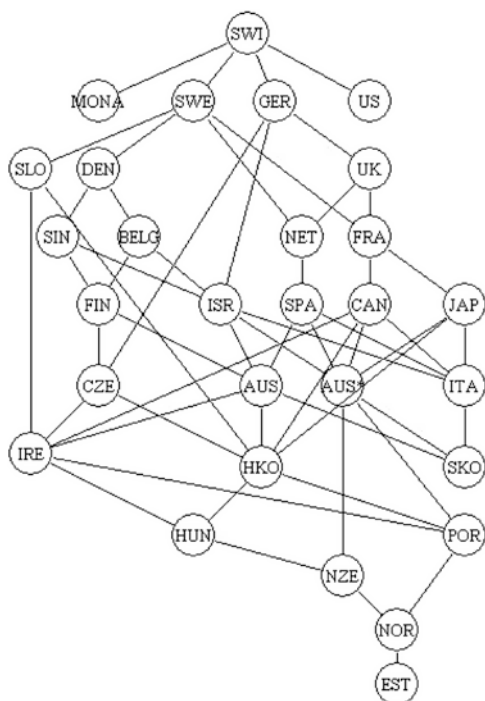


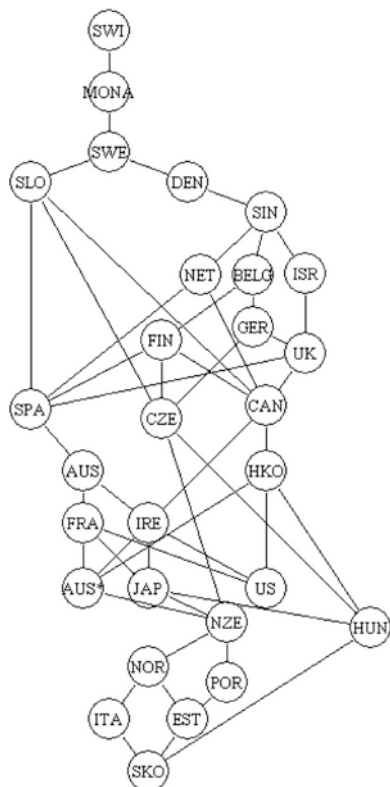
Fig. 5.8 Hasse diagram of 29 countries ordered by two bibliometric properties: NDoc-cap (number of citable articles, reviews, and conference papers) and Hind-cap (H-index per capita) of chemical documents published between 1996 and 2011, with complexity of 0.9593. Country's labels are found in Table 5.1



The three properties are oriented, i.e. high values of them indicate better circulation than low values. The poset of the 29 countries is shown in Fig. 5.6, which was drawn and further analysed with the software WHasse (Brüggemann and Bartel 1999; Brüggemann et al. 1995) available from Rainer Brüggemann. It shows that there is a country with maximum circulation of chemical scientific production, i.e. Switzerland (SWI). There are two countries behaving better than the others except for SWI, i.e. Sweden (SWE) and Germany (GER). Five are the countries with minimum circulation of chemical literature: Monaco (MONA), Hungary (HUN), Estonia (EST), South Korea (SKO), and USA. By inspecting each one of the three properties, we found that SWI always had the maximum score in each of them. Therefore, by removing one or even two properties out of the three discussed, SWI keeps its position as a country with maximum circulation of chemical literature (Figs. 5.7, 5.8 and 5.9).

As $N=29$, H can take 204 different values (Young diagrams). According to Proposition 2, the maximum value H can take is a number close to 1. The poset has 234 comparabilities and 172 incomparabilities, yielding $H=0.9831$, i.e. the poset has a complexity of 98 %, which indicates that it is closer to the maximum allowed complexity of 100 %. The Young diagram depicting this partition would be that with the largest row having 234 boxes and the shortest with 172 ones.

Fig. 5.9 Hasse diagram of 29 countries ordered by two bibliometric properties: NDoc-cap (number of citable articles, reviews, and conference papers, per capita) and ACit-cap (average citations of documents published during 1996–2007, per capita) of chemical documents published between 1996 and 2011, with complexity of 0.5917. Country's labels are found in Table 5.1



The deletion of NDoc-cap and the further ordering of the countries based on the remaining two properties give place to the poset shown in Fig. 5.7 and to its respective complexity ($H=0.9073$). Sweden (SWE) and Germany (GER) keep their position as countries with better circulation than other countries except Switzerland (SWI). Note that USA now accompanies SWE and GER in their behaviour of good circulation. In fact USA, in this poset, is not part of the countries with minimum circulation as it was in the poset of Fig. 5.6. USA is now behaving better than several other countries, e.g. Japan (JAP), Australia (AUS*), Italy (ITA), and Czech Republic (CZE), among others. The countries with minimum circulation are Monaco (MONA), Hungary (HUN), Estonia (EST), and South Korea (SKO). In the diagram of Fig. 5.6 the only thing we knew about USA was that SWI was better than it in terms of chemical scientific circulation. Hence, by ordering countries disregarding the number of citable documents per capita, the complexity of the poset is reduced.

The deletion of ACit-cap and the ordering of the countries based on the remaining two properties yielded the poset shown in Fig. 5.8 with complexity $H=0.9593$. Sweden (SWE) and Germany (GER) keep their good behaviour and USA leaves this group by becoming, again, a country with minimum scientific circulation along

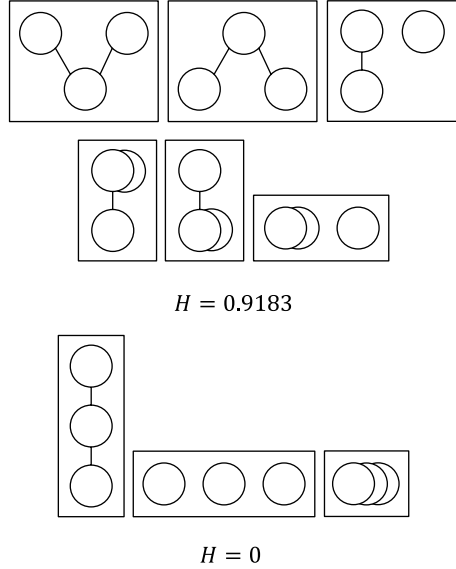
with Monaco (MONA), Estonia (EST), and South Korea (SKO). By comparing this result with that of Fig. 5.6, Hungary (HUN) is not part of the countries with minimum circulation, as it is now better than New Zealand (NZE), Norway (NOR), and Estonia (EST). Hence, by ordering countries disregarding the average citations of documents published per capita, the complexity of the poset is reduced (Fig. 5.8).

The deletion of Hind-cap and the ordering of the countries based on NDoc-cap and ACit-cap yielded the poset shown in Fig. 5.9 with complexity $H=0.5917$. This poset has several changes, e.g. Sweden (SWE) and Germany (GER) leave their good circulation, for there are some other countries behaving better than them. In the case of SWE, now Monaco (MONA) is better than it. In the case of GER, it turns out that Belgium (BELG), Singapore (SIN), Denmark (DEN), and even SWE now have better circulation than GER. South Korea (SKO) and USA are the countries with minimum chemical scientific circulation. In this case MONA and Hungary (HUN) leave the set of countries with minimum circulation as MONA becomes better than many other countries; in fact the only country with better circulation than MONA is Switzerland (SWI). Regarding HUN, in spite of leaving the set of countries with minimum circulation, it is only better than South Korea (SKO). Thus, by ordering countries disregarding the H-index per capita, the complexity of the poset is largely reduced regarding the complexity considering the three properties. This result indicates that this poset is more ordered or possesses more comparabilities than the previous ones. An explanation of such an effect can be seen in the meaning of NDoc-cap and ACit-cap, properties used to obtain the discussed poset. NDoc-cap refers to the citable documents per capita and ACit-cap to the average citations per capita, which depict similar orderings. In fact, $SWI > MONA > SWE$ shows a complete agreement of both properties for these three countries. Thus, it is the inclusion of Hind-cap which gives complexity to the poset.

5.2.3 *Some Possibilities for Posetic Complexity Indicators*

Coming back to the complexity indicator here presented (Definition 1), such an indicator is based on comparabilities and incomparabilities. However, there is another kind of relationship between objects of a poset, namely equivalence, which occurs when $x \leq y$ and $y \leq x$, with x and y being objects to order. This relation between the two objects is written $(x \sim y)$. Hence, a general posetic complexity indicator could include the three relationships. In this case the total number of relationships R , given the number of objects to order, would be partitioned into three parts. The number of different possibilities of doing that (Hardy 1920) is given by $(n+3)^2/12$, which yields the number of corresponding Young diagrams for those partitions. However, not all Young diagrams are attainable by partitioning relationships into comparabilities, incomparabilities and equivalences. This can be seen by a set X of three elements, which yields three possible relationships. For example, the Young diagram with three rows indicating one comparability, one incomparability and one equivalence is not a real possibility, as the equivalence of two elements

Fig. 5.10 Unlabelled posets of three objects, considering three relationships, and their respective values of complexity H . Multiple circles indicate equivalence among circles



implies either the comparability or incomparability with the third one, giving place to a partition 2,1, i.e. a Young diagram of only two rows, one with two boxes and another one with one. A mathematical question to be solved is the calculation of the realisable Young diagrams associated with these particular posetic partitions including three parts (three relationships).

Even if such a relation between relation partitions and Young diagrams is not established in this chapter, the complexity indicator settled in Definition 1 can be generalised to the three order relationships.

Definition 2 Let (X, \leq) be a poset and let $R \subset X \times X$ be the possible relationships between couples of objects to order. Let $\perp \subseteq R$, $\parallel \subseteq R$ and $\sim \subseteq R$ be the comparabilities, incomparabilities, and equivalences of X , respectively, such that $R = \perp \cup \parallel \cup \sim$. As $|R| = N(N-1)/2$ is the cardinality of R , with $|X| = N$, we define $p_{\perp} = |\perp|/|R|$, $p_{\parallel} = |\parallel|/|R|$, and $p_{\sim} = |\sim|/|R|$ as the probabilities of having a comparability, an incomparability and an equivalence in (X, \leq) , respectively. Hence, we say that $H(X, \leq) = -\sum p_i \log_2 p_i$, with $i \in \{\perp, \parallel, \sim\}$, is the *complexity of the poset* (X, \leq) .

The posets with three objects and their respective complexities are shown in Fig. 5.10.

Following a similar analysis of the upper and lower bounds found in Proposition 2, we found that here the respective values are $H_{\min} = 0$ and H_{\max} is equal to 1 if N is even and close to 1 if N is odd. In this latter case, the probabilities of each one of the three relationships are almost equal and are of the form

$$p_i \approx \frac{N(N-1)}{2} \frac{6}{N(N-1)} \approx 0.33, \text{ always satisfying } \sum p_i = 1.$$

5.3 Conclusions and Outlook

The posetic complexity indicator, based on comparabilities and incomparabilities, solves the shortcomings of the complexity indicators reviewed in the chapter, namely the one based on dimension theory and the heuristic one using parameters α and β . When contrasted with the first one, our complexity indicator does not have the problems of dimension calculation for it is based on counting comparabilities and incomparabilities, rather than intersection of linear extensions. These kinds of counting are already included in several statistical packages to treat posets, e.g. WHasse and PyHasse (Brüggemann and Voigt 2009), available from Rainer Brüggemann. Regarding the contrast of our complexity indicator with the heuristic one, our method does not fit any parameter before complexity calculation, which makes it an objective complexity indicator. Additionally, our method does not emphasise the importance of incomparabilities over comparabilities. In the indicator presented in this chapter, both comparabilities and incomparabilities are evenly regarded.

In a recent workshop on posets and their applications,¹¹ Fattore mooted the idea of calculating the complexity through a Kolmogorov's approach. In this case the complexity is not based on counting comparabilities and incomparabilities but on the treatment of a posetic derived matrix, e.g. a cover matrix (showing cover relationships between couples of objects of the poset). This approach is related to the compressibility of the given matrix. Hence, a quite complex poset is one requiring more bits to be represented in a string code, e.g. a binary string. In fact, for a given set of objects, the poset with maximum complexity is the one that after compression has the longest string. In contrast, the less complex poset is the one that, after compression, is represented by a minimum number of bits. It would be interesting to explore this approach and its mathematical properties, as well as its relationship with other complexity indicators.

The example showing the applicability of the complexity indicator introduced here is of particular importance given the current interest on academic rankings. For several reasons, including distribution of funds for research based on research performance, academic rankings have become popular. Examples of these rankings are the Academic Ranking of World Universities (Shanghai ranking), Times Higher Education and the QS World University Rankings, among others. All these rankings, in the end, yield a total order resulting from the weighted aggregation of indicators (properties); the resulting indicator is called a composite indicator. The difference among those rankings lies on the kind of indicators used, i.e. some more oriented to research, some others to education, etc. The additional difference is the importance each ranking gives to the indicators. It turns out that the aggregation of indicators is customarily a linear combination, whose weights are selected upon the importance of the indicators. That is why, contrary to the popularisation of these

¹¹Tenth International Workshop on Partial Order, Theory and Application, Berlin, 27–28 September 2012.

rankings and to their use for decision-making processes, we think they are not “the” best option due to their subjectivities. Changing the subjectivities here mentioned changes the final ranking (Restrepo et al. 2008b). A possible way to overcome the subjectivity on the indicators’ weights is to avoid aggregations, something the Hasse diagram technique allows.

The example of countries’ ordering based on chemical literature circulation was, then, selected as a way to show an alternative option to academic rankings. The analysis of the posetic complexity shows that, among the properties considered, the H-index per capita constitutes the property introducing more incomparabilities when combined with the other two bibliometric properties. This kind of approach—not a ranking—and several other techniques designed to extract information from posets under the shade of the Hasse diagram technique constitute a novel approach worth studying and using by decision-makers. The results on the ranking of countries show that Switzerland (SWI), regardless of the three kinds of descriptors used, is always the best country in chemical circulation of knowledge. There is not “a” worst country regarding this circulation, which constitutes one of the advantages of posets, i.e. if data do not allow it and if aggregations are not performed, several “firsts”, several “seconds”, and several “lasts” may result. Among the 29 countries considered, Monaco (MONA), Hungary (HUN), Estonia (EST), South Korea (SKO), and the USA are countries that need to take measures to address (increase) their circulation of chemical knowledge. It is a matter of surprise to find the USA in the group of countries needing action, as it is common to consider this country as one of the best in this kind of circulation of its research. This idea is true only if the bibliometric data are not considered per capita. For the particular case considered in this chapter, it would be interesting to know the investment on chemical research of the countries studied to see how this information maps the ordering found here or how it affects the ordering. Then, considering the population of a country is important as it is not fair to compare, e.g., the USA with SWI based on raw information, where USA has by far more scientists than SWI. If that is done, the USA appears with the best chemical circulation.

References

- Andrews GE, Eriksson K (2004) Integer partitions. Cambridge University Press, Cambridge
- Brüggemann R, Bartel HG (1999) A theoretical concept to rank environmentally significant chemicals. *J Chem Inf Comput Sci* 39:211–217
- Brüggemann R, Patil GP (2011) Ranking and prioritization for multi-indicator systems. Springer, New York, NY
- Brüggemann R, Voigt K (2009) Analysis of partial orders in environmental systems applying the new software PyHasse. In: Wittmann J, Flechsig M (eds) *Simulation in Umwelt- und Geowissenschaften- workshop Potsdam 2009*. Shaker-Verlag, Aachen, pp 43–55
- Brüggemann R, Halfon E, Bücherl C (1995) Theoretical base of the program “Hasse”. *GSF-Bericht 20/95*: Neuherberg
- Dushnik B, Miller EW (1941) Partially ordered sets. *Am J Math* 63:600–610

- Halfon E (2006) Hasse diagrams and software development. In: Brüggemann R, Carlsen L (eds) *Partial order in environmental sciences and chemistry*. Springer, Berlin, pp 385–392
- Hardy GH (1920) *Some famous problems of the theory of numbers and in particular Waring's problem*. Clarendon, Oxford
- Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 102:16569–16572
- Kelly D, Trotter WT (1982) Dimension theory for ordered sets. In: Rival I (ed) *Ordered sets*. North-Holland, Amsterdam, pp 171–212
- Kierstead HA, Trotter WT (1985) Inequalities for the greedy dimensions of ordered sets. *Order* 2:145–164
- Luther B, Brüggemann R, Pudenz S (2000) An approach to combine cluster analysis with order theoretic tools in problems of environmental pollution. *Match Commun Math Comput Chem* 42:119–143
- McKay BD, Brinkmann G (2002) Posets on up to 16 points. *Order* 19:147–179
- Neggens J, Kim HS (1998) *Basic posets*. World Scientific publications, Singapore
- Restrepo G, Brüggemann R (2008) Dominance and separability in posets, their application to isoelectronic species with equal total nuclear charge. *J Math Chem* 44:577–602
- Restrepo G, Weckert M, Brüggemann R, Gerstmann S, Frank H (2008a) Ranking of refrigerants. *Environ Sci Technol* 42:2925–2930
- Restrepo G, Brüggemann R, Weckert M, Gerstmann S, Frank H (2008b) Ranking patterns, an application to refrigerants. *Match Commun Math Comput Chem* 59:555–584
- Restrepo G, Brüggemann R, Klein D (2011) Partially ordered sets: ranking and prediction of substances' properties. *Curr Comput Aided Drug* 7:133–145
- SCImago (2007) SJR—SCImago Journal & Country Rank. Retrieved 13 Dec 2012 from <http://www.scimagojr.com>
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
- Trotter WJ (1992) *Combinatorics and partially ordered sets, dimension theory*. The Johns Hopkins University Press, Baltimore, MD
- Trotter WJ, Bogart KP (1976) On the complexity of posets. *Discrete Math* 16:71–82
- West DB (1985) Parameters of partial orders and graphs: packing, covering, and representation. In: Rival I (ed) *Graphs and orders*. North-Holland, Amsterdam, pp 267–350
- World Bank (2013) Retrieved 12 Mar 2013 from <http://data.worldbank.org/indicator/SP.POP.TOTL>
- Yáñez J, Montero J (1999) A poset dimension algorithm. *J Algorithm* 30:185–208
- Yannakakis M (1982) The complexity of the partial order dimension problem. *SIAM J Algebraic Discrete Methods* 3:351–358