

# Thinking Ultrametrically, Thinking $p$ -Adically

Fionn Murtagh

**Abstract** We describe the use of ultrametric topology and closely associated  $p$ -adic number theory in a wide range of fields that all share strong elements of common mathematical and computational underpinnings. These include data analysis, including in the “big data” world of massive and high dimensional data sets; physics at very small scales; search and discovery in general information spaces; and in logic and reasoning.

**Keywords** Data analytics • Multivariate data analysis • Pattern recognition • Information storage and retrieval • Clustering • Hierarchy •  $p$ -Adic • Ultrametric topology • Complexity

## 1 Introduction

### 1.1 *Hierarchy and Other Symmetries in Data Analysis*

On one level this chapter is about symmetries in data, such that the data represent complex phenomena, and the symmetries provide a model for understanding these complex phenomena. Hierarchy gives rise to a rich expanse of symmetries and we will be concerned mostly with hierarchy in this article.

Partitioning a set of observations [47, 70, 71] leads to some very simple symmetries. This is one approach to clustering and data mining. But such approaches, often based on optimization, are not of direct interest to us here. Instead we will pursue the theme pointed to by Simon [69], namely that the notion of hierarchy is fundamental for interpreting data and the complex reality which the data expresses. Our work is

---

F. Murtagh (✉)

School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, UK  
e-mail: [fmurtagh@acm.org](mailto:fmurtagh@acm.org)

very different from the marvelous view of the development of mathematical group theory—but viewed in its own right as a complex, evolving system—presented by Foote [24].

Weyl [79] makes the case for the fundamental importance of symmetry in science, engineering, architecture, art and other areas. As a “guiding principle”, “Whenever you have to do with a structure-endowed entity . . . try to determine its group of automorphisms, the group of those element-wise transformations which leave all structural relations undisturbed. You can expect to gain a deep insight in the constitution of [the structure-endowed entity] in this way. After that you may start to investigate symmetric configurations of elements, i.e. configurations which are invariant under a certain subgroup of the group of all automorphisms; . . .” [79, p. 144].

## 1.2 About This Chapter

Theoretical and applied results that are based on ultrametric topology have been studied in fields such as the following:

- In data analysis, both because of the fitting of tree structures and/or visualizations to data sets, to provide a possible way to present a range of partitions to the user, and also to provide for a genealogical model to be fit to data.
- In physics in order to take account of phenomena at very small spatial and time scales, where it is found that discreteness of structures is represented well by  $p$ -adic number systems; and also for any systems that involve movement between discrete states that are characterized by their energy levels.  $p$ -Adic number systems can represent ultrametric topology and vice versa.

It can be added that, as a consequence of applications in physics, the future holds much promise for ultrametric topology-based theory and analysis methods in quantum computing and quantum information theory.

- A further field of use of ultrametric topology arises from being able to show that a considerable number of search and discovery algorithms developed in recent years have an interpretation or vantage point in terms of ultrametric topology.

Computer programming theory also avails of ultrametrics, for example in order to have a framework for non-monotonic reasoning and for multivalued logic.

This chapter will review the state of the art in these fields and will stress the common aspects of methods and applications.

In Sect. 3, we describe ultrametric topology as an expression of hierarchy.

In Sect. 4, we look at the generalized ultrametric context. This is closely linked with analysis based on lattices.

In Sect. 5,  $p$ -adic encoding provides a number theory vantage point on ultrametric topology.

Section 6 deals with application to search and discovery, including work in massive and possibly high dimensional spaces.

## 2 Backgrounders on Hierarchical Clustering, $p$ -Adic Numbers, Ultrametric Topology

### 2.1 A Brief Introduction to Hierarchical Clustering

For the reader new to analysis of data a brief introduction is now provided on hierarchical clustering. Along with other families of algorithm, the objective is automatic classification, for the purpose of data mining, or knowledge discovery. Classification, after all, is fundamental in human thinking and machine-based decision making. But we draw attention to the fact that our objective is *unsupervised*, as opposed to *supervised* classification, also known as discriminant analysis or (in a general way) machine learning. So here we are *not* concerned with generalizing the decision-making capability of training data, nor are we concerned with fitting statistical models to data so that these models can play a role in generalizing and predicting. Instead we are concerned with having “data speak for themselves”. That this unsupervised objective of classifying data (observations, objects, events, phenomena, etc.) is a huge task in our society is unquestionably true. One may think of situations when precedents are very limited, for instance.

Among families of clustering, or unsupervised classification, algorithms, we can distinguish the following: (a) array permuting and other visualization approaches; (b) partitioning to form (discrete or overlapping) clusters through optimization, including graph-based approaches; and—of interest to us in this article—(c) embedded clusters interrelated in a tree-based way.

For the last-mentioned family of algorithm, agglomerative building of the hierarchy from consideration of object pairwise distances has been the most common approach adopted. For comprehensive background texts, see [33, 34, 45, 80].

### 2.2 A Brief Introduction to $p$ -Adic Numbers

The real number system and a  $p$ -adic number system for given prime,  $p$ , are potentially equally useful alternatives.  $p$ -Adic numbers were introduced by Kurt Hensel in 1898.

Whether we deal with Euclidean or with non-Euclidean geometry, we are (nearly) always dealing with reals. But the reals start with the natural numbers, and from associating observational facts and details with such numbers we begin the process of measurement. From the natural numbers, we proceed to the rationals, allowing fractions to be taken into consideration.

The following view of how we do science or carry out other quantitative study was proposed by Volovich in 1987 [76, 77]. See also the surveys in [20, 25]. We can always use rationals to make measurements. But they will be approximate, in general. It is better therefore to allow for observables being “continuous, i.e. endow them with a topology”. Therefore we need a completion of the field  $\mathbb{Q}$  of rationals. To complete the field  $\mathbb{Q}$  of rationals, we need Cauchy sequences and this requires a norm on  $\mathbb{Q}$  (because the Cauchy sequence must converge, and a norm is the tool used to show this). There is the Archimedean norm such that: for any  $x, y \in \mathbb{Q}$ , with  $|x| < |y|$ , then there exists an integer  $N$  such that  $|Nx| > |y|$ . For convenience here, we write:  $|x|_\infty$  for this norm. So if this completion is Archimedean, then we have  $\mathbb{R} = \mathbb{Q}_\infty$ , the reals. That is fine if space is taken as commutative and Euclidean.

What of alternatives? Remarkably all norms are known. Besides the  $\mathbb{Q}_\infty$  norm, we have an infinity of norms,  $|x|_p$ , labelled by primes,  $p$ . By Ostrowski’s theorem [61] these are all the possible norms on  $\mathbb{Q}$ . So we have an unambiguous labelling, via  $p$ , of the infinite set of non-Archimedean completions of  $\mathbb{Q}$  to a field endowed with a topology.

In all cases, we obtain locally compact completions,  $\mathbb{Q}_p$ , of  $\mathbb{Q}$ . They are the fields of  $p$ -adic numbers. All these  $\mathbb{Q}_p$  are continua. Being locally compact, they have additive and multiplicative Haar measures. As such we can integrate over them, such as for the reals.

### 2.3 Brief Discussion of $p$ -Adic and $m$ -Adic Numbers

We will use  $p$  to denote a prime, and  $m$  to denote a non-zero positive integer. A  $p$ -adic number is such that any set of  $p$  integers which are in distinct residue classes modulo  $p$  may be used as  $p$ -adic digits. (Cf. remark below, at the end of Sect. 5.2, quoting from [29]. It makes the point that this opens up a range of alternative notation options in practice.) Recall that a ring does not allow division, while a field does.  $m$ -Adic numbers form a ring; but  $p$ -adic numbers form a field. So a priori, 10-adic numbers form a ring. This provides us with a reason for preferring  $p$ -adic over  $m$ -adic numbers.

We can consider various  $p$ -adic expansions:

1.  $\sum_{i=0}^n a_i p^i$ , which defines positive integers. For a  $p$ -adic number, we require  $a_i \in 0, 1, \dots, p - 1$ . (In practice: just write the integer in binary form.)
2.  $\sum_{i=-\infty}^n a_i p^i$  defines rationals.
3.  $\sum_{i=k}^\infty a_i p^i$  where  $k$  is an integer, not necessarily positive, defines the field  $\mathbb{Q}_p$  of  $p$ -adic numbers.

$\mathbb{Q}_p$ , the field of  $p$ -adic numbers, is (as seen in these definitions) the field of  $p$ -adic expansions.

The choice of  $p$  is a practical issue. Indeed, adelic numbers use all possible values of  $p$  (see [5] for extensive use and discussion of the adelic number framework). A biotechnology example is considered as follows, by Dragovich and

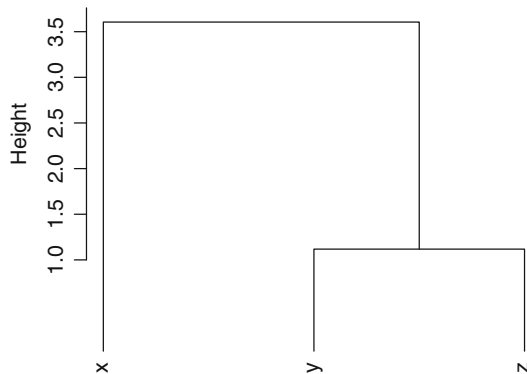
Dragovich [19] and Khrennikov [38]. Desoxyribonucleic acid (DNA) is encoded using four nucleotides: A, adenine; G, guanine; C, cytosine; and T, thymine. In RNA (ribonucleic acid) T is replaced by U, uracil. In [19] a 5-adic encoding is used, since 5 is a prime and thereby offers uniqueness. In [38] a 4-adic encoding is used, and a 2-adic encoding, with the latter based on 2-digit boolean expressions for the four nucleotides (00, 01, 10, 11). A default norm is used, based on a longest common prefix—with  $p$ -adic digits from the start or left of the sequence. (See Sects. 5.3 and 6.3 where a longest common prefix norm or distance is used.)

### 3 Ultrametric Topology

#### 3.1 Ultrametric Space for Representing Hierarchy

Consider Fig. 1 illustrating the ultrametric distance and its role in defining a hierarchy. An early, influential paper is Johnson [37] and an important survey is that of Rammal et al. [63]. Discussion of how a hierarchy expresses the semantics of change and distinction can be found in [58].

The ultrametric topology was introduced by Krasner [39], the ultrametric inequality having been formulated by Hausdorff in 1934. Essential motivation for the study of this area is provided by Schikhof [66] as follows. Real and complex fields gave rise to the idea of studying any field  $K$  with a complete valuation  $|\cdot|$  comparable to the absolute value function. Such fields satisfy the “strong triangle inequality”  $|x + y| \leq \max(|x|, |y|)$ . Given a valued field, defining a totally ordered Abelian (i.e. commutative) group, an ultrametric space is induced through



**Fig. 1** The strong triangular inequality defines an ultrametric: every triplet of points satisfies the relationship:  $d(x, z) \leq \max\{d(x, y), d(y, z)\}$  for distance  $d$ . Cf. by reading off the hierarchy, how this is verified for all  $x, y, z$ :  $d(x, z) = 3.5$ ;  $d(x, y) = 3.5$ ;  $d(y, z) = 1.0$ . In addition the symmetry and positive definiteness conditions hold for any pair of points

$|x - y| = d(x, y)$ . Various terms are used interchangeably for analysis in and over such fields such as  $p$ -adic, ultrametric, non-Archimedean, and isosceles. The natural geometric ordering of metric valuations is on the real line, whereas in the ultrametric case the natural ordering is a hierarchical or rooted tree.

### 3.2 Some Geometrical Properties of Ultrametric Spaces

An ultrametric space is quite different from a metric one. In an ultrametric space everything “lives” on a tree. For various properties that ensue, see [40, Chap. 0, part IV].

In an ultrametric space, all triangles are either isosceles with small base or equilateral. We have here very clear symmetries of shape in an ultrametric topology. These symmetry “patterns” can be used to fingerprint data sets and time series: see [51, 54] for many examples of this.

Some further properties that are studied in [40] are: (a) every point of a circle in an ultrametric space is a centre of the circle. (b) In an ultrametric topology, every ball is both open and closed (termed *clopen*). (c) An ultrametric space is zero-dimensional (see [7, 74]). It is clear that an ultrametric topology is very different from our intuitive, or Euclidean, notions. The most important point to keep in mind is that in an ultrametric space everything “lives” in a hierarchy expressed by a tree.

For an  $n \times n$  matrix of positive reals, symmetric with respect to the principal diagonal, to be a matrix of distances associated with an ultrametric distance on  $X$ , a sufficient and necessary condition is that a permutation of rows and columns satisfies the following form of the matrix:

1. Above the diagonal term, equal to 0, the elements of the same row are non-decreasing.
2. For every index  $k$ , if

$$d(k, k + 1) = d(k, k + 2) = \dots = d(k, k + \ell + 1)$$

then

$$d(k + 1, j) \leq d(k, j) \text{ for } k + 1 < j \leq k + \ell + 1$$

and

$$d(k + 1, j) = d(k, j) \text{ for } j > k + \ell + 1$$

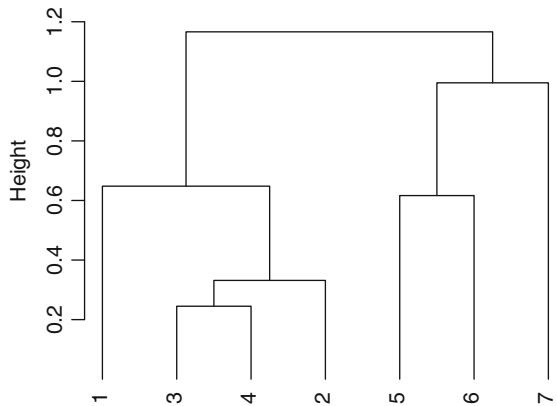
Under these circumstances,  $\ell \geq 0$  is the length of the section beginning, beyond the principal diagonal, the interval of columns of equal terms in row  $k$ .

**Table 1** Input data: eight iris flowers characterized by sepal and petal widths and lengths

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
iris1	5.1	3.5	1.4	0.2
iris2	4.9	3.0	1.4	0.2
iris3	4.7	3.2	1.3	0.2
iris4	4.6	3.1	1.5	0.2
iris5	5.0	3.6	1.4	0.2
iris6	5.4	3.9	1.7	0.4
iris7	4.6	3.4	1.4	0.3

From Fisher’s iris data [23]

**Fig. 2** Hierarchical clustering of seven iris flowers using data from Table 1. No data normalization was used. The agglomerative clustering criterion was the minimum variance or Ward one

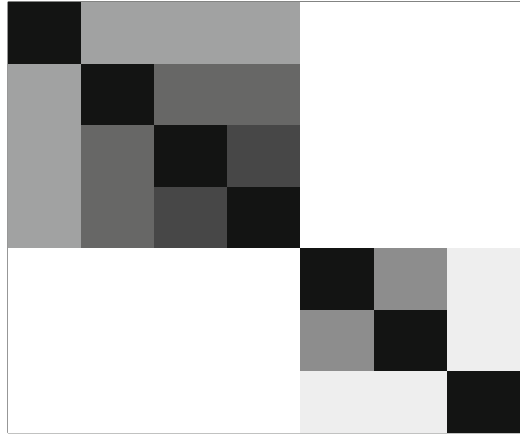


**Table 2** Ultrametric matrix derived from the dendrogram in Fig. 2

	iris1	iris2	iris3	iris4	iris5	iris6	iris7
iris1	0	0.6480741	0.6480741	0.6480741	1.1661904	1.1661904	1.1661904
iris2	0.6480741	0	0.3316625	0.3316625	1.1661904	1.1661904	1.1661904
iris3	0.6480741	0.3316625	0	0.2449490	1.1661904	1.1661904	1.1661904
iris4	0.6480741	0.3316625	0.2449490	0	1.1661904	1.1661904	1.1661904
iris5	1.1661904	1.1661904	1.1661904	1.1661904	0	0.6164414	0.9949874
iris6	1.1661904	1.1661904	1.1661904	1.1661904	0.6164414	0	0.9949874
iris7	1.1661904	1.1661904	1.1661904	1.1661904	0.9949874	0.9949874	0

To illustrate the ultrametric matrix format, consider the small data set shown in Table 1. A dendrogram produced from this is shown in Fig. 2. From the abscissa height of the lowest node or cluster containing the two terminals, the ultrametric distances, also termed cophenetic distances, matrix can be read off this dendrogram. This is shown in Table 2. Finally a visualization of this matrix, illustrating the ultrametric matrix properties discussed above, is shown in Fig. 3.

**Fig. 3** A visualization of the ultrametric matrix of Table 2, where *bright* or *white* = highest value, and *black* = lowest value



### 3.3 Clustering Through Matrix Row and Column Permutation

Direct clustering of the data matrix with no changing of the data matrix values—“non-destructive”, therefore—also comes under the heading of block model clustering.

Figure 3 shows how an ultrametric distance allows a certain structure to be visible (quite possibly, in practice, subject to an appropriate row and column permuting), in a matrix defined from the set of all distances. A generalization opens up for this sort of clustering-by-visualization scheme. An optimized way to do this was pursued in [43, 44]. Comprehensive surveys of clustering algorithms in this area, including objective functions, visualization schemes, optimization approaches, presence of constraints, and applications, can be found in [42, 72]. See also [17, 50].

For all these approaches, underpinning them are row and column permutations that can be expressed in terms of the permutation group,  $S_n$ , on  $n$  elements.

## 4 The Generalized Ultrametric and Formal Concept Analysis

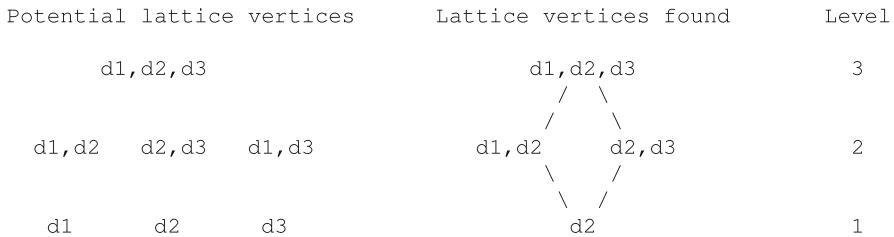
In this section, we consider an ultrametric defined on the power set or join semilattice. Comprehensive background on ordered sets and lattices can be found in [15]. A review of generalized distances and ultrametrics is in [67].

### 4.1 Link with Formal Concept Analysis

Typically hierarchical clustering is based on a distance (which can be relaxed often to a dissimilarity, not respecting the triangular inequality, and *mutatis mutandis* to a



	$v_1$	$v_2$	$v_3$
a	1	0	1
b	0	1	1
c	1	0	1
e	1	0	0
f	0	0	1



**Fig. 4** *Top*: Example data set consisting of five objects, characterized by three boolean attributes. *Bottom*: Lattice corresponding to this data and its interpretation

similarity), defined on all pairs of the object set:  $d : X \times X \rightarrow \mathbb{R}^+$ ; i.e., a distance is a positive real value. Usually we require that a distance cannot be 0-valued unless the objects are identical. That is the traditional approach.

A different form of ultrametrization is achieved from a dissimilarity defined on the power set of attributes characterizing the observations (objects, individuals, etc.)  $X$ . Here we have:  $d : X \times X \rightarrow 2^J$ , where  $J$  indexes the attribute (variables, characteristics, properties, etc.) set.

This gives rise to a different notion of distance that maps pairs of objects onto elements of a join semilattice. The latter can represent all subsets of the attribute set,  $J$ . That is to say, it can represent the power set, commonly denoted  $2^J$ , of  $J$ .

As an example, consider, say,  $n = 5$  objects characterized by three boolean (presence/absence) attributes, shown in Fig. 4 (top). Define dissimilarity between a pair of objects in this table as a *set* of 3 components, corresponding to the 3 attributes, such that if both components are 0, we have 1; if either component is 1 and the other 0, we have 1; and if both components are 1, we get 0. This is the simple matching coefficient [36]. We could use, e.g., Euclidean distance for each of the values sought; but we prefer to treat 0 values in both components as signaling a 1 contribution. We get then  $d(a, b) = 1, 1, 0$  which we will call  $d_1, d_2$ . Then,  $d(a, c) = 0, 1, 0$  which we will call  $d_2$ , etc.

We create lattice nodes shown in Fig. 4, as follows.

The set  $d_1, d_2, d_3$  corresponds to:  $d(b, e)$  and  $d(e, f)$

The subset  $d_1, d_2$  corresponds to:  $d(a, b), d(a, f), d(b, c), d(b, f)$ , and  $d(c, f)$

The subset  $d_2, d_3$  corresponds to:  $d(a, e)$  and  $d(c, e)$

The subset  $d_2$  corresponds to:  $d(a, c)$

Clusters defined by all pairwise linkage at level  $\leq 2$ :

$a, b, c, f$

$a, c, e$

Explanation is as follows:  $a, b, c, f$  all share a 1 for attribute  $v_1$  and a 0 for attribute  $v_2$ , or vice versa. Then  $a, c, e$  share a 1 for attribute  $v_2$  and a 0 for attribute  $v_3$ , or vice versa. See this specification of these clusters in the middle part, “Lattice vertices found”, of Fig. 4.

Finally:

Clusters defined by all pairwise linkage at level  $\leq 3$ :

$a, b, c, e, f$

In Formal Concept Analysis [15, 28], it is the lattice itself which is of primary interest. In [36] there is discussion of, and a range of examples on, the close relationship between the traditional hierarchical cluster analysis based on  $d : I \times I \rightarrow \mathbb{R}^+$ , and hierarchical cluster analysis “based on abstract posets” (a poset is a partially ordered set), based on  $d : I \times I \rightarrow 2^J$ . The latter, leading to clustering based on dissimilarities, was developed initially in [35].

## 4.2 Applications of Generalized Ultrmetrics

As noted in the previous subsection, the usual ultrametric is an ultrametric distance, i.e. for a set  $I$ ,  $d : I \times I \rightarrow \mathbb{R}^+$ . The generalized ultrametric is also consistent with this definition, where the range is a subset of the power set:  $d : I \times I \rightarrow \Gamma$ , where  $\Gamma$  is a partially ordered set. In other words, the *generalized* ultrametric distance is a set. Some areas of application of generalized ultrmetrics will now be discussed.

In the theory of reasoning, a monotonic operator is rigorous application of a succession of conditionals (sometimes called consequence relations). However, negation or multiple valued logic (i.e. encompassing intermediate truth and falsehood) requires support for non-monotonic reasoning.

Thus [31]: “Once one introduces negation ... then certain of the important operators are not monotonic (and therefore not continuous), and in consequence the Knaster-Tarski theorem [i.e. for fixed points; see [15]] is no longer applicable to them. Various ways have been proposed to overcome this problem. One such [approach is to use] syntactic conditions on programs ... Another is to consider different operators ... The third main solution is to introduce techniques from topology and analysis to augment arguments based on order ... [the latter include:] methods based on metrics ... on quasi-metrics ... and finally ... on ultrametric spaces”.

The convergence to fixed points that are based on a generalized ultrametric system is precisely the study of spherically complete systems and expansive automorphisms discussed in Sect. 5.4 below. As expansive automorphisms we see here again an example of symmetry at work. (Cf. too the quotation from Weyl at the end of Sect. 1.1.)

## 5 Hierarchy, Ultrametric Topology and the $p$ -Adic Number System

A dendrogram is widely used in hierarchical, agglomerative clustering, and is induced from observed data. By expressing a dendrogram in  $p$ -adic terms, we open up a wide range of possibilities for seeing symmetries and attendant invariants.

### 5.1 $p$ -Adic Numbers and Their Importance

Rizzi [65] considered ultrametrics and ultramines (i.e. an analogous topology for similarities as opposed to dissimilarities, both of which satisfy the strong triangular inequality). He also discussed the representation of ultrametrics and ultramines using  $p$ -adic numbers.

The importance of  $p$ -adic representation for physics on very small scales has been made by Volovich from the 1980s. See [20, 78]. Such scales are of the order of the Planck length, a fundamental constant ( $1.6 \times 10^{-35}$  m).  $p$ -Adic description of very large scales has similarly been proposed.

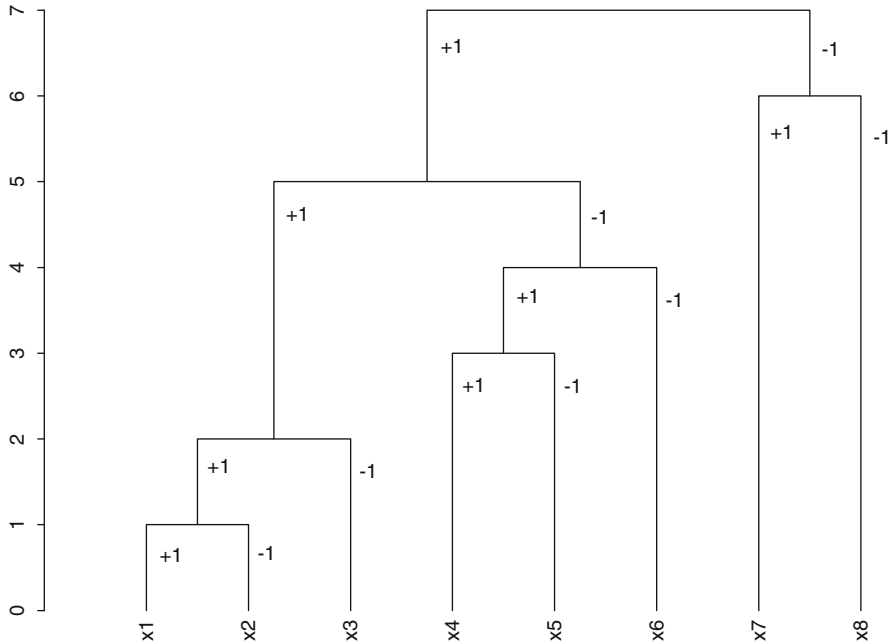
A hierarchy, as a branching process, is a very good means of expressing discrete energy states associated with energy basins requiring at least a requisite quantity of energy to enable a particle to move to another energy basin and possibly, energy level.

Volovich [78] poses the general principle that the fundamental physical laws should be invariant under the change of the number field. Furthermore the  $p$ -adic number fields, it is argued, have great benefit at very small scales and at very large scales. This leads to the following ambitious statement: “If these ideas are true then number theory and the corresponding branches of algebraic geometry are ... the ultimate and unified physical theory”.

### 5.2 $p$ -Adic Encoding of a Dendrogram

We will introduce now the one-to-one mapping of clusters (including singletons) in a dendrogram  $H$  into a set of  $p$ -adically expressed integers (a fortiori, rationals, or  $\mathbb{Q}_p$ ). The field of  $p$ -adic numbers is the most important example of ultrametric spaces. Addition and multiplication of  $p$ -adic integers,  $\mathbb{Z}_p$  (cf. expression in Sect. 2.3), are well defined. Inverses exist and no zero-divisors exist.

A terminal-to-root traversal in a dendrogram or binary rooted tree is defined as follows. We use the path  $x \subset q \subset q' \subset q'' \subset \dots q_{n-1}$ , where  $x$  is a given object specifying a given terminal, and  $q, q', q'', \dots$  are the embedded classes along this path, specifying nodes in the dendrogram. The root node is specified by the class  $q_{n-1}$  comprising all objects.



**Fig. 5** Labelled, ranked dendrogram on eight terminal nodes,  $x_1, x_2, \dots, x_8$ . Branches are labelled +1 and -1. Clusters are:  $q_1 = (x_1, x_2)$ ,  $q_2 = (x_1, x_2, x_3)$ ,  $q_3 = (x_4, x_5)$ ,  $q_4 = (x_4, x_5, x_6)$ ,  $q_5 = (x_1, x_2, x_3, x_4, x_5, x_6)$ ,  $q_6 = (x_7, x_8)$ ,  $q_7 = (x_1, x_2, \dots, x_7, x_8)$

A terminal-to-root traversal is the shortest path between the given terminal node and the root node, assuming we preclude repeated traversal (backtrack) of the same path between any two nodes.

By means of terminal-to-root traversals, we define the following  $p$ -adic encoding of terminal nodes, and hence objects, in Fig. 5.

$$\begin{aligned}
 x_1 &: + 1 \cdot p^1 + 1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7 \\
 x_2 &: - 1 \cdot p^1 + 1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7 \\
 x_3 &: - 1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7 \\
 x_4 &: + 1 \cdot p^3 + 1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7 \\
 x_5 &: - 1 \cdot p^3 + 1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7 \\
 x_6 &: - 1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7 \\
 x_7 &: + 1 \cdot p^6 - 1 \cdot p^7 \\
 x_8 &: - 1 \cdot p^6 - 1 \cdot p^7
 \end{aligned}
 \tag{1}$$

If we choose  $p = 2$ , the resulting decimal equivalents could be the same: cf. contributions based on  $+1 \cdot p^1$  and  $-1 \cdot p^1 + 1 \cdot p^2$ . Given that the coefficients of

the  $p^j$  terms ( $1 \leq j \leq 7$ ) are in the set  $\{-1, 0, +1\}$  (implying for  $x_1$  the additional terms:  $+0 \cdot p^3 + 0 \cdot p^4 + 0 \cdot p^6$ ), the coding based on  $p = 3$  is required to avoid ambiguity among decimal equivalents.

A few general remarks on this encoding follow. For the labelled ranked binary trees that we are considering (for discussion of combinatorial properties based on labelled, ranked and binary trees, see [49]), we require the labels  $+1$  and  $-1$  for the two branches at any node. Of course we could interchange these labels and have these  $+1$  and  $-1$  labels reversed at any node. By doing so we will have different  $p$ -adic codes for the objects,  $x_i$ .

The following properties hold: (a) *Unique encoding*: the decimal codes for each  $x_i$  (lexicographically ordered) are unique for  $p \geq 3$ ; and (b) *Reversibility*: the dendrogram can be uniquely reconstructed from any such set of unique codes.

The  $p$ -adic encoding defined for any object set can be expressed as follows for any object  $x$  associated with a terminal node:

$$x = \sum_{j=1}^{n-1} c_j p^j \text{ where } c_j \in \{-1, 0, +1\} \tag{2}$$

In greater detail we have:

$$x_i = \sum_{j=1}^{n-1} c_{ij} p^j \text{ where } c_{ij} \in \{-1, 0, +1\} \tag{3}$$

Here  $j$  is the level or rank (root:  $n - 1$ ; terminal: 1), and  $i$  is an object index.

In our example we have used:  $c_j = +1$  for a left branch (in the sense of Fig. 5),  $c_j = -1$  for a right branch, and  $c_j = 0$  when the node is not on the path from that particular terminal to the root.

A matrix form of this encoding is as follows, where  $\{\cdot\}^t$  denotes the transpose of the vector.

Let  $\mathbf{x}$  be the column vector  $(x_1, x_2, \dots, x_n)^t$ .

Let  $\mathbf{p}$  be the column vector  $(p^1, p^2, \dots, p^{n-1})^t$ .

Define a characteristic matrix  $C$  of the branching codes,  $+1$  and  $-1$ , and an absent or non-existent branching given by 0, as a set of values  $c_{ij}$  where  $i \in I$ , the indices of the object set; and  $j \in \{1, 2, \dots, n - 1\}$ , the indices of the dendrogram levels or nodes ordered increasingly. For Fig. 5 we therefore have:

$$C = \{c_{ij}\} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix} \tag{4}$$

For given level  $j$ ,  $\forall i$ , the absolute values  $|c_{ij}|$  give the membership function either by node,  $j$ , which is therefore read off columnwise or by object index,  $i$ , which is therefore read off rowwise.

The matrix form of the  $p$ -adic encoding used in Eqs. (2) or (3) is:

$$\mathbf{x} = C \mathbf{p} \quad (5)$$

Here,  $\mathbf{x}$  is the decimal encoding,  $C$  is the matrix with dendrogram branching codes (cf. example shown in expression (4)), and  $\mathbf{p}$  is the vector of powers of a fixed prime  $p$ .

The tree encoding exemplified in Fig. 5, and defined with coefficients in Eqs. (2) or (3), (4) or (5), with labels  $+1$  and  $-1$  was required (as opposed to the choice of 0 and 1, which might have been our first thought) to fully cater for the ranked nodes (i.e. the total order, as opposed to a partial order, on the nodes).

We can consider the objects that we are dealing with to have equivalent integer values. To show that, all we must do is work out decimal equivalents of the  $p$ -adic expressions used above for  $x_1, x_2, \dots$ . As noted in [29], we have equivalence between: a  $p$ -adic number; a  $p$ -adic expansion and an element of  $\mathbb{Z}_p$  (the  $p$ -adic integers). The coefficients used to specify a  $p$ -adic number, [29] notes (p. 69), “must be taken in a set of representatives of the class modulo  $p$ . The numbers between 0 and  $p - 1$  are only the most obvious choice for these representatives. There are situations, however, where other choices are expedient”.

We note that the matrix  $C$  is used in [14]. A somewhat trivial view of how “hierarchical trees can be perfectly scaled in one dimension” (the title and theme of [14]) is that  $p$ -adic numbering is feasible, and hence a one-dimensional representation of terminal nodes is easily arranged through expressing each  $p$ -adic number with a real number equivalent.

In [46], what is termed a nest (i.e. cluster nesting) indicator function is defined, based on the set  $\{a_w, -b_w, 0\}$ ,  $a_w, b_w \in \mathbb{R}^+$  in the same way that the set  $\{1, -1, 0\}$  is used above for the matrix  $C$ . Orthonormality properties of the nest indicator functions are studied.

### 5.3 $p$ -Adic Distance on a Dendrogram

We will now induce a metric topology on the  $p$ -adically encoded dendrogram,  $H$ . It leads to various symmetries relative to identical norms, for instance, or identical tree distances.

We use the following longest common subsequence, starting at the root: we look for the term  $p^r$  in the  $p$ -adic codes of the two objects, where  $r$  is the lowest level such that both sequences have non-zero, i.e.  $+1$  or  $-1$  coefficients, for the  $p^r$  term.

Let us look at the set of  $p$ -adic codes for  $x_1, x_2, \dots$  above (Fig. 5 and relations (1)), to give some examples of this.

For  $x_1$  and  $x_2$ , we find the term we are looking for to be  $p^1$ , and so  $r = 1$ .

For  $x_1$  and  $x_5$ , we find the term we are looking for to be  $p^5$ , and so  $r = 5$ .

For  $x_5$  and  $x_8$ , we find the term we are looking for to be  $p^7$ , and so  $r = 7$ .

Having found the value  $r$ , the distance is defined as  $p^{-r}$  [3, 29].

This longest common prefix metric is also known as the Baire distance. In topology the Baire metric is defined on infinite strings [41]. It is more than just a distance: it is an ultrametric bounded from above by 1, and its *infimum* is 0 which is relevant for very long sequences, or in the limit for infinite-length sequences. The use of this Baire metric is pursued in [60] based on random projections [75], and providing computational benefits over the classical  $O(n^2)$  hierarchical clustering based on all pairwise distances. This is further discussed in Sect. 6.3.

The longest common prefix metric leads directly to a  $p$ -adic hierarchical classification (cf. [4]). This is a special case of the “fast” hierarchical clustering to be discussed in Sect. 6.3.

Compared to the longest common prefix metric, there are other related forms of metric, and simultaneously ultrametric. In [27], the metric is defined via the integer part of a real number. In [3], for integers  $x, y$  we have:  $d(x, y) = 2^{-\text{order}_p(x-y)}$  where  $p$  is prime, and  $\text{order}_p(i)$  is the exponent (non-negative integer) of  $p$  in the prime decomposition of an integer. Furthermore, let  $S(x)$  be a series:  $S(x) = \sum_{i \in \mathbb{N}} a_i x^i$ . ( $\mathbb{N}$  is the set of natural numbers.) The order of  $S(x)$  is the rank of its first non-zero term:  $\text{order}(S) = \inf\{i : i \in \mathbb{N}; a_i \neq 0\}$ . (The series that is all zero is of order infinity.) Then the ultrametric similarity between series is:  $d(S, S') = 2^{-\text{order}(S-S')}$ .

## 5.4 Scale-Related Symmetry

Scale-related symmetry is very important in practice. In this subsection we introduce an operator that provides this symmetry. We also term it a dilation operator, because of its role in the wavelet transform on trees (see [55] for discussion and examples). This operator is  $p$ -adic multiplication by  $1/p$ .

Consider the set of objects  $\{x_i | i \in I\}$  with its  $p$ -adic coding considered above. Take  $p = 2$ . (Non-uniqueness of corresponding decimal codes is not of concern to us now, and taking this value for  $p$  is without any loss of generality.) Multiplication of  $x_1 = +1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^5 + 1 \cdot 2^7$  by  $1/p = 1/2$  gives:  $+1 \cdot 2^1 + 1 \cdot 2^4 + 1 \cdot 2^6$ . Each level has decreased by one, and the lowest level has been lost. Subject to the lowest level of the tree being lost, the form of the tree remains the same. By carrying out the multiplication-by- $1/p$  operation on all objects, it is seen that the effect is to rise in the hierarchy by one level.

Let us call product with  $1/p$  the operator  $A$ . The effect of losing the bottom level of the dendrogram means that either (1) each cluster (possibly singleton) remains the same; or (2) two clusters are merged. Therefore the application of  $A$  to all  $q$  implies a subset relationship between the set of clusters  $\{q\}$  and the result of applying  $A$ ,  $\{Aq\}$ .

Repeated application of the operator  $A$  gives  $Aq, A^2q, A^3q, \dots$ . Starting with any singleton,  $i \in I$ , this gives a path from the terminal to the root node in the tree. Each such path ends with the null element, which we define to be the  $p$ -adic encoding corresponding to the root node of the tree. Therefore the intersection of the paths equals the null element.

Benedetto and Benedetto [1, 2] discuss  $A$  as an expansive automorphism of  $I$ , i.e. form-preserving, and locally expansive. Some implications [2] of the expansive automorphism follow. For any  $q$ , let us take  $q, Aq, A^2q, \dots$  as a sequence of open subgroups of  $I$ , with  $q \subset Aq \subset A^2q \subset \dots$ , and  $I = \bigcup \{q, Aq, A^2q, \dots\}$ . This is termed an inductive sequence of  $I$ , and  $I$  itself is the inductive limit [64, p. 131].

Each path defined by application of the expansive automorphism defines a spherically complete system [27, 66, 74], which is a formalization of well-defined subset embeddedness. Such a methodological framework finds application in multi-valued and non-monotonic reasoning, as noted in Sect. 4.2.

## 6 Exploiting Ultrametric Embeddings for Search and Discovery

### 6.1 Remarkable Symmetries in Very High Dimensional Spaces

In the work of [62, 63] it was shown how as ambient dimensionality increased distances became more and more ultrametric. That is to say, a hierarchical embedding becomes more and more immediate and direct as dimensionality increases. A better way of quantifying this phenomenon was developed in [51]. What this means is that there is inherent hierarchical structure in high dimensional data spaces.

It was shown experimentally in [51, 62, 63] how points in high dimensional spaces become increasingly equidistant with increase in dimensionality. Both [30] and [18] study Gaussian clouds in very high dimensions. The latter finds that “not only are the points [of a Gaussian cloud in very high dimensional space] on the convex hull, but all reasonable-sized subsets span faces of the convex hull. This is wildly different than the behavior that would be expected by traditional low-dimensional thinking”.

That very simple structures come about in very high dimensions is not as trivial as it might appear at first sight. Firstly, even very simple structures (hence with many symmetries) can be used to support fast and perhaps even constant time worst case proximity search [51]. Secondly, as shown in the machine learning framework by Hall et al. [30], there are important implications ensuing from the simple high dimensional structures. Thirdly, [56] shows that very high dimensional clustered data contain symmetries that in fact can be exploited to “read off” the clusters in a computationally efficient way. Fourthly, following [16], what we might want to look for in contexts of considerable symmetry are the “impurities” or small irregularities that detract from the overall dominant picture.



**Table 3** Typical results, based on 300 sampled triangles from triplets of points

No. points	Dimen.	Isosc.	Equil.	UM
Uniform				
100	20	0.10	0.03	0.13
100	200	0.16	0.20	0.36
100	2000	0.01	0.83	0.84
100	20000	0	0.94	0.94
Hypercube				
100	20	0.14	0.02	0.16
100	200	0.16	0.21	0.36
100	2000	0.01	0.86	0.87
100	20000	0	0.96	0.96
Gaussian				
100	20	0.12	0.01	0.13
100	200	0.23	0.14	0.36
100	2000	0.04	0.77	0.80
100	20000	0	0.98	0.98

For uniform, the data are generated on  $[0, 1]^m$ ; hypercube vertices are in  $\{0, 1\}^m$ , and for Gaussian on each dimension, the data are of mean 0, and variance 1. Dimen. is the ambient dimensionality. Isosc. is the number of isosceles triangles with small base, as a proportion of all triangles sampled. Equil. is the number of equilateral triangles as a proportion of triangles sampled. UM is the proportion of ultrametricity-respecting triangles (= 1 for all ultrametric)

See Table 3 exemplifying the change of topological properties as ambient dimensionality increases. It behoves us to exploit the symmetries that arise when we have to process very high dimensional data.

## 6.2 Partial Ultrametric Embedding

In [57], we discuss permutation representations of a data stream. Since hierarchies can also be represented as permutations, there is a ready way to associate data streams with hierarchies. In fact, early computational work on hierarchical clustering used permutation representation to great effect (cf. [68]).

To analyse data streams in this way, in [54] we develop an approach to ultrametric embedding of time-varying signals, including biomedical, meteorological, financial and other. As opposed to the classical way of inducing a hierarchy, through use of an agglomerative hierarchical clustering algorithm, we look for the ultrametric relationship—the strong triangular inequality—and, when found, count such particular cases of adherence to inherent hierarchical properties in the data. The most non-ultrametric time series are found to be chaotic ones. Eyegaze trace data was found to be remarkably high in ultrametricity, which are likely to be due to extreme saccade movements. Some initial questions were raised in that work [54] in regard to the EEG data used, for sleeping, petit mal and irregular epilepsy cases.

This work has been pursued by Khrennikov and his colleagues in modelling multi-agent systems. See [22]. Furthermore this work uses Bose–Einstein and Fermi–Dirac statistical distributions (derived from quantum statistics of energy states of bosons and fermions, i.e. elementary particles with integer, and half odd integer, spin). In [21, 22] multi-agent behaviours are modelled using such energy distributions. The framework is an urn model, where balls can move, with loss of energy over time, and with possibilities to receive input energy, but potentially shared with other balls. See the cited works for a full description of the Monte Carlo system set up. Sequences of actions (and moves), viz. their histories, are coded such that triangle properties can be investigated (cf. also [54]). That leads to a characterization of how ultrametrically embeddable the data is, ab initio (and not through imposing any hierarchical or other structure on the data with retrospective goodness of fit assessment). In [22], the case is presented for such analysis of behavioural histories being important for study of social and economic complexity.

Quantum statistical distributions have been noted in the foregoing work [21, 22]. van Rijsbergen [73] has set out various ways in which a quantum physics formalism makes clearer what is being done in information retrieval and in data analysis generally.

The quantifying of the inherent ultrametric content of text, and finding that some are much more inherently hierarchical than others, was pursued in [59]. As data, the following were used: tales from the Brothers Grimm, Jane Austen novels, dream reports, air accident reports, and James Joyce’s Ulysses.

### 6.3 *Ultrametric Baire Space and Distance*

A Baire space consists of countably infinite sequences with a metric defined in terms of the longest common prefix: the longer the common prefix, the closer a pair of sequences. This longest common prefix metric allows us to define the Baire distance [11, 48, 60]. In this description of the Baire distance, we consider to begin with scalar or univariate values. Below we will generalize to multivariate data such as in the case, for example, of documents with presence or absence, or frequencies of occurrence, on a term set or some other features.

Take the longest common prefixes at issue here as coming from precision of any value. Without loss of generality, take these values as decimal, i.e. base 10, or  $m$ -adic with  $m = 10$ . We take  $x$  and  $y$  to be bounded by 0 and 1. Each of them is of some precision, and we take the integer  $|K|$  to be the maximum precision.

Thus we consider ordered sets  $x_k$  and  $y_k$  for  $k \in K$ , or, we will write, for  $k = 1, 2, \dots, |K|$ . The cardinality of the set  $K$  is the precision with which a number,  $x$ , is measured. So,  $x_k$  with  $k = 1$  is the first decimal place of precision; with  $k = 2$ , we have the second decimal place; . . . ; and with  $k = |K|$  we have the  $|K|$ th decimal place.

Consider as examples  $x = 0.478$ ; and  $y = 0.472$ . Start from the first decimal position. For  $k = 1$ , we find  $x_1 = y_1 = 4$ . For  $k = 2$ ,  $x_2 = y_2 = 7$ . But for  $k = 3$ ,  $x_3 \neq y_3$ .

We now introduce the following distance (case of  $x$  and  $y$ , with 1 attribute, hence unidimensional, and where subscript  $k$  is the digit of precision):

$$d_B(x, y) \equiv d_B(x_K, y_K) = \begin{cases} 1 & \text{if } x_1 \neq y_1 \\ \inf 10^{-k} & x_k = y_k, \quad 1 \leq k \leq |K| \end{cases} \quad (6)$$

We call this  $d_B$  value Baire distance, which can be shown to be an ultrametric [51–54, 60] distance. In the properties of a metric we generally have  $d(x, y) = 0$  iff  $x = y$  whereas for the Baire distance this reflexivity property is relaxed by having the 0 value replaced by the definably minimal value.

When dealing with binary data, 2 is a convenient base. In definition (6) we used a base of 10 for ease of coding when working with real numbers.

It is seen that this distance splits a unidimensional string of decimal values into a 10-way hierarchy, in which each leaf is associated with a grid cell. From Eq. (6) we can read off the distance between points assigned to the same grid cell. All pairwise distances of points assigned to the same cell are the same.

Relative to agglomerative hierarchical clustering, the Baire-based hierarchy is such that each node of this tree is associated with a grid (more strictly, in what we have described, interval) cell. Cell assignments at a particular level can be used to count the number of values  $x, y$  that are associated with that cell, and these counts define local density. As we have described the inducing of a hierarchy, this has been in a top-down manner (cf. how agglomerative hierarchical clustering, in that it is agglomerative, is consequently bottom-up). It follows from this algorithm that we can read the hierarchy off the data in a single scan, by having the target data structure—here, a regular 10-way tree—and assigning each value to all its appropriate nodes in the tree. For ease of characterizing this tree, or hierarchical clustering, we refer to it as a *Baire tree* or *Baire hierarchy*. The minimum Baire distance corresponds to a partial match of the values at each level [13].

For data with higher dimensionality, random projections can be used. Random projection is simple. Forming the random matrix  $R$  and projecting the  $d \times N$  data matrix  $X$  into the  $k$  dimensions is of order  $O(dkN)$ . If  $X$  is sparse with  $c$  non-zero entries per column, the complexity is of order  $O(ckN)$ .

In fact random projection can be seen as a class of hashing function. Hashing is much faster than alternative methods because it avoids the pairwise comparisons required for partitioning and classification. If two points  $(p, q)$  are close, they will have a very small  $\|p - q\|$  (Euclidean metric) value; and they will hash to the same value with high probability; if they are distant, they should collide with small probability.

Clustering using the Baire distance has been successfully applied to areas such as chemoinformatics [60], astronomy [12] and text retrieval [10].

In [60], this principle of binning data is used on a large, high dimensional chemoinformatics data set. The application of merging databases of chemical

compounds is important. In [60] 1.2 million compounds were used, characterized using a particular coding scheme by 1052-valued presence/absence vector. Use of the Baire distance and the associated hierarchical clustering were compared in detail with  $k$ -means partitioning (through partitions derived from the hierarchical clustering).

We studied stability of results, and effectiveness relative to other clustering methods, in particular  $k$ -means partitioning, in [12]. The main domain of application in that work was astronomy, and in particular clustering of redshifts in order to facilitate regression of (more expensively observed but better quality) spectroscopic redshifts on (more easily observed but with less signal resolution) photometric redshifts.

#### ***6.4 Approximating an Ultrametric for Similarity Metric Space Searching***

In [51] we show that, in much work over the years, nearest neighbour searching has been made more efficient through the use of more easily determined feasibility bounds. An early example is Fukunaga and Narendra [26], a chapter review is in [50], and a survey can be found in [9]. Rendering given distances as ultrametric is a powerful way to facilitate nearest neighbour searching. Furthermore “stretching the triangular inequality” (a phrase used by [8]) so that it becomes the strong triangular inequality, or ultrametric inequality, gives a unifying view of some algorithms of this type.

Fast nearest neighbour finding often makes use of pivots to establish bounds on points to be searched, and points to be bypassed as infeasible [6,9].

A full discussion can be found in [51]. Fast nearest neighbour searching in metric spaces often appeals to heuristics. The link with ultrametric spaces gives rise instead to a unifying view.

Hjaltason and Samet [32] discuss heuristic nearest neighbour searching in terms of embedding the given metric space points in lower dimensional spaces. From our discussion in this section, we see that there is evidently another alternative direction for facilitating fast nearest neighbour searching: viz., taking the metric space as an ultrametric one, and if it does not quite fit this perspective, then “stretch” it (transform it locally) so that it does so.

## **7 Conclusions**

There are many exciting perspectives opened up by our work on the theme of symmetry-finding through hierarchy in very large data collections, with insights and perspectives from many application domains that are data-based and motivated, and indeed driven, by complex problem-solving.

“My thesis has been that one path to the construction of a nontrivial theory of complex systems is by way of a theory of hierarchy”. Thus Simon [69, p. 216]. We have noted symmetry in many guises in the representations used, in the transformations applied, and in the transformed outputs. These symmetries are non-trivial too, in a way that would not be the case were we simply to look at classes of a partition and claim that cluster members were mutually similar in some way. We have seen how the  $p$ -adic or ultrametric framework provides significant focus and commonality of viewpoint.

Furthermore we have highlighted the computational scaling properties of our algorithms. They are fully capable of addressing the data and information deluge that we face and providing us with the best interpretative and decision-making tools. The full elaboration of this last point is to be sought in each and every application domain, and face to face with old and new problems.

## References

1. Benedetto, R.L.: Examples of wavelets for local fields. In: Larson, D., Heil, C., Jorgensen, P. (eds.) *Wavelets, Frames, and Operator Theory*, Contemporary Mathematics, American Mathematical Society (Providence, RI) vol. 345, pp. 27–47 (2004)
2. Benedetto, J.J., Benedetto, R.L.: A wavelet theory for local fields and related groups. *J. Geom. Anal.* **14**, 423–456 (2004)
3. Benzécri, J.-P.: *L'Analyse des Données. Tome I. Taxinomie*, 2nd edn. Dunod, Paris (1979)
4. Bradley, P.E.: Mumford dendrograms. *Comput. J.* **53**, 393–404 (2010)
5. Brekke, L., Freund, P.G.O.:  $p$ -Adic numbers in physics. *Phys. Rep.* **233**, 1–66 (1993)
6. Bustos, D., Navarro, G., Chávez, E.: Pivot selection techniques for proximity searching in metric spaces. *Pattern Recognit. Lett.* **24**, 2357–2366 (2003)
7. Chakraborty, P.: Looking through newly to the amazing irrationals. Technical report (2005). arXiv: math.HO/0502049v1
8. Chávez, E., Navarro, G.: Probabilistic proximity search: fighting the curse of dimensionality in metric spaces. *Inf. Process. Lett.* **85**, 39–46 (2003)
9. Chávez, E., Navarro, G., Baeza-Yates, R., Marroquín, J.L.: Searching in metric spaces. *ACM Comput. Surv.* **33**(3), 273–321 (2001)
10. Contreras, P.: *Search and Retrieval in Massive Data Collections*. Ph.D. thesis, Royal Holloway, University of London (2011)
11. Contreras, P., Murtagh, F.: Evaluation of hierarchies based on the longest common prefix, or Baire, metric, 2007. In: *Classification Society of North America (CSNA) Meeting*, University of Illinois, Urbana-Champaign, IL (2007)
12. Contreras, P., Murtagh, F.: Fast, linear time hierarchical clustering using the Baire metric. *J. Classif.* **29**, 118–143 (2012)
13. Contreras, P., Murtagh, F.: Linear time Baire hierarchical clustering for enterprise information retrieval. *Int. J. Softw. Inform.* **6**(3), 363–380 (2012)
14. Critchley, F., Heiser, W.: Hierarchical trees can be perfectly scaled in one dimension. *J. Classif.* **5**, 5–20 (1988)
15. Davey, B.A., Priestley, H.A.: *Introduction to Lattices and Order*, 2nd edn. Cambridge University Press, Cambridge (2002)
16. Delon, F.: Espaces ultramétriques. *J. Symb. Logic* **49**, 405–502 (1984)
17. Deutsch, S.B., Martin, J.J.: An ordering algorithm for analysis of data arrays. *Oper. Res.* **19**, 1350–1362 (1971)

18. Donoho, D.L., Tanner, J.: Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl. Acad. Sci.* **102**, 9452–9457 (2005)
19. Dragovich, B., Dragovich, A.: p-Adic modelling of the genome and the genetic code. *Comput. J.* **53**, 432–442 (2010)
20. Dragovich, B., Khrennikov, A.Yu., Kozyrev, S.V., Volovich, I.V.: On p-adic mathematical physics. *p-Adic Numbers Ultrametric Anal. Appl.* **1**, 1–27 (2009)
21. Ezhov, A.A., Khrennikov, A.Yu.: On ultrametricity and a symmetry between Bose-Einstein and Fermi-Dirac systems. In: *AIP Conferences Proceedings 826, p-Adic Mathematical Physics, 2nd International Conference*, pp. 55–64. American Institute of Physics, Melville (2006)
22. Ezhov, A.A., Khrennikov, A.Yu., Terentyeva, S.S.: Indications of a possible symmetry and its breaking in a many-agent model obeying quantum statistics. *Phys. Rev. E*, **77** (2008). Article number 031126
23. Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936)
24. Foote, R.: Mathematics and complex systems. *Science* **318**, 410–412 (2007)
25. Freund, P.G.O.: p-Adic strings and their applications. In: Rakic, Z., Dragovich, B., Khrennikov, A., Volovich, I. (eds.) *Proceedings of 2nd International Conference on p-Adic Mathematical Physics*, pp. 65–73. American Institute of Physics, Melville (2006)
26. Fukunaga, K., Narendra, P.M.: A branch and bound algorithm for computing k-nearest neighbors. *IEEE Trans. Comput.* **C-24**, 750–753 (1975)
27. Gajić, L.: On ultrametric space. *Novi Sad J. Math.* **31**, 69–71 (2001)
28. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin (1999) [Formale Begriffsanalyse. *Mathematische Grundlagen*. Springer, Berlin (1996)]
29. Gouvêa, F.Q.: *p-Adic Numbers: An Introduction*. Springer, Berlin (2003)
30. Hall, P., Marron, J.S., Neeman, A.: Geometric representation of high dimensional, low sample size data. *J. R. Stat. Soc. B* **67**, 427–444 (2005)
31. Hitzler, P., Seda, A.K.: The fixed-point theorems of Priess-Crampe and Ribenboim in logic programming. *Fields Inst. Commun.* **32**, 219–235 (2002)
32. Hjaltason, G.R., Samet, H.: Properties of embedding methods for similarity searching in metric spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 530–549 (2003)
33. Jain, A.K., Dubes, R.C.: *Algorithms for Clustering Data*. Prentice-Hall, Upper Saddle River (1988)
34. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv.* **31**, 264–323 (1999)
35. Janowitz, M.F.: An order theoretic model for cluster analysis. *SIAM J. Appl. Math.* **34**, 55–72 (1978)
36. Janowitz, M.F.: Cluster analysis based on abstract posets. Technical report (2005–2006)
37. Johnson, S.C.: Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967)
38. Khrennikov, A.Yu.: Gene expression from polynomial dynamics in the 2-adic information space. Technical report (2009)
39. Krasner, M.: Nombres semi-réels et espaces ultramétriques. *C. R. Acad. Sci., Tome II* **219**, 433 (1944)
40. Lerman, I.C.: *Classification et Analyse Ordinale des Données*. Dunod, Paris (1981)
41. Levy, A.: *Basic Set Theory*. Dover, Mineola (2002) [Springer, 1979]
42. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1**, 24–45 (2004)
43. March, S.T.: Techniques for structuring database records. *Comput. Surv.* **15**, 45–79 (1983)
44. McCormick, W.T., Schweitzer, P.J., White, T.J.: Problem decomposition and data reorganization by a clustering technique. *Oper. Res.* **20**, 993–1009 (1982)
45. Mirkin, B.: *Mathematical Classification and Clustering*. Kluwer, Dordrecht (1996)
46. Mirkin, B.: Linear embedding of binary hierarchies and its applications. In: Mirkin, B., McMorris, F., Roberts, F., Rzhetsky, A. (eds.) *Mathematical Hierarchies and Biology*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 37, pp. 331–356. American Mathematical Society, Providence (1997)

47. Mirkin, B.: Clustering for Data Mining. Chapman and Hall/CRC, Boca Raton (2005)
48. Mirkin, B., Fishburn, P.: Group Choice. V.H. Winston, Washington (1979)
49. Murtagh, F.: Counting dendrograms: a survey. *Discrete Appl. Math.* **7**, 191–199 (1984)
50. Murtagh, F.: *Multidimensional Clustering Algorithms*. Physica-Verlag, Heidelberg/Vienna (1985)
51. Murtagh, F.: On ultrametricity, data coding, and computation. *J. Classif.* **21**, 167–184 (2004)
52. Murtagh, F.: Quantifying ultrametricity. In: Antoch, J. (ed.) *COMPSTAT 2004 – Proceedings in Computational Statistics*, pp. 1561–1568. Springer, Berlin (2004)
53. Murtagh, F.: Thinking ultrametrically. In: Banks, D., House, L., McMorris, F.R., Arabie, P., Gaul, W. (eds.) *Classification, Clustering, and Data Mining Applications. Proceedings of the Meeting of the International Federation of Classification Societies (IFCS)*, pp. 3–14. Springer, Berlin (2004)
54. Murtagh, F.: Identifying the ultrametricity of time series. *Eur. Phys. J. B* **43**, 573–579 (2005)
55. Murtagh, F.: The Haar wavelet transform of a dendrogram. *J. Classif.* **24**, 3–32 (2007)
56. Murtagh, F.: The remarkable simplicity of very high dimensional data: application to model-based clustering. *J. Classif.* **26**, 249–277 (2009)
57. Murtagh, F.: Symmetry in data mining and analysis: a unifying view based on hierarchy. *Proc. Steklov Inst. Math.* **265**, 177–198 (2009)
58. Murtagh, F.: The correspondence analysis platform for uncovering deep structure in data and information (6th Annual Boole Lecture). *Comput. J.* **53**, 304–315 (2010)
59. Murtagh, F.: Ultrametric model of mind, II: application to text content analysis. *p-Adic Numbers Ultrametric Anal. Appl.* **4**(3), 207–221 (2012)
60. Murtagh, F., Downs, G., Contreras, P.: Hierarchical clustering of massive, high dimensional data sets by exploiting ultrametric embedding. *SIAM J. Sci. Comput.* **30**, 707–730 (2008)
61. Ostrowski, A.: Über einige Lösungen der Funktionalgleichung  $\phi(x) \cdot \phi(y) = \phi(xy)$ . *Acta Math.* **41**, 271–284 (1918)
62. Rammal, R., Angles d’Auriac, J.C., Doucot, B.: On the degree of ultrametricity. *J. Phys. Lett.* **46**, L-945–L-952 (1985)
63. Rammal, R., Toulouse, G., Virasoro, M.A.: Ultrametricity for physicists. *Rev. Mod. Phys.* **58**(3), 765–788 (1986)
64. Reiter, H., Stegeman, J.D.: *Classical Harmonic Analysis and Locally Compact Groups*, 2nd edn. Oxford University Press, Oxford (2000)
65. Rizzi, A.: Ultrametrics and  $p$ -adic numbers. In: Gaul, W., Opitz, O., Schader, M. (eds.) *Data Analysis: Scientific Modeling and Practical Application*, pp. 325–324. Springer, Berlin (2000)
66. Schikhof, W.H.: *Ultrametric Calculus*. Cambridge University Press, Cambridge (1984) [Chaps. 18–21]
67. Seda, A.K., Hitzler, P.: Generalized distance functions in the theory of computation. *Comput. J.* **53**, 443–464 (2010)
68. Sibson, R.: SLINK: an optimally efficient algorithm for the single link cluster method. *Comput. J.* **16**, 30–34 (1973)
69. Simon, H.A.: *The Sciences of the Artificial*. MIT Press, Cambridge (1996)
70. Steinley, D.: K-means clustering: a half-century synthesis. *Br. J. Math. Stat. Psychol.* **59**, 1–34 (2006)
71. Steinley, D., Brusco, M.J.: Initializing K-means batch clustering: a critical evaluation of several techniques. *J. Classif.* **24**, 99–121 (2007)
72. Van Mechelen, I., Bock, H.-H., De Boeck, P.: Two-mode clustering methods: a structured overview. *Stat. Methods Med. Res.* **13**, 363–394 (2004)
73. van Rijsbergen, C.J.: *The Geometry of Information Retrieval*. Cambridge University Press, Cambridge (2004)
74. Van Rooij, A.C.M.: *Non-Archimedean Functional Analysis*. Dekker, New York (1978)
75. Vempala, S.S.: *The Random Projection Method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 65. American Mathematical Society, Providence (2004)
76. Volovich, I.V.: Number theory as the ultimate physical theory. Technical report (1987) Preprint No. TH 4781/87, CERN, Geneva

77. Volovich, I.V.: p-Adic string. *Class. Quantum Gravity* **4**, L83–L87 (1987)
78. Volovich, I.V.: Number theory as the ultimate physical theory. *p-Adic Numbers Ultrametric Anal. Appl.* **2**, 77–87 (2010)
79. Weyl, H.: *Symmetry*. Princeton University Press, Princeton (1983)
80. Xu, R., Wunsch, D.: Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **16**, 645–678 (2005)