Metrics for Clustering Comparison in Bioinformatics

Giovanni Rossi
Department of Computer Science and Engineering (DISI), University of Bologna
Mura Anteo Zamboni 7, Bologna, 40126, Italy

Keywords: Hamming Distance, Partition Lattice, Hasse Diagram, Weighted Graph, Geodesic Distance, Path.

Abstract: Developing from a concern in bioinformatics, this work analyses alternative metrics between partitions. From both theoretical and applicative perspectives, a useful and interesting distance between any two partitions is HD, which counts the number of atoms finer than either one but not both. While faithfully reproducing the traditional Hamming distance between subsets, HD is very sensible and computable through scalar products between Boolean vectors. It properly deals with complements and axiomatically resembles the entropy-based variation of information VI distance. Entire families of metrics (including HD and VI) obtain as minimal paths in the weighted graph given by the Hasse diagram: submodular weighting functions yield path-based distances visiting the join (of any two partitions), whereas supermodularity leads to visit the meet. This yields an exact (rather than heuristic) approach to the consensus partition (combinatorial optimization) problem.

1 INTRODUCTION

Partitions or clusterings are key instruments in a variety of fields at the interface of computer science, artificial intelligence and engineering, including pattern recognition/learning, web mining and bioinformatics. Quantitative clustering comparison is essential for assessing the proximity between and superiority among diverse partitions of a given set (Meila, 2007; Pinto Da Costa and Rao, 2004).

In bioinformatics, measuring the distance between clusterings of populations, either natural or experimental, is fundamental for sibling relationship reconstruction. Apparently, attention has been placed for the most on a unique distance measure, here denoted by MMD, which relies on *maximum matching* (Konovalov, 2006; Berger-Wolf et al., 2007; Sheikh et al., 2010; Konovalov et al., 2005b; Konovalov et al., 2005a). After its appearance (Almudevar and Field, 1999), MMD was shown (Gusfield, 2002) to be computable via the assignment problem (Korte and Vygen, 2002, p. 236). Another partition distance recently tested in this setting (Brown and Dexter, 2012) is the *variation of information* VI, obtained axiomatically from information theory (Meila, 2007).

In this work, the distance between partitions is measured in quite different ways, since the aim is to have consistency and generalizations in terms of lattice theory. The primary objective is to reproduce the traditional Hamming distance between subsets, given by the cardinality of their symmetric difference (Bollobas, 1986, p. 3). Such a benchmark is extended from Boolean to geometric lattices by focusing on atoms and join-decompositions of lattice elements (Aigner, 1997; Stern, 1999). While every subset admits a unique such a decomposition, involving a number of atoms equal to the cardinality of the subset, a generic partition admits different join-decompositions, most of which redundant. The number of atoms involved in the unique maximal join-decomposition of a partition is here defined to be the *size* of that partition. The size is an integer-valued lattice function, like the rank. In fact, the two coincide for Boolean lattices but differ crucially for geometric lattices. Roughly speaking, replacing the rank with the size yields the Hamming distance HD between partitions proposed below. While achieving combinatorial congruency, HD shares with VI important characterizing axioms and is computed through simple scalar products between Boolean vectors, avoiding any algorithmic issue. Finally, HD also has a large range which provides fine measurement sensitivity.

The traditional Hamming distance between two subsets is also the length of any shortest path between them in the associated Hasse diagram, which is the unit hypercube. This latter is a graph with subsets as vertices and edges linking any two subsets whenever one covers the other (in terms of set inclusion, see (Bollobas, 1986; Godsil and Royle, 2001) and below). In order to have an analog for the Hamming distance between partitions defined here, it is necessary to look at the lattice of partitions of a n-set as the polygon ma-

Rossi, G.
Metrics for Clustering Comparison in Bioinformatics.
DOI: 10.5220/0005707102990308
In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016)*, pages 299-308
ISBN: 978-989-758-173-1

troid of the complete graph K_n on n vertices (Aigner, 1997, pp. 259, 274). In other terms, partitions of a n-set may well be regarded as those graphs on n vertices each of whose components is complete (i.e. a clique). In this way, the partition lattice is seen to be strictly included in the $\binom{n}{2}$-dimensional unit hypercube: the set $\{0,1\}^{\binom{n}{2}}$ of hypercube vertices identifies the $2^{\binom{n}{2}}$ distinct graphs on n vertices, while linear dependence (Whitney, 1935) entails that partitions only span $\mathcal{B}_n < 2^{\binom{n}{2}}$ hypercube vertices, where *Bell number* $\mathcal{B}_n$ is the number of partitions of a n-set (Rota, 1964a; Graham et al., 1994). The convex hull of these $\mathcal{B}_n$ vertices identifies a polytope, and the graph of this polytope (Grünbaum, 2001, pp. 212-16) is in fact the Hasse diagram of the partition lattice. Yet, while the covering relation between subsets assigns a unitary weight to each edge of the hypercube (Sebő and Tannier, 2004, p. 384), edges of the polytope of partitions must be weighted through the size (see above), as this latter quantifies precisely the number of hypercube edges that collapse into a unique edge of the polytope. With such a weighting, the Hamming distance between two partitions (like between two subsets) quantifies the minimum weight of a path connecting them.

The approach allows for generalizations in that the size may be replaced with any alternative order-preserving lattice function, such as the rank (or the entropy of partitions). Then, polytope edges have weights obtained as the difference between the greater and the smaller value taken by the chosen order-preserving function on the associated endpoints. Accordingly, the distance between two lattice elements is the minimum weight of a path connecting them.

2 DISTANCES, LATTICES AND GRAPHS

For a finite set $N = \{1,\dots,n\}$, let $(2^N,\cap,\cup)$ and $(\mathcal{P}^N,\wedge,\vee)$ be the associated subset and partition lattices, ordered by inclusion $\supseteq$ and coarsening $\geqslant$, respectively. Both are atomic and atomistic; the fomer is distributive while the latter is geometric (Aigner, 1997; Stern, 1999). A graph $G = (V,E)$ consists of a vertex set $V = \{v_1,\dots,v_m\}$ and an edge set $E \subseteq V_2$ included in the $\binom{m}{2}$-set $V_2 := \{\{v_i,v_j\} : 1 \leq i < j \leq m\}$ of unordered pairs of vertices. The complete graph on m vertices (see above) is $K_m = (V,V_2)$. The Hamming distance $HD(P,Q)$ between partitions $P,Q \in \mathcal{P}^N$ proposed here aims at reproducing the traditional Hamming distance $|A\Delta B|$ between subsets $A,B \in 2^N$ while keeping into account that partitions of a n-set N are in fact graphs with vertex set $V = N$ and whose components are each a complete subgraph (Aigner, 1997).

Distances within a ordered set must be measured via the order relation, while distances between elements of any set are called 'Hamming distances' when such elements are represented as arrays and the distance between two of them is the number of entries where their array representations differ. The Hamming distance between subsets $A,B \in 2^N$ is

$$|A\Delta B| = |A\backslash B| + |B\backslash A| = r(A\cup B) - r(A\cap B), \quad (1)$$

$r : 2^N \to \mathbb{Z}_+$ being the rank function: $r(A) = |A|$. It counts how many $i \in N$ are included in either A or else B, but not in both. Elements $i \in N$, or 1-cardinal subsets $\{i\} \in 2^N$, are the atoms of lattice $(2^N,\cap,\cup)$, and (1) is a Hamming distance since subsets $A \in 2^N$ are represented as Boolean n-vectors $\chi_A \in \{0,1\}^n$ or vertices of the n-dimensional unit hypercube $[0,1]^n$. This achieves through their characteristic function $\chi_A : N \to \{0,1\}$, defined by $\chi_A(i) = 1$ if $i \in A$ and $\chi_A(i) = 0$ if $i \in N\backslash A$. Thus the distance between any $A,B \in 2^N$ is the number $|A\Delta B|$ of entries where χ_A and χ_B differ (Bollobas, 1986; Aigner, 1997).

A partition $P = \{A_1,\dots,A_{|P|}\} \subset 2^N$ of N is a collection of pairwise disjoint subsets, called blocks (or clusters), whose union yields N. Any subset $A \in 2^N$ has a unique complement $A^c = N\backslash A$. For all partitions $P \in \mathcal{P}^N$ and all non-empty subsets $\emptyset \subset A \subseteq N$, let $P^A = \{B\cap A : B \in P, \emptyset \neq B\cap A\}$ denote the partition of A induced by P. Two *non-Hamming* distances between partitions are now briefly introduced. Maximum matching distance $MMD(P,Q)$ between two partitions $P,Q \in \mathcal{P}^N$ is

$$MMD(P,Q) = \min\{|A^c| : \emptyset \subset A \subseteq N, P^A = Q^A\}. \quad (2)$$

This is the minimum number of elements $i \in N$ that must be deleted in order for the two residual induced partitions to coincide. Also, $MMD(P,Q)$ *"is the minimum number of elements that must be moved between clusters of P so that the resulting partition equals Q"* (Gusfield, 2002, p. 160). It is computable as a maximum matching or assignment problem (Korte and Vygen, 2002, chapter 11). In a graph a matching is a set of pairwise disjoint edges, i.e. the endpoints are all different vertices. Now consider the bipartite graph $G = (P\cup Q,E)$ with $|P| + |Q|$ vertices, one for each block of each partition, and join any two of them $A \in P$ and $B \in Q$ with an edge $\{A,B\} \in E$ if $A\cap B \neq \emptyset$. In addition, let $|A\cap B|$ be the weight of the edge. Then, determining $MMD(P,Q)$ amounts to find a maximum-weight matching E^* in G, that is one where the sum $\sum_{(A,B)\in E^*} |A\cap B|$ of edge weights is maximal. In fact, the minimum number $MMD(P,Q)$ of elements that must be removed for the two residual

partitions to coincide is the sum $\sum_{(A,B)\in E^*} |A\Delta B|$ over all selected edges of the cardinality of the symmetric difference between the associated endpoints.

Another important measure of the distance between two partitions P,Q is the variation of information $VI(P,Q)$, obtained axiomatically from information theory (see (Meila, 2007, expressions (15)-(22), pages 879-80)). Entropy $e(P) = -\sum_{A\in P} \frac{|A|}{n} \log\left(\frac{|A|}{n}\right)$ of a partition P (binary logarithm) enables to measure the distance between $P,Q \in \mathcal{P}^N$ as

$$VI(P,Q) = 2e(P\wedge Q) - e(P) - e(Q), \qquad (3)$$

where $P\wedge Q$ is the coarsest partition finer than both P and Q (while $\vee$ is the *'finest-coarser-than'* operator). Notice that while the range of MMD is $\{0,1,\ldots,n-1\} \subset \mathbb{Z}_+$, VI ranges in a finite subset of interval $[0,\log n] \subset \mathbb{R}_+$.

Apart from MMD and VI, there exist several other partition distance measures (see (Deza and Deza, 2013, sections 10.2 and 10.3, pp. 191-193) and (Day, 1981; Hubert and Arabie, 1985; Warrens, 2008; Mirkin, 1996)). One was proposed as the *Hamming distance between (matrices representing) partitions* (Meila, 2007; Mirkin and Cherny, 1970; Mirkin and Muchnik, 2008), and thus shall be briefly distinguished from the object of this paper. A binary relation $\mathcal{R}$ on N is a subset $\mathcal{R} \subseteq N\times N$ of ordered pairs (i,j) of elements $i,j \in N$. The collection of all such binary relations is subset lattice $(2^{N\times N},\cap,\cup)$. If symmetry $(i,j)\in\mathcal{R} \Rightarrow (j,i)\in\mathcal{R}$ and *transitivity* $(i,j),(j,h)\in\mathcal{R} \Rightarrow (i,h)\in\mathcal{R}$ hold, then $\mathcal{R}$ is an *equivalence* relation, or a partition of N into equivalence classes: maximal subsets $A\in 2^N$ such that $(i,j),(j,i)\in\mathcal{R}$ for all $i,j\in A$ are precisely its blocks. A binary relation $\mathcal{R}$ is represented by a matrix $M^{\mathcal{R}} \in \{0,1\}^{n\times n}$ with entries $M^{\mathcal{R}}_{ij} = 1$ if $(i,j)\in\mathcal{R}$ and $M^{\mathcal{R}}_{ij} = 0$ if $(i,j)\notin\mathcal{R}$. Let two equivalence relations $\mathcal{R},\mathcal{R}'$ have associated partitions P,P' and matrices $M^{\mathcal{R}}, M^{\mathcal{R}'}$. The distance $d(\mathcal{R},\mathcal{R}')$ between subsets $\mathcal{R},\mathcal{R}' \in 2^{N\times N}$ can be computed as $d(\mathcal{R},\mathcal{R}') = |\mathcal{R}\Delta\mathcal{R}'| = |\mathcal{R}\cup\mathcal{R}'| - |\mathcal{R}\cap\mathcal{R}'|$. This is the number of 1s in matrix $M^{\mathcal{R}\Delta\mathcal{R}'} = M^{\mathcal{R}} + M^{\mathcal{R}'}$ modulo 2 (see (Aigner, 1997, p. 338)). While providing a distance between partitions P and P', this is a Hamming distance between certain subsets that correspond to partitions only in quite special cases, as lattice $(2^{N\times N},\cap,\cup)$ contains $2^{n^2} - \mathcal{B}_n$ elements, or binary relations, that do not correspond to partitions, or equivalence relations. The argument also applies when partitions are represented as Boolean $n\times n$-matrices through the complement of equivalence relations, namely apartness relations $\mathcal{R}^c = (N\times N)\backslash\mathcal{R}$ (Ellerman, 2013b; Ellerman, 2013a). This is detailed in subsection 4.1 below.

However regarded, partition lattice $(\mathcal{P}^N,\wedge,\vee)$ is compressed into a larger subset lattice, with which some elements are shared while some others are not. This feature is maintained even when partitions are decomposed as joins of atoms, for they generally admit several such decompositions. Nevertheless, when regarded from this perspective partition lattice $(\mathcal{P}^N,\wedge,\vee)$ is seen to be included in subset lattice $(2^{N_2},\cap,\cup)$, with the two sharing the same $\binom{n}{2}$ atoms. In fact, $(2^{N_2},\cap,\cup)$ is the minimal subset lattice including the partition lattice. Accordingly, the Hamming distance between partitions HD proposed below relies precisely on representing partitions as Boolean $\binom{n}{2}$-vectors, although only $\mathcal{B}_n < 2^{\binom{n}{2}}$ distinct such vectors correspond to partitions. In particular, HD is the traditional Hamming distance $|E\Delta E'|$ between edge sets $E,E' \in 2^{N_2}$, with these latter corresponding to partitions only when in both graphs $G=(N,E), G'=(N,E')$ each component is a complete subgraph.

3 HAMMING DISTANCE

The rank function $r:\mathcal{P}^N\to\mathbb{Z}_+$ for partitions is $r(P) = n - |P|$, where $P_\perp = \{\{1\},\ldots,\{n\}\}$ is the bottom element: $r(P_\perp) = 0$. Atoms are immediately above, with rank 1, in the associated *Hasse diagram*. This latter is ordered by coarsening $\geqslant$, with coarser partitions in upper levels (Meila, 2007; Aigner, 1997; Stern, 1999) and $P\geqslant Q$ meaning that every block of Q is included in some block of P. Hence atoms are those partitions consisting of $n-1$ blocks, namely $n-2$ singletons and one pair. These $\binom{n}{2}$ unordered pairs $\{i,j\}\in N_2$ are the same atoms as in subset lattice $(2^{N_2},\cap,\cup)$. For notational convenience, let $[ij]\in\mathcal{P}^N$ denote the atom where the unique 2-cardinal block is (unordered) pair $\{i,j\}\in[ij]$.

Consider $\chi_N \in \{0,1\}^n$ as the n-vector with all entries equal to 1 and denote by $\langle x,y\rangle$ the scalar product between x and y. The Hamming distance between subsets $A,B\in 2^N$ is $|A\Delta B| = |A| + |B| - 2|A\cap B| =$

$$= \langle\chi_A,\chi_N\rangle + \langle\chi_B,\chi_N\rangle - 2\langle\chi_A,\chi_B\rangle. \qquad (4)$$

In order to replace subsets A with partitions P, let $\mathcal{P}^N_{(1)} = \{[ij] : 1\leq i<j\leq n\}$ be the $\binom{n}{2}$-set of atoms of the partition lattice, i.e. $\mathcal{P}^N_{(1)} \sim N_2$. The analog of characteristic function χ_A is *indicator function* $I_P : \mathcal{P}^N_{(1)} \to \{0,1\}$ defined, for $P\in\mathcal{P}^N, [ij]\in\mathcal{P}^N_{(1)}$, by $I_P([ij]) = \begin{cases} 1 \text{ if } P\geqslant[ij] \\ 0 \text{ if } P\not\geqslant[ij] \end{cases}$. In words, if pair $\{i,j\}$ is included in some block A of P (i.e. $\{i,j\}\subseteq A\in P$),

then partition P is coarser than atom $[ij]$, and the corresponding position $I_P([ij])$ of indicator array I_P has entry 1. Otherwise, $I_P([ij])$ equals 0. Top element $P^\top = \{N\}$ for partitions yields $I_{P^\top}$, i.e. the $\binom{n}{2}$-vector with all entries equal to 1. The number $s(P) = |\{[ij] : [ij] \leqslant P\}|$ of atoms finer than P is (Rossi, 2011) the *size* $s : \Pi_N \to \mathbb{Z}_+$, that is

$$s(P) = \sum_{A \in P} \binom{|A|}{2} = \langle I_P, I_{P^\top} \rangle. \tag{5}$$

While the cardinality of subsets $|A| = \langle \chi_A, \chi_N \rangle$ takes every integer value between 0 and n, the size of partitions $s(P) = \langle I_P, I_{P^\top} \rangle$ does not the same between 0 and $\binom{n}{2}$. Minimally, this may be seen for $N = \{1,2,3\}$, as the $\mathcal{B}_3 = 5$ partitions are the finest $\{\{1\},\{2\},\{3\}\}$ and coarsest $\{1,2,3\}$ ones, together with $\binom{3}{2} = 3$ atoms: $[12] = \{\{1,2\},\{3\}\}$, $[13] = \{\{1,3\},\{2\}\}$ and $[23] = \{\{2,3\},\{1\}\}$. Thus, there is no partition with size equal to 2, in that $\{1,2,3\} = [12] \vee [13] \vee [23]$ and $\{1,2,3\} = [12] \vee [23] = [12] \vee [13] = [13] \vee [23]$. Available sizes for $1 \leq n \leq 7$ are in Table I below.

Table 1: *Available sizes of partitions of a n-set,* $1 \leq n \leq 7$.

$\lvert N\rvert = n$	$\{s(P) : P \in \Pi_N\}$ Available sizes
1	$\{0\}$
2	$\{0,1\}$
3	$\{0,1,3\}$
4	$\{0,1,2,3,6\}$
5	$\{0,1,2,3,4,6,10\}$
6	$\{0,1,2,3,4,6,7,10,15\}$
7	$\{0,1,2,3,4,5,6,7,9,10,11,15,21\}$

In atomistic lattices, such as $2^N, \Pi_N$ and 2^{N_2}, every element admits a decomposition as a join of atoms (Aigner, 1997; Stern, 1999). While subsets $A \in 2^N$ and $E \in 2^{N_2}$ admit a unique such a decomposition, namely $A = \bigcup_{i \in A} \{i\}$ and $E = \bigcup_{\{i,j\} \in E} \{i,j\}$, partitions generally admit several such decompositions. For $n = 3$ as above, the coarsest partition $\{1,2,3\}$ decomposes either as the join of any two atoms, or else as the join of all the three available atoms at once. In particular, the rank $r(P) = n - |P|$ of any partition P is the minimum number of atoms involved in a join-decomposition of P, while the size $s(P)$ is the maximum number of atoms involved in a join-decomposition of P. The coarsest partition $\{1,2,3\}$ of a 3-cardinal set has rank $r(\{1,2,3\}) = 3 - 1 = 2$ and size $s(\{1,2,3\}) = 3 = \binom{3}{2}$.

Proposition 1. *The size is strictly monotone: for all* $P, Q \in \Pi_N$, *if* $P \geqslant Q$, $P \neq Q$, *then* $s(P) > s(Q)$.

Proof. If $P > Q$ (i.e. $P \geqslant Q$, $P \neq Q$), then every block $A \in P$ is the union of some blocks $B_1, \ldots, B_{|Q^A|} \in Q$, with $|Q^A| > 1$ for at least one block $A \in P$. Recall that Q^A is the partition of A induced by Q (see the definition of MMD in section 2 above). The union of any two such blocks $B, B' \in Q$ increases the size by $\binom{|B|+|B'|}{2} - [\binom{|B|}{2} + \binom{|B'|}{2}] = |B||B'|$, which is strictly positive as blocks are non-empty. □

In order to reproduce (1) and (4) above, Hamming distance $HD : \Pi_N \times \Pi_N \to \mathbb{Z}_+$ has to count the number of atoms finer than either one of any two partitions $P, Q \in \Pi_N$ but not finer than both, that is $HD(P,Q) = |\{[ij] : P \geqslant [ij] \not\leqslant Q\}| + |\{[ij] : P \not\geqslant [ij] \leqslant Q\}|$. In view of (5), $HD(P,Q) = s^P + s^Q - 2s(P \wedge Q) =$

$$= \langle I_P, I_{P^\top} \rangle + \langle I_Q, I_{P^\top} \rangle - 2\langle I_P, I_Q \rangle. \tag{6}$$

Also $P \wedge Q = \bigvee_{P \geqslant [ij] \leqslant Q} [ij]$ is the maximal decomposition of $P \wedge Q$ as a join of atoms, thus $HD(P,Q) = \langle I_P, I_{P^\top} \rangle + \langle I_Q, I_{P^\top} \rangle - 2\langle I_{P \wedge Q}, I_{P^\top} \rangle$.

3.1 HD and VI: Axioms

Following (Meila, 2007), attention is now placed on those axioms that characterize both partition distance measures HD and VI.

Proposition 2. *HD is a metric: for all* $P, P', Q \in \Pi_N$,

1. $HD(P,Q) = HD(Q,P)$,
2. $HD(P,Q) \geq 0$, *with equality if and only if* $P = Q$,
3. $HD(P,P') + HD(P',Q) \geq HD(P,Q)$.

Proof. The first condition is obvious. In view of proposition 1, the second one is also immediate as $\min\{s(P), s(Q)\} \geq s(P \wedge Q)$. In fact, $HD(P,Q)$ is the sum $[s(P) - s(P \wedge Q)] + [s(Q) - s(P \wedge Q)]$ of two positive integers, while $\min_{P \neq Q} HD(P,Q) = 1 = \min_{P \in \Pi_N} s(P)$. As for the third condition, known as triangle inequality, difference $HD(P,P') + HD(P',Q) - HD(P,Q) = 2[s(P') - s(P \wedge P') - s(P' \wedge Q) + s(P \wedge Q)]$ must be shown to be ≥ 0 for all $P, P', Q \in \Pi_N$. Since size $s(P \wedge Q)$ is given, $s(P') - [s(P \wedge P') + s(P' \wedge Q)]$ has to be minimized by suitably choosing P'. Now, sum $s(P \wedge P') + s(P' \wedge Q)$ is maximized when $P \wedge P' = P$ (or $P' \geqslant P$) and $P' \wedge Q = Q$ (or $P' \geqslant Q$). If $Q \leqslant P' \geqslant P$, then $P' = P \vee Q$ minimizes the whole difference. Thus HD satisfies triangle inequality as long as the size is supermodular: $s(P \vee Q) - s(P) - s(Q) + s(P \wedge Q) \geq 0$ for all $P, Q \in \Pi_N$. The simplest way to see that this is indeed the case is by focusing on Möbius inversion of lattice (or more generally poset) functions (Rota, 1964b; Aigner, 1997; Stern, 1999). By definition, the size $s : \Pi_N \to \mathbb{Z}_+$ has Möbius inversion $\mu^s : \Pi_N \to \mathbb{Z}$ given by $\mu^s(P) = 1$ if P is an atom (i.e. $P = [ij] \in \Pi^N_{(1)}$), and $\mu^s(P) = 0$ otherwise. In fact,

$s(P) = \sum_{Q \leqslant P} \mu^s(Q)$ for all $P \in \mathcal{P}^N$. Hence the size satisfies a condition which is sufficient (but not necessary) for supermodularity, as its Möbius inversion takes only values ≥ 0. This completes the proof. □

Triangle inequality is satisfied with equality by HD (and VI) when $P' = P \wedge Q$ (for VI, see (Meila, 2007, pp. 883, 888) properties 6, 10(A.2)).

Proposition 3. *HD satisfies horizontal collinearity:* $HD(P, P \wedge Q) + HD(P \wedge Q, Q) = HD(P, Q)$.

Proof. $HD(P, P \wedge Q) + HD(P \wedge Q, Q) =$
$= [s(P) - s(P \wedge Q)] + [s(Q) - s(P \wedge Q)]$ as well as
$HD(P, Q) = s(P) + s(Q) - 2s(P \wedge Q)$. □

Collinearity also applies to distances between partitions P, Q that are comparable, i.e. either $P \geqslant Q$ or $Q \geqslant P$. Firstly consider the case involving the top $P^\top$ and bottom $P_\perp$ elements (for VI, see (Meila, 2007, p. 888) property 10(A.1)).

Proposition 4. *HD satisfies vertical collinearity:* $HD(P_\perp, P) + HD(P, P^\top) = HD(P_\perp, P^\top)$.

Proof. $HD(P_\perp, P) + HD(P, P^\top) =$
$= s(P) + s(P^\top) - s(P) = s(P^\top)$ independently from P, as well as $HD(P_\perp, P^\top) = s(P^\top) = \binom{n}{2}$. □

Vertical collinearity may be generalized for arbitrary comparable partitions $P^\top \geqslant P > Q \geqslant P_\perp$, in that $HD(Q, P') + HD(P', P) = HD(Q, P)$ for all P' satisfying $Q \leqslant P' \leqslant P$.

3.2 Complementation

The distance between the bottom and top elements considered by vertical collinearity leads to regard such lattice elements as complements, thereby focusing on the distance between other, generic complements. Maintaining the traditional Hamming distance between subsets as the fundamental benchmark, it must be taken into account that the subset and partition lattices are very different in terms of complementation. Every subset $A \in 2^N$ has a unique complement A^c, and the distance between any two such complements equals the distance between the bottom and top elements: $|A \Delta A^c| = n = |N \Delta \emptyset|$. Conversely, partitions P generally have several and quite different complements, which are all those Q such that $P \wedge Q = P_\perp$ as well as $P \vee Q = P^\top$. In this respect, MMD measures the distance between any two complements P, Q solely through their cardinalities $|P|, |Q|$, while VI and HD provide a fine distinction between different complements, and also agree on which are closer and which are remoter. The issue may be exemplified as follows: for $N = \{1, \ldots, 7\}$, consider $P = 123|456|7$ and $P^* = 147|2|3|5|6$ and $P_* = 1|2|34|5|67$ (where vertical bar $|$ separates blocks). Both P^* and P_* are complements of P, that is $P \wedge P^* = P \wedge P_* = P_\perp$ and $P \vee P^* = P \vee P_* = P^\top$. Here VI, HD and MMD are:
$VI(P, P_*) \simeq 1.93 < 1.95 \simeq VI(P, P^*)$,
$HD(P, P_*) = 8 < 9 = HD(P, P^*)$,
$MMD(P, P_*) = 4 = MMD(P, P^*)$.
For MMD this examples generalizes as follows.

Proposition 5. *For any two complements* $P, Q \in \mathcal{P}^N$, *it holds* $MMD(P, Q) = \max\{r(P), r(Q)\}$.

Proof. If $P \wedge Q = P_\perp$, then edges $\{A, B\} \in E \subset P \times Q$ of the bipartite graph $G = (P \cup Q, E)$ defined in section 2 have all same weight $1 = |A \cap B|$ (see above). Hence, a maximum-weight matching simply is one including the maximum number of feasible edges. In turn, such a number equals $\sum_{A \in P \vee Q} \min\{|P^A|, |Q^A|\}$, because each block (of either partition) can be the endpoint of at most one edge included in a matching. Also, the number of elements $i \in N$ that must be deleted for the two residual partitions to coincide is $\sum_{A \in P \vee Q} (|A| - \min\{|P^A|, |Q^A|\})$. On the other hand, $P \vee Q = P^\top$ entails $\sum_{A \in P \vee Q} (|A| - \min\{|P^A|, |Q^A|\}) =$
$= n - \min\{|P|, |Q|\} = \max\{r(P), r(Q)\}$. □

The *class* $c : \mathcal{P}^N \to \mathbb{Z}_+^n$ of partitions (Rota, 1964b) identifies the vector $c(P) = (c_1(P), \ldots, c_n(P))$ where $c_k(P)$ is the number of k-cardinal blocks of P, for $1 \leq k \leq n$. As shown by the above example, a partition generally has different complements with different classes. In this view, for all $P \in \mathcal{P}^N$ denote by $\mathcal{CO}(P) = \{Q : P \wedge Q = P_\perp, P \vee Q = P^\top\}$ the set of complements of P.

A modular element of the partition lattice (Aigner, 1997; Stern, 1999; Stanley, 1971) is any $P \in \mathcal{P}^N$ where all blocks are singletons apart from only one, at most, i.e. $\sum_{1<k\leq n} c_k(P) \leq 1$. The sublattice $\mathcal{P}^N_{mod}$ of modular elements contains the bottom and top elements, and all partitions of the form $\{A\} \cup P_\perp^{A^c}$ with $1 < |A| < n$. Hence, $|\mathcal{P}^N_{mod}| = 2^n - n$, while $\mathcal{P}^N_{mod} = \mathcal{P}^N$ for $n \leq 3$ and $\mathcal{P}^N_{mod} \subset \mathcal{P}^N$ for $n > 3$.

Here, the main link between modular elements and complementation is that an element is modular if and only if no two of its complements are comparable (see (Stanley, 1971, Theorem 1)). Therefore, if $P \notin \mathcal{P}^N_{mod}$, then there are $Q, Q' \in \mathcal{CO}(P)$ such that $Q > Q'$. It seems thus important that the distance between P and Q differs from the distance between P and Q'. The following result bounds the Hamming distance HD between a partition and its complements.

Proposition 6. *For all* $P \in \mathcal{P}^N$, *if* $Q \in \mathcal{CO}(P)$, *then* $s(P) + |P| - 1 \leq HD(P, Q) \leq s(P) + \binom{|P|}{2}$, *where the upper bound is always tight, while the lower one is tight only if* $c_1(P) \leq 2 + \sum_{1<k\leq n} (k-2) c_k(P)$.

Proof. If $Q \in \mathcal{CO}(P)$, then $HD(P,Q) = s(P) + s(Q)$. Hence $s(P) + \min\{s(Q) : Q \in \mathcal{CO}(P)\} \leq HD(P,Q)$ and $HD(P,Q) \leq s(P) + \max\{s(Q) : Q \in \mathcal{CO}(P)\}$. A complement of P has join-decompositions minimally involving $|P|-1$ atoms $[ij]_1, \ldots, [ij]_{|P|-1} \in \mathcal{P}^N_{(1)}$, with $|A_m \cap \{i,j\}_m| = 1 = |A_{m+1} \cap \{i,j\}_m|, 1 \leq m < |P|$. Considering the upper bound first, observe that size $s([ij]_1 \vee \cdots \vee [ij]_{|P|-1})$ attains its maximum when $|\{i,j\}_m \cap \{i,j\}_{m+1}| = 1$ for all $1 \leq m < |P|-1$, in which case $s([ij]_1 \vee \cdots \vee [ij]_m) = \binom{m+1}{2}, 1 \leq m < |P|$. This complement $P^* = [ij]_1 \vee \cdots \vee [ij]_{|P|-1}$ always exists, whatever the class $c(P)$ of P, making the bound tight. In fact, $P^* \in \mathcal{P}^N_{mod}$ has $n - |P| + 1$ blocks, out of which $n - |P|$ are singletons, while the remaining one $B \in P^*$ is $|P|$-cardinal and satisfies $|B \cap A| = 1$ for all $A \in P$, i.e. $P^* = \{B\} \cup P_\perp^{B^c} \Rightarrow s(P^*) = \binom{|P|}{2}$. For the lower bound, note that size $s([ij]_1 \vee \cdots \vee [ij]_{|P|-1})$ attains its minimum, ideally, when $\{i,j\}_m \cap \{i,j\}_{m'} = \emptyset$, $1 \leq m < m' < |P|$, i.e. $s([ij]_1 \vee \cdots \vee [ij]_m) = m$ for all $1 \leq m < |P|$. This is not always possible as each $A \in P$ can have non-empty intersection with a number of pair-wise disjoint pairs $\{i,j\}_m, 1 \leq m < |P|$ bounded above by $|A|$, entailing that the constraint is given by the number $c_1(P)$ of singletons $\{i\} \in P$. Specifically, nesting together $\sum_{1<k\leq n} c_k(P)$ non-singleton blocks requires $\sum_{1<k\leq n} c_k(P) - 1$ pairs $\{i,j\}_m$. If these latter have to be pair-wise disjoint, then the maximum number of elements $j \in N$ in non-singleton blocks available to match (into pair-wise disjoint pairs) those elements $\{i\} \in P$ in singleton blocks is precisely $\sum_{1<k\leq n} k c_k(P) - 2\left(\sum_{1<k\leq n} c_k(P) - 1\right)$. □

Proposition 7. *If* $2 + \sum_{1<k\leq n}(k-2)c_k(P) < c_1(P)$, *then* $\min_{P_* \in \mathcal{CO}(P)} s(P_*) = (n - \theta(P)\lfloor \frac{n}{\theta(P)} \rfloor)\binom{\lceil \frac{n}{\theta(P)} \rceil}{2} +$
$+[\theta(P)(\lfloor \frac{n}{\theta(P)} \rfloor + 1) - n]\binom{\lfloor \frac{n}{\theta(P)} \rfloor}{2}$,
where $\theta(P) = 1 + \sum_{1<k\leq n} c_k(P)(k-1)$.

Proof. If $2 + \sum_{1<k\leq n}(k-2)c_k(P) < c_1(P)$, then the above proof of proposition 6 entails that the maximum number $\max\{|Q| : Q \in \mathcal{CO}(P)\}$ of blocks of a complement of P is $\theta(P)$. Among $\theta(P)$-cardinal partitions P_*, the size is minimized when $|B| \in \{\lfloor \frac{n}{\theta(P)} \rfloor, \lceil \frac{n}{\theta(P)} \rceil\}$ for all $B \in P_*$, where the number of $\lfloor \frac{n}{\theta(P)} \rfloor$-cardinal blocks is $\theta(P)(\lfloor \frac{n}{\theta(P)} \rfloor + 1) - n$, while the number of $\lceil \frac{n}{\theta(P)} \rceil$-cardinal blocks is $n - \theta(P)\lfloor \frac{n}{\theta(P)} \rfloor$. □

Proposition 8. *Among complements* $Q \in \mathcal{CO}(P)$, *HD and VI have common minimizers and maximizers, i.e.* $\arg\min_{Q\in\mathcal{CO}(P)} HD(P,Q) = \arg\min_{Q\in\mathcal{CO}(P)} VI(P,Q)$, *and* $\arg\max_{Q\in\mathcal{CO}(P)} HD(P,Q) = \arg\max_{Q\in\mathcal{CO}(P)} VI(P,Q)$.

Proof. If $Q \in \mathcal{CO}(P)$, then $VI(P,Q)$ is minimized or maximized when $e(Q)$ is, respectively, maximized or minimized, as $VI(P,Q) = 2\log n - e(P) - e(Q)$. Given this, if $P \in \mathcal{P}^N_{mod}$, then all $Q \in \mathcal{CO}(P)$ have same rank. Otherwise, there are comparable complements, i.e. with different rank (see above). Thus, in general, among complements $Q \in \mathcal{CO}(P)$ entropy $e(Q)$ is minimized when $|Q|$ is minimized and, in addition, $Q \in \mathcal{P}^N_{mod}$. This is precisely where size $s(Q)$ is maximized. Similarly, $e(Q)$ is maximized when $|Q|$ is maximized and, in addition, $|B| \in \{\lfloor \frac{n}{|Q|} \rfloor, \lceil \frac{n}{|Q|} \rceil\}$ for all $B \in Q$. Again, this is where $s(Q)$ is minimized. □

4 MINIMUM-WEIGHT PATHS

Hamming distance $|E \Delta E'|$ between edge sets $E, E' \in 2^{N_2}$ is the length of a shortest path between vertices $\chi_E, \chi_{E'} \in \{0,1\}^{\binom{n}{2}}$ of the $\binom{n}{2}$-dimensional unit hypercube $[0,1]^{\binom{n}{2}}$, where $\chi_E : N_2 \to \{0,1\}$ is the characteristic function defined above in section 2, i.e. $\chi_E(\{i,j\}) = 1$ if $\{i,j\} \in E$ and 0 otherwise. Recall that a polytope naturally defines a graph with its same vertices and edges (Brøondsted, 1983, p. 93), and the hypercube is perhaps the main example of polytope. In fact, the graph of hypercube $[0,1]^{\binom{n}{2}}$ is the Hasse diagram of Boolean lattice $(2^{N_2}, \cap, \cup)$, for its edges correspond to the covering relation, that is to say $\{E, E'\}$ is an edge of the hypercube if either $E \supset E', |E| = |E'| + 1$ or else the converse, i.e. $E' \supset E, |E'| = |E| + 1$.

Clearly, a shortest path is a minimum-weight path as long as every edge has weight 1. This simple observation is the starting point towards an analog view of the Hamming distance HD between partitions, namely as the weight of a minimum-weight path in the associated Hasse diagram. To this end, define the polytope of partitions $\mathbb{P}$ as the convex hull $\mathbb{P} := conv(\{I_P : P \in \mathcal{P}^N\}) \subset [0,1]^{\binom{n}{2}}$ containing all convex combinations of the $\mathcal{B}_n$ Boolean vectors defined by the indicator functions of partitions. Also denote by $\mathbb{G} = (\mathcal{P}^N, \mathbb{E})$ the graph of polytope $\mathbb{P}$ or, equivalently, the Hasse diagram of partition lattice $(\mathcal{P}^N, \wedge, \vee)$. Edges correspond to the covering relation: $\{P, Q\} \in \mathbb{E}$ if either $\{P' : Q \leqslant P' \leqslant P] = \{P, Q\}$ or else $\{P' : P \leqslant P' \leqslant Q\} = \{P, Q\}$ (see above). Denote this relation by $P \gtrdot Q \Leftrightarrow \{P, Q\} \in \mathbb{E}$. Finally, let $\mathbb{F} \subset \mathbb{R}^{\mathcal{B}_n}$ be the vector space of *symmetric* and *order-preserving* partition functions $f : \mathcal{P}^N \to \mathbb{R}$, that is to say, respectively, for all $P, Q \in \mathcal{P}^N$,
(a) $c(P) = c(Q) \Rightarrow f(P) = f(Q)$, and
(b) $P > Q \Rightarrow f(P) > f(Q)$ or else
(b') $P > Q \Rightarrow f(P) < f(Q)$.

Entropy, rank and size $e, r, s : \mathcal{P}^N \to \mathbb{R}$ are in $\mathbb{F}$, with e satisfying (b') and r, s satisfying (b). For $f \in \mathbb{F}$, let weights $w_f : \mathbb{E} \to \mathbb{R}_{++}$ on edges $\{P,Q\} \in \mathbb{E}$ of $\mathbb{G}$ be $w_f(\{P,Q\}) = \max\{f(P), f(Q)\} - \min\{f(P), f(Q)\}$. For $P, Q \in \mathcal{P}^N$, let $Path(P,Q)$ contain all $P-Q$-paths $p(P,Q)$ in graph $\mathbb{G}$. Any $p(P,Q) \in Path(P,Q)$ is a subgraph $p(P,Q) = (V^p_{P,Q}, E^p_{P,Q}) \subset \mathbb{G}$ with vertex set $V^p_{P,Q} = \{P = P_0, P_1, \ldots, P_m = Q\}$ and edge set $E^p_{P,Q} = \{\{P_0, Q_0\}, \{P_1, Q_1\}, \ldots, \{P_{m-1}, Q_{m-1}\}\}$, where $P_{k+1} = Q_k, 0 \leq k < m$. Graph $\mathbb{G}$ is connected, entailing $Path(P,Q) \neq \emptyset$ for all $P, Q \in \mathcal{P}^N$. The weight of $p(P,Q)$ is $w_f(p(P,Q)) = \sum_{0 \leq k < m} w_f(\{P_k, Q_k\})$.

Definition 9. *For $f \in \mathbb{F}$, minimum-f-weight partition distance $\delta_f : \mathcal{P}_N \times \mathcal{P}_N \to \mathbb{R}_+$ is*

$$\delta_f(P,Q) := \min_{p(P,Q) \in Path(P,Q)} w_f(p(P,Q)). \qquad (7)$$

Proposition 10. *For $f \in \mathbb{F}$ and $P, Q \in \mathcal{P}^N$, every minimum-f-weight $P-Q$-path visits $P \wedge Q$ or $P \vee Q$ or both, i.e. $V^p_{P,Q} \cap \{P \wedge Q, P \vee Q\} \neq \emptyset$ for all $p(P,Q)$ satisfying $w_f(p(P,Q)) = \delta_f(P,Q)$.*

Proof. If $P \geqslant Q$, then $\{P \vee Q, P \wedge Q\} \subseteq V^p_{P,Q}$ for all $p(P,Q) \in Path(P,Q)$, with $P \geqslant Q \Rightarrow \{P,Q\} = V^p_{P,Q}$. Differently, if $P \not\geqslant Q \not\geqslant P$, then any path $p(P,Q)$ visits some vertex P' comparable with both P, Q and, in particular, satisfying either $P' > P, Q$ or else $P, Q > P'$. Accordingly, $p(P,Q) = p(P,P') \cup p(P',Q)$, with $E^p_{P,P'} \cap E^p_{P',Q} = \emptyset$, for some $P-P'$-path $p(P,P')$ and $P'-Q$-path $p(P',Q)$, entailing that $w_f(p(P,Q))$ equals $w_f(p(P,P')) + w_f(p(P',Q))$. Finally, since f is order-preserving and symmetric, $P' = P \vee Q$ minimizes $w_f(p(P,P')) + w_f(p(P',Q))$ over all vertices $P' > P, Q$ as well as $P' = P \wedge Q$ minimizes $w_f(p(P,P')) + w_f(p(P',Q))$ over all $P' < P, Q$. □

Whether a minimum-f-weight path visits the join or else the meet of any two incomparable partitions clearly depends on f. A generic $f \in \mathbb{F}$ may have associated minimum-weight paths visiting the meet of some incomparable partitions P, Q and the join of some others P', Q'. Whether minimum-weight paths awlays visit the meet or else the join depends on whether f or else $-f$ is supermodular. Note that if f is supermodular, then $-f$ is submodular, i.e. $-f(P \wedge Q) - f(P \vee Q) \leq -f(P) - f(Q)$.

Proposition 11. *Let $f \in \mathbb{F}$ satisfy (b), i.e. $P > Q$ entails $f(P) > f(Q)$. If f is supermodular, then the minimum-f-weight partition distance is*

$$\delta_f(P,Q) = f(P) + f(Q) - 2f(P \wedge Q),$$

while if f is submodular, then the minimum-f-weight partition distance is

$$\delta_f(P,Q) = 2f(P \vee Q) - f(P) - f(Q).$$

Proof. Supermodularity entails
$2f(P \vee Q) - f(P) - f(Q) \geq f(P \vee Q) - f(P \wedge Q)$ and $f(P \vee Q) - f(P \wedge Q) \geq f(P) + f(Q) - 2f(P \wedge Q)$, whereas submodularity entails
$2f(P \vee Q) - f(P) - f(Q) \leq f(P \vee Q) - f(P \wedge Q)$ and $f(P \vee Q) - f(P \wedge Q) \leq f(P) + f(Q) - 2f(P \wedge Q)$, for all $P, Q \in \mathcal{P}^N$. □

Since the size s is supermodular (see proposition 2 above), Hamming distance HD is the minimum-s-weight partition distance, i.e. $HD(P,Q) = \delta_s(P,Q)$ for all $P, Q \in \mathcal{P}^N$. The rank r of partitions being submodular (Aigner, 1997, pp. 259, 265, 274), minimum-r-weight distance is $\delta_r(P,Q) = |P| + |Q| - 2|P \vee Q|$. In fact, $w_r(\{P,Q\}) = 1$ for all edges $\{P,Q\} \in \mathbb{E}$, hence δ_r is a shortest-path distance.

Turning to entropy e, a simple example shows that VI distance does not correspond to the e-based minimum-weight distance.

Proposition 12. *There are partitions $P, Q \in \mathcal{P}^N$ such that $2e(P \wedge Q) - e(P) - e(Q) > \delta_e(P,Q)$.*

Proof. For two atoms $[ij], [ij'] \in \mathcal{P}^N_{(1)}$, with non empty intersection $\{i,j\} \cap \{i,j'\} = \{i\}$, VI distance is $VI([ij],[ij']) = 2e([ij] \wedge [ij']) - e([ij]) - e([ij']) = = 2\log n - 2\left(\log n - \frac{2}{n}\right) = \frac{4}{n}$, while minimum-$e$-weight distance is $e([ij]) + e([ij']) - 2e([ij] \vee [ij']) = = 2\left(\log n - \frac{2}{n}\right) - 2\left(\log n - \frac{3}{n}\log 3\right) = = \frac{2}{n}(3\log 3 - 2) = \delta_e([ij],[ij'])$, with $3\log 3 < 4$ entailing $VI([ij],[ij']) > \delta_e([ij],[ij'])$. □

An alternative measure of partition entropy, called logical entropy, has been recently proposed (Ellerman, 2013a) in terms of distinctions or *ordered* pairs $(i,j) \in N \times N$, hence $(i,j) \neq (j,i)$. If distinctions are replaced with unordered pairs $\{i,j\} \in N_2$, then *mutatis mutandis* the non-normalized logical entropy of partitions P is the analog of $\binom{n}{2} - s(P)$, providing a further minimum-weight partition distance. Also, since in information theory partitions are evaluated through functions f such that $P > Q \Rightarrow f(P) < f(Q)$, the approach developed thus far may be applied to the upside-down Hasse diagram of the partition lattice, with co-atoms in place of atoms, as detailed below.

4.1 Distinctions, Co-atoms and Fields

A partition P distinguishes between $i \in N$ and $j \in N \backslash i$ if $i \in A \in P$ while $j \in B \in P$ with $A \neq B$, and the set of such distinctions has been recently proposed as the logical analog of the complement of P, with the (normalized) number of distinctions providing a novel measure of the (logical) entropy of partitions (Ellerman, 2013b; Ellerman, 2013a). This achieves through apartness relations $\mathcal{R}^c$, which are the complement of

equivalence relations $\mathcal{R}$, both being sets of ordered pairs $(i,j) \in N \times N$ (see section 2 above). In terms of atoms $[ij] \in \mathcal{P}^N_{(1)}$, the logical entropy $h : \mathcal{P}^N \to \mathbb{R}_+$ of partitions (Ellerman, 2013a, p. 127) is $h(P) = \frac{2|\{[ij]:P \not\geqslant [ij]\}|}{n^2} = \frac{2\left(\binom{n}{2} - s(P)\right)}{n^2} = \frac{n(n-1)-2s(P)}{n^2}$, with $h(P^\top) = 0 = s(P_\perp)$ and $h(P_\perp) = \frac{n-1}{n} = \frac{2s(P^\top)}{n^2}$.

Proposition 13. *The minimum-h-weight distance is* $\delta_h(P,Q) = 2h(P \wedge Q) - h(P) - h(Q)$.

Proof. Logical entropy h is symmetric and satisfies $P > Q \Rightarrow h(P) < h(Q)$, hence $h \in \mathbb{F}$. Also, apart from constant terms, h varies with $-s$, which is submodular because s is supermodular. That is to say, $h(P) + h(Q) = \frac{2}{n}\left(n - 1 - \frac{s(P)+s(Q)}{n}\right)$ and $h(P \wedge Q) + h(P \vee Q) = \frac{2}{n}\left(n - 1 - \frac{s(P \wedge Q)+s(P \vee Q)}{n}\right)$. Thus $s(P \wedge Q) + s(P \vee Q) \geq s(P) + s(Q)$ entails $h(P \wedge Q) + h(P \vee Q) \leq h(P) + h(Q)$. Also, like in proposition 11 above but with reversed inequalities, $2h(P \wedge Q) - h(P) - h(Q) \leq h(P \wedge Q) - h(P \vee Q)$ and $h(P \wedge Q) - h(P \vee Q) \leq h(P) + h(Q) - 2h(P \vee Q)$. □

Reasoning in terms of ordered pairs results in a double counting, in that $(i,j) \in \mathcal{R}^c \Rightarrow (j,i) \in \mathcal{R}^c$ for all apartness relations $\mathcal{R}^c$ and all $(i,j) \in N \times N$. Hence an analog logical entropy $\hat{h}$ of partitions may be defined in terms of unordered pairs $\{i,j\} \in N_2$ or atoms $[ij] \in \mathcal{P}^N_{(1)}$ by $\hat{h}(P) = \frac{\binom{n}{2} - s(P)}{\binom{n}{2}} = 1 - \frac{s(P)}{\binom{n}{2}}$. Again, $\hat{h} \in \mathbb{F}$ and $\hat{h}(P^\top) = 0$ as well as $\hat{h}(P_\perp) = 1$. Therefore, the minimum-$\hat{h}$-weight distance is $\delta_{\hat{h}}(P,Q) = 2\hat{h}(P \wedge Q) - \hat{h}(P) - \hat{h}(Q)$ for all $P,Q \in \mathcal{P}^N$.

On the other hand, a distance between partitions also obtains by dealing directly with their associated set of distinctions: let $D_P = \{[ij] : P \not\geqslant [ij]\}$ and consider the distance between any two partitions P,Q given by the traditional Hamming distance between their sets of (unordered) distinctions, i.e. $|D_P \Delta D_Q|$. In particular, $|D_P \Delta D_Q| = \frac{n^2}{2}(2h(P \wedge Q) - h(P) - h(Q))$. In view of proposition 13 above, this is the non-normalized minimum-h-weight distance.

A field of subsets is a set system $\mathcal{F} \subseteq 2^N$ which is closed under union, intersection and complementation, hence $A \cap B, A \cup B, A^c \in \mathcal{F}$ for all $A,B \in \mathcal{F}$. Every partition $P \in \mathcal{P}^N$ generates the field $\mathcal{F}_P := 2^P$ containing all subsets $B \in 2^N$ obtained as the union of blocks $A \in P$, with $\mathcal{F}_{P_\perp} = 2^N$ as well as $\mathcal{F}_{P^\top} = \{\emptyset, N\}$. There are $2^{n-1} - 1$ minimal fields that strictly include $\mathcal{F}_{P^\top}$; they are those $\mathcal{F}_A = \mathcal{F}_{A^c} = \{\emptyset, A, A^c, N\}$ with $\emptyset \subset A \subset N$. On the other hand, 2-cardinal partitions $\{A, A^c\} \in \mathcal{P}^N$ are the co-atoms (Aigner, 1997) of partition lattice $(\mathcal{P}^N, \wedge, \vee)$ ordered by coarsening. In fact, in information theory finer partitions are generally more valuable than coarser ones, and thus attention is placed on partition functions f such as entropy e or logical entropy h satisfying $f(P) < f(Q)$ whenever $P > Q$. In this view, the partition lattice is often dealt with as ordered by refinement and thus with the upside-down Hasse diagram. Accordingly, a distance between partitions also obtains by counting co-atoms rather than atoms. Define the co-size $cs : \mathcal{P}^N \to \mathbb{Z}_+$ of partitions by $cs(P) = |\{\{A,A^c\} : P \leqslant \{A,A^c\}\}|$, with $cs(P_\perp) = 2^{n-1} - 1$ and $cs(P^\top) = 0$. In words, $cs(P)$ is the number of co-atoms coarser than P.

Proposition 14. *The minimum-cs-weight partition distance is* $\delta_{cs}(P,Q) = cs(P) + cs(Q) - 2cs(P \vee Q)$.

Proof. Denote by $\hat{\mu}^{cs} : \mathcal{P}^N \to \mathbb{Z}$ the Möbius inversion from above (Rota, 1964b; Aigner, 1997) of the co-size, with $cs(P) = \sum_{Q \geqslant P} \hat{\mu}^{cs}(Q)$ for all P. By definition, $\hat{\mu}^{cs}(P) = 1$ if $|P| = 2$ and 0 otherwise. Like for the size in proposition 2, this entails supermodularity, i.e. $cs(P \wedge Q) + cs(P \vee Q) \geq cs(P) + cs(Q)$. Also, $cs \in \mathbb{F}$ with $cs(P) < cs(Q)$ whenever $P > Q$, thus $cs(P) + cs(Q) - 2cs(P \vee Q) \leq cs(P \wedge Q) - cs(P \vee Q)$, $cs(P \wedge Q) - cs(P \vee Q) \leq 2cs(P \wedge Q) - cs(P) - cs(Q)$ for all $P,Q \in \mathcal{P}^N$. □

Denote by $(\mathfrak{F}, \sqcap, \sqcup)$ the lattice whose elements are the $\mathcal{B}_n$ fields of subsets $\mathcal{F}_P, P \in \mathcal{P}^N$ ordered by inclusion $\supseteq$. The meet and join are, respectively, $\mathcal{F}_P \sqcap \mathcal{F}_Q = \mathcal{F}_{P \vee Q}$ and $\mathcal{F}_P \sqcup \mathcal{F}_Q = \mathcal{F}_{P \wedge Q}$. The set of atoms is the collection $\{\mathcal{F}_{\{A,A^c\}} : \emptyset \subset A \subset N\}$ of minimal fields, i.e. $\mathcal{F}_P = \bigsqcup_{\{A,A^c\} \geqslant P} \mathcal{F}_{\{A,A^c\}}$ for all $\mathcal{F}_P \in \mathfrak{F}$. Thus $\delta_{cs}(P,Q)$ is also an analog of the traditional Hamming distance between subsets: $\delta_{cs}(P,Q) = |\{\{A,A^c\} : \mathcal{F}_{\{A,A^c\}} \subseteq \mathcal{F}_P\}| + |\{\{A,A^c\} : \mathcal{F}_{\{A,A^c\}} \subseteq \mathcal{F}_Q\}| - 2|\{\{A,A^c\} : \mathcal{F}_{\{A,A^c\}} \subseteq (\mathcal{F}_Q \cap \mathcal{F}_P)\}|$. In words, this is the number of minimal fields $\mathcal{F}_{\{A,A^c\}}$ included in either $\mathcal{F}_P$ or else in $\mathcal{F}_Q$, but not in both.

5 THE CONSENSUS PARTITION PROBLEM

Hamming distance between partitions HD firstly appears in the mid '60s (Rénier, 1965) in terms of the *consensus partition* problem, which is important in many applicative scenarios concerned with statistical classification. From a combinatorial optimization perspective, after selecting a metric $\delta : \mathcal{P}^N \times \mathcal{P}^N \to \mathbb{R}_+$, an instance is a m-collection $P_1, \ldots, P_m \in \mathcal{P}^N$, $m \geq 2$, and the objective is to find a partition $\hat{P}$ minimizing the sum of its distances from the m partitions: any $\hat{P}$ satisfying $\sum_{1 \leq k \leq m} \delta(\hat{P}, P_m) \leq \sum_{1 \leq k \leq m} \delta(Q, P_k)$ for all

$Q \in \mathcal{P}^N$ is a consensus partition. For generic δ, finding a solution $\hat{P}$ is tipically hard. If $\delta = MMD$, then each distance $\delta(Q,P_k), 1 \leq k \leq m$ for any $Q \in \mathcal{P}^N$ is computable in $O(n^3)$ time (Korte and Vygen, 2002, p. 236), whereas if $\delta = HD$, then in view of expression (6) above (see section 3) $\delta(Q,P_k)$ is computable more rapidly through scalar products. Independently from the chosen metric δ, the main issue is that the size $\mathcal{B}_n = |\mathcal{P}^N|$ of the search space $\mathcal{P}^N$ makes all approaches relying on direct enumeration simply unviable, at least for relevant values of n. The problem is thus commonly interpreted in terms of heuristics (Pinto Da Costa and Rao, 2004; Celeux et al., 1989).

Although the consensus problem is generally harsh, still the analysis conducted thus far identifies conditions where exact solutions are easy to find. If the chosen metric is a minimum-f-weight partition distance, i.e. $\delta = \delta_f$ with $f \in \mathbb{F}$, and weighting function f is either supermodular or else submodular (but not both, see below), then either the meet $\hat{P} = P_1 \wedge \cdots \wedge P_m$ or else the join $\hat{P} = P_1 \vee \cdots \vee P_m$ of instance elements are consensus partitions. The former case applies to Hamming distance or size-based $\delta_s = HD$ and to logical entropy-based δ_h, while the latter applies to rank-based δ_r and to co-size-based δ_{cs}. The computational burden thus reduces solely to assessing the m distances between instance elements and their meet (or their join), with no search need.

Proposition 15. *If distances between partitions are measured by HD, then the meet of instance elements achieves consensus: for $Q,P_1,\ldots,P_m \in \mathcal{P}^N$,*

$$\sum_{1\leq k\leq m} HD(P_1 \wedge \cdots \wedge P_m, P_k) \leq \sum_{1\leq k\leq m} HD(Q,P_k).$$

Proof. Firstly note that for $m = 2$ this consensus condition is in fact a restatement of horizontal collinearity and triangle inequality (see propositions 2 and 3 above). Hence, in order to use induction, assume that the condition holds for some $m \geq 2$, and denote by $\hat{P}$ the solution or consensus partition of a $m+1$-instance $P_1 \ldots, P_m, P_{m+1}$. By assumption, $P_1 \wedge \cdots \wedge P_m$ is a solution of instance $P_1,\ldots,P_m$, thus novel solution $\hat{P}$ minimizes the sum of its distances from the previous solution $P_1 \wedge \cdots \wedge P_m$ and from the novel instance element P_{m+1}, i.e. for all $Q \in \mathcal{P}^N$,
$HD(P_1 \wedge \cdots \wedge P_m, \hat{P}) + HD(\hat{P}, P_{m+1}) \leq$
$\leq HD(P_1 \wedge \cdots \wedge P_m, Q) + HD(Q, P_{m+1})$ Then, horizontal collinearity and triangle inequality entail
$HD(P_1 \wedge \cdots \wedge P_m, \hat{P}) + HD(\hat{P}, P_{m+1}) \geq$
$\geq HD(P_1 \wedge \cdots \wedge P_m, P_{m+1})$, with equality if
$\hat{P} = P_1 \wedge \cdots \wedge P_m \wedge P_{m+1}$. □

The consensus partition problem may be also framed in a novel manner through fuzzy modeling. A fuzzy subset of N is a function $q : N \to [0,1]$ or, geometrically, a point $q = (q_1,\ldots,q_n) \in [0,1]^n$ in the n-dimensional unit hypercube, where $q_i = q(i), i \in N$. A fuzzy partition is thus commonly intended as a partition P with associated $|P|$ points $q^A \in [0,1]^n, A \in P$ in the hypercube such that $q^A_i \in (0,1]$ for all $i \in A$ and all $A \in P$. Also, a fuzzy or random graph with vertex set N may be seen as one whose edge set is a fuzzy subset of N_2, i.e. a function $t : N_2 \to [0,1]$ or, geometrically, a point in the $\binom{n}{2}$-dimensional unit hypercube, i.e. $t = \left(t_{\{i,j\}_1},\ldots,t_{\{i,j\}_{\binom{n}{2}}}\right) \in [0,1]^{\binom{n}{2}}$.

By looking at partitions of N as graphs with vertex set N each of whose components is complete (see above), the fuzzy consensus partition t_I associated with instance $I \subseteq \mathcal{P}^N$ may be defined as the point in the interior of the polytope $\mathbb{P}$ of partitions (see above) corresponding to the center of the convex hull $conv(\{I_P : P \in I\})$ consisting of all convex combinations of the indicator functions $I_P, P \in I$ of instance elements. Then, the fuzzy consensus partition is a function ranging in the unit interval $[0,1]$ and taking values on the atoms of $\mathcal{P}^N$, i.e. $t_I : \mathcal{P}^N_{(1)} \to [0,1]$ and $t_I([ij]) = \frac{1}{|I|}\sum_{P\in I} I_P([ij])$ for all atoms $[ij] \in \mathcal{P}^N_{(1)}$. In this framework, the *strong patterns* of instance I considered in (Pinto Da Costa and Rao, 2004) are the blocks of partition $P(t_I)$ obtained through defuzzification of t_I as follows: $P(t_I) = \bigvee_{t_I([ij])=1} [ij]$. In words, $P(t_I)$ obtains as the join of all atoms where the fuzzy consensus partition t_I attains its maximum, i.e. 1.

6 CONCLUSIONS

Measuring the distance between partitions is an important topic in statistical classification since the '60s (Lerman, 1981; Rénier, 1965). This works considers the analog of the traditional Hamming distance between subsets by counting unordered pairs of partitioned elements. Counting ordered and/or unordered pairs is not new (see (Meila, 2007, Section 2.1)), but the Hamming distance HD is here analyzed from a novel geometric perspective. Special attention is placed on complements in comparison with two distances proposed in recent years, namely MMD and VI. Given its low computational complexity and fine measurement sensitivity, HD seems interesting for applications, especially in bioinformatics.

HD relies on the size, which counts the atoms finer than partitions. While the cardinality (or rank) of subsets is a valuation (i.e. supermodular and submodular), the size is supermodular. In fact, if f is a valuation of the partition lattice, then it is constant, i.e. $f(P) = f(Q)$ for all partitions P,Q (Aigner, 1997).

Since the Hamming distance between $A, B \in 2^N$ is $|A\Delta B| = |A \cup B| - |A \cap B|$, it may seem reasonable to consider distances $\delta(P,Q)$ between $P,Q \in \mathcal{P}^N$ of the form $\delta(P,Q) = f(P \vee Q) - f(P \wedge Q)$ with $f \in \mathbb{F}$. Yet, this clearly does not distinguish between different complements $Q, Q' \in \mathcal{CO}(P)$ when $P \neq P_\perp, P^\top$.

The geometric approach enables to analyze further partition distances obtained by replacing the size with alternative partition functions such as entropy, rank and logical entropy, where these latter two are submodular. In general, any symmetric and order-preserving partition function f provides a distance between partitions P,Q by considering $f(P), f(Q)$ and the values taken on their meet $f(P \wedge Q)$ or else on their join $f(P \vee Q)$. Specifically, f defines weights on edges of the Hasse diagram of partitions such that the corresponding partition distance between any P,Q is the weight of a lightest $P - Q$-path.

REFERENCES

Aigner, M. (1997). *Combinatorial Theory*. Springer.

Almudevar, A. and Field, C. (1999). Estimation of single-generation sibling relationships based on DNA markers. *Journal of Agricultural, Biological and Environmental Statistics*, 4(2):136–165.

Berger-Wolf, T. Y., Sheikh, S. I., DasGupta, B., Ashley, M. V., Caballero, I. C., Chaovalitwongse, W., and Putrevu, S. L. (2007). Reconstructing sibling relationship in wild populations. *Bioinf.*, 23(13):i49–i56.

Bollobas, B. (1986). *Combinatorics. Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability*. Cambridge University Press.

Brøondsted, A. (1983). *An introduction to convex polytopes*. Springer.

Brown, D. G. and Dexter, D. (2012). Sibjoin: a fast heuristic for half-sibling reconstruction. *Algorithms in Bioinformatics*, LNCS 7534:44–56.

Celeux, G., Diday, E., Govaert, G., Lechevalier, G., and Ralambondrainy, H. (1989). *Classification Automatique Des Données*. Dunod.

Day, W. (1981). The complexity of computing metric distances between partitions. *Math. Soc. Sc.*, 1(3):269–287.

Deza, M. M. and Deza, E. (2013). *Encyclopedia of Distances - Second Edition*. Springer.

Ellerman, D. (2013a). An introduction to logical entropy and its relation to Shannon entropy. *International Journal of Semantic Computing*, 7(2):121–145.

Ellerman, D. (2013b). An introduction to partition logic. *Logic Journal of the IGPL*, 22(1):94–125.

Godsil, C. and Royle, G. F. (2001). *Algebraic Graph Theory*. Springer.

Graham, R., Knuth, D., and Patashnik, O. (1994). *Concrete Mathematics*. Addison-Wesley.

Grünbaum, B. (2001). *Convex Polytopes*. Springer.

Gusfield, D. (2002). Partition-distance: A problem and class of perfect graphs arising in clustering. *Information Processing Letters*, 82:159–164.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.

Konovalov, D. A. (2006). Accuracy of four heuristics for the full sibship reconstruction problem in the presence of genotype errors. *Adv. Bioinf. Comp. Bio.*, 3:7–16.

Konovalov, D. A., Bajema, N., and Litow, B. (2005a). Modified Simpson $O(n^3)$ algorithm for the full sibship reconstruction problem. *Bioinf.*, 21(20):3912–3917.

Konovalov, D. A., Litow, B., and Bajema, N. (2005b). Partition-distance via the assignment problem. *Bioinf.*, 21(10):2463–2468.

Korte, B. and Vygen, J. (2002). *Combinatorial Optimization: Theory and Algorithms (2nd edition)*. Springer.

Lerman, I. C. (1981). *Classification et Analyse Ordinale des Données*. Dunod.

Meila, M. (2007). Comparing clusterings - an information based distance. *J. of Mult. Ananysis*, 98(5):873–895.

Mirkin, B. G. (1996). *Mathematical Classification and Clustering*. Kluwer Academic Press.

Mirkin, B. G. and Cherny, L. B. (1970). Measurement of the distance between distinct partitions of a finite set of objects. *Aut. and Rem. Con.*, 31(5):786–792.

Mirkin, B. G. and Muchnik, I. (2008). Some topics of current interest in clustering: Russian approaches 1960-1985. *Electronic Journal for History of Probability and Statistics*, 4(2):1–12.

Pinto Da Costa, J. F. and Rao, P. R. (2004). Central partition for a partition-distance and strong pattern graph. *REVSTAT - Statistical Journal*, 2(2):127–143.

Rénier, S. (1965). Sur quelques aspects mathématiques des problémes de classification automatique. *ICC Bulletin*, 4:175–191. Reprinted in Mathématiques et Sciences Humaines 82:13-29, 1983.

Rossi, G. (2011). *Partition distances*. arXiv:1106.4579v1.

Rota, G.-C. (1964a). The number of partitions of a set. *American Mathematical Monthly*, 71:499–504.

Rota, G.-C. (1964b). On the foundations of combinatorial theory I: theory of Möbius functions. *Z. Wahrscheinlichkeitsrechnung u. verw. Geb.*, 2:340–368.

Sebő, A. and Tannier, E. (2004). On metric generators of graphs. *Math. of Op. Res.*, 29(2):383–393.

Sheikh, S. I., Berger-Wolf, T. Y., Khokhar, A. A., Caballero, I. C., Ashley, M. V., Chaovalitwongse, W., Chou, C.-A., and DasGupta, B. (2010). Combinatorial reconstruction of half-sibling groups from microsatellite data. *J. Bioinf. Comp. Biol.*, 8(2):337–356.

Stanley, R. (1971). Modular elements of geometric lattices. *Algebra Universalis*, (1):214–217.

Stern, M. (1999). *Semimodular Lattices. Theory and Applications. Encyclopedia of Mathematics and its Applications 73*. Cambridge University Press.

Warrens, M. J. (2008). On the equivalence of Chen's Kappa and the Hubert-Arabie adjusted Rand index. *Journal of Classification*, 25(1):177–183.

Whitney, H. (1935). On the abstract properties of linear dependence. *Amer. J. of Math.*, 57:509–533.