Hamming distance between partitions, clustering comparison and information

Giovanni Rossi

Abstract—Measuring the distance between partitions is useful for clustering comparison in different fields. For example, in bioinformatics the measuring mostly obtains through a maximum matching distance MMD, although this is algorithmically demanding and hardly fits certain instances. In fact, another distance measure is being tested, namely one based on information theory and termed variation of information VI. Alternatively, this paper proposes the Hamming distance HD, displaying large range and great measurement sensitivity, while also relying on a neat binary string representation of partitions. Novel distance measure is being tested, namely one based on information theory and termed maximum matching MMD. In this view, a notable exception relies on the assignment problem (see [20, p. 236] and [12, chapter 4]). Secondly, the Hamming distance HD between partitions provided here obtains simply by means of scalar products between (input) vectors, no proper computational issue arises, while determining MMD and VI is algorithmically demanding (see [15], [19], [18] and [12, chapter 4]).

For a finite set \( N = \{1, \ldots, n\} \), let \( (2^N, \cap, \cup) \) and \( (\mathcal{P}(N), \wedge, \vee) \) denote the corresponding subset and partition lattices, with inclusion \( \supseteq \) and coarsening \( \supseteq \) as order relations, respectively. Both are atomic (and atomistic), but the former is distributive while the latter is geometric [1], [32]. For populations or data sets, \( n \) is the number of individuals or data points being partitioned.

The distance between elements of a ordered set is to be measured in terms of the order relation. Also, measures of the distance between elements of any set are called Hamming distances when these 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 \( d(A, B) \) between two subsets \( A, B \in 2^N \) is

\[
d(A, B) = |A \Delta B| = |A \setminus B| + |B \setminus A| = r(A \cup B) - r(A \cap B),
\]

where \( r : 2^N \to \mathbb{Z}_+ \) being the rank function: \( r(A') = |A'| \) for all \( A' \in 2^N \). In words, it counts how many \( i \in N \) are included in either \( A \) or else \( B \), but not in both. Elements \( i \in N \), when regarded as 1-cardinal subsets \( \{i\} \in 2^N \), are atoms in subset lattice \( (2^N, \cap, \cup) \). Expression (1) defines indeed a Hamming distance in that subsets \( A \subset 2^N \) are firstly represented as binary \( n \)-vectors \( \chi_A \in \{0,1\}^n \) (or vertexes of the \( n \)-dimensional unit hypercube \([0,1]^n\)) 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 \( i \in N \setminus A \). Next, the distance between any two of them \( A, B \subset 2^N \) is the number of entries where \( \chi_A \) and \( \chi_B \) differ. That is, the cardinality of their symmetric difference \( |A \Delta B| \) [7], [1].

A partition \( P = \{A_1, \ldots, A_P\} \subset 2^N \) of \( N \) is a collection of pairwise disjoint subsets, called blocks (or clusters), whose union yields \( N \). Any subset \( A \subset 2^N \) has a unique complement \( A^c = N \setminus A \). For all partitions \( P \in \mathcal{P}(N) \) and non-empty
subsets $\emptyset \subset A \subseteq N$, denote the partition of $A$ induced by $P$ by $P^A = \{B \cap A : B \in P, \emptyset \neq B \cap A\}$. The maximum matching distance $MMD : \mathcal{P}^N \times \mathcal{P}^N \to \{0, 1, \ldots, n - 1\}$ between any two partitions $P, Q$ is

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

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 $Q$ ([15, p. 160]).

Consider the bipartite graph $G = (P \cup Q, E)$ with $|P| + |Q|$ vertexes ($\cup$ is the union of disjoint sets), 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 weighted 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 (minimal) number $MMD(P, Q)$ of elements that must be removed for the two residual partitions to coincide is the sum $\sum_{(A, B) \in E^*} |\Delta B|$ 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 $VI(P, Q)$. Firstly obtained axiomatically from information theory [22], this distance is now useful in bioinformatics as well [9]. Entropy $H(P) = -\sum_{A \in P} \frac{|A|}{n} \log \left(\frac{|A|}{n}\right)$ of a partition $P$ and mutual information $I(P, Q) = \sum_{A \in P, B \in Q} \frac{|A \cap B|}{n} \log \left(\frac{n|A \cap B|}{|A||B|}\right)$ between $P$ and $Q$ (binary logarithm) enable to measure the distance between these latter as variation of information

$$VI(P, Q) = H(P) + H(Q) - 2I(P, Q). \quad (3)$$

While MMD ranges over all integer values $\{0, 1, \ldots, n - 1\}$ in $[0, n - 1] \subset \mathbb{R}_+$, VI ranges in $[0, \log n]$.

Apart from MMD and VI, there exist several other partition distance measures (see [13, sections 10.2 and 10.3, pp. 191-193] and [11, 16, 33, 23]). One was proposed as the Hamming distance between (matrices representing) partitions [22], [24], [25], and thus shall be briefly distinguished from the object of this paper. A binary relation $R$ on $N$ is a subset $R \subseteq N \times N$ 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 $(i, j) \in R \Rightarrow (j, i) \in R$ and transitivity $(i, j), (j, h) \in R \Rightarrow (i, h) \in R$ hold, then $R$ is an equivalence relation, or a partition of $N$ into equivalence classes; maximal subsets $A \in 2^N$ such that $(i, j), (j, i) \in R$ for all $i, j \in A$ are precisely its blocks. A binary relation $R$ may be represented as a $n \times n$ matrix $M_R \in \{0, 1\}^{n \times n}$ with entries $M^R_{ij} = 1$ if $(i, j) \in R$ and $M^R_{ij} = 0$ if $(i, j) \notin R$. Now let two equivalence relations $R, R'$ have associated partitions $P, P'$ and representing matrices $M_R, M_{R'}$. The distance $d(R, R')$ between subsets $R, R' \in 2^{N \times N}$ can be computed through expression (1) above:

$$d(R, R') = |\Delta AR'\cap R\setminus R'\cap R| + |R \cap R'\setminus R\cap R'| \quad (1)$$

While providing a distance $d(P, P')$ between partitions, this factually is a Hamming distance between subsets. In particular, $|R\Delta R'|$ is the number of 1s in matrix $M_{R\Delta R'} = M_R + M_{R'}$ modulo 2 (see [1, p. 338]). The point is that $2^{N \times N}$ contains many lattice elements (or binary relations) that do not correspond to partitions (or equivalence relations).

Roughly speaking, 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. Apart from binary relations just described, another example comes from noticing that partitions $P$ are collections of subsets, i.e. $P \in 2^N$, and thus the distance between $P$ and $Q$ may be computed as the Hamming distance $|P \Delta Q|$ between elements of subset lattice $(2^N, \setminus, \cup)$, i.e. the number of subsets $A \in 2^N$ that are blocks of either one but not both. Again, there are many set systems (or collections $S \in 2^{2^N}$ of subsets) that do not correspond to partitions. This feature is maintained even when $P$ and $Q$ are represented as joins of atoms, for they generally admit several such representations. The Hamming distance HD between partitions proposed below relies on representing partitions as $\{0, 1\}$-valued and $\binom{n}{2}$-dimensional arrays, because $\binom{n}{2}$ is the number of atoms. But not all $\binom{n}{2}$ such arrays correspond to partitions.

### III. Atoms, Rank and Size

The rank function $r : \mathcal{P}^N \rightarrow \mathbb{Z}_+$ for the partition lattice is $r(P) = n - |P|$, that is, the maximum conceivable number of blocks for a partition of a $n$-set (which is $n$, of course) minus the actual number of blocks of any partition $P$. The unique partition $P_\bot = \{\{1\}, \ldots, \{n\}\}$ with rank $r(P_\bot) = 0$ is the bottom element. By definition, atoms are immediately above, with rank 1. That is, atoms populate level 1 (above level 0) of the associated Hasse diagram ([1], [32], [22]), with coarsening $\geq$ as the order relation (i.e. coarser partitions in upper levels). This means that an atom, in the partition lattice, is any $P$ with $n - 1$ blocks, out of which $n - 2$ must be singletons, while the remaining one is a pair, and there are $\binom{n}{2}$ unordered pairs.

For notational convenience, let $[ij] \in \mathcal{P}^N$ denote the atom where the unique 2-cardinal block is (unordered) pair $(i, j) \in \{ij\}$, noting that $2(\binom{n}{2})$ is the number of simple graphs with $n$ vertexes. In fact, partitions are the transitive closure of such graphs: the former obtain from the latter by adding all edges needed to render complete (or fully connected) each component [1, Point 2.31, p. 54]. The order relation among partitions is coarsening $\geq$, where $P \geq Q$ means that every block of $Q$ is included in some block of $P$. Let $1 \in \{0, 1\}^n$ denote the $n$-vector with all entries equal to 1 and $(x, y)$ be the scalar product of any two vectors $x, y$. For any subset $A \in 2^N$, clearly $|A| = r(A) = \langle x_A, 1 \rangle$. Also, given any $A, B \in 2^N$,

$$d(A, B) = |\Delta AB| = |A| + |B| - 2|A \cap B| = \langle x_A, 1 \rangle + \langle x_B, 1 \rangle - 2\langle x_A, x_B \rangle. \quad (5)$$

Let $P^N = \{[ij] : 1 \leq i < j \leq n\}$ denote the $\binom{n}{2}$-set of atoms of the partition lattice. The analog of the characteristic
function \( \chi_A \) of subsets \( A \in 2^n \) for partitions \( P \in \mathcal{P}^N \) is the indicator function \( I_P : \mathcal{P}^N \rightarrow \{0, 1\} \) defined by

\[
I_P([ij]) = \begin{cases} 
1 & \text{if } P \supseteq [ij] \\
0 & \text{if } P \nsubseteq [ij]
\end{cases}
\]

for all \( P \in \mathcal{P}^N, [ij] \in \mathcal{P}^N_1 \).

In words, if pair \([i, j]\) is included in some block \( A \) of \( P \) (i.e. \([i, j] \subseteq A \subseteq P\)), then partition \( P \) is coarser than atom \([i, j]\), and the corresponding position \( I_P([ij]) \) of indicator array \( I_P \) has entry 1. Otherwise, that position is 0.

By reversing the order relation, \( P \supseteq [ij] \) turns into \([ij] \subseteq P\), i.e. atom \([ij]\) is finer than \( P \). The number of atoms finer than \( P = \{A_1, \ldots, A_{|P|}\} \) is the size \( s : \mathcal{P}^N \rightarrow \mathbb{Z}_+ \) (see [27]), i.e.

\[
s^P = \sum_{1 \leq k \leq |P|} \binom{|A_k|}{2} = (I_P, 1)
\]

with \( s^P = s(P) \) and \( 1 \in \{0, 1\}^2 \) now denoting the unitary \( (\frac{3}{2}) \)-vector. That is to say, \( s^P = |\{[ij] \in \mathcal{P}^N_1 : [ij] \subseteq P\}| \).

| \(|N| = n\) | \(|s^P : P \in \mathcal{P}^N\) | 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 | \(\{0, 1, 2, 3, 4, 6, 7, 9, 10, 11, 15, 21\}\) | |

While the cardinality \(|A|\) of subsets \( A \in 2^n \) takes every integer value between 0 and \( n \), the size \( s^P = \langle I_P, 1 \rangle \) of partitions \( P \in \mathcal{P}^N \) does not take the same between 0 and \( \binom{n}{2} \). In fact, a main difference between characteristic function \( \chi_A \) of subsets \( A \in 2^n \) and indicator function \( I_P \) of partitions \( P \in \mathcal{P}^N \), is that the former spans the whole vertex set of the unit \( n \)-dimensional hypercube (that is \( \chi_A : A \in 2^n \) = \( \{0, 1\}^n \)), while the latter only spans a (rather small) proper subset of the vertex set of the unit \( (\frac{3}{2}) \)-dimensional hypercube (that is \( I_P : P \in \mathcal{P}^N \) \subseteq \( \{0, 1\}^2 \), where \( \subseteq \) denotes strict inclusion \( \subset \); again, see [1, Point 2.3.1, p. 54]). This is due to linear dependence [34], characterizing geometric lattices in general. The number of non-spanned vertexes is \( 2^\binom{n}{2} - B_n \), where \( B_n = |\mathcal{P}^N| \) is the \( n \)-th Bell number or number of partitions of a \( n \)-set [29], [14] [18, Section 2.2]. The smallest value of \( n \) where linear dependence may be appreciated is \( n = 3 \), in that there are \( B_5 = 5 \) partitions, namely the finest \( \{\{1\}, \{2\}, \{3\}\} \) and coarsest \( \{1, 2, 3\} \) ones, and \( \binom{3}{2} = 3 \) atoms: \( [12] = \{\{1, 2\}, \{3\}\} \), \( [13] = \{\{1, 3\}, \{2\}\} \) and \( [23] = \{\{2, 3\}, \{1\}\} \). Thus, there is no partition with size equal to 2, as \( [12] \cup [23] = [12] \cup [13] \cup [13] \cup [23] = \{1, 2, 3\} \neq [12] \cup [13] \cup [23] \), where for any two partitions \( P, Q \in \mathcal{P}^N \), the join \( P \cup Q \) is the finest partition coarser than both \( P \) and \( Q \), while the meet \( P \cap Q \) is the coarsest partition finer than both \( P \) and \( Q \) (see [1], [32]). Available sizes for \( 1 \leq n \leq 7 \) are in Table I above.

A. Representations as joins of atoms: rank and size

In atomic (or atomistic, see [1], [32]) lattices every element admits some representation as a join of atoms. Yet, while subsets \( A \in 2^n \) admit a unique such a representation, namely \( A = \bigcup_{i \in A} \{i\} \), partitions generally admit several such representations. Again, this results from linear dependence and is observable already when \( n = 3 \) (see above), in that the coarsest partition \( \{\{1, 2, 3\}\} \) may be represented either as the join of any two atoms, or even 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 needed for representing \( P \) as a join of atoms, while the size \( s^P = \sum_{A \in P} \binom{|A|}{2} \) is the maximum number of atoms available for representing \( P \) as a join of atoms. Hence, keeping the coarsest partition of a 3-cardinal set as the simplest example, the rank is \( r(\{\{1, 2, 3\}\}) = 3 - 1 = 2 \), while the size is \( s(\{\{1, 2, 3\}\}) = 3 \).

The rank or cardinality of subsets is a peculiar lattice function, namely a valuation of subset lattice \( (2^n, \cap, \cup) \), [1, p. 187]), meaning that \( |A \cup B| + |A \cap B| = |A| + |B| \) for all \( A, B \in 2^n \). On the other hand, neither the rank nor the size of partitions are valuations of partition lattice \( (\mathcal{P}^N, \land, \lor) \). In fact, valuations of the partition lattice are constant (partition functions), taking the same value on each and every partition [1, Exercise 12 (ii) p. 195]. The following definitions enable to appreciate the features displayed by the rank and the size of partitions, regarded as lattice functions \( r, s : \mathcal{P}^N \rightarrow \mathbb{Z}_+ \).

For any two partitions \( P, Q \in \mathcal{P}^N \), the strict coarsening relation \( > \) is \( P > Q \), meaning \( P \supseteq Q \), \( P \neq Q \). The rank is easily seen to be a strictly monotone partition function, that is to say \( P > Q \) entails \( r(P) > r(Q) \). The same is established hereafter for the size.

**Proposition 1:** The size is a strictly monotone partition function:

\[
s^P > s^Q \quad \text{for all } P, Q \in \mathcal{P}^N \text{ such that } P > Q.
\]

**Proof:** If \( P > Q \), 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 \( A \in P \). The union of any two such blocks \( B, B' \in P \) increases the size (toward \( s^P \), i.e. while reaching an higher level of the Hasse diagram toward \( P \)) by

\[
\left( \frac{|B| + |B'|}{2} \right) - \left( \frac{|B|}{2} + \frac{|B'|}{2} \right) = |B||B'|,
\]

which is strictly positive as blocks are non-empty. \( \blacksquare \)

If \( f(P) + f(Q) \geq f(P \lor Q) \bigcup f(P \land Q) \) for all \( P, Q \in \mathcal{P}^N \), then partition function \( f : \mathcal{P}^N \rightarrow \mathbb{R} \) is sub-modular. Super-modularity obtains by reversing the inequality. As partitions are the transitive closure of graphs, partition lattice \( \mathcal{P}^N \) is the polygon matroid (see again [1, pp. 54 and 274-5]), and matroids have sub-modular rank [1, Rank axioms 6.14 p. 265]. Conversely, the size is a super-modular partition function. To see this, firstly consider the covering relation \( >^* \), with \( P >^* Q \) meaning: \( P > Q \) and there is no \( P' \) such that \( P > P' > Q \).

\( ^1 \)Recall that \( Q^A \) is the partition of \( A \in 2^n \) induced by \( Q \in \mathcal{P}^N \), see section 2 above.
Proposition 2: The size is a super-modular partition function:

\[ s^{P \lor Q} + s^{P \land Q} \geq s^P + s^Q \]

for all \( P, Q \in \mathcal{P}_N \).

Proof: If the two partitions are comparable, say \( P \succeq Q \), then \( P = P \lor Q \) and \( Q = P \land Q \), which makes the statement satisfied with equality. Otherwise, if \( P \not\succeq Q \not\succeq P \), then there are two maximal chains of partitions (see [1], [32]), one meets \( P \land Q \) and \( P \lor Q \) as well as \( P \lor Q \) as the other meets \( P \land Q \) and \( Q \) as well as \( P \land Q \). Focusing on the relevant part or segment of the former chain, there are \( P_t > \cdots > P_1 > P_0 \), with \( r = r(P \lor Q) - r(P \land Q) \), such that \( P_0 = P \land Q \) and \( P_t = P \lor Q \) as well as \( P_k = P \) for some \( k > k_0 < r \). Similarly, focusing on the relevant segment of the latter chain, there are \( Q_r > \cdots > Q_1 > Q_0 \) such that \( Q_0 = P \land Q \) and \( Q_r = P \lor Q \) as well as \( Q_k = Q \) for some \( k_0, k < k_0 < r \). Note that \( r(P) = r(Q) \Leftrightarrow k = k_0 \).

Difference \( s^{P \lor Q} + s^{P \land Q} - (s^P + s^Q) \) may be counted by summing across (Hasse diagram) levels for the two segments. The fact is that frequently many atoms are both \((a)\) finer than \( P \lor Q \), and \((b)\) \(\geq\)-incomparable with respect to both \( P \) and \( Q \). Atoms \( [ij] \in P \land Q \) may be disregarded because they are counted in the size of all the four involved partitions, namely \( P, Q, P \land Q, P \lor Q \). As for the remaining ones, firstly observe that sets

\[ S_P = \{ [ij] \in \mathcal{P}_N^1 : P \geq [ij] \not\subseteq P \land Q \} \]

and

\[ S_Q = \{ [ij] \in \mathcal{P}_N^1 : Q \geq [ij] \not\subseteq P \land Q \} \]

are disjoint, i.e. have empty intersection \( S_P \cap S_Q = \emptyset \). In fact, if an atom \([ij] \not\subseteq P \land Q \) satisfied \( P \geq [ij] \not\subseteq Q \), then \((P \land Q) \lor [ij] \) (and not \( P \land Q \)) would be the coarsest partition finer than both \( P, Q \), i.e. a contradiction.

Now consider going upwards through the Hasse diagram from \( P \land Q \) to \( P \lor Q \) twice, starting with all atoms finer than \( P \lor Q \) apart from those also finer than \( P \land Q \). Specifically, a first route is through segment \( P_0, \ldots, P_t \) of the former maximal chain, and at each partition reached up to \( P_{k_P} = P \) all atoms finer than the current partition but not also finer than the preceding one are removed. A second, subsequent route starts with only the residual atoms and is through segment \( Q_0, \ldots, Q_r \) of the latter maximal chain. Again, up to \( Q_{k_Q} = Q \) at each reached level all atoms finer than the current partition but not also finer than the preceding are removed. The above intersection being empty, it is not possible that the same atom is found to be removed twice, and at the end of the second route there still remains a non-empty (generally large) collection of atoms, namely all those for reaching \( P \lor Q \) from either \( P \) or \( Q \).

IV. HAMMING DISTANCE BETWEEN PARTITIONS

In order to parallel the traditional Hamming distance between subsets given by (1) above, HD has to count the number of atoms finer than either one of any two partitions but not finer than both. Thus, in terms of cardinalities of subsets of atoms, distance \( HD : \mathcal{P}_N \times \mathcal{P}_N \rightarrow \mathbb{Z}_+ \) is given, for all \( P, Q \in \mathcal{P}_N \), by

\[
HD(P, Q) = |\{ [ij] \in \mathcal{P}_N^1 : P \geq [ij] \not\subseteq Q \}| + \\
+ |\{ [ij] \in \mathcal{P}_N^1 : P \not\subseteq [ij] \subseteq Q \}| + \\
- |\{ [ij] \in \mathcal{P}_N^1 : P \not\subseteq [ij] \subseteq Q \}|.
\]

(7)

(8)

where OR in second expression (8) means precisely that the corresponding subset contains all atoms finer than at least one of the two partitions, while the other (second) subset clearly contains all atoms finer than both.

The size and the indicator functions allow to reproduce these expressions in a computationally efficient manner. Considering the size first, note that

\[
P \land Q = \bigvee_{\substack{[ij] \in \mathcal{P}_N^1 \; \vdash \; \not\subseteq \; P \lor Q \;} \atop {\not\subseteq \; P \land Q \; \not\subseteq \; P \lor Q }} [ij],
\]

and this is the maximal representation of \( P \land Q \) as a join of atoms, namely that utilizing \( s^{P \lor Q} \) atoms. Accordingly, a further expression of \( HD \) obtains immediately as follows

\[
HD(P, Q) = s^P + s^Q - 2s^{P \land Q} \text{ for all } P, Q \in \mathcal{P}_N.
\]

(10)

In words, this subtracts twice the number \( s^{P \land Q} \) of atoms finer than both \( P \) and \( Q \) from the sum of the number \( s^P \) of atoms finer than \( P \) and the number \( s^Q \) of atoms finer than \( Q \).

Remark 3: In proposition 2 above, subtracting \( 2s^{P \land Q} \) from both sides yields

\[
s^{P \lor Q} + s^{P \land Q} \geq s^P + s^Q - 2s^{P \land Q}.
\]

If the two partitions are comparable (say \( P \succeq Q \)), then equality holds, but the converse is not true: there exist incomparable partitions \( P, Q \), that is \( P \not\succeq Q \not\succeq P \), where equality holds as well. For example, \( N = \{1, 2, 3, 4\} \) and \( P = \{12\}, Q = \{34\} \). Also, the left hand side is the size-based distance [28], classifiable as a modular partition distance measure.

Turning to the indicator function, there are two distinct ways of using it for the current computational needs. Let \( \mathbf{1} \) denote the \( \binom{N}{2} \)-vector each of whose entries equals 1, as before. Then,

\[
HD(P, Q) = \sum_{[ij] \in \mathcal{P}_N^1} (IP([ij]) - IQ([ij]))^2 = \\
= \langle IP - IQ \rangle^2, \mathbf{1}
\]

(11)

(12)

where \( \langle IP, IQ \rangle = \langle IP \land IQ, \mathbf{1} \rangle \) (see expression (9) above), hence this is the analogue of expression (5) above.

From another perspective, for all \([ij] \in \mathcal{P}_N^1 \) define

\[
I_{P, Q}^\text{max}([ij]) = \max \{ IP([ij]), IQ([ij]) \},
\]

\[
I_{P, Q}^\text{min}([ij]) = \min \{ IP([ij]), IQ([ij]) \}.
\]

Then, in terms of the indicator function, expression (8) yields

\[
HD(P, Q) = \langle I_{P, Q}^\text{max}, \mathbf{1} \rangle - \langle IP, IQ \rangle = \langle I_{P, Q}^\text{max} - I_{P, Q}^\text{min}, \mathbf{1} \rangle = \\
= \langle IP + IQ \mod \mathbf{2}, \mathbf{1} \rangle.
\]
Remark 4: While \( I_{PQ}^{\text{max}} = I_{P+Q} \), in general both \( I_{PQ}^{\text{min}} \) and \( (I_P + I_Q) \) modulo 2 may well fail to be available sizes (see section 3 above), as
\[
s^{P\lor Q} = (I_{P\lor Q}, 1) \geq (I_{PQ}^{\text{max}}, 1) = s\left(\{ij\} \in P^N \mid P \supseteq \{ij\} \text{ OR } Q \supseteq \{ij\}\right).
\]

Attention is now placed on comparing HD with both MMD and VI described above. While VI displays a precise axiomatic characterization [22], its applicative possibilities and computational complexity are not yet extensively investigated [12]. Conversely, MMD is widely used in bioinformatics applications and tightly bounded in terms of required computations [15, 19, 18].

V. HD AND VI: AXIOMS

**PARTITION** distance measures HD and VI display very similar axiomatic behaviors [22], as detailed hereafter.

**Proposition 5:** HD is a metric: for all \( P, P', Q \in \mathcal{P}^N \),
1) \( HD(P, Q) \geq 0 \), with equality if and only if \( P = Q \),
2) \( HD(P, Q) = HD(Q, P) \),
3) \( HD(P, P') + HD(P', Q) \geq HD(P, Q) \).

**Proof:** The second condition is obvious. The first one is also immediate as \( \min\{s^P, s^Q\} \geq s^{P\lor Q} \). Concerning the third one, known as triangle inequality,
\[
HD(P, P') + HD(P', Q) - HD(P, Q) =
\]
\[
2(s^{P'} - s^{P\lor P'} - s^{P'\land Q} + s^{P \land Q} \). \text{ It must be shown that this is positive for all } P, P', Q \in \mathcal{P}^N. \text{ For any such a triplet, the quantity is minimized when } s^{P\lor P'} + s^{P'\land Q} \text{ is maximized, which in turn occurs for } P' \subseteq P, Q \text{ or } P \land P' = P' \land Q. \text{ This entails } s^{P'} = s^{P\lor P'} = s^{P'\land Q} \leq s^{P\land Q}. \]

Triangle inequality is satisfied with equality by both HD and VI when \( P' = P \land Q \).

**Proposition 6:** HD satisfies horizontal collinearity:
\[
HD(P, P \land Q) + HD(P \land Q, Q) = HD(P, Q)
\]
for all \( P, Q \in \mathcal{P}^N \).

**Proof:** Observe that summing \( HD(P, P \land Q) = s^P - s^{P\land Q} \) and \( HD(P \land Q, Q) = s^Q - s^{P\land Q} \) yields \( HD(P, Q) \).

A further axiom characterizing both HD and VI is expressed in terms of two partitions \( P_\perp = \{\{1\}, \ldots, \{n\}\} \) and \( P^\top = \{\{N\}\} \) occupying the bottom and top levels of the Hasse diagram, with \( r(P^\top) = n - 1 \).

**Proposition 7:** HD satisfies vertical collinearity:
\[
HD(P_\perp, P) + HD(P, P^\top) = HD(P_\perp, P^\top)
\]
for all \( P \in \mathcal{P}^N \).

**Proof:** Summands \( HD(P_\perp, P) \) and \( HD(P, P^\top) \) are \( s^P \) and \( s^{P^\top} = s^P \). N\ } \). That is, the distance between any two complements is the same as the distance between the bottom and top elements. Analogously, a partition distance measure \( \delta : \mathcal{P}^N \times \mathcal{P}^N \rightarrow \mathbb{R}_+ \) satisfies such a complement maximality condition [28] if for any two complements \( P', Q' \), it holds \( \delta(P', Q') = \max_{(P', Q') \in \mathcal{P}^N \times \mathcal{P}^N} \delta(P, Q) = \delta(P_\perp, P^\top) \).

For generic \( P = \{A_1, \ldots, A_{|P|}\} \), consider \( P_* \) with size \( |P_*| = |P| - 1 \) and maximal representation as a join of atoms (see above) \( P_* = \{\{j\}_1 \cup \cdots \cap \{j\}_k\}_{|P|-1} \). Also, for \( 1 \leq k < |P| \), let \( P \lor \{j\}_k = \{A_1, \ldots, A_{k-1}, A_k \cup A_{k+1}, A_{k+2}, \ldots, A_{|P|}\} \) (assume \( |A| \geq 2 \) for all \( A \in P \)). In words, \( P_* \) is the join of \( |P| - 1 \) pair wise disjoint atoms each merging two distinct blocks of \( P \). Such a \( P_* \) is a complement of \( P \). A further complement of \( P \) is \( P^\top = B \cup P_\perp^\top \) with \( |B \cap A| = 1 \) for every block \( A \in P \). This partition \( P^\top \) has \( n - |P| + 1 \) blocks: one is \( B \) containing \( |P| \) elements, while all remaining ones are singletons. A sensible partition distance measure should distinguish between such complements. That is, it should not satisfy complement maximality. Indeed, this is the behavior displayed by all distances MMD, VI and HD, and is easily checked by means of an example: the partitioned set is \( N = \{1, \ldots, 7\} \) while partitions are \( P = 123/456/7 \) and \( P_* = 12/[2]/3/45/[6] \) as well as \( P^\top = 147/[2]/3/5/6, \) with \( |P| \) separating blocks. Then, \( V1(P, P^\top) = \frac{1}{2} \log 6 - \frac{1}{2} \log 6 + \frac{1}{2} \log 4 + \frac{1}{2} \log 3 - \frac{1}{2} = V1(P, P^\top) \) as well as \( HD(P, P_\perp) = 8 < 9 = HD(P, P^\top) \). On the other hand, \( MMD(P, P_\perp) = 4 > 2 = MMD(P, P^\top) \).

VI. HD AND MMD: COMPUTATIONS AND EXAMPLES

A computational procedure determining HD accepts as inputs two \((\binom{n}{2})\)-dimensional and \([0, 1]\)-valued arrays as inputs. These are the indicator functions \( I_P, I_Q \) of the two partitions \( P, Q \) between which the distance is to be computed. In fact, the indicator function alone constitutes a novel, neat and mostly natural way of coding partitions (while how to code partitions as inputs when computing MMD is not straightforward [15, 19, 18]). Most importantly, HD does not rise any algorithmic issue. In fact, consider expression (12) above. The number of elementary operations (or running time [20]) needed to compute HD is easily counted. Each of the three scalar products requires \( \binom{n}{2} \) multiplications (between simple numbers, namely 0s or 1s) and \( \binom{n}{2} - 1 \) additions (between positive integers never exceeding \( \binom{n}{2} \) itself). Next, another multiplication and two further additions provide total \( 3 \binom{n}{2} - 1 \). This is evidently better than the best running time \( O(n^3) \) when computing MMD in bioinformatics [18].

Beside computations, HD and MMD can also be compared through examples showing the different sensitivity of their normalized versions (see below). This seems achievable even for the small value \( n = 7 \) applying to Table 2 below. Consider that the partition lattice has \( n \) levels. Maintaining the notation used thus far for level 1 (above level 0), populated by \( \binom{n}{2} \) atoms, let \( \mathcal{P}^N_{(n-k)} \) denote level \( n - k \), for \( k = 1, \ldots, n \). Each level has a number of elements given by the Stirling numbers of the second kind \( S_{n,k} \) [11, 14]. For \( k = 1, \ldots, n \),
\[
S_{n,k} = \frac{1}{k!} \sum_{0 \leq h \leq n} (-1)^{k-h} \binom{k}{h} h^n = |\mathcal{P}^N_{(n-k)}|.
\]
where \( h > k \Rightarrow \binom{k}{h} = 0 \). These numbers are known to be such that most populated levels are central ones [1, Proposition 3.30 p. 91]. Also, upper levels are comparatively more populated than lower ones. For \( n = 7 \),

\[
\begin{align*}
S_{7,7} &= 1 = S_{7,1}, \\
S_{7,6} &= 21 = \binom{7}{2}, \\
S_{7,5} &= 140, \\
S_{7,4} &= 350, \\
S_{7,3} &= 301, \\
S_{7,2} &= 63.
\end{align*}
\]

Examples 1-20 in Table II above compare measurement sensitivity between MMD and HD with partitions picked initially from less populated levels, and subsequently from more populated ones.

It seems worth noting again that while MMD takes all \( n \) integer values \( 0, 1, \ldots, n - 1 \) in its interval \([0, n - 1]\), neither VI nor HD have a range displaying the same even distribution over associated intervals \([0, \log n]\) and \([0, \binom{n}{2}]\), respectively. Concerning HD, the size of partitions does not take all values more populated ones.

### VII. Conclusion

In statistical classification, measuring the distance between partitions is an important combinatorial problem, attracting attention since the ‘60s [26], [21], [11]. Solution HD provided here parallels the (traditional) Hamming distance between subsets by counting atoms in the corresponding lattices. From an applicable viewpoint, HD is added to MMD and VI as tools for clustering comparison in bioinformatics. Easily computed, HD has fine measurement sensitivity and displays suitable axiomatic features.

### Acknowledgments

The authors are grateful to Michael Petreczky for providing the original Sibjoin software and to the anonymous reviewers for their comments and suggestions, which significantly improved this version of the paper.

### References


