Comparing Partitions: Shortest Path Length Metrics and Submodularity

Jyrko Correa-Morris

***Abstract*—With the recent impetus in the development of generic properties and formal frameworks for understanding and organizing the different clustering methods at a technical level, the interest in measures to compare partitions has risen, specially motivated by the applications these have to average-based consensus methods, and the various notions of clusterability. In this regard, Shortest Path Length metrics (also known as Minimum Number of Structural Transformations metrics) have been established as one of the great paradigms for the comparison of not only partitions, but of structured data in general. It has been proven that these metrics can encode many of the properties of the primary notion of proximity that the *refinement relation* endows the lattice of partitions with. On the other hand, another property that has naturally emerges in many mathematical model in combinatorial optimization, economics, machine learning, among others, is submodularity, which has proven to be quite useful from the algorithmic and computational point of view. Motivated by these facts, a question arose: Are there Shortest Path Length metric which are submodular in any of its arguments? In this paper, we prove that there is no shortest path length metric on the lattice of partitions which is submodular in any of their arguments, thus demonstrating that measures such as Mirkin metric and Variation of Information fail to meet this property. We also prove that there are dissimilarity measures that are nonnegative; symmetric; satisfy the triangle inequality; for a chain of partitions, respects the nearness among partitions in the chain (which basically represent the aforementioned primary notion of proximity); and, in addition, are submodular in each of their arguments. These constitute a novel family of measure for comparing partitions with promising attributes.**

***Index Terms*—Short Path Length Metrics; Sub-modularity; Clustering; Lattice of Partitions**

I. Introduction

The development of technology and computing has enabled the processing of large datasets and every day it becomes more necessary to rely on tools to carry out this task automatically. The exploratory analysis of data, seeking for an underlying structure that allows to understand the intrinsic interrelations among them, is often an obligatory task that precedes any further analysis or processing. In this regard, clustering methods play a crucial role. Clustering algorithms receive a finite dataset X and produce a partition P of X attempting to ensure that data in the same cluster of P are closer to each other than those in different clusters of P. Applications of clustering methods can be find in all the fields of knowledge, including mathematics, computer science, biology, social sciences, machine learning, artificial intelligence, and engineering. However, in spite of its popularity, until the past decade, the study of clustering methods was centered at a very general level of description [1].

Recent contributions have mainly addressed this issue from two different viewpoints. The first is devoted to the development of formal theories for clustering methods and the search for generic concepts and rules that allow us to understand the behavior of clustering algorithms [1], [2], [3], while the second is dedicated to the development of more sophisticated algorithms whose modus operandi consists of combining the results of independent algorithms (i.e., the ensemble) to produce a final partition that is, according to a certain criterion, better than the originals [4], [5]. Such algorithms are called *consensus algorithms*.

Measures for quantifying the distance between partitions have gained significant attention with the progress of consensus methods, playing a fundamental role in the process of fusing the partitions in the ensemble into a holistic solution. In other formal approaches to clustering methods, such as the analysis of clusterability [6], these measures are used to evaluate the robustness of the algorithms by means of generic properties. In the specify scope of partitions comparison, studies at a more technical level have adopted a perspective based on the theory of partially ordered sets [7], [8]. The rationale lies in the fact that the space $\mathbb{P}_X$ where the partitions of finite data set X exist has a lattice structure induced by the *refinement* relation, which establishes a primary notion of proximity on $\mathbb{P}_X$. This notion requires that distance measures satisfy nonnegativeness; symmetry; triangle inequality; for a chain of partitions, nearness among partitions in the chain is respected; and a sort of predominance of common sub/superstructures. Measures for comparing partitions have been analyzed regarding their suitability to represent such notion of proximity.

Most of the famous measures for comparing partitions, like the lattice metric, Mirkin metric (Symmetric Difference), Dual Symmetric Difference, and Variation of Information, are based on the shortest path length paradigm [4], [9]. Such measures have been rigorously studied by several authors [10], [11], [12] from a generic and formal point of view, and they have proven to be considerably in compliance with the natural organization of the space of partitions. However, with the immense number of applications that submodularity [13] has found in machine learning (e.g., data summarization, including documents and speech [14], [15], [16]; influence in social network [17]; deep learning [18]), and clustering being one of the main tasks in this and other related fields, and motivated by the positive impact that submodularity would have on the minimization of

Jyrko Correa-Morris is with the Department of Mathematics and Natural Sciences, Miami Dade College, InterAmerican Campus, Miami, FL, 33135 USA e-mail: jcorrea7@mdc.edu

consensus functions, we ask: Is there any Shortest Path Length metric which is submodular in one of its arguments? (Notice that if a metric is submodular in one of its arguments, then, by symmetry, this metric would be submodular in each of its arguments.) To our knowledge, this question has not been addressed so far.

In this paper, we prove that there is no Shortest Path Length Metric for comparing partitions which is submodular in each of its arguments. In fact, our result goes further: we prove that there is no metric on the lattice of partitions which is submodular in each of its arguments and, at the same time, for a chain of partitions, respects the nearness among partitions in the chain. In addition to this impossibility result, this paper introduces a novel family of measures for quantifying the distance between partitions which are symmetric, satisfy triangle inequality, for a chain of partitions, respects the nearness among partitions in the chain, and, in addition, are submodular in each of its arguments.

II. The Lattice of Partitions of a Finite Set X

Let X be a finite set of data with n elements. A *partition* $\text{P} = \{C_1, C_2, \ldots, C_s\}$ of X is a collection of non-empty subsets of X, called the clusters of P, such that $X = \bigcup_{i=1}^{s} C_i$ and $C_i \bigcap C_j = \emptyset$ whenever $i \neq j$. From now on, $\mathbb{P}_X$ denotes the set of all the partitions of X with any number of clusters (from 1 to n). Throughout this paper, $|S|$ denotes the cardinality of the set S, whatever the set S is. In particular, $|\text{P}|$ denotes the number of clusters in the partition P. The space of partitions is endowed with a lattice structure that induces a primary notion of proximity between the partitions encoded in the topology of its Hasse diagram (see below), and any measure intended to quantify the distance between partitions is expected to be consistent with such a proximity.

The *refinement of partitions* —P refines P′ iff every cluster of P is contained in a cluster of P′, in notation, $\text{P} \preceq \text{P}'$— is a partial order on $\mathbb{P}_X$. When $\text{P} \preceq \text{P}'$, it is usually said that P is finer than P′, or equivalently, that P′ is coarser than P. Partition P and P′ such that either $\text{P} \preceq \text{P}'$ or $\text{P}' \preceq \text{P}$ are called *comparable*. If $\text{P} \preceq \text{P}'$ and $\text{P} \neq \text{P}'$, then we write $\text{P} \prec \text{P}'$. In addition, we will say that P′ covers P, in notation $\text{P} \sqsubset \text{P}'$, iff $\text{P} \prec \text{P}'$ and the set $\{\text{P}'' \in \mathbb{P}_X : \text{P} \prec \text{P}'' \prec \text{P}'\}$ is empty. Thus, $\text{P} \sqsubset \text{P}'$ iff P′ is obtained from P by merging exactly two of its clusters. Given two arbitrary partitions, P and P′, it is always possible to find a partition that refines both of them. The coarsest partition satisfying this property is called the *meet* of P and P′ and denoted by $\text{P} \wedge \text{P}'$. Two elements of X are placed in a same cluster of $\text{P} \wedge \text{P}'$ iff they both are placed in a same cluster of P and in a same cluster of P′. Analogously, it is always possible to find a partition that is simultaneously refined by P and P′. The finest partition satisfying this property is called the *join* of P and P′ and denoted by $\text{P} \vee \text{P}'$. Two elements of X, say x and x', are placed in a same cluster of $\text{P} \vee \text{P}'$ iff there is a sequence $x = x_1, x_2, \ldots, x_k = x'$ such that two consecutive elements of this sequence are either placed in the same cluster of P or in the same cluster of P′. These operations can be inductively extended to any finite number of partitions. From now on, m_X denotes the meet of all the partitions in $\mathbb{P}_X$, while g_X will denote their join. Notice that m_X has as many clusters as elements X has, all of which are singletons, while g_X has only one cluster which is X.

The undirected graph whose node set is $\mathbb{P}_X$ and an edge connects the partitions P and P′ if either $\text{P} \sqsubset \text{P}'$ or $\text{P}' \sqsubset \text{P}$ is called the Hasse diagram of $\mathbb{P}_X$ and denoted by $H(X)$. This graph encodes topological relationships between the partitions and hence it induces a primary notion of proximity between them. Any measure D for comparing partitions is expected to be in compliance with this spatial organization/proximity notion. In other words, if a set X is endowed with both, the structure of a metric space (i.e., there is a metric defined on X) and with the structure of a partial ordered set, then it would be of great benefit for these structures to be in compliance with each other [19]. One main aspect here lies in the fact that the basic connections in this space occur between comparable partitions, which forces all the paths from one partition P to another partition non-comparable with P to pass across either a common substructure (e.g., their meet) or a common superstructure (e.g., the join). Accordingly, nonnegativeness; symmetry; triangle inequality; for a chain of partitions, nearness among partitions in the chain is respected; and the rule for substructures and superstructures, are desirable for any measure intended to quantify the distance between partitions. As we will see below, Shortest Path Length metrics obey the rules established by this primary notion of proximity in a great extend, and that the reason we focus our attention on this class of distance measures.

Submodularity, which can be thought as a sort of discrete convexity, states the following: A function $f : L \to \mathbb{R}$ defined on a lattice L is said to be submodular iff, for all $a, b \in L$, $f(a) + f(b) \geq f(a \wedge b) + f(a \vee b)$, where $\wedge$ and $\vee$ denote the meet and join operators in L, respectively. Thus, a distance measure $D : \mathbb{P}_X \times \mathbb{P}_X \to \mathbb{R}$ is submodular in its second argument iff, for all partitions $\text{P}, \text{P}', \text{P}'' \in \mathbb{P}_X$,

$$D(\text{P}, \text{P}') + D(\text{P}, \text{P}'') \geq D(\text{P}, \text{P}' \wedge \text{P}'') + D(\text{P}, \text{P}' \vee \text{P}''). \quad (1)$$

We can interpret this property geometrically in the following sense: the average distance from a partition P to other partitions P′ and P″ is never shorter than the average distance from P to the substructure $\text{P}' \wedge \text{P}''$ and to the superstructure $\text{P}' \vee \text{P}''$. Thus, submodularity in the arguments of a metric also tends to favor substructures and superstructures.

III. Shortest Path Length Metrics

To define a Shortest Path Length metric $d : \mathbb{P}_X \times \mathbb{P}_X \to \mathbb{R}$ on $\mathbb{P}_X$, we start by endowing each edge $\{\text{P}_i, \text{P}_j\}$ of $H(X)$ with a length $\ell(\text{P}_i, \text{P}_j) > 0$ and later we extend this length function to the set of the paths of $H(X)$ by defining the length of a path $\mathbf{p} = \text{P}_1, \text{P}_2, \ldots, \text{P}_s$ in $H(X)$ as zero if $s = 1$ ($\mathbf{p}$ consists of a single vertex), otherwise as the sum of the lengths of its edges: $\ell(\mathbf{p}) = \sum_{k=1}^{s-1} \ell(\text{P}_i, \text{P}_{i+1})$. Then,

$$D(\text{P}, \text{P}') = \min\{\ell(\mathbf{p}) : \mathbf{p} \text{ connects P and P}'\}.$$

Two special cases of Shortest Path Length metrics exist in which the length of the edges in $H(X)$ are defined by means of an order-preserving function $\nu : \mathbb{P}_X \to \mathbb{R}$ by setting $\ell(\mathrm{P}_i, \mathrm{P}_j) = |\nu(\mathrm{P}_i) - \nu(\mathrm{P}_j)|$. To make explicit its dependence on the function ν, we will denote the corresponding shortest path metric by D_ν.

The following Theorem due to [22] provides an analytical expression for the metric d_ν for submodular and supermodular function ν, respectively.

Theorem 1: Let ν be an order-preserving function on $\mathbb{P}_X$.

1) ν is a supermodular function if, and only if, for any partitions P and P',
$$D_\nu(\mathrm{P}, \mathrm{P}') = \nu(\mathrm{P}) + \nu(\mathrm{P}') - 2\nu(\mathrm{P} \wedge \mathrm{P}').$$
2) ν is a submodular function if, and only if, for any partitions P and P',
$$D_\nu(\mathrm{P}, \mathrm{P}') = 2\nu(\mathrm{P} \wedge \mathrm{P}') - \nu(\mathrm{P}) - \nu(\mathrm{P}').$$

Mirkin metric M, which is defined for two arbitrary partitions $\mathrm{P} = \{C_1, \dots, C_k\}$ and $\mathrm{P}' = \{C'_1, \dots, C'_{k'}\}$ by
$$M(\mathrm{P}, \mathrm{P}') = \frac{1}{n}\left(\sum_{i=1}^{k} n_i^2 + \sum_{j=1}^{k'} (n'_j)^2 - 2\sum_{i=1}^{k}\sum_{j=1}^{k'} n_{ij}^2\right),$$
where n_i and n'_j denotes the number of elements in the ith cluster, C_i, of P and jth cluster, C'_j, of P, respectively, and n_{ij} stands for the number of elements in their intersection, $C_i \cap C'_j$, and Variation of Information, given by
$$VI(\mathrm{P}, \mathrm{P}') := -\sum_{i=1}^{k}\sum_{j=1}^{k'} \frac{n_{ij}}{n}\left[\log\left(\frac{n_{ij}}{n_i}\right) + \log\left(\frac{n_{ij}}{n'_j}\right)\right],$$
belong to the second family [4]; while Dual Symmetric Difference, $DSD(\mathrm{P}, \mathrm{P}') = 1/2\left(2^{|\mathrm{P}|} + 2^{|\mathrm{P}'|}\right) - 2^{|\mathrm{P} \vee \mathrm{P}'|}$ and the lattice metric, $\delta(\mathrm{P}, \mathrm{P}') = |\mathrm{P}| + |\mathrm{P}'| - 2|\mathrm{P} \vee \mathrm{P}'|$, fall in the first class.

Now we focus on the main known properties that all shortest path length metrics meet, in addition to the metric requirements: non-negativity, identity of indiscernibles, symmetry, and triangle inequality. These properties encode the suitability of these metrics to represent the basic notion of proximity induced by the refinement relation in the lattice of partitions.

Proposition 1: Let d be an arbitrary Shortest Path Length metric on $\mathbb{P}_X$. Then:

P1 For every chain $\mathrm{P} \preceq \mathrm{P}' \preceq \mathrm{P}''$ in $\mathbb{P}_X$,
 (a) $D(\mathrm{P}, \mathrm{P}') \leq D(\mathrm{P}, \mathrm{P}'')$; the equality holds if and only if $\mathrm{P}' = \mathrm{P}''$.
 (b) $D(\mathrm{P}'', \mathrm{P}') \leq D(\mathrm{P}'', \mathrm{P})$; the equality holds if and only if $\mathrm{P} = \mathrm{P}'$.

P2 For any partitions P and P', either
$$D(\mathrm{P}, \mathrm{P}') \geq D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}') \text{ or } D(\mathrm{P}, \mathrm{P}') \geq D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}').$$

SUP If ν is a supermodular function, then
$$D_\nu(\mathrm{P}, \mathrm{P}') = D_\nu(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}') + D_\nu(\mathrm{P} \wedge \mathrm{P}', \mathrm{P}').$$
In particular,
$$D_\nu(\mathrm{P}, \mathrm{P}') \geq D_\nu(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}').$$

SUB If ν is a supermodular function, then
$$D_\nu(\mathrm{P}, \mathrm{P}') = D_\nu(\mathrm{P}, \mathrm{P} \vee \mathrm{P}') + D_\nu(\mathrm{P} \vee \mathrm{P}', \mathrm{P}').$$
In particular,
$$D_\nu(\mathrm{P}, \mathrm{P}') \geq D_\nu(\mathrm{P}, \mathrm{P} \vee \mathrm{P}').$$

IV. Including Submodularity

In this section, we investigate the compatibility and interrelations between submodularity and the other desirable properties such as P1. We also analyze how these properties govern together the behavior of those distance measures that meet all of them simultaneously.

Proposition 2: Let $D : \mathbb{P}_X \times \mathbb{P}_X \to \mathbb{R}$ be a symmetric dissimilarity measure satisfying P1 and, for every partition P, $D(\mathrm{P}, .) : \mathbb{P}_X \to \mathbb{R}$ is submodular. Then, for any partition $\mathrm{P}' \in \mathbb{P}_X$, $\mathrm{P} \neq \mathrm{P}'$,
 (i) $D(\mathrm{P}, \mathrm{P}') \geq \max\{D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}'), D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')\}$.
 (ii) $D(\mathrm{P}, \mathrm{P}) < D(\mathrm{P}, \mathrm{P}')$.

Proof. The submodularity of $D(\mathrm{P}, .)$ ensures that
$$D(\mathrm{P}, \mathrm{P}) + D(\mathrm{P}, \mathrm{P}') \geq D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}') + D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}'). \quad (2)$$
Consider the chain $\mathrm{P} \wedge \mathrm{P}' \preceq \mathrm{P} \preceq \mathrm{P} \vee \mathrm{P}'$. P1(a) assures that $D(\mathrm{P}, \mathrm{P}) \leq D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')$, whereas P1(b) guarantees that $D(\mathrm{P}, \mathrm{P}) \leq D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}')$. Thus
$$D(\mathrm{P}, \mathrm{P}) \leq \min\{D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}'), D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')\}$$
and therefore inequality (2) forces
$$D(\mathrm{P}, \mathrm{P}') \geq \max\{D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}'), D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')\},$$
which gives (i).

In addition, notice that since $\mathrm{P} \neq \mathrm{P}'$, either $\mathrm{P} \neq \mathrm{P} \wedge \mathrm{P}'$ or $\mathrm{P} \neq \mathrm{P} \vee \mathrm{P}'$, which ensures that $D(\mathrm{P}, \mathrm{P}) < D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}')$ or $D(\mathrm{P}, \mathrm{P}) < D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')$. Consequently,
$$\begin{aligned} D(\mathrm{P}, \mathrm{P}) &\leq \min\{D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}'), D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')\} \\ &< \max\{D(\mathrm{P}, \mathrm{P} \wedge \mathrm{P}'), D(\mathrm{P}, \mathrm{P} \vee \mathrm{P}')\} \leq D(\mathrm{P}, \mathrm{P}'), \end{aligned}$$
which yields (ii). □

The first of these statements basically tells us that the distance traveled from a partition P to a non-comparable partition P' is greater than or equal to the distance traveled from P to either the optimal common substructure (the meet) or the optimal common superstructure (the join). Thus, the inclusion of submodularity restricts the behavior of the dissimilarity D in relation to the conditions that Shortest Path Length metrics demand, since, as shown in Proposition 1, these metrics favor one of these structures, but not necessarily both. In turn, the second statement establishes what is considered the most basic fact about distance: no partition is closer to a partition P than P itself.

This basic proposition enables us to prove that Mirkin metric and Dual Symmetric Difference metric, among other alike metrics, fail to satisfy submodularity. Indeed, if we consider $X = \{a, b, c, d\}$, then for the partitions $\mathrm{P} = \{\{a\}, \{b, c, d\}\}$ and $\mathrm{P}' = \{\{a, c\}, \{b, d\}\}$, whose meet and join are $\mathrm{P} \wedge \mathrm{P}' = \{\{a\}, \{c\}, \{b, d\}\}$ and $\mathrm{P} \vee \mathrm{P}' = \{\{a, b, c, d\}\}$, then Mirkin metric satisfies $M(\mathrm{P},\mathrm{P}')=0.75$ and $M(\mathrm{P}',\mathrm{P}\vee\mathrm{P}')=1$. If instead, we consider the partitions $\mathrm{P}=\{\{a,b\},\{c,d\}\}$ and $\mathrm{P}'=\{\{a,c\},\{b,d\}\}$, whose meet and join are $\mathrm{P}\wedge\mathrm{P}'=\{\{a\},\{c\},\{b\},\{d\}\}$ and $\mathrm{P}\vee\mathrm{P}'=\{\{a,b,c,d\}\}$, then, for Dual Symmetric Difference, we get $DSD(\mathrm{P},\mathrm{P}')=2$ and $DSD(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')=6$. However, we have not been able to find a counterexample for Variation of Information, for instance. A natural question then arises: which are the characteristic features of the metrics that are submodular in each of its arguments? The following theorem provides an answer to this question.

Theorem 2: Let D be a metric on $\mathbb{P}_X$ which is submodular in each of its arguments. Then, for all $\mathrm{P},\mathrm{P}'\in\mathbb{P}$,

(i) $D(\mathrm{P},\mathrm{P}')=D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P},\mathrm{P}\vee\mathrm{P}')$.

(ii) $D(\mathrm{P},\mathrm{P}')=D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\wedge\mathrm{P}')$.

(iii) $D(\mathrm{P},\mathrm{P}')=D(\mathrm{P},\mathrm{P}\vee\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\vee\mathrm{P}')$.

In addition,

(iv) For any nonmodular sublattice $\mathcal{L}$ of $\mathbb{P}_X$ with five elements, say $\mathrm{P}_a,\mathrm{P}_b,\mathrm{P}_c,\mathrm{P}_e$ and P_f, such that $\mathrm{P}_b\prec\mathrm{P}_c$ and

$$\mathrm{P}_a\wedge\mathrm{P}_b=\mathrm{P}_a\wedge\mathrm{P}_c=\mathrm{P}_e,\ \ \mathrm{P}_a\vee\mathrm{P}_b=\mathrm{P}_a\vee\mathrm{P}_c=\mathrm{P}_f, \quad (3)$$

the functions $D(\mathrm{P}_a,.)$, $D(\mathrm{P}_e,.)$ and $D(\mathrm{P}_f,.)$ are positive constant functions on the interval $[\mathrm{P}_b,\mathrm{P}_c]$. In other words, there are $\alpha,\beta,\gamma\in\mathbb{R}$, $\alpha\cdot\beta\cdot\gamma\neq 0$, such that $D(\mathrm{P}_a,\mathrm{P})=\alpha$, $D(\mathrm{P}_e,\mathrm{P})=\beta$ and $D(\mathrm{P}_f,\mathrm{P})=\gamma$, for every partition $\mathrm{P}\in[\mathrm{P}_b,\mathrm{P}_c]$.

(v) $\beta+\gamma=\alpha$.

(vi) $D(\mathrm{P},\mathrm{P}')\leq\alpha$, for any partitions $\mathrm{P},\mathrm{P}'\in[\mathrm{P}_b,\mathrm{P}_c]$.

(vii) $D(\mathrm{P}_e,\mathrm{P}_f)\leq\alpha$.

Proof. Let D be a metric which satisfies our hypothesis, and let P and P' be arbitrary partitions of X. On the one hand, submodularity of D combined with its symmetry and the fact that $D(\mathrm{P},\mathrm{P})=D(\mathrm{P}',\mathrm{P}')=0$, yields

$$D(\mathrm{P},\mathrm{P}') \geq D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P},\mathrm{P}\vee\mathrm{P}');\ \text{and} \quad (4)$$
$$D(\mathrm{P},\mathrm{P}') \geq D(\mathrm{P}',\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\vee\mathrm{P}'). \quad (5)$$

On the other hand, triangular inequality ensures

$$D(\mathrm{P},\mathrm{P}') \leq D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\wedge\mathrm{P}');\ \text{and} \quad (6)$$
$$D(\mathrm{P},\mathrm{P}') \leq D(\mathrm{P},\mathrm{P}\vee\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\vee\mathrm{P}'). \quad (7)$$

Subtracting (5) from the sum of (6) and (7), we get

$$D(\mathrm{P},\mathrm{P}')\leq D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P},\mathrm{P}\vee\mathrm{P}'), \quad (8)$$

and by virtue of (4)

$$D(\mathrm{P},\mathrm{P}')=D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P},\mathrm{P}\vee\mathrm{P}'), \quad (9)$$

which proves (i).

Similarly, subtracting (7) from the sum of (4) and (5), we obtain

$$D(\mathrm{P},\mathrm{P}')\geq D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\wedge\mathrm{P}'), \quad (10)$$

and by virtue of (6)

$$D(\mathrm{P},\mathrm{P}')=D(\mathrm{P},\mathrm{P}\wedge\mathrm{P}')+D(\mathrm{P}',\mathrm{P}\wedge\mathrm{P}'), \quad (11)$$

which yields (ii). Statement (iii) follows analogously by subtracting (6) from the sum of (4) and (5) and then applying (7).

Now, we shall proceed analogously to [23]. As $\mathbb{P}_X$ is not a modular lattice, there is a nonmodular sublattice with five elements, say $\mathrm{P}_a,\mathrm{P}_b,\mathrm{P}_c,\mathrm{P}_e$ and P_f such that $\mathrm{P}_b\prec\mathrm{P}_c$ and

$$\mathrm{P}_a\wedge\mathrm{P}_b=\mathrm{P}_a\wedge\mathrm{P}_c=\mathrm{P}_e,\ \ \mathrm{P}_a\vee\mathrm{P}_b=\mathrm{P}_a\vee\mathrm{P}_c=\mathrm{P}_f. \quad (12)$$

Note now that, for any partition $\mathrm{P}\in[\mathrm{P}_b,\mathrm{P}_c]$, $\mathrm{P}_e=\mathrm{P}_a\wedge\mathrm{P}_c\preceq\mathrm{P}_a\wedge\mathrm{P}\preceq\mathrm{P}_a\wedge\mathrm{P}_c=\mathrm{P}_e$ and $\mathrm{P}_f=\mathrm{P}_a\vee\mathrm{P}_c\preceq\mathrm{P}_a\vee\mathrm{P}\preceq\mathrm{P}_a\vee\mathrm{P}_c=\mathrm{P}_e$, and hence, in view of (i), we get

$$\begin{aligned} D(\mathrm{P}_a,\mathrm{P}) &= D(\mathrm{P}_a,\mathrm{P}_a\wedge\mathrm{P})+D(\mathrm{P}_a,\mathrm{P}_a\vee\mathrm{P}) \\ &= D(\mathrm{P}_a,\mathrm{P}_e)+D(\mathrm{P}_a,\mathrm{P}_f), \end{aligned} \quad (13)$$

which does not depend on P. Thus $D(\mathrm{P}_a,.)$ is constant (and equals to some $\alpha>0$) in the interval $[\mathrm{P}_b,\mathrm{P}_c]$. Moreover, (i) also yields

$$\begin{aligned} D(\mathrm{P}_a,\mathrm{P}) &= D(\mathrm{P},\mathrm{P}_a\wedge\mathrm{P})+D(\mathrm{P},\mathrm{P}_a\vee\mathrm{P}) \\ &= D(\mathrm{P},\mathrm{P}_e)+D(\mathrm{P},\mathrm{P}_f). \end{aligned} \quad (14)$$

Now using (ii), it can be concluded that

$$D(\mathrm{P}_a,\mathrm{P})=D(\mathrm{P}_a,\mathrm{P}_a\wedge\mathrm{P})+D(\mathrm{P},\mathrm{P}_a\wedge\mathrm{P}), \quad (15)$$

which becomes, in view of (14),

$$D(\mathrm{P},\mathrm{P}_e)+D(\mathrm{P},\mathrm{P}_f)=D(\mathrm{P}_a,\mathrm{P}_e)+D(\mathrm{P},\mathrm{P}_e), \quad (16)$$

and therefore $D(\mathrm{P},\mathrm{P}_f)=D(\mathrm{P}_a,\mathrm{P}_e)$, which means that $D(\mathrm{P},\mathrm{P}_f)$ does not depend on the partition P and is equal to some constant γ.

Condition (iii), on the other hand, allows us to see that

$$D(\mathrm{P}_a,\mathrm{P})=D(\mathrm{P}_a,\mathrm{P}_a\vee\mathrm{P})+D(\mathrm{P},\mathrm{P}_a\vee\mathrm{P}), \quad (17)$$

which can be transformed, by virtue of (14), into

$$D(\mathrm{P},\mathrm{P}_e)+D(\mathrm{P},\mathrm{P}_f)=D(\mathrm{P}_a,\mathrm{P}_f)+D(\mathrm{P},\mathrm{P}_f), \quad (18)$$

and hence $D(\mathrm{P},\mathrm{P}_e)=D(\mathrm{P}_a,\mathrm{P}_f)$. Thus, $D(\mathrm{P},\mathrm{P}_f)$ to be equal to some constant β. This proves (iv), and (v) follows by replacing in (14) $D(\mathrm{P}_a,\mathrm{P})$, $D(\mathrm{P}_e,\mathrm{P})$ and $D(\mathrm{P}_f,\mathrm{P})$ by α, β and γ, respectively.

Finally, for any $\mathrm{P},\mathrm{P}'\in[\mathrm{P}_b,\mathrm{P}_c]$, triangle inequality ensures

$$D(\mathrm{P},\mathrm{P}') \leq D(\mathrm{P},\mathrm{P}_e)+D(\mathrm{P}_e,\mathrm{P}');\ \text{and} \quad (19)$$
$$D(\mathrm{P},\mathrm{P}') \leq D(\mathrm{P},\mathrm{P}_f)+D(\mathrm{P}_f,\mathrm{P}'). \quad (20)$$

Adding (19) and (20), we obtain

$$2D(\mathrm{P},\mathrm{P}')\leq D(\mathrm{P},\mathrm{P}_e)+D(\mathrm{P},\mathrm{P}_f)+D(\mathrm{P}_e,\mathrm{P}')+D(\mathrm{P}_f,\mathrm{P}'), \quad (21)$$

but the right term in (21) is equal to 2α, so $D(\mathrm{P},\mathrm{P}')\leq\alpha$. This proves (vi).

Furthermore, triangle inequality assures that

$$D(\mathrm{P}_e,\mathrm{P}_f)\leq D(\mathrm{P}_e,\mathrm{P})+D(\mathrm{P},\mathrm{P}_f)=\alpha,$$

which completes the proof. □

Corollary 1: If D is a metric which is submodular in each of its arguments, then D does not satisfy neither P1(a) nor P1(b).

Proof. Since $\mathrm{P}_e \prec \mathrm{P}_b \prec \mathrm{P}_c$ and, by virtue of Theorem 2, Statement 4., $D(\mathrm{P}_e, .)$ is constant on the interval $[\mathrm{P}_b, \mathrm{P}_c]$, $D(\mathrm{P}_e, \mathrm{P}_b) = D(\mathrm{P}_e, \mathrm{P}_c)$, which contradicts P1(a).

Similarly, the refinements $\mathrm{P}_b \prec \mathrm{P}_c \prec \mathrm{P}_f$ and the fact that $D(\mathrm{P}_e, .)$ is constant on the interval $[\mathrm{P}_b, \mathrm{P}_c]$, also consequence of Theorem 2, Statement 4., yield $D(\mathrm{P}_f, \mathrm{P}_c) = D(\mathrm{P}_f, \mathrm{P}_b)$, which contradicts P1(b). □

Corollary 2: There is no Shortest Path Length metric which is submodular in each of its arguments. In particular, Variation of Information and Lattice Metric fail to be submodular in each of its arguments.

V. A Novel Family of Measures

In contrast with the results above, in this section we prove that there are distance measures which are "almost" metrics and preserve the main properties of the Shortest Path Length metrics.

Theorem 3: Let $\omega, \omega' : \mathbb{P}_X \times \mathbb{P}_X \to \mathbb{R}$ be decreasing monotonic functions such that ω is submodular and ω' is supermodular. Suppose in addition that, for all $\mathrm{P}, \mathrm{P}' \in \mathbb{P}_X$ such that $\mathrm{P} \preceq \mathrm{P}'$, $\omega(\mathrm{P}) \geq \omega'(\mathrm{P}')$. Then, for arbitrary partitions P, P' and P'' of X, the measure

$$D_{\omega\omega'}(\mathrm{P}, \mathrm{P}') = \omega(\mathrm{P} \wedge \mathrm{P}') - \omega'(\mathrm{P} \vee \mathrm{P}')$$

satisfies:

(i) P1(a) and P1(b).

(ii) Submodularity in each of its arguments.

(iii) Non-negativity: for any partitions $\mathrm{P}, \mathrm{P}' \in \mathbb{P}_X$,

$$D_{\omega\omega'}(\mathrm{P}, \mathrm{P}') \geq 0.$$

(iv) Identity of Indiscernibles: for any partitions $\mathrm{P}, \mathrm{P}' \in \mathbb{P}_X$, $D_{\omega\omega'}(\mathrm{P}, \mathrm{P}) \leq D_{\omega\omega'}(\mathrm{P}, \mathrm{P}')$, and the equality holds if, and only if, $\mathrm{P} = \mathrm{P}'$.

(v) Symmetry: for any partitions $\mathrm{P}, \mathrm{P}' \in \mathbb{P}_X$,

$$D_{\omega\omega'}(\mathrm{P}, \mathrm{P}') = D_{\omega\omega'}(\mathrm{P}', \mathrm{P}).$$

(vi) Triangle Inequality: for any partitions $\mathrm{P}, \mathrm{P}', \mathrm{P}'' \in \mathbb{P}_X$,

$$D_{\omega\omega'}(\mathrm{P}, \mathrm{P}') \leq D_{\omega\omega'}(\mathrm{P}, \mathrm{P}'') + D_{\omega\omega'}(\mathrm{P}'', \mathrm{P}').$$

Proof. (To simplify the notation, we shall use D instead of $D_{\omega\omega'}$ throughout the proof.)

Let $\mathrm{P} \preceq \mathrm{P}' \preceq \mathrm{P}''$. Then $\mathrm{P} \wedge \mathrm{P}' = \mathrm{P} \wedge \mathrm{P}'' = \mathrm{P}$, whereas $\mathrm{P} \vee \mathrm{P}' = \mathrm{P}'$ and $\mathrm{P} \vee \mathrm{P}'' = \mathrm{P}''$. Thus, $D(\mathrm{P}, \mathrm{P}') = \omega(\mathrm{P}) - \omega'(\mathrm{P}')$ and $D(\mathrm{P}, \mathrm{P}'') = \omega(\mathrm{P}) - \omega'(\mathrm{P}'')$. Since ω' is a decreasing function, $\omega'(\mathrm{P}') \geq \omega'(\mathrm{P}'')$ and hence $D(\mathrm{P}, \mathrm{P}') \leq D(\mathrm{P}, \mathrm{P}'')$. This shows that D satisfies the axiom P1(a). Analogously, $\mathrm{P}'' \wedge \mathrm{P} = \mathrm{P}$ and $\mathrm{P}'' \wedge \mathrm{P}' = \mathrm{P}'$, while $\mathrm{P}'' \vee \mathrm{P} = \mathrm{P}'' \vee \mathrm{P}' = \mathrm{P}''$. So, $D(\mathrm{P}'', \mathrm{P}) = \omega(\mathrm{P}) - \omega'(\mathrm{P}'')$ and $D(\mathrm{P}'', \mathrm{P}') = \omega(\mathrm{P}') - \omega'(\mathrm{P}'')$. By virtue of the decreasing monotonicity of ω, $\omega(\mathrm{P}) \geq \omega(\mathrm{P}')$ and therefore, $D(\mathrm{P}', \mathrm{P}'') \leq D(\mathrm{P}, \mathrm{P}'')$. This proves that D satisfies P1(b).

Let us prove now that $D(\mathrm{P}, .)$ is a submodular function. Indeed, let P, P' and P'' be arbitrary partitions in $\mathbb{P}_X$. Then,

$$D(\mathrm{P}, \mathrm{P}') + D(\mathrm{P}, \mathrm{P}'') = \omega(\mathrm{P} \wedge \mathrm{P}') - \omega'(\mathrm{P} \vee \mathrm{P}') + \omega(\mathrm{P} \wedge \mathrm{P}'') - \omega'(\mathrm{P} \vee \mathrm{P}''),$$

or, equivalently,

$$D(\mathrm{P}, \mathrm{P}') + D(\mathrm{P}, \mathrm{P}'') = \omega(\mathrm{P} \wedge \mathrm{P}') + \omega(\mathrm{P} \wedge \mathrm{P}'') - \omega'(\mathrm{P} \vee \mathrm{P}') - \omega'(\mathrm{P} \vee \mathrm{P}''). \tag{22}$$

Also,

$$D(\mathrm{P}, \mathrm{P}' \wedge \mathrm{P}'') + D(\mathrm{P}, \mathrm{P}' \vee \mathrm{P}'') = \omega(\mathrm{P} \wedge (\mathrm{P}' \wedge \mathrm{P}'')) - \omega'(\mathrm{P} \vee (\mathrm{P}' \wedge \mathrm{P}'')) + \omega(\mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')) - \omega'(\mathrm{P} \vee \mathrm{P}' \vee \mathrm{P}''),$$

which can be regrouped to be

$$D(\mathrm{P}, \mathrm{P}' \wedge \mathrm{P}'') + D(\mathrm{P}, \mathrm{P}' \vee \mathrm{P}'') = \omega(\mathrm{P} \wedge (\mathrm{P}' \wedge \mathrm{P}'')) + \omega(\mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')) - \omega'(\mathrm{P} \vee (\mathrm{P}' \wedge \mathrm{P}'')) - \omega'(\mathrm{P} \vee \mathrm{P}' \vee \mathrm{P}''). \tag{23}$$

On the one hand, submodularity of ω forces

$$\omega(\mathrm{P} \wedge \mathrm{P}') + \omega(\mathrm{P} \wedge \mathrm{P}'') \geq \omega(\mathrm{P} \wedge \mathrm{P}' \wedge \mathrm{P}'') + \omega((\mathrm{P} \wedge \mathrm{P}') \vee (\mathrm{P} \wedge \mathrm{P}'')). \tag{24}$$

On the other hand, since $\mathrm{P} \wedge \mathrm{P}'$ refines both P and $\mathrm{P}' \vee \mathrm{P}''$, we get $\mathrm{P} \wedge \mathrm{P}' \preceq \mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')$; and similarly we deduce that $\mathrm{P} \wedge \mathrm{P}'' \preceq \mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')$. Hence, $(\mathrm{P} \wedge \mathrm{P}') \vee (\mathrm{P} \wedge \mathrm{P}'') \preceq \mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')$. Using now the monotonicity of ω, we can conclude that

$$\omega((\mathrm{P} \wedge \mathrm{P}') \vee (\mathrm{P} \wedge \mathrm{P}'')) \geq \omega(\mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')). \tag{25}$$

Combining (24) with (25), we can assert that

$$\omega(\mathrm{P} \wedge \mathrm{P}') + \omega(\mathrm{P} \wedge \mathrm{P}'') \geq \omega(\mathrm{P} \wedge \mathrm{P}' \wedge \mathrm{P}'') + \omega(\mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')),$$

and by (22)

$$D(\mathrm{P}, \mathrm{P}') + D(\mathrm{P}, \mathrm{P}'') \geq \omega(\mathrm{P} \wedge \mathrm{P}' \wedge \mathrm{P}'') + \omega(\mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')) - (\omega'(\mathrm{P} \vee \mathrm{P}') + \omega'(\mathrm{P} \vee \mathrm{P}'')). \tag{26}$$

Similarly as we did before with ω, we can use the supermodularity of ω' to get

$$\omega'(\mathrm{P} \vee \mathrm{P}') + \omega'(\mathrm{P} \vee \mathrm{P}'') \leq \omega'((\mathrm{P} \vee \mathrm{P}') \wedge (\mathrm{P} \vee \mathrm{P}'')) + \omega'(\mathrm{P} \vee \mathrm{P}' \vee \mathrm{P}''). \tag{27}$$

Since $(\mathrm{P} \vee \mathrm{P}') \wedge (\mathrm{P} \vee \mathrm{P}'') \succeq \mathrm{P}$ and $(\mathrm{P} \vee \mathrm{P}') \wedge (\mathrm{P} \vee \mathrm{P}'') \succeq \mathrm{P}' \wedge \mathrm{P}''$, $(\mathrm{P} \vee \mathrm{P}') \wedge (\mathrm{P} \vee \mathrm{P}'') \succeq \mathrm{P} \vee (\mathrm{P}' \wedge \mathrm{P}'')$, and view of the monotonicity of ω',

$$\omega'((\mathrm{P} \vee \mathrm{P}') \wedge (\mathrm{P} \vee \mathrm{P}'')) \leq \omega'(\mathrm{P} \vee (\mathrm{P}' \wedge \mathrm{P}'')). \tag{28}$$

Jointly using (27) and (28), we obtain

$$\omega'(\mathrm{P} \vee \mathrm{P}') + \omega'(\mathrm{P} \vee \mathrm{P}'') \leq \omega'(\mathrm{P} \vee (\mathrm{P}' \wedge \mathrm{P}'')) + \omega'(\mathrm{P} \vee \mathrm{P}' \vee \mathrm{P}''), \tag{29}$$

which, substituted in (26), gives

$$D(\mathrm{P}, \mathrm{P}') + D(\mathrm{P}, \mathrm{P}'') \geq \omega(\mathrm{P} \wedge \mathrm{P}' \wedge \mathrm{P}'') + \omega(\mathrm{P} \wedge (\mathrm{P}' \vee \mathrm{P}'')) - \omega'(\mathrm{P} \vee (\mathrm{P}' \wedge \mathrm{P}'')) + \omega'(\mathrm{P} \vee \mathrm{P}' \vee \mathrm{P}'')). \tag{30}$$

Now comparing the right members in (23) and (30), we get

$$D(\mathrm{P}, \mathrm{P}') + D(\mathrm{P}, \mathrm{P}'') \geq D(\mathrm{P}, \mathrm{P}' \wedge \mathrm{P}'') + D(\mathrm{P}, \mathrm{P}' \vee \mathrm{P}''),$$

which proves 2..

Since $\mathrm{P} \wedge \mathrm{P}' \preceq \mathrm{P} \vee \mathrm{P}'$, in view of the hypothesis, $\omega(\mathrm{P} \wedge \mathrm{P}') \geq \omega'(\mathrm{P} \vee \mathrm{P}')$. Hence, $D(\mathrm{P}, \mathrm{P}') = \omega(\mathrm{P} \wedge \mathrm{P}') - \omega'(\mathrm{P} \vee \mathrm{P}')$, which proves 3..

Identity of indiscernibles is an immediate consequence of Statement 2. and Proposition 2.

In turn, Symmetry is trivially derived from the fact that the meet and join are symmetric operators, which gives 5..

Finally, let us verify triangle inequality. Let P, P' and P'' be arbitrary partitions of X.

$$\begin{aligned} &D(\text{P}, \text{P}'') + D(\text{P}'', \text{P}') = \\ &\omega(\text{P} \wedge \text{P}'') + \omega(\text{P}'' \wedge \text{P}') - \omega'(\text{P} \vee \text{P}'') - \omega'(\text{P}'' \vee \text{P}'). \end{aligned} \tag{31}$$

Using first the submodularity of ω and then its monotonicity, we get

$$\begin{aligned} &\omega(\text{P} \wedge \text{P}'') + \omega(\text{P}'' \wedge \text{P}') \geq \\ &\omega(\text{P} \wedge \text{P}' \wedge \text{P}'') + \omega((\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}')) \geq \\ &\omega(\text{P} \wedge \text{P}') + \omega((\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}')). \end{aligned}$$

Then, according to (31),

$$\begin{aligned} &D(\text{P}, \text{P}'') + D(\text{P}'', \text{P}') \geq \omega(\text{P} \wedge \text{P}') + \\ &+\omega((\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}')) - \omega'(\text{P} \vee \text{P}'') - \omega'(\text{P}'' \vee \text{P}'). \end{aligned} \tag{32}$$

Now, applying the submodularity of $-\omega'$ and thereupon its monotonicity, we obtain

$$\begin{aligned} &-\omega'(\text{P} \vee \text{P}'') - \omega'(\text{P}'' \vee \text{P}') \geq \\ &-\omega'((\text{P} \vee \text{P}'') \wedge (\text{P}'' \vee \text{P}')) - \omega'(\text{P}'' \vee \text{P}' \vee \text{P}) \geq \\ &-\omega'((\text{P} \vee \text{P}'') \wedge (\text{P}'' \vee \text{P}')) - \omega'(\text{P}' \vee \text{P}). \end{aligned}$$

Thus, (32) ensures

$$\begin{aligned} &D(\text{P}, \text{P}'') + D(\text{P}'', \text{P}') \geq \omega(\text{P} \wedge \text{P}') + \omega((\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}')) - \\ &-\omega'((\text{P} \vee \text{P}'') \wedge (\text{P}'' \vee \text{P}')) - \omega'(\text{P}' \vee \text{P}), \end{aligned}$$

and therefore,

$$\begin{aligned} &D(\text{P}, \text{P}'') + D(\text{P}'', \text{P}') \geq D(\text{P}, \text{P}') + \\ &+\omega((\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}')) - \omega'((\text{P} \vee \text{P}'') \wedge (\text{P}'' \vee \text{P}')). \end{aligned} \tag{33}$$

Since $(\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}') \preceq (\text{P} \vee \text{P}'') \wedge (\text{P}'' \vee \text{P}')$, we can conclude, in view of the hypothesis, that

$$\omega((\text{P} \wedge \text{P}'') \vee (\text{P}'' \wedge \text{P}')) - \omega'((\text{P} \vee \text{P}'') \wedge (\text{P}'' \vee \text{P}')) \geq 0,$$

and hence

$$D(\text{P}, \text{P}'') + D(\text{P}'', \text{P}') \geq D(\text{P}, \text{P}').$$

□

It is worthy to emphasize that Theorem 3 does not claim that $D_{\omega\omega'}$ is metric (this would contradicts Theorem 2). The impossibility lies in the fact that $D_{\omega\omega'}(\text{P}, \text{P})$, even though is the minimum distance traveled from P to another partition (including P), is not necessarily zero, and may vary if we change the initial partition P, which are requirements for a distance measure to be a metric. However, this condition does not seem to be essential for most of the tasks where measures to compare partitions play a crucial role. In contrast, submodularity has proven to be a property that favors the performance of the functions inserted in some mathematical frameworks intended to model certain machine learning problems, such as data summarization [14] and image processing [18], to only mention a few.

An example of a dissimilarity function that satisfies the premises of Theorem 3 is $D_{I\frac{|.|}{\sigma}} : \mathbb{P}_X \times \mathbb{P}_X \to \mathbb{R}$, $\sigma \geq n$, given by

$$D_{I\frac{|.|}{\sigma}}(\text{P}, \text{P}') = I(\text{P} \wedge \text{P}') - \frac{|\text{P} \wedge \text{P}'|}{\sigma},$$

where $I : \mathbb{P}_X \to \mathbb{R}$ is Information Measure (sifted up 1 unit), given by

$$I(\text{P}) = 1 + \sum_{i=1}^{s} \log\left(\frac{n}{n_i}\right).$$

The values of this measure are compiled in Table I for $X = \{a, b, c\}$. Here $\text{m}_X = \{\{a\}, \{b\}, \{c\}\}$, $\text{P}_{ab} = \{\{a, b\}, \{c\}\}$, $\text{P}_{ac} = \{\{a, c\}, \{b\}\}$, $\text{P}_{bc} = \{\{a\}, \{b, c\}\}$, and $\text{g}_X = \{\{a, b, c\}\}$.

	m_X	P_{ab}	P_{ac}	P_{bc}	g_X
m_X	1.43	1.76	1.76	1.76	2.09
P_{ab}		0.99	2.09	2.09	1.32
P_{ac}			0.99	2.09	1.32
P_{bc}				0.99	1.32
m_X					0.66

TABLE I
VALUES OF $D_{I\frac{|.|}{\sigma}}$, $\sigma = 3$.

VI. CONCLUSIONS

In this paper we addressed the question about the existence of measures for quantifying the distance between partitions that are submodular in each of their arguments and, at the same time, are consistent with the primary notion of proximity of which the refinement relation naturally endows the lattice of partitions of a finite set X. In this regard, we proved in first instance that there is no metric on the lattice of partition which satisfy these conditions. In particular, we concluded that no Shortest Path Length metric on the lattice of partitions is submodular in each of its arguments. Contrasting with this impossibility result, we introduced we proved that there are nonnegative and symmetric distance measures that satisfy triangle inequality; for a chain of partitions, respects the nearness among partitions in the chain; and, in addition, are submodular in each of their arguments.

A potential application of this work is the construction of average-based consensus functions. Such functions would inherit the submodularity property from the associated measure for quantifying the distance between partitions, which could provide several algorithmic and computational advantages. With the incrementally increase in the applications, interest in effective methods to minimize submodular functions has increased tremendously in recent years. This impetus has led to the emergence of sophisticated and promising algorithms to solve the problem of minimizing a submodular function, mostly in distributive lattices. Among the most important references are the pioneering contribution of A. Schrijver [24], providing one of the two first strongly polynomial algorithms intended to minimize arbitrary submodular set functions, the subsequent relevant contributions from S. Fijishige and S. Itawa (see, for instance,[25], [26], [27], [28]), and another strongly polynomial algorithm, probably nowadays the fastest one, from J. B. Orlin [29]. More recently [30] [31].

References

[1] J. Kleinberg, “An impossibility theorem for clustering,” in *NIPS' 02: In proceedings of Advances in Neural Information Procesing Systems*, vol. 15, 2002.

[2] M. Ackerman and S. Ben-David, “Clusterability: A theoretical study,” in *AISTATS*, 2009, pp. 1–8.

[3] J. Correa-Morris, “An indication of unification for different clustering approaches,” *Pattern Recognition*, vol. 46, no. 9, pp. 2548–2561, 2013.

[4] J.-P. Barthélemy and B. Leclerc, “The median procedure for partitions,” in *Partitioning Data Sets*. American Mathematics Society, Series in Discrete Math, 1995, pp. 3–34.

[5] A. Strehl and J. Ghosh, “Cluster ensembles — a knowledge reuse framework for combining multiple partitions,” *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2002.

[6] S. Ben-David and M. Ackerman, “Measures of clustering quality: A working set of axioms for clustering,” in *NIPS*, 2008, pp. 121–128.

[7] J.-P. Barthélemy, B. Leclerc, and B. Monjardet, “On the use of ordered sets in problems of comparison and consensus classifications,” *Journal of Classification*, vol. 3, pp. 187–224, 1986.

[8] M. Meilă, “Comparing clusterings—an information based distance,” *J. Multivar. Anal.*, vol. 98, no. 5, pp. 873–895, 2007.

[9] G. Rossi, “Weighted paths between partitions,” *CoRR*, vol. abs/1509.01852, 2015.

[10] B. Monjardet, “Metrics on partially ordered sets - A survey,” *Discrete Mathematics*, vol. 35, no. 1-3, pp. 173–184, 1981.

[11] M. Meilă, “Comparing clusterings: an axiomatic view,” in *ICML '05: Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 577–584.

[12] M. Meilă, “Local equivalences of distances between clusterings - a geometric perspective,” *Machine Learning*, vol. 86, no. 3, pp. 369–389, 2012.

[13] S. Fujishige, *Submodular Functions and Optimization*. Elsevier (2nd Edition), 2005.

[14] B. Mirzasoleiman, A. Karbasi, A. Badanidiyuru, and A. Krause, “Distributed submodular cover: Succinctly summarizing massive data,” in *Advances in Neural Information Processing Systems 28*, 2015, pp. 2881–2889.

[15] H. Lin and J. Bilmes, “A class of submodular functions for document summarization,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, ser. HLT '11, 2011, pp. 510–520.

[16] H. Lin, J. Bilmes, and S. Xie, “Graph-based submodular selection for extractive summarization,” in *2009 IEEE Workshop on Automatic Speech Recognition Understanding*, 2009, pp. 381–386.

[17] E. Mossel and S. Roch, “On the submodularity of influence in social networks,” in *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, ser. STOC '07. New York, NY, USA: ACM, 2007, pp. 128–134.

[18] H. Oh Song, S. Jegelka, V. Rathod, and K. Murphy, “Deep metric learning via facility location,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[19] L. R. Wilcox and M. F. Smiley, “Metric lattices,” *Annals of Mathematics*, vol. 40, no. 2, pp. 309–327, 1939.

[20] G. Rossi, “Hamming distance between partitions, clustering comparison and information,” in *Recent Advances in Mathematics*, 2015, pp. 101–107.

[21] D. A. Simovici and S. Jaroszewicz, “Generalized conditional entropy and a metric splitting criterion for decision trees,” in *Adv. in Knowl. Disc. and Data Min., PAKDD 2006,*, 2006, pp. 35–44.

[22] G. Birkhoff, *Amer. Math. Soc., Providence*. Lattice Theory (3rd Edition), 1967.

[23] J. Hashimoto, “On a lattice with a valuation,” *Proceeding of the American Mathematics Society*, vol. 3, pp. 1–2, 1952.

[24] A. Schrijver, “A combinatorial algorithm minimizing submodular functions in strongly polynomial time,” *J. Comb. Theory, Ser. B*, vol. 80, no. 2, pp. 346–355, 2000.

[25] S. Fujishige and S. Iwata, “A descent method for submodular function minimization,” *Math. Program.*, vol. 92, no. 2, pp. 387–390, 2002.

[26] S. Iwata, L. Fleischer, and S. Fujishige, “A combinatorial strongly polynomial algorithm for minimizing submodular functions,” *J. ACM*, vol. 48, no. 4, pp. 761–777, 2001.

[27] S. Iwata, “A faster scaling algorithm for minimizing submodular functions,” *SIAM J. Comput.*, vol. 32, no. 4, pp. 833–840, 2003.

[28] ——, “Submodular function minimization,” *Math. Program.*, vol. 112, no. 1, pp. 45–64, 2008.

[29] J. B. Orlin, “A faster strongly polynomial time algorithm for submodular function minimization,” *Math. Program.*, vol. 118, no. 2, pp. 237–251, 2009.

[30] V. Kolmogorov, “Minimizing a sum of submodular functions,” *Discrete Applied Mathematics*, vol. 160, no. 15, pp. 2246 – 2258, 2012.

[31] S. Ramalingam, C. Russell, L. Ladický, and P. H. Torr, “Efficient minimization of higher order submodular functions using monotonic boolean functions,” *Discrete Applied Mathematics*, vol. 220, pp. 1 – 19, 2017.