

# Monotonic Entropies

Dan SIMOVICI<sup>1</sup>

## Abstract

We introduce an axiomatization of entropy that generates, as special cases, novel entropy types. These entropies generalize Shannon’s entropy and allow the introduction of entropy for partitions of sets of objects located in metric spaces, and for partitions of sets of vertices in undirected graphs. Corresponding metrics on the sets of partitions are introduced. Also, we hint to applications of these metrics for evaluation of clustering quality.

**Keywords:** partition, inertia, undirected graphs, generalized entropy

## 1 Introduction

The notion of entropy, the cornerstone of information theory, was introduced by Claude Shannon in his 1948 double paper [16, 17], as a limit of lossless data compression in a noiseless data transmission channel. There exists an ample literature containing axiomatizations of the notion of entropy for probability distributions. Some of these axiomatizations involve the Shannon entropy [6, 8, 11, 14]. Others, such as [24, 23, 7, 9, 20, 22], focus on generalizations of entropy.

This note presents an axiomatization of entropy that leverages algebraic properties of sets of partitions of finite sets in order to produce a simpler system of axioms for entropy, and to extend this notion to a diverse collection of data types.

---

This work is licensed under the [Creative Commons Attribution Licence \(CC BY\)](#)

<sup>1</sup>Computer Science Department, University of Massachusetts Boston, 125 Morrissey Blvd. Boston, Massachusetts 02125, United States of America, Email: [Dan.Simovici@umb.edu](mailto:Dan.Simovici@umb.edu)

Partitions are fundamental for clustering algorithms which aim to detect groupings of objects that have similar properties or are geometrically close to each other. There is a vast literature (see [18]) that focuses on clustering algorithms and a great diversity of approaches to clustering. Also, evaluating cluster quality is an important and challenging task for comparing appropriateness of clustering algorithms for various object configurations.

Partitions of finite sets constitute a natural framework for studying non-overlapping clusterings. The metrics of the partition space of a set of objects generated by various types of entropies offer an instrument for assessing the quality of clusterings. In many cases, data that is subjected to clusterings is labeled and one natural way of grouping object is using these labels and place objects with the same label in a cluster. On the other hand, objects could be grouped using their attributes using a variety of clustering techniques and the metric space of partitions offers a methodology of comparing naturally defined partitions (generated by object labels) with partitions produced by clustering algorithms and, thus, assess the efficacy of these algorithms.

The notion of entropy is usually defined for probability distributions. A finite probability distribution presents itself as a  $n$ -tuple of non-negative numbers  $\mathbf{p} = (p_1, \dots, p_n)$  that satisfies the condition  $p_1 + \dots + p_n = 1$ . Its Shannon entropy is then defined as  $\mathcal{H}(\mathbf{p}) = \sum_{i=1}^n p_i \log \frac{1}{p_i}$  and plays a fundamental role in the study of information transmission.

In this note we adopt a related but distinct approach to entropy, by defining several types of entropies for partitions of finite sets instead of probability distributions. The advantage of this approach is the possibility of using the properties of the partial ordered set of partitions of a finite set. Thus, we are able to formulate analogues of entropy suitable for partitions of metric spaces, or partitions of the sets of vertices of finite graphs.

An central concept in this paper is the notion of monotonic function. If  $(P, \leq)$  and  $(Q, \leq)$  are two partial ordered sets, a function  $f : P \rightarrow Q$  is *monotonic* if  $x \leq y, x, y \in P$ , imply  $f(x) \leq f(y)$ . Monotonic functions will serve to define three distinct types of entropies.

In Section 2 we review some elementary definitions and properties of partitions. Section 3 focuses on using numerical monotonic functions defined on sets on partitions to introduce a set of three axioms that characterize three different types of partition entropies involving partitions of unstructured sets, partitions of subsets of metric spaces and graph partitions. Conditional monotonic entropies are discussed in Section 4. These entropies induce

metrics on various types of partitions, which we examine in Section 5. The final Section 6 presents our conclusions and some ideas for future work.

## 2 Partitions

The set of subsets of a set  $S$  is denoted by  $\mathcal{P}(S)$ .

Formally, a *partition* of a set  $S$  is a collection of non-empty subsets of  $S$  referred to as *blocks*,  $\pi = \{B_i \mid i \in I, B_i \subseteq S\}$  such that  $i, j \in I$  and  $i \neq j$  implies  $B_i \cap B_j = \emptyset$  and  $\bigcup_{i \in I} B_i = S$ . If  $\pi$  consists of two blocks,  $\pi = \{B_1, B_2\}$ , we refer to  $\pi$  as a *bipartition*. The set of partitions of a set  $S$  is denoted by  $PART(S)$ . The notation  $PART_{fin}(S)$  is reserved for the set of finite partitions of  $S$ . Of course, when  $S$  is finite,  $PART_{fin}(S) = PART(S)$ .

If  $\pi \in PART(S)$  and  $x, y \in S$  belong to the same block of  $\pi$  we write  $x \equiv y(\pi)$ . The relation “ $\equiv$ ” is reflexive, symmetric, and transitive and, therefore, it is an equivalence relation on  $S$ . Conversely, if  $\rho$  is an equivalence of  $S$ , the sets of the form  $[x]_\rho = \{u \in S \mid (x, u) \in \rho\}$  constitute a partition  $\pi_\rho$  of  $S$ .

A partial order “ $\leq$ ” is defined on partitions in  $PART(S)$  by setting  $\pi \leq \sigma$  if each block of  $\pi$  is included in a block of  $\sigma$ . The partition  $\alpha_S = \{\{x\} \mid x \in S\}$  is the least element of the partially ordered set  $(PART(S), \leq)$ , while the single-block partition  $\omega_S = \{S\}$  is the largest element of  $(PART(S), \leq)$ .

If  $\pi, \sigma \in PART(S)$ ,  $\pi \leq \sigma$ , and there is no partition  $\tau \in PART(S) - \{\pi, \sigma\}$  such that  $\pi \leq \tau \leq \sigma$ , then we say that  $\sigma$  *covers*  $\pi$  and we write  $\pi \triangleleft \sigma$ . It is easy to show that  $\pi \triangleleft \sigma$  if and only if  $\sigma$  is obtained from  $\pi$  by fusing two of the blocks of  $\pi$  (see [19]).

Let  $U, V$  be two non-empty, disjoint sets, and let  $\sigma \in PART(U)$ , and  $\tau \in PART(V)$ , where  $\sigma = \{B_1, \dots, B_m\}$  and  $\tau = \{C_1, \dots, C_n\}$ . The *sum of the partitions  $\sigma$  and  $\tau$*  is the partition  $\sigma + \tau \in PART(U \cup V)$  defined as:

$$\sigma + \tau = \{B_1, \dots, B_m, C_1, \dots, C_n\}.$$

For every two non-empty disjoint sets  $U$  and  $V$  we have:

$$\begin{aligned} \alpha_U + \alpha_V &= \alpha_{U \cup V}, \\ \omega_U + \omega_V &= \{U, V\} \in PART(U \cup V). \end{aligned}$$

Furthermore, if  $U, V, W$  are non-empty disjoint sets,  $\sigma \in PART(U)$ ,  $\tau \in PART(V)$  and  $v \in PART(W)$ , we have

$$\sigma + (\tau + v) = (\sigma + \tau) + v,$$

a property referred to as the *restricted associativity* of partition addition. The term “restricted” refers to the fact that the underlying sets  $U, V, W$  are supposed to be disjoint.

If  $\sigma = \{B_1, \dots, B_m\} \in \text{PART}(S)$  then we have:

$$\sigma = \omega_{B_1} + \dots + \omega_{B_m}.$$

If the set  $S$  consists of a single element,  $S = \{s\}$ , then  $\alpha_S = \omega_S = \{s\}$ .

The algebraic structure of sets of partitions as semi-modular lattices is discussed in the classical reference [4].

### 3 Axiomatization of Monotonic Partition Entropy

Our axiomatization of partition entropies starts with monotonic functions defined on sets of partitions. We present three examples of monotonic functions defined on specialized collections of sets that will allow us to generate a variety of entropy types.

Let  $\mu : \mathcal{P}(S) \rightarrow \mathbb{R}_{\geq 0}$  be a *non-negative monotonic* function of sets, that is, a function such that  $U \subseteq V$  implies  $\mu(U) \leq \mu(V)$  for  $U, V \in \mathcal{P}(S)$ , and  $|U| > 1$  implies  $\mu(U) > 0$ .

Next, we consider examples of non-negative monotonic functions that generate corresponding entropies.

**Example 1.** Let  $S$  be a finite set and let  $\mu : \mathcal{P}(S) \rightarrow \mathbb{R}_{\geq 0}$  be given by  $\mu(B) = |B|^\beta$  for  $B \in \mathcal{P}(S)$  and some  $\beta > 0$ . The function is clearly monotonic and  $B \neq \emptyset$  implies  $\mu(B) > 0$ .

Furthermore, if  $|B| = 1$ , then  $\mu(B) = 1$ .

**Example 2.** Let  $W = \{\mathbf{x}_1, \dots, \mathbf{x}_m\} \subseteq \mathbb{R}^n$  be a finite set and let  $d$  be a metric on  $\mathbb{R}^n$ . Define the centroid of  $W$  as  $\mathbf{c}_W = \frac{1}{|W|} \sum_{\mathbf{x} \in W} \mathbf{x}$ .

The sum of square errors of the set  $W$  is defined as:

$$\text{sse}(W) = \sum_{i=1}^m d^2(\mathbf{x}_i, \mathbf{c}_W) = \sum_{\mathbf{x} \in W} \|\mathbf{x}\|^2 - |W| \|\mathbf{c}_W\|^2.$$

If  $W$  is a finite subset of  $\mathbb{R}^n$  and  $\sigma = \{U, V\}$  is a bipartition of  $W$  a straightforward computation yields:

$$\text{sse}(W) = \text{sse}(U) + \text{sse}(V) + \frac{|U||V|}{|W|} \|\mathbf{c}_U - \mathbf{c}_V\|^2,$$

which implies

$$sse(U) + sse(V) \leq sse(W). \quad (1)$$

Note also that  $U, W$  are two finite subsets of  $\mathbb{R}^n$  such that  $U \subseteq W$ , we have  $sse(U) \leq sse(W)$ , which shows that  $sse : \mathcal{P}_{fin}(\mathbb{R}^n) \rightarrow \mathbb{R}_{\geq 0}$  is a monotonic function. Furthermore, if  $|W| = 1$ , then  $sse(W) = 0$ .

Another function that can be defined on finite subsets of  $(\mathbb{R}^n, d)$  is the diameter  $diam : \mathcal{P}(\mathbb{R}^n) \rightarrow \mathbb{R}_{\geq 0}$ , given by  $diam(W) = \max\{d(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in W\}$ . It is immediate that  $diam$  is monotonic.

**Example 3.** Let  $G = (V, E)$  be a connected loop-free finite graph having  $V$  as its set of vertices and  $E$  as its set of edges. For a set of vertices  $B$  define  $int(B)$ , the set of internal edges of  $B$  as

$$int(B) = \{\{x, y\} \in E \mid \{x, y\} \subseteq B\}.$$

This definition is extended to partitions of sets of vertices by defining

$$int(\pi) = \bigcup_{B \in \pi} int(B).$$

The set  $int(\pi)$  is the set of internal edges of  $\pi$ .

If  $\pi, \sigma \in PART(V)$  then  $int(\pi \wedge \sigma) = int(\pi) \cap int(\sigma)$ .

The set  $ext(\pi)$  of *external edges* of  $\pi$  (also known as *cut edges* of  $\pi$ ) consists of edges that join vertices in distinct blocks and is given by:

$$ext(\pi) = E - int(\pi).$$

Thus, we have:

$$\begin{aligned} ext(\pi \wedge \sigma) &= E - int(\pi \wedge \sigma) \\ &= E - (int(\pi) \cap int(\sigma)) \\ &= (E - int(\pi)) \cup (E - int(\sigma)) \\ &= ext(\pi) \cup ext(\sigma). \end{aligned}$$

Note that

$$\begin{aligned} int(\alpha_V) &= \emptyset, & ext(\alpha_V) &= E, \\ int(\omega_V) &= E, & ext(\omega_V) &= \emptyset \end{aligned}$$

for every graph  $G = (V, E)$ .

It follows from the above discussion that the function  $int : PART(V) \rightarrow \mathcal{P}(E)$  is monotonic, while  $ext : PART(V) \rightarrow \mathcal{P}(E)$  is dually monotonic.

Therefore, the function  $\mu_{int} : PART(V) \rightarrow \mathbb{R}_{\geq 0}$  defined as  $\mu_{int}(B) = |\text{int}(B)|$  is a monotonic function.

Starting from monotonic functions of sets we introduce a set of three axioms that define an entropy associated to these functions.

**Definition 4.** Let  $S$  be a non-empty set and let  $\mu : \mathcal{P}(S) \rightarrow \mathbb{R}_{\geq 0}$  be a non-negative monotonic function defined on the subsets of  $S$ . A  $\mu$ -entropy is a function  $\mathcal{H}_\mu : PART(S) \rightarrow \mathbb{R}_{\geq 0}$  that satisfies the following conditions:

- ( $\mathbf{A}_0$ )-initialization axiom: For any set  $S$ ,  $\mathcal{H}_\mu(\omega_S) = 0$ .
- ( $\mathbf{A}_1$ )-monotonicity axiom: If  $\pi, \sigma \in PART(S)$  and  $\pi \leq \sigma$ , then  $\mathcal{H}_\mu(\pi) \geq \mathcal{H}_\mu(\sigma)$ .
- ( $\mathbf{A}_2$ )-addition axiom: For every finite disjoint subsets  $U, V$  of a set  $S$  such that  $S = U \cup V$ ,  $\sigma \in PART(U)$  and  $\tau \in PART(V)$  we have:

$$\mathcal{H}_\mu(\sigma + \tau) = \frac{\mu(U)}{\mu(U \cup V)} \mathcal{H}_\mu(\sigma) + \frac{\mu(V)}{\mu(U \cup V)} \mathcal{H}_\mu(\tau) + \mathcal{H}_\mu(\{U, V\}).$$

Note that if  $\mathcal{H}_\mu$  is a function on the partitions of  $S$  that satisfies the above axioms then for any positive  $a$ ,  $a\mathcal{H}_\mu$  also satisfies the axioms.

**Lemma 5.** If  $|S| = 1$ , then  $\mathcal{H}_\mu(\alpha_S) = 0$ .

*Proof.* This follows from the fact that for a singleton set  $S = \{a\}$  we have  $\alpha_S = \omega_S$ .  $\square$

**Lemma 6.** Let  $U, V$  be two non-empty, finite disjoint sets,  $\mu : \mathcal{P}(U \cup V) \rightarrow \mathbb{R}_{\geq 0}$  be a positive monotonic function of sets, and let  $\sigma$  be a partition of the set  $U$ . Then,

$$\mathcal{H}_\mu(\sigma + \alpha_V) = \mathcal{H}_\mu(\sigma + \omega_V) + \frac{\mu(V)}{\mu(U \cup V)} \mathcal{H}_\mu(\alpha_V).$$

*Proof.* By Definition 4 we can write:

$$\begin{aligned} \mathcal{H}_\mu(\sigma + \alpha_V) &= \frac{\mu(U)}{\mu(U \cup V)} \mathcal{H}_\mu(\sigma) + \frac{\mu(V)}{\mu(U \cup V)} \mathcal{H}_\mu(\alpha_V) + \mathcal{H}_\mu(\{U, V\}), \\ \mathcal{H}_\mu(\sigma + \omega_V) &= \frac{\mu(U)}{\mu(U \cup V)} \mathcal{H}_\mu(\sigma) + \mathcal{H}_\mu(\{U, V\}). \end{aligned}$$

The equalities imply the desired result.  $\square$

**Theorem 7.** *Let  $S$  be a set such that  $|S| \geq 2$  and let  $\pi = \{B_1, \dots, B_m\}$  be a partition of  $S$ . For any non-negative monotonic function  $\mu : \mathcal{P}(S) \rightarrow \mathbb{R}_{\geq 0}$  we have:*

$$\mathcal{H}_\mu(\pi) = \mathcal{H}_\mu(\alpha_S) - \sum_{i=1}^m \frac{\mu(B_i)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_i}). \quad (2)$$

*Proof.* Since  $\pi = \omega_{B_1} + \omega_{B_2} + \dots + \omega_{B_m}$ , we can consider the descending sequence of partitions of the set  $S$ :

$$\begin{aligned} \pi_0 &= \omega_{B_1} + \omega_{B_2} + \dots + \omega_{B_m} = \pi \\ \pi_1 &= \alpha_{B_1} + \omega_{B_2} + \dots + \omega_{B_m} \\ \pi_2 &= \alpha_{B_1} + \alpha_{B_2} + \dots + \omega_{B_m} \\ &\vdots \\ \pi_m &= \alpha_{B_1} + \alpha_{B_2} + \dots + \alpha_{B_m} = \alpha_S. \end{aligned}$$

Define  $\sigma_i = \alpha_{B_1} + \dots + \alpha_{B_i} + \omega_{B_{i+2}} + \dots + \omega_{B_m} \in \text{PART}(S - B_{i+1})$  for  $1 \leq i \leq m-1$ . Note that

$$\pi_i = \sigma_i + \omega_{B_{i+1}} \text{ and } \pi_{i+1} = \sigma_i + \alpha_{B_{i+1}}$$

are both partitions of the set  $S$ . By Lemma 6 we have:

$$\begin{aligned} \mathcal{H}_\mu(\pi_1) &= \mathcal{H}_\mu(\pi_0) + \frac{\mu(B_1)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_1}), \\ \mathcal{H}_\mu(\pi_2) &= \mathcal{H}_\mu(\pi_1) + \frac{\mu(B_2)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_2}), \\ &\vdots \\ \mathcal{H}_\mu(\pi_m) &= \mathcal{H}_\mu(\pi_{m-1}) + \frac{\mu(B_m)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_m}). \end{aligned}$$

Therefore,

$$\mathcal{H}_\mu(\pi_m) = \mathcal{H}_\mu(\pi_0) + \sum_{i=1}^m \frac{\mu(B_i)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_i})$$

Equivalently, since  $\pi_m = \alpha_S$ , we gave

$$\mathcal{H}_\mu(\pi) = \mathcal{H}_\mu(\alpha_S) - \sum_{i=1}^m \frac{\mu(B_i)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_i}).$$

□

**Corollary 8.** *Let  $S$  be set such that  $|S| \geq 2$ . For any non-negative monotonic function  $\mu : \mathcal{P}(S) \rightarrow \mathbb{R}_{\geq 0}$  and any partition  $\pi = \{B_1, \dots, B_n\} \in \text{PART}(S)$  we have:*

$$\mathcal{H}_\mu(\alpha_S) \geq \sum_{i=1}^m \frac{\mu(B_i)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_i}). \quad (3)$$

*Proof.* By the initialization and monotonicity axioms  $\pi \leq \omega_S$  imply  $\mathcal{H}_\mu(\pi) \geq \mathcal{H}_\mu(\omega_S) = 0$ , hence the  $\mu$ -entropy of any partition is non-negative. This fact combined with Theorem 7 yields the desired result.  $\square$

**Example 9.** *Let  $\mu(S) = |S|^\beta$  for any finite and non-empty set  $S$  and  $\beta > 0$  and let*

$$\mathcal{H}_\mu(\alpha_S) = \frac{1 - |S|^{1-\beta}}{1 - 2^{1-\beta}}.$$

*By Theorem 7 this choice of  $\mathcal{H}_\mu(\alpha_S)$  implies:*

$$\mathcal{H}_\mu(\pi) = \frac{1}{1 - 2^{1-\beta}} \left( 1 - \sum_{B \in \pi} \frac{|B|^\beta}{|S|^\beta} \right),$$

*which is the Havrda-Charvát generalized entropy obtained in [7].*

*Note that*

$$\lim_{\beta \rightarrow 1^+} \frac{1 - |S|^{1-\beta}}{1 - 2^{1-\beta}} = \ln |S|,$$

*by a straightforward application of l'Hospital rule.*

*If  $\pi \leq \sigma$  the axiom  $(\mathbf{A}_1)$  is satisfied. It suffices to show that  $\pi \triangleleft \sigma$  implies  $\mathcal{H}_\mu(\pi) \geq \mathcal{H}_\mu(\sigma)$ , so let  $\pi = \{B_1, \dots, B_{m-2}, B_{m-1}, B_m\}$  and let  $\sigma = \{B_1, \dots, B_{m-2}, B_{m-1} \cup B_m\}$ . These choices imply:*

$$\begin{aligned} \mathcal{H}_\mu(\pi) &= \frac{1}{1 - 2^{1-\beta}} \left( 1 - \sum_{i=1}^m \frac{|B_i|^\beta}{|S|^\beta} \right), \\ \mathcal{H}_\mu(\sigma) &= \frac{1}{1 - 2^{1-\beta}} \left( 1 - \sum_{i=1}^{m-2} \frac{|B_i|^\beta}{|S|^\beta} \right) - \frac{|B_{m-1} \cup B_m|^\beta}{|S|^\beta}, \end{aligned}$$

*and the axiom  $(\mathbf{A}_1)$  is satisfied because*

$$|B_{m-1}|^\beta + |B_m|^\beta \leq |B_{m-1} \cup B_m|^\beta.$$

*The special case  $\beta = 2$  yields*

$$\mathcal{H}_\mu(\pi) = 2 \left( 1 - \sum_{i=1}^m \frac{|B_i|^2}{|S|^2} \right),$$



which is the double of the Gini index.

By applying l'Hospital rule we obtain:

$$\lim_{\beta \rightarrow 1} \mathcal{H}_\mu(\pi) = - \sum_{i=1}^m \frac{|B_i|}{|S|} \ln \frac{|B_i|}{|S|}$$

which is the Shannon entropy.

**Example 10.** Let  $\mu$  be the positive monotonic function introduced in Example 2,  $\mu(B) = sse(B)$ , where  $B$  is a finite subset of  $\mathbb{R}^n$ . Choose  $\mathcal{H}_\mu(\alpha_U) = 1$  for every finite set  $U \in \mathcal{P}(S)$ . Then, the  $\mu$ -entropy is:

$$\mathcal{H}_\mu(\pi) = 1 - \sum_{i=1}^m \frac{sse(B_i)}{sse(S)},$$

which is the expression of the inertial entropy of a partition studied in [21].

The satisfaction of axiom  $(A_1)$  follows from Inequality (1).

With the alternative choice,  $\mu(B) = \text{diam}(B)$  we obtain the entropy

$$\mathcal{H}_\mu(\pi) = 1 - \sum_{i=1}^m \frac{\text{diam}(B_i)}{\text{diam}(S)},$$

**Example 11.** Let  $G = (V, E)$  be a connected loop-free finite graph and let  $\mu_{gr}(B) = |\text{int}(B)|$  for every set of vertices  $B$  be the function defined in Example 3. By choosing  $\mathcal{H}_\mu(\alpha_B) = 1$ , the expression of the  $\mu$ -entropy of a partition  $\pi = \{B_1, \dots, B_m\} \in \text{PART}(V)$  is:

$$\begin{aligned} \mathcal{H}_{\mu_{gr}}(\pi) &= 1 - \sum_{i=1}^m \frac{\mu_{\text{int}}(B_i)}{\mu_{\text{int}}(V)} \\ &= 1 - \sum_{i=1}^m \frac{|\text{int}(B_i)|}{|\text{int}(V)|} \\ &= \frac{|\text{ext}(\pi)|}{|E|}. \end{aligned}$$

**Example 12.** A subset  $C$  of the set of vertices  $V$  of a graph  $G = (V, E)$  is a clique if  $\text{int}(C) = C \times C$ .

The clique-partitioning problem of a graph  $G = (V, E)$  seeks to find a partition  $\pi \in \text{PART}(V)$  that consists of cliques such that each vertex is

contained in exactly one clique. Let  $\kappa = \{C_1, \dots, C_k\}$  be a clique partitioning of  $G = (V, E)$ . The maximum number of edges of  $G$  is  $\frac{|V|(|V|-1)}{2}$ . Therefore,

$$|E| \leq \frac{|V|(|V|-1)}{2} - \frac{k(k-1)}{2}$$

because for each pair of cliques  $C, C'$  at least one edge should be absent between the vertices of these cliques in order to avoid eliminating that pair of cliques by consolidating  $C$  and  $C'$  into one clique. This inequality was established in [3], where it is noted that the upper bound  $\theta_{upper}$  of the minimum number of cliques is

$$\theta_{upper} = \left\lceil \frac{1 + \sqrt{4|V|^2 - 4|V| - 8|E| + 1}}{2} \right\rceil.$$

It is shown that  $\theta_{upper}$  is an optimal bound, which means that for each  $|V|$  and  $|E|$  there exists a graph that has  $\theta_{upper}$  cliques.

The size of the set of internal edges of the cliques is:

$$\begin{aligned} |int(\kappa)| &= \sum_{i=1}^k |int(C_i)| = \sum_{i=1}^k \frac{|C_i|(|C_i|-1)}{2} \\ &= \frac{1}{2} \left( \sum_{i=1}^k |C_i|^2 - |V| \right). \end{aligned}$$

Thus, if  $\pi = \{C_1, \dots, C_k\}$  is a clique partition of the graph  $G$ , its entropy is:

$$\begin{aligned} \mathcal{H}_{\mu_{gr}}(\kappa) &= \frac{|ext(\kappa)|}{|E|} = \frac{1}{|E|} \left( |E| - \sum_{i=1}^k \frac{|C_i|(|C_i|-1)}{2} \right) \\ &= \frac{1}{|E|} \left( |E| - \sum_{i=1}^k \frac{|C_i|^2}{2} + \frac{|V|}{2} \right). \end{aligned}$$

Since  $\sum_{i=1}^k |C_i| = |V|$ , the entropy  $\mathcal{H}_{\mu_{gr}}(\kappa)$  is maximal, when the sizes of the cliques are approximatively equal.

Elementary properties of partition cut-sets of graphs allow us to obtain the necessity of axiom **A**<sub>2</sub> for graph entropies. Indeed, let  $\kappa = \{U, W\}$  be a cut in the graph  $G$  and let  $\sigma \in PART(U)$  and  $\tau \in PART(W)$  be two partitions of the sets  $U$  and  $W$ . The partition  $\sigma + \tau$  of  $V$  consists of all blocks of  $\sigma$  and all blocks of  $\tau$ .

An external edge  $e$  of partition  $\sigma + \tau$  may fall in one of the following pairwise disjoint sets:

- $e$  is an external edge of  $\sigma$  but an internal edge of  $\kappa$ ;
- $e$  is an external edge of  $\tau$  but an internal edge of  $\kappa$ ;
- $e$  is an external edge of  $\kappa$ .

Since the sets  $\text{ext}(\sigma)$ ,  $\text{ext}(\tau)$ , and  $\text{ext}(\kappa)$  are disjoint we have:

$$\text{ext}(\sigma + \tau) = \text{ext}(\sigma) \cup \text{ext}(\tau) \cup \text{ext}(\kappa).$$

The last equality implies

$$\frac{|\text{ext}(\sigma + \tau)|}{|V|} = \frac{|U|}{|V|} \frac{|\text{ext}(\sigma)|}{|U|} + \frac{|W|}{|V|} \frac{|\text{ext}(\tau)|}{|W|} + \frac{|\text{ext}(\kappa)|}{|V|}.$$

When this equality is expressed using the graph entropy we recover axiom **A<sub>2</sub>**, namely:

$$\mathcal{H}_\mu(\sigma + \tau) = \frac{\mu(U)}{\mu(U \cup W)} \mathcal{H}_\mu(\sigma) + \frac{\mu(W)}{\mu(U \cup W)} \mathcal{H}_\mu(\tau) + \mathcal{H}_\mu(\{U, W\}).$$

A graph  $G = (V, E)$  is bipartite if there exists a bipartition  $\pi = \{V_1, V_2\}$  such that  $\text{ext}(\pi) = E$ . This is equivalent to the existence of a bipartition  $\pi$  such that  $\mathcal{H}_\mu(\pi) = 1$ . In general, a graph  $G = (V, E)$  is  $k$ -colorable, if it has a partition  $\pi = \{B_1, \dots, B_k\}$  such that if  $\{x, y\} \in E$ , then  $x$  and  $y$  belong to two distinct blocks of  $\pi$ . In other words,  $G$  is  $k$ -colorable if and only if there exists a partition of  $V$  having  $k$  blocks such that  $\mathcal{H}_\mu(\pi) = 1$ .

Since the graph  $k$ -coloring problem is known to be NP-complete (see [10]), it follows by direct transformation, that the problem of the existence of a partition with  $k$  blocks of the set of vertices of a graph and has monotonic entropy equal to 1 is NP-complete.

## 4 Conditional Monotonic Entropy

Let  $\pi = \{B_1, \dots, B_m\} \in \text{PART}(S)$  and let  $C \subseteq S$ . The *trace* of  $\pi$  on  $C$  is the partition

$$\pi_C = \{B \cap C \mid B \in \pi \text{ and } B \cap C \neq \emptyset\} \in \text{PART}(C).$$

**Definition 13.** Let  $\pi, \sigma \in PART(S)$ , where  $\sigma = \{C_1, \dots, C_n\}$ . The  $\mu$ -conditional entropy of  $\pi$  and  $\sigma$  is given by:

$$\mathcal{H}_\mu(\pi|\sigma) = \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_j}).$$

Note that  $\mathcal{H}(\pi|\omega_S) = \mathcal{H}(\pi)$ ,

$$\mathcal{H}(\omega_S|\sigma) = \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(C_j),$$

and  $\mathcal{H}_\mu(\pi|\alpha_S) = 0$  for every  $\pi \in PART(S)$ .

**Theorem 14.** For any two partitions  $\pi, \sigma \in PART(S)$  we have:

$$\mathcal{H}_\mu(\pi \wedge \sigma) = \mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma).$$

*Proof.* For  $\pi = \{B_1, \dots, B_m\}$  and  $\sigma = \{C_1, \dots, C_n\}$  in  $PART(S)$  the conditional entropy can be written as:

$$\begin{aligned} \mathcal{H}_\mu(\pi|\sigma) &= \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_j}) \\ &= \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \left( \mathcal{H}_\mu(\alpha_{C_j}) - \sum_{i=1}^m \frac{\mu(B_i \cap C_j)}{\mu(C_j)} \mathcal{H}_\mu(\alpha_{B_i \cap C_j}) \right) \\ &= \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\alpha_{C_j}) - \sum_{i=1}^m \frac{\mu(B_i \cap C_j)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_i \cap C_j}) \\ &= \mathcal{H}_\mu(\alpha_S) - \sum_{i=1}^m \frac{\mu(B_i \cap C_j)}{\mu(S)} \mathcal{H}_\mu(\alpha_{B_i \cap C_j}) \\ &\quad - \left( \mathcal{H}_\mu(\alpha_S) - \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\alpha_{C_j}) \right) \\ &= \mathcal{H}_\mu(\pi \wedge \sigma) - \mathcal{H}_\mu(\sigma). \end{aligned}$$

□

**Corollary 15.** Let  $\pi, \sigma \in PART(S)$ , where  $S$  is a finite set. We have

$$\mathcal{H}_\mu(\pi \wedge \sigma) = \mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma) = \mathcal{H}_\mu(\sigma|\pi) + \mathcal{H}_\mu(\pi).$$

*Proof.* This is a direct consequence of Theorem 14.  $\square$

The following corollary is immediate:

**Corollary 16.** For  $\pi = \{B_1, \dots, B_m\}$  and  $\sigma = \{C_1, \dots, C_n\}$  in  $PART(S)$  we have

$$\mathcal{H}_\mu(\pi|\sigma) \leq \mathcal{H}_\mu(\pi \wedge \sigma)$$

and

$$\mathcal{H}_\mu(\pi) \leq \mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}(\sigma).$$

**Theorem 17.** Let  $\pi, \sigma \in PART(S)$  be two partitions of a finite set  $S$ . We have  $\mathcal{H}_\mu(\pi|\sigma) = 0$  if and only if  $\sigma \leq \pi$ .

*Proof.* Suppose that  $\sigma = \{C_1, \dots, C_n\}$ . If  $\sigma \leq \pi$ , then  $\pi_{C_j} = \omega_{C_j}$  and, therefore,

$$\mathcal{H}_\mu(\pi|\sigma) = \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\omega_{C_j}) = 0.$$

Conversely, suppose that

$$\mathcal{H}_\mu(\pi|\sigma) = \sum_{j=1}^n \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_j}) = 0.$$

This implies  $\mathcal{H}_\mu(\pi_{C_j}) = 0$  for  $1 \leq j \leq n$ , which means that  $\pi_{C_j} = \omega_{C_j}$  for  $1 \leq j \leq n$ . Therefore, each block  $C_j$  of  $\sigma$  is included in a block of  $\pi$ , so  $\sigma \leq \pi$ .  $\square$

We will show that the conditional monotonic entropy is dually monotonic with respect to its first argument and monotonic with respect to its second argument.

**Theorem 18.** Let  $\pi, \sigma, \sigma' \in PART(S)$ , where  $S$  is a finite subset of  $\mathbb{R}^p$ . If  $\sigma \leq \sigma'$ , then  $\mathcal{H}_\mu(\sigma|\pi) \geq \mathcal{H}_\mu(\sigma'|\pi)$ , and  $\mathcal{H}_\mu(\pi|\sigma) \leq \mathcal{H}_\mu(\pi|\sigma')$ .

*Proof.* Since  $\sigma \leq \sigma'$  we have  $\pi \wedge \sigma \leq \pi \wedge \sigma'$ , so  $\mathcal{H}_\mu(\pi \wedge \sigma) \geq \mathcal{H}_\mu(\pi \wedge \sigma')$ . Therefore,  $\mathcal{H}_\mu(\sigma|\pi) + \mathcal{H}_\mu(\pi) \geq \mathcal{H}_\mu(\sigma'|\pi) + \mathcal{H}_\mu(\pi)$ , which implies  $\mathcal{H}_\mu(\sigma|\pi) \geq \mathcal{H}_\mu(\sigma'|\pi)$ .

For the second part, it suffices to prove the inequality for partitions  $\sigma$  and  $\sigma'$  such that  $\sigma$  is covered by  $\sigma'$ . Without loss of generality assume that

$$\sigma = \{C_1, \dots, C_{n-2}, C_{n-1}, C_n\} \text{ and } \sigma' = \{C_1, \dots, C_{n-2}, C_{n-1} \cup C_n\}.$$

We have

$$\begin{aligned}
\mathcal{H}_\mu(\pi|\sigma') &= \sum_{j=1}^{n-2} \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_j}) + \frac{\mu(C_{n-1} \cup C_n)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_{n-1} \cup C_n}) \\
&\geq \sum_{j=1}^{n-2} \frac{\mu(C_j)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_j}) + \frac{\mu(C_{n-1})}{\mu(S)} \mathcal{H}_\mu(\pi_{C_{n-1}}) + \frac{\mu(C_n)}{\mu(S)} \mathcal{H}_\mu(\pi_{C_n}) \\
&\quad \text{(by the Addition Axiom)} \\
&= \mathcal{H}_\mu(\pi|\sigma).
\end{aligned}$$

□

**Theorem 19.** *Let  $\pi, \sigma, \tau$  be three partitions of the finite set  $S$ , where  $S \subseteq \mathbb{R}^p$ . We have:*

$$\mathcal{H}_\mu(\pi|\sigma \wedge \tau) + \mathcal{H}_\mu(\sigma|\tau) = \mathcal{H}_\mu(\pi \wedge \sigma|\tau).$$

*Proof.* By Corollary 15 we have

$$\begin{aligned}
\mathcal{H}_\mu(\pi|\sigma \wedge \tau) &= \mathcal{H}_\mu(\pi \wedge \sigma \wedge \tau) - \mathcal{H}_\mu(\sigma \wedge \tau), \\
\mathcal{H}_\mu(\sigma|\tau) &= \mathcal{H}_\mu(\sigma \wedge \tau) - \mathcal{H}_\mu(\tau).
\end{aligned}$$

By adding these equalities we have

$$\mathcal{H}_\mu(\pi|\sigma \wedge \tau) + \mathcal{H}_\mu(\sigma|\tau) = \mathcal{H}_\mu(\pi \wedge \sigma \wedge \tau) - \mathcal{H}_\mu(\tau).$$

A further application of Corollary 15 yields the desired equality. □

**Theorem 20.** *Let  $\pi, \sigma, \tau$  be three partitions of the finite set  $S$ , where  $S \subseteq \mathbb{R}^p$ . We have*

$$\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau) \geq \mathcal{H}_\mu(\pi|\tau).$$

*Proof.* The monotonicity of the conditional inertial entropy in its second argument and the anti-monotonicity of the same in its first argument allows us to write:

$$\begin{aligned}
\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau) &\geq \mathcal{H}_\mu(\pi|\sigma \wedge \tau) + \mathcal{H}_\mu(\sigma|\tau) \\
&= \mathcal{H}_\mu(\pi \wedge \sigma|\tau) \geq \mathcal{H}_\mu(\pi|\tau),
\end{aligned}$$

which is the desired inequality. □

**Corollary 21.** *Let  $\pi, \sigma$  be two partitions of the finite set  $S$ , where  $S \subseteq \mathbb{R}^p$ . We have*

$$\mathcal{H}_\mu(\pi \vee \sigma) + \mathcal{H}_\mu(\pi \wedge \sigma) \leq \mathcal{H}_\mu(\pi) + \mathcal{H}_\mu(\sigma).$$

*Proof.* By Theorem 20 we have  $\mathcal{H}_\mu(\pi|\sigma) \leq \mathcal{H}_\mu(\pi|\tau) + \mathcal{H}_\mu(\tau|\sigma)$ . Replacing the conditional inertial entropies we obtain

$$\mathcal{H}_\mu(\pi \wedge \sigma) - \mathcal{H}_\mu(\sigma) \geq \mathcal{H}_\mu(\pi \wedge \tau) - \mathcal{H}_\mu(\tau) + \mathcal{H}_\mu(\tau \wedge \sigma) - \mathcal{H}_\mu(\sigma),$$

which implies

$$\mathcal{H}_\mu(\tau) + \mathcal{H}_\mu(\pi \wedge \sigma) \leq \mathcal{H}_\mu(\pi \wedge \tau) + \mathcal{H}_\mu(\tau \wedge \sigma).$$

Choosing  $\tau = \pi \vee \sigma$  yields the desired inequality.  $\square$

## 5 Metrics on Partitions Induced by Monotonic Entropies

Conditional monotonic entropy induce metrics on spaces of partitions as we show next. These properties of these metrics generalize previous results obtained by López de Mántaras in [5].

**Theorem 22.** *The mapping  $d_\mu : PART(S)^2 \rightarrow \mathbb{R}_{\geq 0}$  defined as*

$$d_\mu(\pi, \sigma) = \mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\pi)$$

*is a metric on  $PART(S)$ .*

*Proof.* A double application of Theorem 20 yields

$$\begin{aligned} \mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau) &\geq \mathcal{H}_\mu(\pi|\tau), \\ \mathcal{H}_\mu(\sigma|\pi) + \mathcal{H}_\mu(\tau|\sigma) &\geq \mathcal{H}_\mu(\tau|\pi). \end{aligned}$$

Adding these inequalities gives the triangular inequality for  $d_\mu$ :

$$d_\mu(\pi, \sigma) + d_\mu(\sigma, \tau) \geq d_\mu(\pi, \tau).$$

The symmetry of  $d_\mu$  is immediate and it is clear that  $d_\mu(\pi, \pi) = 0$  for every  $\pi \in PART(S)$ .

Suppose now that  $d_\mu(\pi, \sigma) = 0$ . Since the values of  $\mathcal{H}_\mu$  are non-negative, this implies  $\mathcal{H}_\mu(\pi|\sigma) = \mathcal{H}_\mu(\sigma|\pi) = 0$ . By Theorem 17, we have both  $\sigma \leq \pi$  and  $\pi \leq \sigma$ , so  $\pi = \sigma$ . Thus,  $d_\mu$  is a metric on  $PART(S)$ .  $\square$

**Example 23.** The Rand distance of two partitions  $\pi, \sigma \in \text{PART}(S)$  is the number  $\text{rd}(\pi, \sigma)$  of unordered pairs  $\{x, y\}$  of elements of  $S$  such that there exists a block in one partition containing both  $x$  and  $y$  but  $x$  and  $y$  are in different blocks in the other partition (see [15]). For example, if  $S = \{1, 2, 3\}$ ,  $\pi = \{\{1, 2\}, \{3\}\}$  and  $\sigma = \{\{1\}, \{2, 3\}\}$ , then  $\text{rd}(\pi, \sigma) = 2$  (the pairs involved being  $\{1, 2\}$  and  $\{2, 3\}$ ). In particular, if  $\pi \in \text{PART}(S)$ , then  $\text{rd}(\pi, \omega_S)$  equals the number of pairs  $(x, y)$  such that  $x \neq y(\pi)$ .

Actually, the Rand distance is a multiple of the metric  $d_\mu$ , where  $\mu(U) = |U|^2$  for  $U \in \mathcal{P}(S)$ . Indeed, let  $\pi = \{B_1, \dots, B_m\}$  and  $\sigma = \{C_1, \dots, C_n\}$  be two partitions in  $\text{PART}(S)$ . We have

$$\begin{aligned} \mathcal{H}_\mu(\pi|\sigma) &= \mathcal{H}_\mu(\pi \wedge \sigma) - \mathcal{H}_\mu(\sigma) \\ &= \frac{2}{|S|^2} \left( \sum_{j=1}^n |C_j|^2 - \sum_{i=1}^m \sum_{j=1}^n |B_i \cap C_j|^2 \right). \end{aligned}$$

The expression  $\sum_{j=1}^n |C_j|^2 - \sum_{i=1}^m \sum_{j=1}^n |B_i \cap C_j|^2$  equals the number of pairs of elements that belong to the same block of  $\sigma$  but to distinct blocks of  $\pi$ . The similar expression  $\sum_{i=1}^m |B_i|^2 - \sum_{i=1}^m \sum_{j=1}^n |B_i \cap C_j|^2$  gives the number of pairs of elements that belong to the same class of  $\pi$ , but to two distinct classes of  $\sigma$ , and, therefore,  $\text{rd}(\pi, \sigma)$  is a multiple of  $d_\mu$ .

**Lemma 24.** Let  $\pi, \sigma, \tau$  be three partitions in  $\text{PART}(S)$ . We have:

$$\frac{\mathcal{H}_\mu(\pi|\sigma)}{\mathcal{H}_\mu(\pi \wedge \sigma)} + \frac{\mathcal{H}_\mu(\sigma|\tau)}{\mathcal{H}_\mu(\sigma \wedge \tau)} \geq \frac{\mathcal{H}_\mu(\pi|\tau)}{\mathcal{H}_\mu(\pi \wedge \tau)}.$$

*Proof.* By applying the definition of conditional entropy we can write:

$$\begin{aligned} & \frac{\mathcal{H}_\mu(\pi|\sigma)}{\mathcal{H}_\mu(\pi \wedge \sigma)} + \frac{\mathcal{H}_\mu(\sigma|\tau)}{\mathcal{H}_\mu(\sigma \wedge \tau)} \\ &= \frac{\mathcal{H}_\mu(\pi|\sigma)}{\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma)} + \frac{\mathcal{H}_\mu(\sigma|\tau)}{\mathcal{H}_\mu(\sigma|\tau) + \mathcal{H}_\mu(\tau)} \\ & \quad \text{(by Theorem 14)} \\ & \geq \frac{\mathcal{H}_\mu(\pi|\sigma)}{\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau) + \mathcal{H}_\mu(\tau)} + \frac{\mathcal{H}_\mu(\sigma|\tau)}{\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau) + \mathcal{H}_\mu(\tau)} \\ & \quad \text{(by Corollary 16)} \\ &= \frac{\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau)}{\mathcal{H}_\mu(\pi|\sigma) + \mathcal{H}_\mu(\sigma|\tau) + \mathcal{H}_\mu(\tau)} \\ & \geq \frac{\mathcal{H}_\mu(\pi|\tau)}{\mathcal{H}_\mu(\pi|\tau) + \mathcal{H}_\mu(\tau)} = \frac{\mathcal{H}_\mu(\pi|\tau)}{\mathcal{H}_\mu(\pi \wedge \tau)}, \end{aligned}$$



which is the desired inequality.  $\square$

**Theorem 25.** *The mapping  $\delta_\mu : PART(S)^2 \rightarrow \mathbb{R}_{\geq 0}$  defined as*

$$\delta_\mu(\pi, \sigma) = \frac{d_\mu(\pi, \sigma)}{\mathcal{H}_\mu(\pi \wedge \sigma)}$$

*is a metric on  $PART(S)$  such that  $0 \leq \delta_\mu(\pi, \sigma) \leq 1$  for  $\pi, \sigma \in PART(S)$ .*

*Proof.* The non-negativity and the symmetry of  $\delta_\mu$  are immediate. To prove the triangular axiom we write:

$$\begin{aligned} \delta_\mu(\pi, \tau) &= \frac{d_\mu(\pi, \tau)}{\mathcal{H}_\mu(\pi \wedge \tau)} \\ &= \frac{\mathcal{H}_\mu(\pi|\tau) + \mathcal{H}_\mu(\tau|\pi)}{\mathcal{H}_\mu(\pi \wedge \tau)} \\ &\leq \frac{\mathcal{H}_\mu(\pi|\sigma)}{\mathcal{H}_\mu(\pi \wedge \sigma)} + \frac{\mathcal{H}_\mu(\sigma|\tau)}{\mathcal{H}_\mu(\sigma \wedge \tau)} \\ &\quad + \frac{\mathcal{H}_\mu(\tau|\sigma)}{\mathcal{H}_\mu(\tau \wedge \sigma)} + \frac{\mathcal{H}_\mu(\sigma|\pi)}{\mathcal{H}_\mu(\sigma \wedge \pi)} \\ &\quad \text{(by Lemma 24)} \\ &= \delta_\mu(\pi, \sigma) + \delta_\mu(\sigma, \pi), \end{aligned}$$

Furthermore, since  $\mathcal{H}_\mu(\pi|\sigma) \leq \mathcal{H}_\mu(\pi \wedge \sigma)$  it follows that  $0 \leq \delta_\mu(\pi, \sigma) \leq 1$ .  $\square$

For  $\pi, \sigma \in PART(S)$  we have:

$$\delta_\mu(\pi, \sigma) = 2 - \frac{\mathcal{H}_\mu(\pi) + \mathcal{H}_\mu(\sigma)}{\mathcal{H}_\mu(\pi \wedge \sigma)}.$$

**Example 26.** *For the graph-related entropy introduced in Example 11 the distance  $\delta_\mu$  is given by:*

$$\delta_\mu(\pi, \sigma) = 2 - \frac{|\text{ext}(\pi)| + |\text{ext}(\sigma)|}{|\text{ext}(\pi \wedge \sigma)|}.$$

**Example 27.** *We consider an analogue of the Rand distance between partitions of sets of edges in undirected graphs that can be introduced using our approach. Using the notations from Example 3, let  $\pi, \sigma \in PART(V)$ ,*

where  $V$  is the set of vertices of a graph  $G = (V, E)$ ; the graph Rand distance  $\delta(\pi, \sigma)$  between these partitions is:

$$\begin{aligned} \delta(\pi, \sigma) &= 2 - \frac{|\text{ext}(\pi)| + |\text{ext}(\sigma)|}{|\text{ext}(\pi \wedge \sigma)|} \\ &= \frac{|\text{int}(\pi) \cap \text{ext}(\sigma)| + |\text{int}(\sigma) \cap \text{ext}(\pi)|}{|\text{ext}(\pi \wedge \sigma)|}, \end{aligned}$$

because

$$\begin{aligned} \text{ext}(\pi \wedge \sigma) - \text{ext}(\pi) &= \text{int}(\pi) \cap \text{ext}(\sigma) \\ \text{ext}(\pi \wedge \sigma) - \text{ext}(\sigma) &= \text{int}(\sigma) \cap \text{ext}(\pi). \end{aligned}$$

## 6 Conclusion

We introduced monotonic entropy as a generalization of Shannon entropy and formulated a system of three axioms that depend on a numeric monotonic function defined on the set of partitions of a finite set. By specializing this function we recapture as a special case the Shannon entropy. Furthermore, this axiomatization allows us to extend the notion of entropy to partitions of sets of objects that possess special properties (such as being embedded in a metric space, or being defined by partitions of undirected graphs).

We show that the notion of conditional entropy defined for the newly axiomatized types of entropy allows us to introduce certain metrics on partitions generalizing the results obtained in [5].

Since clustering can be regarded as partitions, the new metrics structure will allow us to apply our results in the study of stability of clustering algorithms (see [1, 2, 12, 13]) and in the external validation of clusterings, where an a priori data labeling can be compared with the results of clustering algorithms.

## References

- [1] Shai Ben-David, Dávid Pál, and Hans Ulrich Simon. Stability of  $k$ -means clustering. In Nader H. Bshouty and Claudio Gentile, editors, *Learning Theory, 20th Annual Conference on Learning Theory, COLT 2007*, volume 4539 of *Lecture Notes in Computer Science*, pages 20–34. Springer, 2007. doi:10.1007/978-3-540-72927-3\_4.

- 
- [2] Shai Ben-David, Ulrike von Luxburg, and Dávid Pál. A sober look at clustering stability. In Gábor Lugosi and Hans Ulrich Simon, editors, *Learning Theory, 19th Annual Conference on Learning Theory, COLT 2006*, volume 4005 of *Lecture Notes in Computer Science*, pages 5–19. Springer, 2006. doi:[10.1007/11776420\\_4](https://doi.org/10.1007/11776420_4).
- [3] J. Bhasker and Tariq Samad. The clique-partitioning problem. *Computers & Mathematics with Applications*, 22(6):1–11, 1991. doi:[10.1016/0898-1221\(91\)90001-K](https://doi.org/10.1016/0898-1221(91)90001-K).
- [4] Garrett Birkhoff. *Lattice Theory*, volume XXV of *American Mathematical Society Colloquim Publications*. Providence, RI, 3rd edition, 1973.
- [5] Ramón López de Mántaras. A distance-based attribute selection measure for decision tree induction. *Machine Learning*, 6:81–92, 1991. doi:[10.1023/A:1022694001379](https://doi.org/10.1023/A:1022694001379).
- [6] Dmitry Konstantinovich Faddeev. On the notion of entropy of finite probability distributions (in russian). *Uspekhi Matematicheskikh Nauk*, 11:227–231, 1956.
- [7] Jan Havrda and František Charvát. Quantification methods of classification processes: Concepts of structural  $\alpha$ -entropy. *Kybernetika*, 3(1):30–35, 1967.
- [8] Roman S. Ingarden and Kazimierz Urbanik. Information without probability. *Colloquium Mathematicum*, 9(1):131–150, 1962.
- [9] Szymon Jaroszewicz and Dan A. Simovici. On axiomatization of conditional entropy of functions between finite sets. In *29th IEEE International Symposium on Multiple-Valued Logic, ISMVL 1999*, pages 24–28. IEEE Computer Society, 1999. doi:[10.1109/ISMVL.1999.779690](https://doi.org/10.1109/ISMVL.1999.779690).
- [10] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972. doi:[10.1007/978-1-4684-2001-2\\_9](https://doi.org/10.1007/978-1-4684-2001-2_9).
- [11] Aleksandr Yakovlevich Khinchin. *Mathematical Foundations of Information Theory*. Dover, New York, 1957.

- [12] Tianmou Liu, Han Yu, and Rachael Hageman Blair. Stability estimation for unsupervised clustering: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(6):e1575, 2022. doi:[10.1002/wics.1575](https://doi.org/10.1002/wics.1575).
- [13] Ulrike Von Luxburg. Clustering stability: An overview. *Foundations and Trends in Machine Learning*, 2(3):235–274, 2010. doi:[10.1561/2200000008](https://doi.org/10.1561/2200000008).
- [14] Alfréd Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 10:193–215, 1959. doi:[10.1007/BF02063299](https://doi.org/10.1007/BF02063299).
- [15] Frank Ruskey and Jennifer Woodcock. The Rand and block distances of pairs of set partitions. In Costas S. Iliopoulos and William F. Smyth, editors, *22nd International Workshop on Combinatorial Algorithms, IWOCA 2011*, volume 7056 of *Lecture Notes in Computer Science*, pages 287–299. Springer, 2011. doi:[10.1007/978-3-642-25011-8\\_23](https://doi.org/10.1007/978-3-642-25011-8_23).
- [16] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. doi:[10.1002/J.1538-7305.1948.TB01338.X](https://doi.org/10.1002/J.1538-7305.1948.TB01338.X).
- [17] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(4):623–656, 1948. doi:[10.1002/J.1538-7305.1948.TB00917.X](https://doi.org/10.1002/J.1538-7305.1948.TB00917.X).
- [18] Dan A. Simovici. *Clustering - Theoretical and Practical Aspects*. World-Scientific, Singapore, 2021.
- [19] Dan A. Simovici and Chabane Djeraba. *Mathematical Tools for Data Mining - Set Theory, Partial Orders, Combinatorics*. Advanced Information and Knowledge Processing. Springer, 2nd edition, 2014. doi:[10.1007/978-1-4471-6407-4](https://doi.org/10.1007/978-1-4471-6407-4).
- [20] Dan A. Simovici and Szymon Jaroszewicz. An axiomatization of partition entropy. *IEEE Transactions on Information Theory*, 48(7):2138–2142, 2002. doi:[10.1109/TIT.2002.1013159](https://doi.org/10.1109/TIT.2002.1013159).
- [21] Dan A. Simovici and Joshua Yee. Inertial entropy and external validation of clusterings. under review, June 2024.

- 
- [22] Constantino Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988. doi:[10.1007/BF01016429](https://doi.org/10.1007/BF01016429).
- [23] Kenjiro Yanagi, Shigeru Furuichi, and Ken Kuriyama. Fundamental properties of Tsallis relative entropy. *Journal of Mathematical Physics*, 45(12):4868–4877, 2004. doi:[10.1063/1.1805729](https://doi.org/10.1063/1.1805729).
- [24] Zoltán Daróczy. Generalized information functions. *Information and Control*, 16(1):36–51, 1970. doi:[10.1016/S0019-9958\(70\)80040-7](https://doi.org/10.1016/S0019-9958(70)80040-7).