Axiomatic Characterization of the Hamming and Jaccard Distances

Stanisław Szufa[a,*] **and Tomasz Wąs**[b]

[a]CNRS, Université Paris Dauphine-PSL
[b]University of Oxford
ORCID (Stanisław Szufa): https://orcid.org/0000-0001-6301-6227, ORCID (Tomasz Wąs): https://orcid.org/0000-0003-3492-6584

Abstract. Measures of dissimilarity between a pair of objects can play a pivotal role in many machine learning objectives such as clustering, outlier detection, or data visualization. We focus on data in the form of binary vectors and analyze several methods of measuring dissimilarity between them. We introduce several properties, *axioms*, that a measure of dissimilarity can satisfy and characterize the *Hamming* and *Jaccard* distances as the only measures satisfying particular subsets of our axioms. Based on our analysis, we identify shortcomings of both distances, and propose novel approaches that are better suited for certain applications. We complement our theoretical findings by extensive empirical study. Our primary motivation is the analysis of election data, in which the votes have the form of binary approval of alternatives, but the applicability of our results reaches far beyond that.

1 Introduction

The Hamming and Jaccard distances find various applications throughout and beyond machine learning. For example, they can play a key role in clustering [18, 17], keywords similarity analysis [25], and recommendation systems [5]. Jaccard distance also serves as a loss function for object detection in computer vision [12, 28, 33], while Hamming is often used in multi-label classification problems [31, 30]. However, in many applications, it is unclear why this particular metric should be used out of many possible measures of dissimilarity between binary vectors. This question is of dire importance as the choice of the metric can lead to drastic differences in the produced results [9, 28, 17].

To provide a deeper understanding of the nature of each measure and to offer guidance in selecting an appropriate one for a given application, we adopt the *axiomatic method*. This approach involves introducing simple and intuitive properties, called *axioms*, which characterize particular measures, showcasing the distinctive behavior of each measure and highlighting the similarities and differences between them. Axioms have been already employed in the theoretical analysis of machine learning tools such as clustering [1, 4, 20], or classification [11]. Furthermore, they are cornerstones of the related fields of computational social choice [7] and game theory [29].

Specifically, we introduce five *invariance* axioms (namely, *Anonymity*, *Independent Symmetry*, *Add Zero*, *Zero-One Symmetry*, and *Scaling*) each requiring that a particular simple operation applied to a pair of binary vectors does not affect the dissimilarity between them. Additionally, we consider *Triangle Inequality*—a standard distance axiom; *Convergence* which asserts that the distance between the vectors decreases as we introduce agreement between them; and *Normalization* that fixes the dissimilarity value in a certain corner case.

We prove that the Hamming distance is the unique dissimilarity measure that satisfies Anonymity, Independent Symmetry, Zero-One Symmetry, Scaling, Triangle Inequality, Convergence, and Normalization axioms. Moreover, if in this characterization we exchange Zero-One Symmetry for Add Zero axiom, then such set of axioms uniquely characterize the Jaccard distance. This axiomatic approach thus offers a comprehensive comparison of both measures highlighting the key difference between them. Satisfaction of our axioms by the rules we consider is summarized in Table 1.

Furthermore, based on our analysis, we identify a common feature of the Hamming and Jaccard distances that may be undesirable in many scenarios. Indeed, both distances satisfy Independent Symmetry, which in practice implies that the vectors of different saturation will always be far from each other. We propose the modifications of both measures that escape this problem and show that they can be more suited for certain applications.

We complement our theoretical findings with extensive experiments on approval election data, in which a set of voters express binary preferences over a set of alternatives. Such data is one of the primary objects of study in the computational social choice (for more details see the work of Boehmer et al. [6]) and shows a high level of asymmetry: A voters' approval for an alternative is a much stronger signal than the lack thereof. We compare the behavior of the studied dissimilarity measures in several canonical classes of such elections and finish with the case study on real-world election instances from participatory budgeting. Our analysis confirms that the newly proposed measures can indeed be more efficient in certain settings. All missing proofs are available in Section 9.

2 Preliminaries

For every $n \in \mathbb{N}$, we denote $[n] = \{1, \ldots, n\}$. For two sets A, B, we denote their *symmetric difference* by $A \triangle B = (A \cup B) \setminus (A \cap B)$.

* Corresponding Author. Email: s.szufa@gmail.com

Table 1: Which dissimilarity measure satisfies which axiom.

	Anonymity	Scaling	Independent Sym.	Zero-One Sym.	Convergence	Triangle Ineq.	Normalization	Add 0
Hamming	✓	✓	✓	✓	✓	✓	✓	✗
Jaccard	✓	✓	✓	✗	✓	✓	✓	✓

2.1 *Binary Vectors*

We consider a space of *binary vectors* of arbitrary (but finite) length. Each such vector, $x \in \{0,1\}^n$, can equivalently be viewed as a subset of coordinates on which there are ones in the vector, which we denote by $A_x = \{i \in [n] : x_i = 1\}$. For a permutation $\pi : [n] \to [n]$, by $\pi(x)$ we denote the vector x with changed order of coordinates, i.e., $\pi(x)_i = x_{\pi(i)}$ for every $i \in [n]$. Drawing from game theory notation, for vector $x \in \{0,1\}^n$, coordinate $i \in [n]$, and $b \in \{0,1\}$ we write $y = (x_{-i}, b)$ to denote a vector such that $y_i = b$ and $y_j = x_j$ for every $j \in [n] \setminus \{i\}$. For vector $x \in \{0,1\}^n$ by $\bar{x}$ we denote its bitwise negation, i.e., $\bar{x} = (1 - x_1, \ldots, 1 - x_n)$. By $\circ$ we denote *concatenation* of vectors, i.e., for $x \in \{0,1\}^n$ and $y \in \{0,1\}^m$, if $z = x \circ y$, then $z_i = x_i$ for $i \in [n]$ and $z_i = y_{i-n}$ for every $i \in [m+n] \setminus [n]$. Finally, by x^k we denote the result of concatenating k copies of x one after another. For convenience, we will allow for x^0 to be the empty vector, which concatenated with any other vector y, results in y.

2.2 *Dissimilarity Measures*

A *dissimilarity measure* is a function that takes a pair of binary vectors of the same size as arguments and outputs some real nonnegative values, i.e., $f : \left(\{0,1\}^n \times \{0,1\}^n\right)_{n\in\mathbb{N}} \to \mathbb{R}_{\geq}$.

Examples include the (normalized) *Hamming distance*, also known as the ℓ_1 distance, which outputs the fraction of positions at which the corresponding elements in two vectors differ, i.e.,

$$H(x,y) = |A_x \triangle A_y|/n.$$

Jaccard distance is defined similarly, but instead of normalizing by the vector length, it divides the symmetric difference by the number of positions in which at least one vector has a positive entry, i.e.,

$$J(x,y) = |A_x \triangle A_y|/|A_x \cup A_y|.$$

Other examples include the *Euclidean distance* (or ℓ_2), which is just a square root of the Hamming distance, or the *discrete distance* (or ℓ_∞) that returns 0 if the vectors are identical and 1 otherwise.

3 Axiomatic Characterization of the Hamming Distance

We begin by introducing seven axioms that uniquely characterize the Hamming distance, arguably the most popular notion of dissimilarity between binary vectors.

The first four axioms are so-called *invariance axioms*, which describe certain operations on a pair of vectors and assert that this operation should not affect the dissimilarity value. This first such axiom is called *Anonymity* and intuitively states that the order of coordinates does not matter. It is natural in settings where there is no spatial or temporal relation between the bits of information that the vectors represent. Rather, they come from independent sources, as, for example, when the vectors represent support for a given candidate by various voters (indeed, the version of this axiom often appears in the social choice literature [24]).

Definition 3.1 (Anonymity)**.** *A dissimilarity measure, f, satisfies* Anonymity *if for every vectors $x, y \in \{0,1\}^n$ and permutation $\pi : [n] \to [n]$ it holds that*

$$f(\pi(x), \pi(y)) = f(x,y).$$

For example, Anonymity would imply that $f((1,0,1,1),(1,1,0,0)) = f((0,1,1,1),(1,0,1,0))$.

The second axiom, *Scaling*, captures the intuition that the scale of the input should not be relevant for the dissimilarity. Hence, we can copy the entries of the vectors several times and this will not affect the value. This property was noted as crucial for some objectives such as measuring accuracy in object detection tasks in computer vision [28].

Definition 3.2 (Scaling)**.** *A dissimilarity measure, f, satisfies* Scaling *if for every vectors $x, y \in \{0,1\}^n$ and $k \in \mathbb{N}$ it holds that*

$$f(x^k, y^k) = f(x,y).$$

For example, Scaling would imply that $f((1,0),(1,1)) = f((1,0,1,0),(1,1,1,1))$.

Our next axiom is a strengthening of a standard distance axiom of Symmetry. *Independent Symmetry* states that on every coordinate we can switch the values between the two vectors, and this operation should preserve the dissimilarity value.

Definition 3.3 (Independent Symmetry)**.** *A dissimilarity measure, f, satisfies* Independent Symmetry *if for every vectors $x, y \in \{0,1\}^n$ and index $i \in [n]$ it holds that*

$$f((x_{-i}, y_i), (y_{-i}, x_i)) = f(x,y).$$

For example, Independent Symmetry would imply that $f((1,0,1),(1,1,0)) = f((1,1,1),(1,0,0))$.

Our final invariance axiom, *Zero-One Symmetry*, states that the roles of 0 and 1 in the vectors are symmetric, hence if we exchange all 0s for 1s in both vectors and vice-versa, we should get the same dissimilarity. Indeed, in some applications the roles of zeros and ones are arbitrary.

Definition 3.4 (Zero-One Symmetry)**.** *A dissimilarity measure, f, satisfies* Zero-one Symmetry *if for every vectors $x, y \in \{0,1\}^n$ it holds that*

$$f(\bar{x}, \bar{y}) = f(x,y).$$

For example, Zero-One Symmetry would imply that $f((1,0,1),(1,1,0)) = f((0,1,0),(0,0,1))$.

Our next axiom is *Convergence*. It captures the intuition that as we add new information to the input on which both vectors agree, then the dissimilarity between them should move towards zero, and that this convergence to zero should be quick enough. Formally, we say that if we add a new entry to both vectors with 1 in both of them, then the distance should decrease at least proportionally to the size of the vector.

Definition 3.5 (Convergence)**.** *A dissimilarity measure, f, satisfies* Convergence *if for every vectors $x, y \in \{0,1\}^n$ it holds that*

$$f(x \circ (1), y \circ (1)) \leq \frac{n}{n+1} f(x,y).$$

For example, Convergence would imply that $f((1,0,1),(1,1,1)) \leq 2/3 \cdot f((1,0),(1,1))$.

Next, we include a standard distance metric axiom of *Triangle Inequality*.

Definition 3.6 (Triangle Inequality)**.** *A dissimilarity measure, f, satisfies* Triangle Inequality *if for every vectors $x, y, z \in \{0,1\}^n$ it holds that*

$$f(x,y) + f(y,z) \geq f(x,z).$$

Finally, *Normalization* specifies the dissimilarity in a basic case of length-1 vectors.

Definition 3.7 (Normalization)**.** *A dissimilarity measure, f, satisfies* Normalization *if*

$$f((0),(1)) = 1.$$

The above seven axioms uniquely characterize the Hamming distance.

Theorem 3.8. *A dissimilarity measure, f, satisfies Anonymity, Scaling, Independent Symmetry, Zero-One Symmetry, Convergence, Triangle Inequality, and Normalization if and only if f is the Hamming distance.*

Proof sketch. The proofs that Hamming satisfies each of the axioms are relatively straightforward. Thus, in this sketch we focus on showing that if a dissimilarity measure f satisfies all of the axioms, then it is the Hamming distance (the full proof can be found in Section 9). The proof proceeds by considering growing subsets of axioms and characterizing the classes of dissimilarity measures that satisfy these axioms (in Section 9, we present each such characterization as a separate lemma).

We begin by observing that dissimilarity measure satisfying Anonymity must be in fact a function of four arguments, which are the sizes of the intersection, both differences, and the intersection of complements of sets A_x and A_y for each pair of vectors x, y. Then, if we add Independent Symmetry, this function must depend on only three arguments, as instead of looking at $|A_x \setminus A_y|$ and $|A_y \setminus A_x|$, we can simply look at the size of their symmetric difference. Adding Zero-One symmetry prevents distinguishing the intersection form the intersection of complements, so we end up with two arguments: $|A_x \triangle A_y|$ and $n - |A_x \triangle A_y|$, or, equivalently, $|A_x \triangle A_y|$ and n. Then, Scaling implies that $f(x,y)$ is actually a function of a single argument, i.e., there exists g such that $f(x,y) = g(|A_x \triangle A_y|/n) = g(H(x,y))$.

We use the remaining axioms to establish properties of g. Triangle Inequality means that g is subadditive. Next, we prove that with the addition of Convergence, g has to be a linear homogeneous function. In other words, there is $a \in \mathbb{R}$ such that $f(x,y) = a \cdot H(x,y)$. Then, adding Normalization concludes the proof. □

We note that the set of axioms used for our characterization of the Hamming distance is minimal. In other words, the axioms are *independent*, as no subset of our axioms implies the remaining ones.

Theorem 3.9. *For every axiom in the set Anonymity, Scaling, Independent Symmetry, Zero-One Symmetry, Convergence, Triangle Inequality, and Normalization there is a dissimilarity measure that satisfies all other axioms in this set except for this one.*

4 Axiomatic Characterization of the Jaccard Distance

As mentioned in the previous section, *Zero-One Symmetry*, an axiom characterizing Hamming, requires that 0s and 1s have symmetric roles in all vectors. This is the case in some scenarios, however in many applications this is clearly not the case. For example, if the vectors represent the support for a candidate from different voters, then the support seems much more meaningful than the lack of it (as the lack of support may be a result of actual disapproval, but at the same time it can come from neutrality, or lack of knowledge about a candidate). Then, the fact that two candidates are supported by the same voters can be much more important for their perceived similarity than the fact that they agree on more entries.

Consider the following four vectors:

$$x = (1,1,1,0,0,0,0,0,0,1,1,0),$$
$$y = (1,1,1,0,0,0,0,0,0,0,0,1),$$
$$a = (1,0,0,0,0,0,0,0,0,0,0,0),$$
$$b = (0,0,0,0,0,0,0,0,0,0,0,1).$$

Under the Hamming distance, the vectors a and b are more similar than the other two. While, if we treat ones as approvals and zeros as disapprovals, the sets of approved items by a and b are disjoint.

In many types of data there is a significant disproportion between the number of 1s and 0s, which suggests that they play very different roles. In such cases, the following *Add Zero* axiom might be desired, which basically says that coordinates in which both vectors have 0s might be disregarded.

Definition 4.1 (Add Zero)**.** *A dissimilarity measure, f, satisfies* Add Zero *if for every vectors $x, y \in \{0,1\}^n$ it holds that*

$$f(x \circ (0), y \circ (0)) = f(x,y).$$

For example, Add Zero would imply that $f((1,0,1),(1,1,0)) = f((1,0,1,0),(1,1,0,0))$.

It turns out that if in the characterization of the Hamming distance we exchange Zero-One Symmetry for Add Zero axiom, then we obtain a unique characterization of the Jaccard distance.

Theorem 4.2. *A dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Add Zero, Scaling, Convergence, Triangle Inequality, and Normalization if and only if f is the Jaccard distance.*

Proof Sketch. The structure of the proof of the characterization of Jaccard resembles the one from the proof of the characterization of Hamming. Again, verifying that Jaccard satisfies all axioms is relatively straightforward, thus we focus on showing that a dissimilarity measure f satisfying the axioms is equal to the Jaccard distance.

From the proof of Theorem 3.8 we know that f satisfying Anonymity and Independent symmetry for each pair of vectors x, y is a function of three arguments, $|A_x \cap A_y|$, $|A_x \triangle A_y|$, and $n - |A_x \cup A_y|$. Add Zero allows us to disregard the last argument, thus we end up with a function of two arguments $|A_x \cap A_y|$ and $|A_x \triangle A_y|$, or, equivalently, $|A_x \triangle A_y|$ and $|A_x \cup A_y|$. Then, Scaling implies that $f(x,y)$ is actually a function of a single argument, i.e., there exists g such that $f(x,y) = g(|A_x \triangle A_y|/|A_x \cup A_y|) = g(J(x,y))$. Finally, using Triangle Inequality, Convergence, and Normalization, we prove that g has to be an identity. □

Again, our axioms characterizing Jaccard are independent.

Theorem 4.3. *For every axiom in the set Anonymity, Scaling, Independent Symmetry, Add Zero, and Convergence, Triangle Inequality, and Normalization there is a dissimilarity measure that satisfies all other axioms in this set except for this one.*

Finally, in Section 10, we offer an additional axiomatic characterization of the discrete distance.

5 Rejecting Independent Symmetry

In this section, we argue that Independent Symmetry, which is used in characterization of both Hamming and Jaccard, might be actually not desirable in many situations. To see this, consider the following four vectors:

$$x = (0,0,1,1,0,0,0,0), \qquad a = (1,1,1,1,0,0,0,0),$$
$$y = (1,1,1,1,1,1,0,0), \qquad b = (0,0,1,1,1,1,0,0).$$

Every dissimilarity measure f that satisfies Independent Symmetry must give $f(x,y) = f(a,b)$ by the definition of the axiom. However, observe that $A_x \subseteq A_y$, which is not true for a and b. If the vectors correspond to candidates and their coordinates to voters that can support them, this means that every supporter of candidate x also voted for y. This is a very strong signal that projects x and y are related. On the other hand, only 50% of supporters of a also voted for b and vice versa. Therefore, we might want to consider x and y as more similar than a and b. Alternatively, it is possible that in some settings actually a and b can be considered more similar, as they are of similar size. Either way, Independent Symmetry prevents distinguishing these two cases.

What could be a sensible dissimilarity metric that does not satisfy Independent Symmetry? Let us observe that the Hamming distance can be written as

$$H(x,y) = 2 \cdot \frac{|A_x \setminus A_y| + |A_y \setminus A_x|}{2} \cdot \frac{1}{n}.$$

I.e., it is proportional to the arithmetic mean of $|A_x \setminus A_y|$ and $|A_y \setminus A_x|$. Similarly, we can write

$$J(x,y) = 2 \cdot \frac{|A_x \setminus A_y| + |A_y \setminus A_x|}{2} \cdot \frac{1}{|A_x \cup A_y|}.$$

The arithmetic mean is invariant to subtracting a value from one argument and adding it to the other, which is why both distances satisfy Independent Symmetry. However, other means, like quadratic or geometric mean, lack this property. For $p \in \mathbb{R}_{\geq 0}$, we define a *generalized mean* of two numbers as

$$M_p(k,\ell) = \begin{cases} \sqrt[p]{(k^p + \ell^p)/2}, & \text{for } p \in (0,\infty), \\ \sqrt{k \cdot \ell}, & \text{for } p = 0. \end{cases}$$

Then, let us define p-Hamming and p-Jaccard as follows

$$H_p(x,y) = \frac{2M_p(|A_x \setminus A_y|, |A_y \setminus A_x|))}{n},$$

$$J_p(x,y) = \frac{2M_p(|A_x \setminus A_y|, |A_y \setminus A_x|))}{2M_p(|A_x \setminus A_y|, |A_y \setminus A_x|) + |A_x \cap A_y|}.$$

Clearly $H_1(x,y) = H(x,y)$ and $J_1(x,y) = J(x,y)$. Observe that whenever $|A_x \setminus A_y| = |A_y \setminus A_x|$, we have that $H_p(x,y) = H_q(x,y)$, for every $p, q \in \mathbb{R}_{\geq 0}$. However, if $|A_x \setminus A_y| \neq |A_y \setminus A_x|$, then $H_p(x,y) > H_q(x,y)$ for all $p > q$ (and similar properties hold for p-Jaccard). Therefore, assuming constant $|A_x \triangle A_y|$, whenever there is an imbalance in the sizes of the sets A_x and A_y, $H_p(x,y)$ will give higher distances than the Hamming distance, when $p > 1$, or lower distance when $p < 1$. In particular, in the extreme, $H_0(x,y)$ and $J_0(x,y)$, which we call *geometric Hamming* and *geometric Jaccard*, returns dissimilarity 0, whenever $A_x \subseteq A_y$ or $A_y \subseteq A_x$ (but observe that this is incompatible with Triangle Inequality).

Example 5.1. *Assume that we have three projects a, b, and c, where $|A_a| = |A_b| = 20, |A_c| = 6$, and where $|A_a \cap A_b| = 10, |A_a \cap A_c| = 0, |A_b \cap A_c| = 5$. And the question is whether b is more similar to a or c. On the one hand, the size of the intersection between A_a and A_b is twice as large as the intersection between A_b and A_c. On the other hand, almost all voters who approve c also approve b. Under the J_1 distance b is more similar to a, while under the J_0 the b is more similar to c.*

6 Experiments

In this section, we illustrate the intuitive difference between the six following dissimilarity measures: H_0 (geometric Hamming), H_1 (standard Hamming), H_2 (quadratic Hamming), J_0 (geometric Jaccard), J_1 (standard Jaccard), and J_2 (quadratic Jaccard). Since one of the important potential applications of dissimilarity measures is the field of computational social choice, we use statistical models and real-life data from that field. In particular, we focus on approval elections, which consist of a set of candidates and a set of voters, where each voter approves a subset of the candidates [23]. This structure can naturally be represented as a binary matrix, where rows correspond to voters and columns to candidates. Occasionally, we refer to such a matrix as a profile.

We begin by evaluating the measures on synthetic data and later we move to the real-life one. To generate synthetic data we use the Euclidean statistical cultures [6] (i.e., popular statistical models used in social choice). To sample d-Euclidean profile, we proceed as follows: first, sample ideal points for each voter and each candidate from d-dimensional Euclidean metric space; second, for each candidate, sample a radius (i.e., strength) from $\mathbb{R}_+$; third, a voter approves a given candidate if the Euclidean distance between their ideal points is less than or equal to the candidate's radius.

6.1 Ordering Accuracy

In the following experiment, we analyze how different metrics behave on 1D-Euclidean profiles. We consider two variants of the experiment, each variant involving 100 sampled profiles, with 100 candidates and 100 voters sampled uniformly at random from $[0,1]$ interval. In the first variant, the radius of each candidate is the same $r_1 = r_2 = 0.1$, while in the second variant, we sample radii uniformly at random from the $[0.015, 0.15]$ interval.

Let p_a denote the position of a candidate a on the interval. Given three candidates a, b and c such that $p_a < p_b < p_c$, we expect a well-behaved distance measure to satisfy the following two inequalities: $f(a,b) \leq f(a,c)$ and $f(b,c) \leq f(a,c)$. However, this is not necessarily the case. Therefore, for each of the studied distance measures, we compute the proportion of such inequalities that are satisfied. We refer to this proportion as the *ordering accuracy*.

The results are presented in Table 2. The Jaccard distance variants significantly outperform the Hamming distance variants in both cases. Notably, when all the candidates have equal radii, the Jaccard distance always satisfies the ordering inequalities.

Table 2: Average ordering accuracy.

Metric	$r = 0.1$	$r \in [0.015, 0.15]$
H_0	0.690 ± 0.067	0.719 ± 0.063
H_1	0.692 ± 0.066	0.724 ± 0.061
H_2	0.690 ± 0.066	0.719 ± 0.061
J_0	1.000 ± 0.000	0.997 ± 0.001
J_1	1.000 ± 0.000	0.997 ± 0.001
J_2	1.000 ± 0.000	0.997 ± 0.001

Proposition 6.1. *Under Jaccard distance, if r is constant, then the ordering accuracy is equal to* 1.

Table 3: Quality of clustering.

Metric	1D	2D
H_0	$\mathbf{0.0013} \pm 0.0014$	$\mathbf{0.0028} \pm 0.0091$
H_1	0.1178 ± 0.0875	0.0901 ± 0.0372
H_2	0.1888 ± 0.0903	0.0998 ± 0.0301
J_0	$\mathbf{0.0015} \pm 0.0019$	$\mathbf{0.0003} \pm 0.0010$
J_1	0.0028 ± 0.0033	0.0009 ± 0.0025
J_2	0.0042 ± 0.0070	0.0009 ± 0.0020

6.2 Clustering

Given a dissimilarity measure f, and two clusters $C = \{c_1, \ldots, c_k\}$ and $D = \{d_1, \ldots, d_\ell\}$, we define the distance between them as $\text{dist}(C, D) = 1/(|C| \cdot |D|) \sum_{c_i \in C} \sum_{d_i \in D} f(c_i, d_j)$.

Using this definition, we perform clustering on candidates in a Euclidean profile via the hierarchical method. It works as follows. Initially, each candidate forms its own cluster. Then, iteratively, at each step, we merge the two most similar clusters, i.e., such that the average distance between their members is minimal. We continue the procedure until exactly five clusters remain.

We begin with 1D-Euclidean space. We consider profiles with 1000 voters, and 100 candidates, sampled uniformly at random from $[0, 1]$ interval with radii sampled uniformly at random from $[0.015, 0.15]$ interval. The results are shown in Figure 1. Each candidate is depicted as a dot and a vertical line, which is proportional to its radius. The geometric variants (i.e., H_0 and J_0) successfully divide the space into five separate clusters. In contrast, for H_1 and H_2, the green cluster overlaps significantly with other clusters, which is an undesirable outcome in this context. Finally, J_1 and J_2 are outperforming their Hamming counterparts, but still yield less interpretable clusters than the geometric variants.

Next, we consider a 2D-Euclidean space, with a similar setup: 1000 voters, and 100 candidates sampled uniformly at random from $[0, 1]^2$ square, but with candidate radii drawn from $[0.1, 0.33]$ interval. The results are shown in Figure 2. The grey circles denote the voters' positions. The colorful circles denote the candidates' positions, with the size being proportional to the radius. Each color marks a different cluster.

Overall, the results in the 2D case are similar to those in the 1D case. As expected, the Hamming distance performs poorly. It typically produces one dominant cluster that spans much of the space, grouping together a large number of disparate candidates. In contrast, the Jaccard variants (with the exception of the quadratic one) yield significantly better clustering quality. However, the standard Jaccard tends to produce one or two extremely small clusters, typically composed of candidates with very limited voter support.

We can verify the intuitive observations from the plots, by quantifying the number of incorrectly embedded points. For the 1D profiles, let A_3 denote the set of all triplets of candidates such that the first two belong to the same cluster and the third one does not. A triplet is considered invalid if the third candidate lies between the other two. Similarly, for the 2D profiles, let A_4 denote the set of all quadruplets of candidates, such that the first three of them belong to the same cluster and the fourth one does not. A quadruplet is considered invalid if the fourth candidate lies inside the triangle formed by the other three. In Table 3 we present averaged results over 100 profiles. As previously shown in the plots, for 1D profiles, geometric variants outperform the other methods. In the case of 2D profiles, the best performance is achieved by the geometric Jaccard, which results in three times fewer invalid quadruplets than the standard Jaccard.

6.3 Real-Life Data

Finally, we examine real-life data, which comes from Participatory Budgeting (PB) elections. PB elections are democratic processes in which some municipalities distribute a portion of their budget [27]. They usually involve citizens expressing approval preferences over potential projects, a selection of which is later implemented in practice.

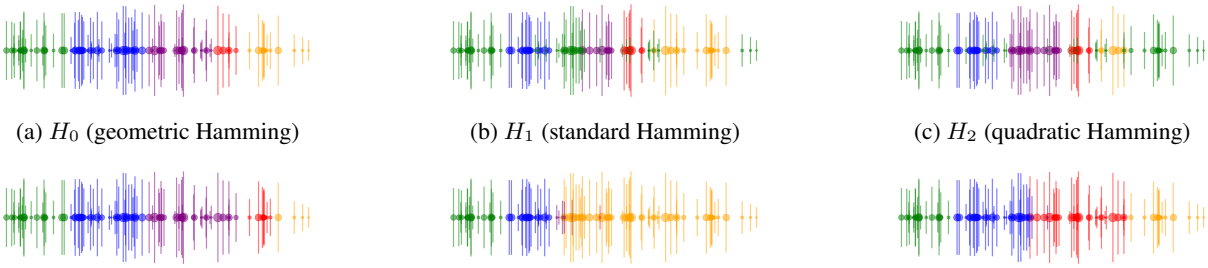
We use participatory budgeting instances from Pabulib [13], focusing on district elections held in 2023 and 2024 in three major Polish cities: Warsaw, Kraków, and Łódź. (Detailed parameters of these instances are available in Table 5 in Section 11). In these cities, projects are labeled with categories. In Kraków and Łódź, each project belongs to exactly one category, while in Warsaw, projects can have multiple categories. Although these categories are assigned manually and may not always be accurate, we nonetheless expect that projects within the same category tend to be more similar to each other on average.

For each category, we compute two values: the intra distance d_{intra}, defined as the average pairwise distance between all projects within the same category, and the inter distance d_{inter}, defined as the average pairwise distance between projects in the category and those outside of it. For each instance, we compute the proportion of categories for which it holds that $d_{\text{inter}}/d_{\text{intra}} > 1.01$ (i.e., the inter distances are nontrivially larger than the intra ones). We then average these proportions across all districts (i.e., instances) for a given year and city. The results are presented in Table 4. Interestingly, geometric Jaccard consistently outperforms all other distance measures.

Besides categories, several of the Pabulib instances include geographic coordinates of some of the projects. We focus on district elections held between 2019 and 2024 in Warsaw. We conjectured that there might be a correlation between the Euclidean distance between two projects (based on location) and their distance under our measures.

Averaging across all 108 district elections held in Warsaw during that period (18 per year), the average Pearson correlation coefficient between the standard Jaccard distance and the Euclidean one is 0.219, and for geometric Jaccard it is slightly higher at 0.233. Interestingly, in 90% of the instances, the correlation was stronger for geometric Jaccard (detailed results, including comparison with other distance measures, are provided in Table 6 in Section 11).

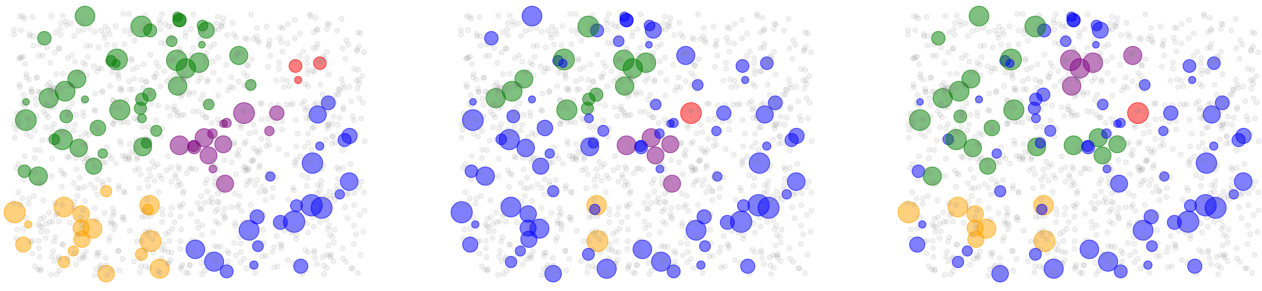
Beyond Warsaw, we also analyze a unique election, the green participatory budgeting held in 2023 in Wieliczka (a small Polish city). This case is particularly interesting as it was the first city to use in practice the recently developed method of equal shares [26]. Here, the average Pearson correlation between Jaccard and Euclidean dis-



(a) H_0 (geometric Hamming) (b) H_1 (standard Hamming) (c) H_2 (quadratic Hamming)

(d) J_0 (geometric Jaccard) (e) J_1 (standard Jaccard) (f) J_2 (quadratic Jaccard)

Figure 1: Clustering of 1D-Euclidean profiles.



(a) H_0 (geometric Hamming) (b) H_1 (standard Hamming) (c) H_2 (quadratic Hamming)

(d) J_0 (geometric Jaccard) (e) J_1 (standard Jaccard) (f) J_2 (quadratic Jaccard)

Figure 2: Clustering of 2D-Euclidean profiles.

tances is 0.369, while for geometric Jaccard it is 0.441. Wieliczka municipality consists of the central city and several small villages. It appears that voters in the villages tend to vote for their local projects, which may explain the higher correlation between geographic proximity and our distance measures in this case.

7 Related Work

As mentioned in the introduction, axiomatic approach has been already successfully employed to enhance understanding of various machine learning concepts. For example, a large body of work has concentrated on axioms for clustering. In particular, Kleinberg [20] proved that there is no clustering function that satisfies certain three desired properties. However, this impossibility could be avoided if we focus on functions that produce a specific, exogenously given number of clusters [1] or slightly modify the axioms [10, 21]. Furthermore, several works propose axiomatic characterizations of particular clustering rules [8, 32] or families of rules [2, 3].

Several distances between objects have also been studied axiomatically. A good example is an axiomatization of Kendall Tau distance by Kemeny [19]. The closest to our work, is an axiomatic characterization of Jaccard distance by Gerasimou [14]. However, this characterization considers only three axioms. Two of them are standard distance metric axioms, namely Identity and Triangle Inequality, and one axiom which specifies the marginal contribution of an additional bit of information on which two vectors do not agree. This last axiom is very specific to Jaccard, and there are no other widely known dissimilarity measures that satisfy it. As a result, this characterization does not really offer a view of the Jaccard distance as a consequence of several, not directly related assumptions. In contrast, we propose a larger family of simple axioms, each one of which is satisfied by many natural and popular dissimilarity measures. The Hamming distance was not considered in this paper.

Similar work was also done by Gösgens et al. [16]. However, the main premise of their paper was to compare a large number of indices in order to indicate the most suitable ones for quite specific applications of evaluating classification performance. In contrast, in our work we strive to provide a better understanding of two very specific dissimilarity measures, but widely popular across many different applications, i.e., the Hamming and Jaccard distances.

Table 4: Averaged share of categories with higher inter- than intra-category distance.

	Warsaw		Kraków		Łódź	
Metric	2023	2024	2023	2024	2023	2024
H_0 (geometric Hamming)	63.8%	47.2%	45.5%	40.9%	66.4%	69.6%
H_1 (standard Hamming)	68.6%	56.9%	49.7%	37.4%	65.7%	62.8%
H_2 (quadratic Hamming)	68.5%	62.5%	52.1%	41.9%	63.8%	63.8%
J_0 (geometric Jaccard)	82.6%	88.2%	79.4%	85.9%	82.8%	91.6%
J_1 (standard Jaccard)	67.3%	65.3%	56.9%	58.1%	75.4%	73.3%
J_2 (quadratic Jaccard)	70.2%	79.2%	66.3%	67.8%	78.4%	76.3%

8 Conclusions

We have developed an axiomatic characterization of the Hamming and Jaccard distances, which enables us to clearly capture the key properties of each measure. In particular, the two characterizations differ by a single axiom: Zero-One Symmetry for Hamming vs Add Zero for Jaccard, which highlights the most important difference between them.

Moreover, our axiomatic analysis allowed us to identify a potential shared shortcoming of the Hamming and Jaccard distances, as both satisfy Independent Symmetry, which may be undesirable in certain applications. To address this limitation, we introduced a family of novel measures based on generalized means. Empirical evaluation on both synthetic and real-life data showed that one of the new measures, i.e., geometric Jaccard, consistently outperforms other measures on various tasks, and is better at capturing some nuances. Future work may analyze the new measures in more detail.

Acknowledgments

We are thankful to Jérôme Lang and Florian Yger for many insightful discussions regarding this paper. We also thank anonymous reviewers for constructive feedback. Tomasz Wąs was supported by UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/X038548/1.

References

[1] M. Ackerman and S. Ben-David. Measures of clustering quality: A working set of axioms for clustering. *Advances in Neural Information Processing Systems*, 21, 2008.

[2] M. Ackerman and S. Ben-David. Discerning linkage-based algorithms among hierarchical clustering methods. In *IJCAI*, pages 1140–1145, 2011.

[3] M. Ackerman, S. Ben-David, and D. Loker. Characterization of linkage-based clustering. In *COLT*, volume 2010, pages 270–281, 2010.

[4] M. Ackerman, S. Ben-David, S. Brânzei, and D. Loker. Weighted clustering: Towards solving the user's dilemma. *Pattern Recognition*, 120: 108152, 2021.

[5] S. Bag, S. K. Kumar, and M. K. Tiwari. An efficient recommendation generation using relevant jaccard similarity. *Information Sciences*, 483: 53–64, 2019.

[6] N. Boehmer, P. Faliszewski, L. Janeczko, A. Kaczmarczyk, G. Lisowski, G. Pierczynski, S. Rey, D. Stolicki, S. Szufa, and T. Was. Guide to numerical experiments on elections in computational social choice. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 7962–7970, 2024.

[7] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[8] G. E. Carlsson, F. Mémoli, et al. Characterization, stability and convergence of hierarchical clustering methods. *J. Mach. Learn. Res.*, 11 (Apr):1425–1470, 2010.

[9] S.-H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical models and Methods in Applied Sciences*, 1(4):300–307, 2007.

[10] V. Cohen-Addad, V. Kanade, and F. Mallmann-Trenn. Clustering redemption–beyond the impossibility of kleinberg's axioms. *Advances in Neural Information Processing Systems*, 31, 2018.

[11] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226, 2012.

[12] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2147–2154, 2014.

[13] P. Faliszewski, J. Flis, D. Peters, G. Pierczyński, P. Skowron, D. Stolicki, S. Szufa, and N. Talmon. Participatory budgeting: Data, tools and analysis. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2667–2674, 8 2023.

[14] G. Gerasimou. Characterization of the jaccard dissimilarity metric and a generalization. *Discrete Applied Mathematics*, 355:57–61, 2024.

[15] G. Gilbert. Distance between sets. *Nature*, 239(5368):174–174, 1972.

[16] M. M. Gösgens, A. Tikhonov, and L. Prokhorenkova. Systematic analysis of cluster similarity indices: How to validate validation measures. In *International Conference on Machine Learning*, pages 3799–3808. PMLR, 2021.

[17] Q. Huang, P. Luo, and A. K. H. Tung. A new sparse data clustering method based on frequent items. *Proceedings of the ACM on Management of Data*, 1(1):1–28, 2023.

[18] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.

[19] J. G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

[20] J. Kleinberg. An impossibility theorem for clustering. *Advances in neural information processing systems*, 15, 2002.

[21] M. A. Kłopotek and R. A. Kłopotek. In-the-limit clustering axioms. In *International Conference on Artificial Intelligence and Soft Computing*, pages 199–209. Springer, 2020.

[22] S. Kosub. A note on the triangle inequality for the jaccard distance. *Pattern Recognition Letters*, 120:36–38, 2019.

[23] M. Lackner and P. Skowron. *Multi-winner voting with approval preferences*. Springer Nature, 2023.

[24] H. Moulin. Implementing efficient, anonymous and neutral social choice functions. *Journal of Mathematical Economics*, 7(3):249–269, 1980.

[25] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu. Using of jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, pages 380–384, 2013.

[26] D. Peters, G. Pierczyński, and P. Skowron. Proportional participatory budgeting with additive utilities. *Advances in Neural Information Processing Systems*, 34:12726–12737, 2021.

[27] S. Rey, F. Schmidt, and J. Maly. The (computational) social choice take on indivisible participatory budgeting. *arXiv preprint arXiv:2303.00621*, 2023.

[28] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019.

[29] L. S. Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28):307, 1953.

[30] A. N. Tarekegn, M. Giacobini, and K. Michalak. A review of methods for imbalanced multi-label classification. *Pattern Recognition*, 118: 107965, 2021.

[31] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*, pages 64–74, 2008.

[32] R. B. Zadeh and S. Ben-David. A uniqueness theorem for clustering. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial*

Intelligence, pages 639–646, 2009.
[33] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan. Focal and efficient iou loss for accurate bounding box regression. *Neurocomputing*, 506:146–157, 2022.

9 Missing Proofs

9.1 Proof of Theorem 3.8

Proof. Let us first show that the Hamming distance indeed satisfies the respective axioms, by considering them one by one.

Anonymity
Clearly, for every $x, y \in \{0,1\}^n$ and permutation $\pi : [n] \to [n]$ it holds that $|A_x \triangle A_y| = |A_{\pi(x)} \triangle A_{\pi(y)}|$. Thus,

$$H(\pi(x), \pi(y)) = H(x, y).$$

Scaling
Observe that for every $x, y \in \{0,1\}^n$ and $k \in \mathbb{N}$ it holds that

$$H(x^k, y^k) = \frac{k \cdot |A_x \triangle A_y|}{k \cdot n} = \frac{|A_x \triangle A_y|}{n} = H(x, y).$$

Independent Symmetry
For every $x, y \in \{0,1\}^n$ and $i \in [n]$ it holds that $|A_x \triangle A_y| = |A_{(x_{-i}, y_i)} \triangle A_{(y_{-i}, x_i)}|$. Thus,

$$H((x_{-i}, y_i), (y_{-i}, x_i)) = H(x, y).$$

Zero-One Symmetry
Note that for every $x, y \in \{0,1\}^n$ it holds that $|A_x \triangle A_y| = |A_x \setminus A_y| + |A_y \setminus A_x| = |A_{\bar{y}} \setminus A_{\bar{x}}| + |A_{\bar{x}} \setminus A_{\bar{y}}| = |A_{\bar{x}} \triangle A_{\bar{y}}|$. Hence,

$$H(\bar{x}, \bar{y}) = H(x, y).$$

Convergence
Fix arbitrary $x, y \in \{0,1\}^n$. Observe that after concatenating an additional bit to x and y with 1s on both sides, the symmetric difference will not change, i.e., $|A_x \triangle A_y| = |A_{x \circ (1)} \triangle A_{y \circ (1)}|$. Thus, we have

$$H(x \circ (1), y \circ (1)) = \frac{|A_x \triangle A_y|}{n+1} = \frac{n}{n+1} H(x, y),$$

which implies the weak inequality.
Triangle Inequality[1]
Take arbitrary $x, y, z \in \{0,1\}^n$. Also, consider arbitrary index $i \in [n]$ such that $i \in A_x \setminus A_z$. If $i \in A_y$ as well, then it implies that $i \in A_y \setminus A_z$. Conversely, if $i \notin A_y$, then $i \in A_x \setminus A_y$. Thus, $A_x \setminus A_z \subseteq (A_x \setminus A_y) \cup (A_y \setminus A_z)$, which means that $|A_x \setminus A_z| \leq |A_x \setminus A_y| + |A_y \setminus A_z|$ since $A_x \setminus A_y$ and $A_y \setminus A_z$ are disjoint (as one is a subset of A_y and the other of $A_{\bar{y}}$). Analogously, we get that $|A_z \setminus A_x| \leq |A_z \setminus A_y| + |A_y \setminus A_x|$. Adding the inequalities sidewise together and dividing by n we obtain

$$\begin{aligned} H(x, z) = \frac{|A_x \triangle A_z|}{n} &\leq \frac{|A_x \triangle A_y| + |A_y \triangle A_z|}{n} \\ &= H(x, y) + H(y, z). \end{aligned}$$

Normalization
Clearly, $H((0), (1)) = 1$.

In the remainder of the proof, let us focus on the converse statement, i.e., that any dissimilarity measure that satisfies these axioms must necessarily be the Hamming distance. We will prove that in a series of lemmas that will consider arbitrary dissimilarity measure f satisfying increasing subset of our axioms. In each lemma, we

[1] It is a widely known fact that the Hamming distance is a proper distance metric, thus it satisfies Triangle Inequality. We provide a proof for completeness.

will prove what form has to have such a function f. In the last one, Lemma 9.10, we show that if f satisfies all of our axioms except for Normalization, then f has to be equal to the Hamming distance multiplied by a nonnegative scalar, i.e., there exists $a \in \mathbb{R}_{\geq 0}$ such that $f(x,y) = a \cdot H(x,y)$, for every $x, y \in \{0,1\}$. Then, Normalization implies the thesis.

Lemma 9.1. *If a dissimilarity measure, f, satisfies Anonymity, then there exists a function, $g : \mathbb{N}^4 \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = g(|A_x \cap A_y|, |A_x \setminus A_y|, |A_y \setminus A_x|, n - |A_x \cup A_y|).$$

Proof. Assume otherwise, i.e., there exist four vectors $x, x', y, y' \in \{0,1\}^n$ such that $|A_x \cap A_y| = |A_{x'} \cap A_{y'}|$, $|A_x \setminus A_y| = |A_{x'} \setminus A_{y'}|$, $|A_y \setminus A_x| = |A_{y'} \setminus A_{x'}|$, and $n - |A_x \cup A_y| = n - |A_{x'} \cup A_{y'}|$ and $f(x,y) \neq f(x',y')$.

Observe that because of the equal sizes of the corresponding sets, we can find following bijections:

$$\begin{aligned}
&\pi_1 : A_x \cap A_y \to A_{x'} \cap A_{y'},\\
&\pi_2 : A_x \setminus A_y \to A_{x'} \setminus A_{y'},\\
&\pi_3 : A_y \setminus A_x \to A_{y'} \setminus A_{x'}, \text{ and}\\
&\pi_4 : [n] \setminus (A_x \cup A_y) \to [n] \setminus A_{x'} \cup A_{y'}.
\end{aligned}$$

Since their domains are disjoint and sum up to the whole set $[n]$ and the same is true for their codomains, we can take a disjoint union of these bijections $\pi = \pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 : [n] \to [n]$. Then observe that $\pi(x) = x'$ and $\pi(y) = y'$. Thus, by Anonymity $f(x,y) = f(\pi(x), \pi(y)) = f(x',y')$, which is a contradiction. □

Lemma 9.2. *If a dissimilarity measure, f, satisfies Anonymity, and Independent Symmetry, then there exists a function, $g : \mathbb{N}^3 \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = g(|A_x \cap A_y|, |A_x \triangle A_y|, n - |A_x \cup A_y|).$$

Proof. Since f satisfies Anonymity, from Lemma 9.1 we know that there exists a function, $g : \mathbb{N}^4 \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that

$$f(x,y) = g(|A_x \cap A_y|, |A_x \setminus A_y|, |A_y \setminus A_x|, n - |A_x \cup A_y|).$$

It remains to show that for arbitrary $x, y \in \{0,1\}^n$, it holds that

$$\begin{aligned}
g(|A_x \cap A_y|, |A_x \setminus A_y|, |A_y \setminus A_x|, n - |A_x \cup A_y|) = \\
g(|A_x \cap A_y|, |A_x \triangle A_y|, 0, n - |A_x \cup A_y|). \quad (1)
\end{aligned}$$

To this end, take $x', y' \in \{0,1\}^n$ such that $x'_i = 1$ and $y'_i = 0$, for every $i \in A_y \triangle A_x$, and $x'_i = x_i$ and $y'_i = y_i$, otherwise. Then, by Independent Symmetry used $|A_y \setminus A_x|$ times, we obtain that $f(x,y) = f(x',y')$. Since $|A_{x'} \setminus A_{y'}| = |A_x \triangle A_y|$ and $|A_{y'} \setminus A_{x'}| = 0$ this proves Equation (1). □

Lemma 9.3. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, and Scaling, then there exists a function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ with $|A_x \cup A_y| = n$ it holds that*

$$f(x,y) = g(|A_x \triangle A_y|/n).$$

Proof. Assume otherwise, i.e., there exist four vectors $x, y \in \{0,1\}^n$ and $a, b \in \{0,1\}^m$ such that $|A_x \cup A_y| = n$, $|A_a \cup A_b| = m$ and $|A_x \triangle A_y|/n = |A_a \triangle A_b|/m$, but $f(x,y) \neq f(a,b)$. Let k be the least common multiple of n and m and let $p = k/n$ and $q = k/m$. By Scaling, we know that $f(x,y) = f(x^p, y^p)$ and $f(a,b) = f(a^q, b^q)$. Thus, to get the contradiction, it remains to show that $f(x^p, y^p) = f(a^q, b^q)$.

Since f satisfies Anonymity and Independent Symmetry, from Lemma 9.1 we know that there exists a function, $g' : \mathbb{N}^3 \to \mathbb{R}_{\geq 0}$, such that

$$f(x^p, y^p) = g'(|A_{x^p} \cap A_{y^p}|, |A_{x^p} \triangle A_{y^p}|, n - |A_{x^p} \cup A_{y^p}|).$$

Observe that the assumption that $|A_x \cup A_y| = n$ implies also that $|A_{x^p} \cup A_{y^p}| = k$ (if there is no index $i \in [n]$ such that $x_i = y_i = 0$, then there is no such index for x^p and y^p as well). Thus, we get

$$f(x^p, y^p) = g'(|A_{x^p} \cap A_{y^p}|, |A_{x^p} \triangle A_{y^p}|, 0).$$

Analogously, we get that

$$f(a^q, b^q) = g'(|A_{a^q} \cap A_{b^q}|, |A_{a^q} \triangle A_{b^q}|, 0).$$

Now, observe that

$$\begin{aligned}
\frac{|A_{x^p} \triangle A_{y^p}|}{k} &= \frac{p|A_x \triangle A_y|}{p \cdot n} = \frac{|A_x \triangle A_y|}{n}\\
&= \frac{|A_a \triangle A_b|}{m} = \frac{q|A_a \triangle A_b|}{q \cdot m} = \frac{|A_{a^q} \triangle A_{b^q}|}{k}.
\end{aligned}$$

Let us denote this ratio by r. Then, we get

$$f(x^p, y^p) = g'(k - r \cdot k, r \cdot k, 0) = f(a^q, b^q),$$

which concludes the proof. □

Lemma 9.4. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Scaling, and Triangle Inequality, then there exists a subadditive function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ with $|A_x \cup A_y| = n$ it holds that*

$$f(x,y) = g(|A_x \triangle A_y|/n).$$

Proof. From Lemma 9.3 we know that there exists a function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ with $|A_x \cup A_y| = n$ it holds that $f(x,y) = g(|A_x \triangle A_y|/n)$. It remains to show that g is subadditive, i.e., for arbitrary $p, q \in (\mathbb{Q} \cap [0,1])$ such that $p + q \leq 1$, it holds that $g(p) + g(q) \geq g(p+q)$. To this end, let $r = 1 - p - q$ and let k be such that $pk, qk, rk \in \mathbb{N}$. Then, consider the following three vectors

$$\begin{aligned}
x &= (0)^{pk} \circ (1)^{rk} \circ (1)^{qk},\\
y &= (1)^{pk} \circ (1)^{rk} \circ (1)^{qk}, \text{ and}\\
z &= (1)^{pk} \circ (1)^{rk} \circ (0)^{qk}.
\end{aligned}$$

Then, from Triangle Inequality we get $f(x,y) + f(y,z) \geq f(x,z)$, which in terms of g means $g(\frac{kp}{k(p+q+r)}) + g(\frac{kq}{k(p+q+r)}) \geq g(\frac{k(p+q)}{k(p+q+r)})$. Since $p+q+r = 1$, we get that indeed $g(p) + g(q) \geq g(p+q)$, which concludes the proof. □

Lemma 9.5. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Scaling, Convergence, and Triangle Inequality then there exists constant $a \in \mathbb{R}_{\geq 0}$ such that for every $x, y \in \{0,1\}^n$ with $|A_x \cup A_y| = n$ it holds that*

$$f(x,y) = a \cdot H(x,y).$$

Proof. From Lemma 9.9 we know that there exists a subadditive function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that $f(x,y) = g(|A_x \triangle A_y|/n) = g(H(x,y))$. Let $a = g(1)$. We will show that for every $r \in \mathbb{Q} \cap [0,1]$ it holds that $g(r) = ar$, which will imply the thesis.

We begin by showing that for every $r \in \mathbb{Q} \cap [0,1]$ it holds that

$$g(r/2) = g(r)/2. \tag{2}$$

Let $p, q \in \mathbb{N}$ be such that $r = p/q$. Then, by Convergence used n times we get that

$$\begin{aligned} g(p/q) &= f((1)^q, (0)^p \circ (1)^{q-p}) \\ &\geq 2 \cdot f((1)^q \circ (1)^q, (0)^p \circ (1)^{q-p} \circ (1)^q) \\ &= 2 \cdot g(p/(2q)). \end{aligned}$$

Hence, $g(r) \geq 2 \cdot g(r/2)$. On the other hand, $g(r) \leq 2 \cdot g(r/2)$ from subadditivity. Thus, indeed Equation (2) holds.

Next, we generalize Equation (2) and prove that for every $r \in \mathbb{Q} \cap [0,1]$ and $q \in \mathbb{N}$ it holds that

$$g(r/q) = g(r)/q. \tag{3}$$

Observe that it is enough to prove this equality for prime q, as for composite qs we can obtain the thesis by combining the results for all prime factors of q. Thus, without loss of generality, let us assume that q is prime. Clearly, from subadditivity, we have that $g(r/q) \geq g(r)/q$. For a contradiction assume that $g(r/q) = g(r)/q + \varepsilon$ for some $\varepsilon > 0$. By Little Fermat's Theorem, we know that there exists $p \in \mathbb{N}$ such that $p \cdot q = 2^{q-1} - 1$. Then, we have that $1 + p \cdot q = 2^{(q-1)}$, which means that $1 = (1 + p \cdot q)/2^{(q-1)}$. In turn, this implies that

$$\frac{1}{q} = \frac{1 + p \cdot q}{2^{q-1} q} = \frac{1}{2^{q-1} q} + \frac{p}{2^{q-1}}. \tag{4}$$

Thus, we get

$$\begin{aligned} g\left(\frac{r}{q}\right) &= g\left(\frac{r}{2^{k(q-1)} q} + \frac{r \cdot p}{2^{k(q-1)}}\right) && \text{(from Equation (4))} \\ &\leq g\left(\frac{r}{2^{q-1} q}\right) + p \cdot g\left(\frac{r}{2^{q-1}}\right) && \text{(from subadditivity)} \\ &= \frac{1}{2^{q-1}} g\left(\frac{r}{q}\right) + \frac{p}{2^{q-1}} g(r). \\ & && \text{(from Equation (2) used } q-1 \text{ times)} \end{aligned}$$

Substituting $g(r/q) = g(r)/q + \varepsilon$, we get

$$\begin{aligned} g(r)/q + \varepsilon &\leq \frac{g(r)/q + \varepsilon}{2^{q-1}} + \frac{p}{2^{q-1}} g(r) \\ &= g(r) \frac{1 + p \cdot q}{2^{q-1} q} + \varepsilon/2^{q-1} \\ &= g(r)/q + \varepsilon/2^{q-1}. && \text{(from Equation (4))} \end{aligned}$$

Thus, $\varepsilon \leq \varepsilon/2^{q-1}$, which is a contradiction for $\varepsilon > 0$. Therefore, Equation (3) indeed holds, from which we immediately obtain that

$$\begin{aligned} g(0) &= 0, \text{ and} \\ g(1/q) &= a/q, \text{ for every } q \in \mathbb{N}. \end{aligned} \tag{5}$$

Finally, let us take $r = p/q$ for arbitrary $p, q \in \mathbb{N}$ such that $1 \leq p \leq q$. From subadditivity and Equation (5) we get that $g(p/q) \leq p \cdot g(1/q) = ap/q$. On the other hand, again from subadditivity and Equation (5), we get that $g(p/q) \geq g(1) - (q-p) g(1/q) = a - a(q-p)/q = a \cdot p/q$. Thus, indeed $g(r) = ar$, for every $r \in \mathbb{Q} \cap [0,1]$, which concludes the proof. □

Lemma 9.6. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Zero-One Symmetry, Scaling, Convergence, and Triangle Inequality then there exists constant $a \in \mathbb{R}_{\geq 0}$ such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = a \cdot H(x,y).$$

Proof. We will prove the thesis by induction on $n - |A_x \cup A_y|$. Let a be a constant from Lemma 9.5. For the basis of induction, we know that for every $x, y \in \{0,1\}^n$ such that $n - |A_x \cup A_y| = 0$, i.e., $|A_x \cup A_y| = n$ by Lemma 9.5 it holds that

$$f(x,y) = a \cdot H(x,y).$$

Now, assume that the thesis holds for all $n \in \mathbb{N}$ and $x, y \in \{0,1\}^n$ such that $n - |A_x \cup A_y| = N$ and fix arbitrary $x, y \in \{0,1\}^n$ with $n - |A_x \cup A_y| = N + 1$. By Anonymity, without loss of generality we can assume that $x_n = y_n = 0$. Let x' and y' denote vectors x and y without the last coordinate, i.e, $x' \circ (0) = x$ and $y' \circ (0) = y$. Then, from Zero-One Symmetry and Convergence we have

$$\begin{aligned} f(x,y) = f(\bar{x}, \bar{y}) &= f(\bar{x}' \circ (1), \bar{y}' \circ (1)) \\ &\leq \frac{n-1}{n} f(\bar{x}', \bar{y}') = \frac{n-1}{n} f(x', y'). \end{aligned}$$

On the other hand, $(n-1) - |A_{x'} \cup A_{y'}| = (n-1) - |A_x \cup A_y| = N$, thus from inductive assumption we get

$$f(x,y) \leq \frac{n-1}{n} f(x', y') = a \frac{n-1}{n} H(x', y') = a H(x,y).$$

It remains to show that $f(x,y)$ cannot be strictly smaller than $aH(x,y)$. To this end, we first prove that $f(y, y' \circ (1)) \leq a/n$. Since y and $y' \circ (1)$ agree on $n-1$ positions this can be shown from reordering their coordinates and using Convergence or Convergence with Zero-One Symmetry $n-1$ times. Then, from Triangle Inequality we obtain

$$f(x,y) + f(y, y' \circ (1)) \geq f(x, y' \circ (1)),$$

which is equivalent to

$$f(x,y) \geq f(x, y' \circ (1)) - f(y, y' \circ (1))$$

Using inductive assumption on $f(x, y' \circ (1))$ (as $n - |A_x \cup A_{y' \circ (1)}| = N$) and substituting $f(y, y' \circ (1)) \leq 1/n$ we get

$$\begin{aligned} f(x,y) &\geq aH(x, y' \circ (1)) - a/n \\ &= a(H(x,y) + 1/n) - a/n = aH(x,y), \end{aligned}$$

which concludes the proof. □

Combining Lemma 9.6 with Normalization we obtain the thesis. □

9.2 Proof of Theorem 3.9

Proof. For each of the axioms we provide a dissimilarity measure f that satisfies all but that axiom.

Anonymity: $f(x,y) = \sum_{i=1}^{n} |x_i - y_i| \cdot 2^i / (2^{n+1} - 2)$
Scaling: $f(x,y) = (|A_x \triangle A_y| + 1)/(2n)$
Independent Symmetry: $f(x,y) = \max(|A_x \setminus A_y|, |A_y \setminus A_x|)/n$
Zero-One Symmetry: $f(x,y) = J(x,y)$
Convergence: $f(x,y) = \max_{i \in [n]} |x_i - y_i|$ (discrete distance)
Triangle Inequality: $f(x,y) = H(x,y)^2$
Normalization: $f(x,y) = 2 \cdot H(x,y)$

□

9.3 Proof of Theorem 4.2

Proof. Let us first show that the Jaccard distance indeed satisfies the respective axioms, by considering them one by one.

Anonymity

Clearly, for every $x, y \in \{0,1\}^n$ and permutation $\pi : [n] \to [n]$ it holds that $|A_x \triangle A_y| = |A_{\pi(x)} \triangle A_{\pi(y)}|$ as well as $|A_x \cup A_y| = |A_{\pi(x)} \cup A_{\pi(y)}|$. Thus,

$$J(\pi(x), \pi(y)) = J(x, y).$$

Scaling

Observe that for every $x, y \in \{0,1\}^n$ and $k \in \mathbb{N}$ it holds that

$$J(x^k, y^k) = \frac{k \cdot |A_x \triangle A_y|}{k \cdot |A_x \cup A_y|} = \frac{|A_x \triangle A_y|}{|A_x \cup A_y|} = J(x, y).$$

Independent Symmetry

For every $x, y \in \{0,1\}^n$ and $i \in [n]$ it holds that $|A_x \triangle A_y| = |A_{(x_{-i}, y_i)} \triangle A_{(y_{-i}, x_i)}|$ as well as $|A_x \cup A_y| = |A_{(x_{-i}, y_i)} \cup A_{(y_{-i}, x_i)}|$. Thus,

$$J((x_{-i}, y_i), (y_{-i}, x_i)) = J(x, y).$$

Add Zero

Note that for every $x, y \in \{0,1\}^n$ it holds that $|A_x \triangle A_y| = |A_{x \circ (0)} \triangle A_{y \circ (0)}|$ and $|A_x \cup A_y| = |A_{x \circ (0)} \cup A_{y \circ (0)}|$ Hence,

$$J(x \circ (0), y \circ (0)) = J(x, y).$$

Convergence

Fix arbitrary $x, y \in \{0,1\}^n$. Observe that after concatenating an additional bit to x and y with 1s on both sides, the symmetric difference will not change, i.e., $|A_{x \circ (1)} \triangle A_{y \circ (1)}| = |A_x \triangle A_y|$. On the other hand, the union increases by 1, i.e., $|A_{x \circ (1)} \cup A_{y \circ (1)}| = |A_x \cup A_y| + 1$. Thus, we have

$$\begin{aligned} J(x \circ (1), y \circ (1)) &= \frac{|A_x \triangle A_y|}{|A_x \cup A_y| + 1} \\ &\leq \frac{n}{n+1} \cdot \frac{|A_x \triangle A_y|}{|A_x \cup A_y|} = \frac{n}{n+1} \cdot J(x, y), \end{aligned}$$

where the inequality comes from the fact that $|A_x \cup A_y| \leq n$, so $|A_x \cup A_y| / (|A_x \cup A_y| + 1) \leq n/(n+1)$.

Triangle Inequality[2]

Take arbitrary $x, y, z \in \{0,1\}^n$. Since $J(x, z) = |A_x \triangle A_z| / |A_x \cup A_z| \leq 1$, we have

$$\begin{aligned} J(x, z) &= \frac{|A_x \triangle A_z| + |A_y \setminus (A_x \cup A_z)|}{|A_x \cup A_z| + |A_y \setminus (A_x \cup A_z)|} \\ &= \frac{|A_x \triangle A_z| + |A_y \setminus (A_x \cup A_z)|}{|A_x \cup A_y \cup A_z|}. \end{aligned}$$

On the other hand,

$$J(x, y) = \frac{|A_x \triangle A_y|}{|A_x \cup A_y|} \geq \frac{|A_x \triangle A_y|}{|A_x \cup A_y \cup A_z|}$$

and similarly $J(y, z) \geq |A_y \triangle A_z| / |A_x \cup A_y \cup A_z|$. Now, observe that $(A_x \triangle A_y) \cup (A_y \triangle A_z)$ is actually equal to $(A_x \cup A_y \cup A_z) \setminus (A_x \cap A_y \cap A_z)$. Thus, it holds that $(A_x \triangle A_z) \subseteq (A_x \triangle A_y) \cup (A_y \triangle A_z)$ as well as $(A_y \setminus (A_x \cup A_z)) \subseteq (A_x \triangle A_y) \cup (A_y \triangle A_z)$. Since $(A_x \triangle A_z)$ and $(A_y \setminus (A_x \cup A_z))$ are disjoint, this gives us

$$\begin{aligned} |A_x \triangle A_z| + |A_y \setminus (A_x \cup A_z)| &\leq |(A_x \triangle A_y) \cup (A_y \triangle A_z)| \\ &\leq |A_x \triangle A_y| + |A_y \triangle A_z|. \end{aligned}$$

Combining all inequalities we get

$$\begin{aligned} J(x, y) + J(y, z) &\geq \frac{|A_x \triangle A_y| + |A_y \triangle A_z|}{|A_x \cup A_y \cup A_z|} \\ &\geq \frac{|A_x \triangle A_z| + |A_y \setminus (A_x \cup A_z)|}{|A_x \cup A_y \cup A_z|} \geq J(x, z). \end{aligned}$$

Normalization

Clearly, $J((0), (1)) = 1$.

Now, let us focus on the converse statement, i.e., that any dissimilarity measure that satisfies these axioms must necessarily be the Jaccard distance. We will prove that in a series of lemmas that will consider arbitrary dissimilarity measure f satisfying increasing subset of our axioms. In each lemma, we will prove what form has to have such a function f. In the last one, Lemma 9.10, we show that if f satisfies all of our axioms except for Normalization, then f has to be equal to the Jaccard distance multiplied by a nonnegative scalar, i.e., there exists $a \in \mathbb{R}_{\geq 0}$ such that $f(x, y) = a \cdot J(x, y)$, for every $x, y \in \{0,1\}$. Then, Normalization implies the thesis.

From the proof of the characterization of the Hamming distance we already know how do dissimilarity measures that satisfy Anonymity and Independent Symmetry look like. Let us add Add Zero to that.

Lemma 9.7. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, and Add Zero, then there exists a function, $g : \mathbb{N}^2 \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x, y) = g(|A_x \triangle A_y|, |A_x \cup A_y|).$$

Proof. Since f satisfies Anonymity and Independent Symmetry, from Lemma 9.1 we know that there exists a function, $g : \mathbb{N}^3 \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ we have

$$f(x, y) = g(|A_x \cap A_y|, |A_x \triangle A_y|, n - |A_x \cup A_y|).$$

It remains to show that for arbitrary $x, y \in \{0,1\}^n$ it holds that

$$g(|A_x \cap A_y|, |A_x \triangle A_y|, n - |A_x \cup A_y|) = g(|A_x \cap A_y|, |A_x \triangle A_y|, 0). \tag{6}$$

By Anonymity, without loss of generality, we can assume that all indices $i \in [n]$ such that $x_i = y_i = 0$ are greater than all other indices. Then, let $n' = |A_x \cup A_y|$ and $x', y' \in \{0,1\}^{n'}$ be vectors such that $x'_i = x_i$ and $y'_i = y_i$ for every $i \in [n']$. By Add Zero used $n - n'$ times, we obtain that $f(x, y) = f(x', y')$. Since $n' - |A_{x'} \cup A_{y'}| = 0$, this proves Equation (6). □

Lemma 9.8. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Add Zero, and Scaling then there exists a function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x, y) = g(|A_x \triangle A_y| / |A_x \cup A_y|).$$

Proof. Assume otherwise, i.e., there exist four vectors $x, y, a, b \in \{0,1\}^n$ such that $|A_x \triangle A_y| / |A_x \cup A_y| = |A_a \triangle A_b| / |A_a \cup A_b|$ and $f(x, y) \neq f(a, b)$. Let m be the least common multiple of $|A_x \cup A_y|$ and $|A_a \cup A_b|$ and let $p = m / |A_x \cup A_y|$ and $q = m / |A_a \cup A_b|$. By

[2] It is a known fact that the Jaccard distance satisfies Triangle Inequality (see e.g., [15, 22]). We provide a proof for completeness.

Scaling, we know that $f(x,y) = f(x^p, y^p)$ and $f(a,b) = f(a^q, b^q)$. Thus, to get the contradiction, it remains to show that $f(x^p, y^p) = f(a^q, b^q)$.

To this end, observe that

$$\begin{aligned}
|A_{x^p} \triangle A_{y^p}| &= p \cdot |A_x \triangle A_y| \\
&= p \cdot |A_x \cup A_y| \cdot \frac{|A_a \triangle A_b|}{|A_a \cup A_b|} \\
&= \frac{p \cdot |A_x \cup A_y|}{q \cdot A_a \cup A_b|} \cdot q \cdot |A_a \triangle A_b| \\
&= \frac{m}{m} \cdot q \cdot |A_a \triangle A_b| \\
&= |A_{a^q} \triangle A_{b^q}|.
\end{aligned}$$

Thus, both sum and the symmetric difference of A_{x^p}, A_{y^p} and A_{a^q}, A_{b^q} are of equal size. This implies that also their intersections are of equal size. Since f satisfies Anonymity, Independent Symmetry, and Add Zero, by Lemma 9.7, this means that $f(x^p, y^p) = f(a^q, b^q)$. This concludes the proof. □

Lemma 9.9. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Add Zero, Scaling, and Triangle Inequality, then there exists a subadditive function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = g(|A_x \triangle A_y| / |A_x \cup A_y|).$$

Proof. From Lemma 9.8 we know that there exists a function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that $f(x,y) = g(|A_x \triangle A_y| / |A_x \cup A_y|)$. It remains to show that g is subadditive, i.e., for arbitrary $p, q \in (\mathbb{Q} \cap [0,1])$ such that $p + q \leq 1$, it holds that $g(p) + g(q) \geq g(p+q)$. To this end, let $r = 1 - p - q$ and let k be such that $pk, qk, rk \in \mathbb{N}$. Then, consider the following three vectors

$$\begin{aligned}
x &= (1)^{pk} \circ (1)^{rk} \circ (0)^{qk}, \\
y &= (0)^{pk} \circ (1)^{rk} \circ (0)^{qk}, \text{ and} \\
z &= (0)^{pk} \circ (1)^{rk} \circ (1)^{qk}.
\end{aligned}$$

Then, from Triangle Inequality we get $f(x,y) + f(y,z) \geq f(x,z)$, which in terms of g means $g(\frac{kp}{k(p+q+r)}) + g(\frac{kq}{k(p+q+r)}) \geq g(\frac{k(p+q)}{k(p+q+r)})$. Since $p+q+r = 1$, we get that indeed $g(p) + g(q) \geq g(p+q)$, which concludes the proof. □

Lemma 9.10. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Add Zero, Scaling, Convergence, and Triangle Inequality then there exists constant $a \in \mathbb{R}_{\geq 0}$ such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = a \cdot J(x,y).$$

Proof. From Lemma 9.9 we know that there exists a subadditive function, $g : (\mathbb{Q} \cap [0,1]) \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that $f(x,y) = g(|A_x \triangle A_y| / |A_x \cup A_y|) = g(J(x,y))$. Let $a = g(1)$. We will show that for every $r \in \mathbb{Q} \cap [0,1]$ it holds that $g(r) = ar$, which will imply the thesis.

We begin by showing that for every $r \in \mathbb{Q} \cap [0,1]$ it holds that

$$g(r/2) = g(r)/2. \tag{7}$$

Let $p, q \in \mathbb{N}$ be such that $r = p/q$. Then, by Convergence we get that

$$\begin{aligned}
g(p/q) &= f((1)^q, (0)^p \circ (1)^{q-p}) \\
&\geq 2 \cdot f((1)^q \circ (1)^q, (0)^p \circ (1)^{q-p} \circ (1)^q) \\
&= 2 \cdot g(p/(2q)).
\end{aligned}$$

Hence, $g(r) \geq 2 \cdot g(r/2)$. On the other hand, $g(r) \leq 2 \cdot g(r/2)$ from subadditivity. Thus, indeed Equation (7) holds.

Next, we generalize Equation (7) and prove that for every $r \in \mathbb{Q} \cap [0,1]$ and $q \in \mathbb{N}$ it holds that

$$g(r/q) = g(r)/q. \tag{8}$$

Observe that it is enough to prove this equality for prime q, as for composite qs we can obtain the thesis by combining the results for all prime factors of q. Thus, without loss of generality, let us assume that q is prime. Clearly, from subadditivity, we have that $g(r/q) \geq g(r)/q$. For a contradiction assume that $g(r/q) = g(r)/q + \varepsilon$ for some $\varepsilon > 0$. By Little Fermat's Theorem, we know that there exists $p \in \mathbb{N}$ such that $p \cdot q = 2^{q-1} - 1$. Then, we have that $1 + p \cdot q = 2^{(q-1)}$, which means that $1 = (1 + p \cdot q)/2^{(q-1)}$. In turn, this implies that

$$\frac{1}{q} = \frac{1 + p \cdot q}{2^{q-1} q} = \frac{1}{2^{q-1} q} + \frac{p}{2^{q-1}}. \tag{9}$$

Thus, we get

$$\begin{aligned}
g\left(\frac{r}{q}\right) &= g\left(\frac{r}{2^{k(q-1)} q} + \frac{r \cdot p}{2^{k(q-1)}}\right) && \text{(from Equation (9))} \\
&\leq g\left(\frac{r}{2^{q-1} q}\right) + p \cdot g\left(\frac{r}{2^{q-1}}\right) && \text{(from subadditivity)} \\
&= \frac{1}{2^{q-1}} g\left(\frac{r}{q}\right) + \frac{p}{2^{q-1}} g(r). \\
& && \text{(from Equation (7) used } q-1 \text{ times)}
\end{aligned}$$

Substituting $g(r/q) = g(r)/q + \varepsilon$, we get

$$\begin{aligned}
g(r)/q + \varepsilon &\leq \frac{g(r)/q + \varepsilon}{2^{q-1}} + \frac{p}{2^{q-1}} g(r) \\
&= g(r) \frac{1 + p \cdot q}{2^{q-1} q} + \varepsilon / 2^{q-1} \\
&= g(r)/q + \varepsilon / 2^{q-1}. && \text{(from Equation (9))}
\end{aligned}$$

Thus, $\varepsilon \leq \varepsilon / 2^{q-1}$, which is a contradiction for $\varepsilon > 0$. Therefore, Equation (8) indeed holds, from which we immediately obtain that

$$\begin{aligned}
g(0) &= 0, \text{ and} \\
g(1/q) &= a/q, \text{ for every } q \in \mathbb{N}.
\end{aligned} \tag{10}$$

Finally, let us take $r = p/q$ for arbitrary $p, q \in \mathbb{N}$ such that $1 \leq p \leq q$. From subadditivity and Equation (10) we get that $g(p/q) \leq p \cdot g(1/q) = ap/q$. On the other hand, again from subadditivity and Equation (10), we get that $g(p/q) \geq g(1) - (q-p) g(1/q) = a - a(q-p)/q = a \cdot p/q$. Thus, indeed $g(r) = ar$, for every $r \in \mathbb{Q} \cap [0,1]$, which concludes the proof. □

Combining Lemma 9.10 with Normalization we obtain the thesis. □

9.4 Proof of Theorem 4.3

Proof. For each of the axioms we provide a dissimilarity measure f that satisfies all but that axiom.

Anonymity: $f(x,y) = \sum_{i=1}^{n} |x_i - y_i| \cdot 2^i / (\sum_{i=1}^{n} \max(x_i, y_i) \cdot 2^i)$
Scaling: $f(x,y) = (|A_x \triangle A_y| + 1)/(2|A_x \cup A_y|)$
Independent Symmetry: $f(x,y) = \max(|A_x \setminus A_y|, |A_y \setminus A_x|)/|A_x \cup A_y|$
Add Zero: $f(x,y) = H(x,y)$
Convergence: $f(x,y) = \max_{i \in [n]} |x_i - y_i|$ (discrete distance)
Triangle Inequality: $f(x,y) = J(x,y)^2$
Normalization: $f(x,y) = 2 \cdot J(x,y)$

□

9.5 Proof of Proposition 6.1

Proof. Consider three arbitrary candidates x, y, z and without loss of generality assume that their positions are such that $p_x < p_y < p_z$. With a slight abuse of notation by $x, y, z \in \{0,1\}^n$ we will also denote the approval vectors for candidates x, y, z, respectively. Let us use the following notation:

$$\begin{aligned} n_x &= |A_x \setminus (A_y \cup A_z)|, \\ n_y &= |A_y \setminus (A_x \cup A_z)|, \\ n_z &= |A_z \setminus (A_x \cup A_y)|, \\ n_{xy} &= |(A_x \cap A_y) \setminus A_z|, \\ n_{yz} &= |(A_y \cap A_z) \setminus A_x|, \\ n_{xyz} &= |A_x \cap A_y \cap A_z|. \end{aligned}$$

Note that all of the above sets are pairwise disjoint. Since y is between x and z and we assumed equal radii, then $(A_x \cap A_z) \setminus A_y$ has to be empty (the intersection of the x's and z's interval, if it exists, must be entirely contained in the y's interval).

Now, we need to show that $J(x,y) \leq J(x,y)$, i.e.,

$$\frac{|A_x \triangle A_y|}{|A_x \cap A_y|} \leq \frac{|A_x \triangle A_z|}{|A_x \cap A_z|}.$$

This is equivalent to

$$\frac{n_x + n_y + n_{yz}}{n_x + n_y + n_{xy} + n_{yz} + n_{xyz}} \leq \frac{n_x + n_z + n_{xy} + n_{yz}}{n_x + n_z + n_{xy} + n_{yz} + n_{xyz}}.$$

After simplifying, we get

$$n_y n_{xyz} \leq n_{xy}(n_x + n_z + n_{xy} + n_{yz} + n_{xyz}) + n_z n_{xyz}.$$

However, since the radii of all intervals are the same, it must hold that either the y's interval is completely contained in the sum of x's and z's intervals, and then $n_y = 0$, or x's and z's intervals do not intersect, and then $n_{xyz} = 0$. Thus, the left hand side of the last inequality is always equal to 0 and since the right hand is always nonnegative, this concludes the proof. □

10 Axiomatic Characterization of the Discrete Distance

The axioms on which the Hamming and Jaccard distances differ are Zero-One Symmetry and Add Zero. A reader may wonder if it is possible to satisfy both of this axioms at the same time. In this section, we show that these two axioms together with some axioms that are shared by Hamming and Jaccard uniquely characterize the discrete distance, which for every pair of vectors $x, y \in \{0,1\}^n$, is defined as

$$D(x,y) = \max_{i \in [n]} |x_i - y_i| = \begin{cases} 0, & \text{if } x = y, \\ 1, & \text{otherwise.} \end{cases}$$

To this end, we first introduce one more axiom, which is a standard distance metric axiom and is satisfied by both Hamming and Jaccard.

Definition 10.1 (Identity)**.** *A dissimilarity measure, f, satisfies* Identity *if for every vectors $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = 0 \iff x = y.$$

Now, let us state the main result of this appendix.

Theorem 10.2. *A dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Add Zero, Zero-One Symmetry, Scaling, Identity, and Normalization if and only if f is the discrete distance.*

Proof. Let us first show that the discrete distance indeed satisfies the respective axioms, by considering them one by one.

Anonymity
Clearly, for every $x, y \in \{0,1\}^n$ and permutation $\pi : [n] \to [n]$, it holds that $x = y \Leftrightarrow \pi(x) = \pi(y)$. Thus,

$$D(\pi(x), \pi(y)) = D(x,y).$$

Scaling
Observe that for every $x, y \in \{0,1\}^n$ and $k \in \mathbb{N}$, it holds that $x = y \Leftrightarrow x^k = y^k$. Hence,

$$D(x^k, y^k) = D(x,y).$$

Independent Symmetry
For every $x, y \in \{0,1\}^n$ and $i \in [n]$, it holds that $x = y \Leftrightarrow (x_{-i}, y_i) = (y_{-i}, x_i)$. Thus,

$$D((x_{-i}, y_i), (y_{-i}, x_i)) = D(x,y).$$

Zero-One Symmetry
Note that for every $x, y \in \{0,1\}^n$, it holds that $x = y \Leftrightarrow \bar{x} = \bar{y}$. Hence,

$$D(\bar{x}, \bar{y}) = D(x,y).$$

Add Zero
Note that for every $x, y \in \{0,1\}^n$, it holds that $x = y \Leftrightarrow x \circ (0), y \circ (0)$. Thus,

$$D(x \circ (0), y \circ (0)) = D(x,y).$$

Identity
We get the Identity directly from the definition of the discrete distance.

Normalization
Clearly, $D((0),(1)) = 1$.

Now, let us focus on the converse statement, i.e., that any dissimilarity measure that satisfies these axioms must necessarily be the discrete distance.

From the proof of the Jaccard distance we already know how does the class of dissimilarity measures satisfying Anonymity, Independent Symmetry, and Add Zero behaves. Now, let us add to it Zero-One Symmetry.

Lemma 10.3. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, Add Zero, and Zero-One Symmetry then there exists a function, $g : \mathbb{N} \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that*

$$f(x,y) = g(|A_x \triangle A_y|).$$

Proof. From Lemma 9.7 we know that there exists a function, $g : \mathbb{N}^2 \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that $f(x,y) = g(|A_x \triangle A_y|, |A_x \cup A_y|)$. Hence, it is enough to show that $g(|A_x \triangle A_y|, |A_x \cup A_y|) = g(|A_x \triangle A_y|, |A_x \triangle A_y|)$ for every pair $x, y \in \{0,1\}^n$.

To this end, fix arbitrary $x, y \in \{0,1\}^n$. Let $x', y' \in \{0,1\}^{n'}$ be vectors obtained from x, y with all coordinates on which x and y have both 0s removed. Since $|A_{x'} \triangle A_{y'}| = |A_x \triangle A_y|$ and $|A_{x'} \cup A_{y'}| = |A_x \cup A_y|$, we know that $g(|A_x \triangle A_y|, |A_x \cup A_y|) = f(x', y')$. Then, by Zero-One Symmetry we obtain $g(|A_x \triangle A_y|, |A_x \cup A_y|) = f(x', y') = f(\bar{x}', \bar{y}') = g(|A_{\bar{x}'} \triangle A_{\bar{y}'}|, |A_{\bar{x}'} \cup A_{\bar{y}'}|)$. However, since there is no index $i \in [n']$ such that $x'_i = y'_i = 0$, there is also none for which $\bar{x}'_i = \bar{y}'_i = 1$, i.e., $A_{\bar{x}'} \cap A_{\bar{y}'} = \varnothing$. This means that $A_{\bar{x}'} \cup A_{\bar{y}'} = A_{\bar{x}'} \triangle A_{\bar{y}'}$. On the other hand, $A_{\bar{x}'} \cup A_{\bar{y}'} = A_x \cup A_y$. Thus, indeed

$$\begin{aligned} g(|A_x \triangle A_y|, |A_x \cup A_y|) &= f(x', y') = f(\bar{x}', \bar{y}') \\ &= g(|A_x \triangle A_y|, |A_x \triangle A_y|) \end{aligned}$$

□

Next, we consider dissimilarity measures satisfying Scaling as well.

Lemma 10.4. *If a dissimilarity measure, f, satisfies Anonymity, Scaling, Independent Symmetry, Add Zero, and Zero-One Symmetry, then there exist constants $a, b \in \mathbb{R}_{\geq 0}$, such that*

$$f(x,y) = \begin{cases} a & \text{if } x = y, \\ b & \text{otherwise.} \end{cases}$$

Proof. From Lemma 10.3 we know that there exists a function, $g : \mathbb{N} \to \mathbb{R}_{\geq 0}$, such that for every $x, y \in \{0,1\}^n$ it holds that $f(x,y) = g(|A_x \triangle A_y|)$. Let us denote $a = g(0)$ and $b = g(1)$. It remains to show that $g(k) = b$ also for every $k \in \{2, 3, \dots\}$. From Scaling, we know that $f((1)^k, (0)^k) = f((1), (0))$. Thus, indeed

$$g(k) = f((1)^k, (0)^k) = f((1), (0)) = g(1) = b.$$

□

Finally, as from Lemma 10.4 we know that

$$f(x,y) = \begin{cases} a & \text{if } x = y, \\ b & \text{otherwise,} \end{cases}$$

for every $x, y \in \{0,1\}^n$, arriving at the thesis is straightforward. Indeed, we know that $a = 0$ from Identity and $b = 1$ from Normalization. □

Once again, we can show that the axioms characterizing the discrete distance are independent.

Theorem 10.5. *For every axiom in the set Anonymity, Independent Symmetry, Add Zero, Zero-One Symmetry, Scaling, Identity, and Normalization there is a dissimilarity measure that satisfies all other axioms in this set except for this one.*

Proof. For each of the axioms we provide a dissimilarity measure f that satisfies all but that axiom.

Anonymity: $f(x,y) = \max_{i \in [n]} \frac{|x_i - y_i|}{2^{i-1}}$
Scaling: $f(x,y) = \sum_{i \in [n]} |x_i - y_i| = n \cdot H(x,y)$
Independent Symmetry: $f(x,y) = \begin{cases} 0, & \text{if } x = y, \\ \max(|A_x \setminus A_y|, |A_y \setminus A_x|)/|A_x \triangle A_y|, & \text{otherwise.} \end{cases}$
Add Zero: $f(x,y) = H(x,y)$
Zero-One Symmetry: $f(x,y) = J(x,y)$
Identity: $f(x,y) = (D(x,y) + 1)/2$
Normalization: $f(x,y) = 2 \cdot D(x,y)$

□

11 Real-Life Data Details

See Tables 5 and 6.

Table 5: Parameters of real-life participatory budgeting instances.

	Warsaw		**Kraków**		**Łódź**	
Metric	2023	2024	2023	2024	2023	2024
Number of instances	18	18	18	18	36	36
Avg. num. of voters	4562.61	4128.00	3605.94	3616.72	2549.86	2470.94
Avg. num. of projects	59.17	60.94	31.06	31.72	18.33	19.03
Avg. num. of categories	7.94	8.0	5.67	5.72	3.5	3.61
Avg. num. of proj. per cat.	19.24	19.96	4.47	4.33	3.76	3.83

Table 6: Average Pearson correlation between our six measures and the Euclidean distance.

	Warsaw						**Wieliczka**
Metric	2019	2020	2021	2022	2023	2024	2023
Number of districts	18	18	18	18	18	18	1
H_0 geometric Hamming	0.123	0.077	0.067	0.093	0.058	0.024	0.228
H_1 Hamming	0.101	0.049	0.057	0.074	0.041	0.012	0.186
H_2 quadratic Hamming	0.081	0.034	0.045	0.056	0.031	0.005	0.154
J_0 geometric Jaccard	**0.283**	**0.219**	**0.226**	**0.175**	**0.236**	**0.260**	**0.441**
J_1 Jaccard	0.268	0.208	0.210	0.162	0.222	0.244	0.369
J_2 quadratic Jaccard	0.262	0.205	0.205	0.156	0.219	0.238	0.387
$J_0 > J_1$	0.94	0.89	0.94	0.83	0.89	0.89	1.0