A Missing Proofs

A.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

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

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 will prove what form has to have such a function f. In the last one, Lemma A.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.

[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.

Lemma A.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 A.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 A.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 A.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 A.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

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

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. $\square$

Lemma A.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 A.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. $\square$

Lemma A.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 A.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 A.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 A.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 A.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

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

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

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

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

$$f(x,y) \geq aH(x, y' \circ (1)) - a/n = a(H(x,y) + 1/n) - a/n = aH(x,y),$$

which concludes the proof. □

Combining Lemma A.6 with Normalization we obtain the thesis. □

A.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: $f(x,y) = 2 \cdot H(x,y)$

□

A.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

$$J(x \circ (1), y \circ (1)) = \frac{|A_x \triangle A_y|}{|A_x \cup A_y| + 1} \le \frac{n}{n+1} \cdot \frac{|A_x \triangle A_y|}{|A_x \cup A_y|} = \frac{n}{n+1} \cdot J(x,y),$$

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

Triangle Inequality[2]

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

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

On the other hand,

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

and similarly $J(y,z) \ge |A_y \triangle A_z|/|A_x \cup A_y \cup A_z|$. 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

$$|A_x \triangle A_z| + |A_y \setminus (A_x \cup A_z)| \le |(A_x \triangle A_y) \cup (A_y \triangle A_z)| \le |A_x \triangle A_y| + |A_y \triangle A_z|.$$

Combining all inequalities we get

$$J(x,y) + J(y,z) \ge \frac{|A_x \triangle A_y| + |A_y \triangle A_z|}{|A_x \cup A_y \cup A_z|} \ge \frac{|A_x \triangle A_z| + |A_y \setminus (A_x \cup A_z)|}{|A_x \cup A_y \cup A_z|} \ge J(x,z).$$

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 A.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 $a \in \mathbb{R}_{\ge 0}$ 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 A.7. *If a dissimilarity measure, f, satisfies Anonymity, Independent Symmetry, and Add Zero, then there exists a function, $g : \mathbb{N}^2 \to \mathbb{R}_{\ge 0}$, 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 A.1 we know that there exists a function, $g : \mathbb{N}^3 \to \mathbb{R}_{\ge 0}$, 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). $\square$

[2]It is a known fact that the Jaccard distance satisfies Triangle Inequality (see e.g., [15, 21]). We provide a proof for completeness.

Lemma A.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 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 A.7, this means that $f(x^p, y^p) = f(a^q, b^q)$. This concludes the proof. □

Lemma A.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 A.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 A.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 A.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 A.10 with Normalization we obtain the thesis. □

A.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: $f(x, y) = 2 \cdot J(x, y)$

$\square$

A.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 $J(x, y) \le J(x, y)$, i.e.,

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

This is equivalent to

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

After simplifying, we get

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

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. $\square$

B 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 B.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 B.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 B.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 A.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

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

□

Next, we consider dissimilarity measures satisfying Scaling as well.

Lemma B.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 B.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 $k \in \{2, 3, \ldots\}$. 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 B.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 B.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: $f(x, y) = 2 \cdot D(x, y)$

□

C Real-Life Data Details

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

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, S. Brânzei, and D. Loker. Weighted clustering: Towards solving the user's dilemma. *Pattern Recognition*, 120:108152, 2021.

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

[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] 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.

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

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

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

[20] 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.

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

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

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

[24] 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.

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

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

[27] 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.

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

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

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

[31] 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.

[32] 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.