# String Set Metrics

## 1 Introduction

In this sketch we are interested in studying metric spaces between sets of strings.

## 2 Preliminaries

**Definition 1** (Metric Space)**.** *A metric space* $(M, d)$ *is a set* $M$ *along with a distance function* $d \colon M \times M \to \mathbb{R}_{\geq 0}$ *such that for any* $x, y, z \in M$*:*

*(1)* $d(x, y) \geq 0$

*(2)* $d(x, y) = 0 \iff x = y$

*(3)* $d(x, y) = d(y, x)$

*(4)* $d(x, z) \leq d(x, y) + d(y, z)$

**Definition 2** (Alphabet)**.** *An alphabet* $\Sigma$ *is a finite set of unique symbols.*

**Definition 3** (String)**.** *Given an alphabet* $\Sigma$*, a string* $\sigma$ *is a finite sequence of symbols from* $\Sigma$*.*

**Definition 4** (Alphabet Strings)**.** *Let* $\Sigma^\star$ *denote the set of all possible strings from* $\Sigma$*.*

**Definition 5** (String Metric)**.** *A function* $\delta \colon \Sigma^\star \times \Sigma^\star \to \mathbb{R}_{\geq 0}$ *that satisfies metric space axioms.*

## 3 String Set Metric Spaces

### 3.1 Hausdorff Metric

We first consider the following problem. Given a single string $\sigma$, and a set of string $A$, how might we calculate a distance from $\sigma$ to $A$? Let $\delta$ be a string metric, then one idea is as follows:

**Definition 6** (Merge Distance)**.** *The merge distance* $\delta_M\colon \Sigma \times \Sigma^\star \to \mathbb{R}_{\geq 0}$ *is defined as:*

$$\delta_M\left(\sigma, A\right) = \inf_{a \in A} \left\{\delta\left(\sigma, a\right) \ : \ a \in A\right\}$$

The idea here is that we take the string in $A$ that most closely resembles $\sigma$ with respect to the string metric $\delta$, and consider that the distance between $\sigma$ and $A$.

We can take this idea further. Suppose we have two sets of strings $A$ and $B$. The Hausdorff distance is then defined as follows:

**Definition 7** (Hausdorff Distance)**.** *The Hausdorff distance* $d_H\colon \Sigma^\star \times \Sigma^\star \to \mathbb{R}_{\geq 0}$ *is defined as:*

$$d_H\left(A, B\right) = \max\left\{\sup\left\{\delta_M\left(a, B\right) \ : \ a \in A\right\}, \sup\left\{\delta_M\left(b, A\right) \ : \ b \in B\right\}\right\}$$

The Hausdorff distance is essentially the promotion of a metric space $(X, d)$ to another metric space $\left(2^X, d_H\right)$ where $d_H$ is defind with respect to $d$. The definition is not unique to strings here.