Fast Approximations and Coresets for (k, ℓ)-Median Under Dynamic Time Warping

Jacobus Conradi
University of Bonn, Germany

Benedikt Kolbe
Hausdorff Center for Mathematics, University of Bonn, Germany

Ioannis Psarros
Archimedes, Athena Research Center, Greece

Dennis Rohde
University of Bonn, Germany

Abstract

We present algorithms for the computation of ε-coresets for k-median clustering of point sequences in $\mathbb{R}^d$ under the p-dynamic time warping (DTW) distance. Coresets under DTW have not been investigated before, and the analysis is not directly accessible to existing methods as DTW is not a metric. The three main ingredients that allow our construction of coresets are the adaptation of the ε-coreset framework of sensitivity sampling, bounds on the VC dimension of approximations to the range spaces of balls under DTW, and new approximation algorithms for the k-median problem under DTW. We achieve our results by investigating approximations of DTW that provide a trade-off between the provided accuracy and amenability to known techniques. In particular, we observe that given n curves under DTW, one can directly construct a metric that approximates DTW on this set, permitting the use of the wealth of results on metric spaces for clustering purposes. The resulting approximations are the first with polynomial running time and achieve a very similar approximation factor as state-of-the-art techniques. We apply our results to produce a practical algorithm approximating (k, ℓ)-median clustering under DTW.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms

Keywords and phrases Dynamic time warping, coreset, median clustering, approximation algorithm

Digital Object Identifier 10.4230/LIPIcs.SoCG.2024.42

Related Version *Full Version*: `https://doi.org/10.48550/arXiv.2312.09838`

Funding *Jacobus Conradi*: Partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 313421352 (FOR 2535 Anticipating Human Behavior) and the iBehave Network: Sponsored by the Ministry of Culture and Science of the State of North Rhine-Westphalia. Affiliated with Lamarr Institute for Machine Learning and Artificial Intelligence.
Benedikt Kolbe: Partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 459420781.
Ioannis Psarros: This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.
Dennis Rohde: Partially supported by the Hausdorff Center for Mathematics. Partially funded by the iBehave Network: Sponsored by the Ministry of Culture and Science of the State of North Rhine-Westphalia.

Acknowledgements We thank Anne Driemel of the University of Bonn for detailed discussion and guidance.

40th International Symposium on Computational Geometry (SoCG 2024).
Editors: Wolfgang Mulzer and Jeff M. Phillips; Article No. 42; pp. 42:1–42:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

One of the core challenges of contemporary data analysis is the handling of massive data sets. A powerful approach to clustering problems involving such sets is data reduction, and ε-coresets offer a popular approach that has received substantial attention. An ε-coreset is a problem-specific condensate of the given input set of reduced size which captures its core properties towards the problem at hand and can be used as a proxy to run an algorithm on, producing a solution with a relative error of $(1 \pm \varepsilon)$.

Clustering and especially k-median represent fundamental tasks in classification problems, where they have been extensively studied for various spaces. With the growing availability of e.g. geospatial tracking data, clustering problems for time series or curves have received growing attention both from a theoretical and applied perspective. In practice, time series classification largely relies on the dynamic time warping (DTW) distance and is widely used in the area of data mining. Simple nearest neighbor classifiers under DTW are considered hard to beat [26, 38] and much effort has been put into making classification using DTW computationally efficient [25, 30, 31, 34]. In contrast to its cousin the Fréchet distance, DTW is less sensitive to outliers, but its algorithmic properties are also less well understood, owing to the fact that it is not a metric. In particular, the wealth of research surrounding k-median clustering for metric spaces does not directly apply to clustering problems under DTW.

For time series and curves, k-median takes the shape of the (k, ℓ)-median problem, where the sought-for center curves are restricted to have a complexity (number of vertices) of at most ℓ, with a two-fold motivation. First, the otherwise NP-hard problem becomes tractable, and second, it suppresses overfitting.

The construction of ε-coresets for the (k, ℓ)-median problem for DTW is precisely what this paper will address. To this end, we adapt the framework of *sensitivity sampling* by Feldman and Landberg [22] to our setting, derive bounds on the VC dimension of approximate range spaces of balls under DTW, develop fast approximation algorithms solving (k, ℓ)-median clustering, and use coresets to improve existing (k, ℓ)-median algorithms, for curves under DTW. We rely on approximations of nearly all objects involved in our inquiry, thereby improving the bounds we obtain for the VC dimension of the range spaces in question and broadening the scope of our approach.

Our analysis of the VC dimension is possibly of independent interest. The VC dimension exhibits a near-linear dependency on the complexity of the sequences used as centers of the ranges, yet it depends only logarithmically on the size of the curves within the ground set. This distinction holds significant implications in the analysis of real datasets, where queries may involve simple, short sequences, but the dataset itself may consist of complex, lengthy sequences. Note that our results hold for range spaces that are defined by small perturbations of DTW distances. This means that for any given set of input sequences requiring DTW-based analysis, there is slight perturbation of DTW with associated range space of bounded VC dimension. This is sufficient to enable a broad array of algorithmic techniques that leverage the VC dimension, particularly in scenarios where approximate computations are allowed.

Related Work. Among different practical approaches for solving the k-median problem, a very influential heuristic is the DTW Barycentric Average (DBA) method [32]. While it has seen much success in practice [1, 23, 33], it does not have any theoretical guarantees and indeed may converge to a local configuration that is arbitrarily far from the optimum. Recently, theoretical results for average series problems under DTW have been obtained. The problem is NP-hard for rational-valued time series and W[1]-hard in the number of input time series [10, 18]. Furthermore, it can not be solved in time $O(f(n)) \cdot m^{o(n)}$ for any computable function f unless the Exponential Time Hypothesis (ETH) fails. There is an exponential time exact algorithm for rational-valued time series [8] and polynomial time exact algorithms for binary time series [8, 36]. There is an exact algorithm for the related problem of finding a single mean curve of given complexity for time series over $\mathbb{Q}^d$, minimizing the sum of squares of DTW distances to input curves, which runs in polynomial time if the number of points of the average series is constant [16]. Furthermore, approximation algorithms were recently developed [16], and some of these can be slightly modified to work within the median clustering approximation framework of [13, 15]. Unfortunately, known median clustering approximation algorithms either have running time exponential in the length of the average series, or a very large approximation factor.

Approximation Algorithms for Series Clustering. In the last decade, the problems of (k,ℓ)-median and (k,ℓ)-center clustering for time series in $\mathbb{R}^d$ under the Fréchet distance have gained significant attention. The problem is NP-hard [9, 11, 21], even if $k = 1$ and $d = 1$ (in these works, time series are real-valued sequences), and the (k,ℓ)-center problem is even NP-hard to approximate within a factor of $(2.25 - \varepsilon)$ for $d \geq 2$ [9] ($(1.5 - \varepsilon)$, if $d = 1$). For the (k,ℓ)-median problem, all presently known $(1+\varepsilon)$-approximation algorithms are based on an approximation scheme [14, 20, 21] which has been generalized several times [2, 15, 27]. The most recent version of this scheme [15, Theorem 7.2] can be utilized to approximate any k-median type problem in an arbitrary space X with a distance function. All that it needs is a plugin-algorithm that, when given a set T of elements from some (problem-specific) subset $Y \subseteq X$, returns a set of candidates C that contains, for any set $T' \subseteq T$ with roughly $|T'| \geq |T|/k$, with a previously fixed probability, an approximate median. The resulting approximation quality and running time depend on the approximation factor of the plugin and $|C|$, respectively, with a factor of $O(|C|^k)$ in the running time.

For the Fréchet distance, plugin-algorithms exist that yield $(1+\varepsilon)$-approximations [14, 20]. For DTW however, the best plugin-algorithm [16] has runing time exponential in k – roughly with a dependency of $\widetilde{O}((32k^2\varepsilon^{-1})^{k+2}n)$ – and approximation guarantee of $(8+\varepsilon)(m\ell)^{1/p}$ with constant success probability. Here, the $\widetilde{O}$ notation hides polylogarithmic factors. In principle, some of the ideas from plugins for the Fréchet distance could be adapted, but the more involved plugins, i.e., the ones yielding $(1+\varepsilon)$-approximations, crucially make use of the metric properties of the distance function.

In practice, an adaption of Gonzalez algorithm for (k,ℓ)-center clustering under the Fréchet distance performs well [12]. Similar ideas have also been used for clustering under (a continuous variant of) DTW [5], but there are no approximation guarantees, and the usual analysis is based on repetitive use of the triangle inequality. To the best of our knowledge, all (k,ℓ)-median $(1+\varepsilon)$-approximation algorithms for Fréchet and DTW are impractical due to large constants and an exponential dependency on ℓ in the running time.

For the Fréchet distance, ε-coresets can be constructed [6, 17] that help facilitate the practicability of available algorithms. Using ε-coresets, a $(5+\varepsilon)$-approximation algorithm for the 1-median problem was recently analyzed [17], yielding a running time of roughly $nm^{O(1)} + (m/\varepsilon)^{O(\ell)}$, in contrast to a running time of $n(m/\varepsilon)^{O(\ell)}$ without the use of coresets [15].

For DTW, no coreset construction is known to this point. This is at least partially due to prominent coreset frameworks assuming a normed or at least a metric space [22, 28]. Also, recently a coreset construction relying solely on uniform sampling was developed that greatly simplifies existing coreset constructions [7], including the aforementioned coresets under the Fréchet distance. Unfortunately, the construction again relies on different incarnations of the triangle inequality, limiting its use for DTW.

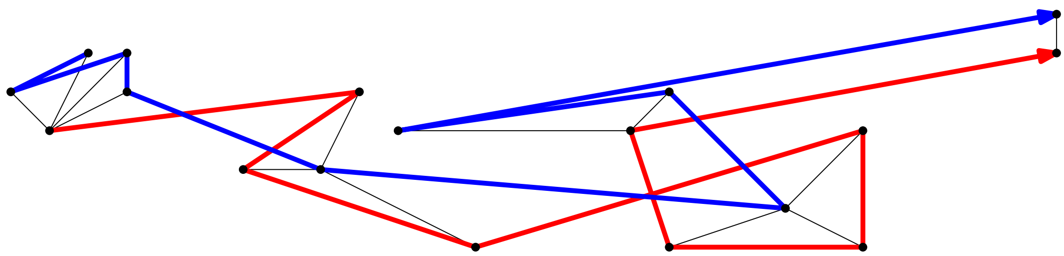

Figure 1 Example of a traversal between the red and blue curve realizing the dynamic time warping distance. The sum of the black distances is minimized.

Results. To construct ε-coresets, we use approximations of the range space defined by balls under p-DTW and bound their VC dimension. Assuming that the input is a set of n curves of complexity at most m, we present an approximation algorithm (Theorem 35) for k-median with running time in $O(n)$ (hiding other factors), that improves upon existing work in terms of running time, with comparable approximation guarantees. Our approach relies on curve simplifications and approximating p-DTW by a path metric. This allows us to apply state-of-the-art k-median techniques in this nearby path metric space, circumventing the use of heavy k-median machinery in non-metric spaces which would incur exponential dependence on k and the success probability. Our main ingredient is a new insight into the notion of relaxed triangle inequalities for p-DTW (Lemma 18). We then construct a coreset based on the approximation algorithm. For this, we bound the so-called sensitivity of the elements of the given data set, as well as their sum. The sensitivities are a measure of the data elements' importance and determine the sample probabilities in the coreset construction. We construct an ε-coreset for (k, ℓ)-median clustering of size quadratic in $1/\varepsilon$ and k, logarithmic in n, and depending on $(m\ell)^{1/p}$ and ℓ (Corollary 37). We achieve this by upper bounding the VC dimension of the approximate range space with logarithmic dependence on m (Theorem 17).

2 Preliminaries

We think of a sequence $(p_1, \ldots, p_m) \in (\mathbb{R}^d)^m$ of points in $\mathbb{R}^d$ as a (polygonal) *curve*, with complexity m. We denote by $\mathbb{X}^d_{=m}$ the space of curves in $\mathbb{R}^d$ with complexity exactly m and by $\mathbb{X}^d_m$ the space of curves with complexity at most m.

▶ **Definition 1** (p-Dynamic Time Warping). *For given $m, \ell > 0$ we define the space $\mathcal{T}_{m,\ell}$ of* (m, ℓ)-traversals *as the set of sequences $((a_1, b_1), (a_2, b_2), \ldots, (a_l, b_l))$, such that*

- *$a_1 = 1$ and $b_1 = 1$; and $a_l = m$ and $b_l = \ell$,*
- *for all $i \in [l-1] := \{1, \ldots, l-1\}$ it holds that $(a_{i+1}, b_{i+1}) - (a_i, b_i) \in \{(1, 0), (0, 1), (1, 1)\}$.*

For $p \in [1, \infty)$ and two curves $\sigma = (\sigma_1, \ldots, \sigma_m) \in \mathbb{X}^d_{=m}, \tau = (\tau_1, \ldots, \tau_\ell) \in \mathbb{X}^d_{=\ell}$ the (p-)Dynamic Time Warping distance (p-DTW) is defined as

$$\mathrm{dtw_p}(\sigma, \tau) = \min_{T \in \mathcal{T}_{m,\ell}} \left(\sum_{(i,j) \in T} \|\sigma_i - \tau_j\|_2^p \right)^{1/p}.$$

The central focus of the paper is the following clustering problem.

▶ **Definition 2** (Problem definition). *The (k, ℓ)-median problem for $\mathbb{X}^d_m$ and $k \in \mathbb{N}$ is the following: Given a set of $n \in \mathbb{N}$ input curves $T = \{\tau_1, \ldots, \tau_n\} \subset \mathbb{X}^d_m$, identify k center curves $C = \{c_1, \ldots, c_k\} \subset \mathbb{X}^d_\ell$ that minimize $\mathrm{cost}(T, C) = \sum_{\tau \in T} \min_{c \in C} \mathrm{dtw}(\tau, c)$.*

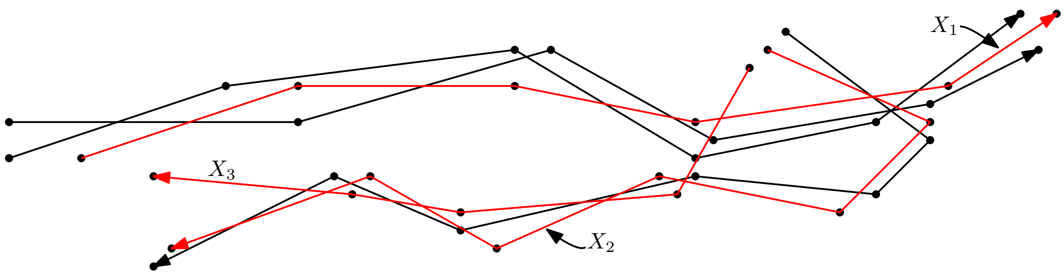

Figure 2 Illustration of a coreset (red), i.e. a weighted sparse representation of the original set of curves (in red and black). The weights in this case are $w(X_1) = 3$, $w(X_2) = 2$ and $w(X_3) = 1$.

An influential approach to solving k-median problems is to construct a point set that acts as proxy on which to run computationally more expensive algorithms that yield solutions with approximation guarantees. The condensed input set is known as a coreset.

▶ **Definition 3** (ε-coreset)**.** *Let $T \subset \mathbb{X}^d_m$ be a finite set and $\varepsilon \in (0,1)$. Then a weighted multiset $S \subset \mathbb{X}^d_m$ with weight function $w : S \to \mathbb{R}_{>0}$ is a weighted ε-coreset for (k,ℓ)-median clustering of T under* $\mathrm{dtw_p}$ *if for all $C \subset \mathbb{X}^d_\ell$ with $|C| = k$*

$$(1-\varepsilon)\,\mathrm{cost}(T,C) \le \sum_{s\in S} w(s) \min_{c\in C} \mathrm{dtw_p}(s,c) \le (1+\varepsilon)\,\mathrm{cost}(T,C).$$

▶ **Definition 4** ((α,β)-approximation)**.** *Let a set of $n \in \mathbb{N}$ input curves $T = \{\tau_1,\ldots,\tau_n\} \subset \mathbb{X}^d_m$ be given. A set $\hat{C} \subset \mathbb{X}^d_\ell$ is called an (α,β)-approximation of (k,ℓ)-median, if $|\hat{C}| \le \beta k$ and $\sum_{\tau\in T} \min_{c\in\hat{C}} \mathrm{dtw}(\tau,c) \le \alpha \sum_{\tau\in T} \min_{c\in C} \mathrm{dtw}(\tau,c)$ for any $C \subset \mathbb{X}^d_\ell$ of size k.*

Relaxing the problem to (α,β)-approximations allows us to pass through so called simplifications of the input curves.

▶ **Definition 5** ($(1+\varepsilon)$-approximate ℓ-simplifications)**.** *Let $\sigma \in \mathbb{X}^d_m$, $\ell \in \mathbb{N}$ and $\varepsilon > 0$. We call $\sigma^* \in \mathbb{X}^d_\ell$ an* $(1+\varepsilon)$-approximate ℓ-simplification *if*

$$\inf_{\sigma_\ell\in\mathbb{X}^d_\ell} \mathrm{dtw_p}(\sigma_\ell,\sigma) \le \mathrm{dtw_p}(\sigma^*,\sigma) \le (1+\varepsilon) \inf_{\sigma_\ell\in\mathbb{X}^d_\ell} \mathrm{dtw_p}(\sigma_\ell,\sigma).$$

A *range space* is defined as a pair of sets $(X,\mathcal{R})$, where X is the *ground set* and $\mathcal{R}$ is the *range set* which is a natural subset of the power set $\mathcal{P}(X) = \{X' | X' \subset X\}$. Let $(X,\mathcal{R})$ be a range space. For $Y \subseteq X$, we denote: $\mathcal{R}_{|Y} = \{R \cap Y \mid R \in \mathcal{R}\}$. If $\mathcal{R}_{|Y} = \mathcal{P}(Y)$, then Y is *shattered* by $\mathcal{R}$. A key property of range spaces is the so called Vapnik-Chernovenkis dimension [35, 37, 39] (VC dimension) which for a range space $(X,\mathcal{R})$ is the maximum cardinality of a shattered subset of X.

We are interested in range spaces defined by balls by the p-DTW distance: We define the (p-)DTW ball, of given complexity $m \in \mathbb{N}$ and radius $r \ge 0$, of a curve $\sigma \in \mathbb{X}^d_\ell$ as $B^p_{r,m}(\sigma) = \{\tau \in \mathbb{X}^d_m \mid \mathrm{dtw_p}(\sigma,\tau) \le r\}$. Define the range set of p-DTW balls as $\mathcal{R}^p_{m,\ell} = \{B^p_{r,m}(\sigma) \mid \sigma \in \mathbb{X}^d_\ell, r > 0\}$. The *$p$-DTW range space* is the range space $\mathcal{X}_{m,\ell} = \left(\mathbb{X}^d_m, \mathcal{R}^p_{m,\ell}\right)$.

3 VC Dimension of DTW

In this section, we derive bounds on the VC dimension of a range space that approximates the DTW range space. Our reasoning exclusively relies on establishing the prerequisites of Theorem 7 below. Missing proofs of this section can be found in the full version.

▶ **Definition 6** ([3]). *Let H be a class of $\{0,1\}$-valued functions defined on a set X, and F a class of real-valued functions defined on $\mathbb{R}^d \times X$. We say that H is a k-combination of* $\operatorname{sign}(F)$ *if there is a function $g : \{-1,1\}^k \to \{0,1\}$ and functions $f_1,\ldots,f_k \in F$ so that for all $h \in H$ there is a parameter vector $\alpha \in \mathbb{R}^d$ such that for all x in X,*

$$h(x) = g(\operatorname{sign}(f_1(\alpha,x)),\ldots,\operatorname{sign}(f_k(\alpha,x))).$$

The definition for the sign function we use is that $\operatorname{sign}(x) = 1$ for $\mathbb{R} \ni x \geq 0$ and $\operatorname{sign}(x) = -1$ for $x < 0$. Observe that the class H of functions corresponds to a system of subsets of X.

▶ **Theorem 7** (Theorem 8.3 [3]). *Let F be a class of maps from $\mathbb{R}^s \times X$ to $\mathbb{R}$, so that for all $x \in X$ and $f \in F$, the function $\alpha \mapsto f(\alpha,x)$ is a polynomial on $\mathbb{R}^s$ of degree δ. Let H be a κ-combination of* $\operatorname{sign}(F)$. *Then the VC dimension of H is less than $2s\log_2(12\delta\kappa)$.*

Theorem 7 implies a bound of $O(d\ell^2 \log(mp))$ on the VC dimension of range spaces defined by p-DTW for even values of p, as the decision of whether p-DTW exceeds a given threshold can be formulated as a $|\mathcal{T}_{m,\ell}|$-combination of signs of polynomial functions; each one realizing the cost of a traversal. A detailed proof can be found in the full version. The situation becomes more intriguing in the general case, since for any odd p, the cost of each traversal is no longer a polynomial. To overcome this, we investigate range spaces defined by approximate p-DTW balls and show that we can get bounds that do not depend on p.

The following lemma shows that one can determine (approximately) the p-DTW between two sequences, based solely on the signs of certain polynomials, that are designed to provide a sketchy view of all point-wise distances.

▶ **Lemma 8.** *Let $\tau \in \mathbb{X}^d_{=\ell}$, $\sigma \in \mathbb{X}^d_{=m}$, $r > 0$ and $\varepsilon \in (0,1]$. For each $i \in [\ell]$, $j \in [m]$ and $z \in [\lfloor \varepsilon^{-1} + 1 \rfloor]$ define*

$$f_{i,j,z}(\tau,r,\sigma) = \|\tau_i - \sigma_j\|^2 - (z \cdot \varepsilon r)^2.$$

There is an algorithm that, given as input the values of $\operatorname{sign}(f_{i,j,z}(\tau,r,\sigma))$*, for all $i \in [\ell], j \in [m]$ and $z \in [\lfloor \varepsilon^{-1} + 1 \rfloor]$, outputs a value in $\{0,1\}$ such that:*

- *if* $\mathrm{dtw_p}(\tau,\sigma) \leq r$ *then it outputs* 1,
- *if* $\mathrm{dtw_p}(\tau,\sigma) > (1 + (m+\ell)^{1/p}\varepsilon)r$ *then it outputs* 0.

The algorithm of Lemma 8 essentially defines a function that implements approximate p-DTW balls membership, and satisfies the requirements set by Theorem 7.

▶ **Lemma 9.** *Let $\varepsilon \in (0,1]$, and let $m,\ell \in \mathbb{N}$ be given. There are injective functions $\pi_\ell : \mathbb{X}^d_\ell \to \mathbb{R}^{(d+1)\ell}$ and $\pi_m : \mathbb{X}^d_m \to \mathbb{R}^{(d+1)m}$ and a class of functions F_ε mapping from $\left(\mathbb{R}^{(d+1)\ell} \times \mathbb{R}\right) \times \mathbb{R}^{(d+1)m}$ to $\mathbb{R}$, such that for any $f \in F_\varepsilon$, the function $\alpha \mapsto f(\alpha,x)$ is a polynomial function of degree 2. Furthermore, there is a function $g \colon \{-1,1\}^k \to \{0,1\}$ and functions $f_1,\ldots,f_k \in F_\varepsilon$, with $k = m\ell\lfloor \varepsilon^{-1} + 1 \rfloor + m + \ell$, such that for any $\tau \in \mathbb{X}^d_\ell$, $r > 0$ and $\sigma \in \mathbb{X}^d_m$, it holds that if* $\mathrm{dtw_p}(\sigma,\tau) \leq r$ *then* $g(\operatorname{sign}(f_1(\pi_\ell(\tau),r,\pi_m(\sigma))),\ldots) = 1$*, and if* $\mathrm{dtw_p}(\sigma,\tau) > (1 + (m+\ell)^{1/p}\varepsilon)r$ *then* $g(\operatorname{sign}(f_1(\pi_\ell(\tau),r,\pi_m(\sigma))),\ldots) = 0$.

We use the previous lemmas to define a distance function $\widetilde{\mathrm{dtw}}_{\mathrm{p}}$ between elements of $\mathbb{X}^d_m$ and $\mathbb{X}^d_\ell$, which we will use throughout the paper as an approximate function of $\mathrm{dtw_p}$. To get an estimate of the VC dimension of the range space induced by balls under $\widetilde{\mathrm{dtw}}_{\mathrm{p}}$ and decide membership of points to these balls, the approximate distance will only take discrete values.

▶ **Definition 10.** *Let $\varepsilon \in (0,1]$ and define the set of radii $R_\varepsilon = \{(1+\varepsilon)^z \mid z \in \mathbb{Z}\}$. Lemma 9 defines an approximation of $\mathrm{dtw_p}(\sigma,\tau)$ for any $\sigma \in \mathbb{X}^d_\ell$ and $\mathbb{X}^d_m$, by virtue of the functions g and $f_1,...,f_k$ for $F_{\varepsilon/(m+\ell)^{1/p}}$, as*

$$\widetilde{\mathrm{dtw_p}}(\sigma,\tau) = (1+\varepsilon)\cdot \sup\{r \in R_\varepsilon \mid g(\mathrm{sign}(f_1(\pi_\ell(\tau), r, \pi_m(\sigma))),\ldots) = 1\}.$$

Overall, $\widetilde{\mathrm{dtw_p}}$ corresponds to the first value in R_ε, for which the function g of Definition 10 outputs a 0, and for all larger values in R_ε the algorithm also outputs 0. Notably, the function g of Definition 10 outputs 1 for $r/(1+\varepsilon)$. In the following lemma, we formally show that $\widetilde{\mathrm{dtw_p}}(\sigma,\tau)$ approximates p-DTW between σ and τ within a factor of $1+\varepsilon$.

▶ **Lemma 11.** *Let $0 < \varepsilon \leq 1$. For any $\sigma \in \mathbb{X}^d_m$ and $\tau \in \mathbb{X}^d_\ell$ it holds that*

$$\mathrm{dtw_p}(\sigma,\tau) < \widetilde{\mathrm{dtw_p}}(\sigma,\tau) \leq (1+\varepsilon)\,\mathrm{dtw_p}(\sigma,\tau).$$

Proof. Let $r = \widetilde{\mathrm{dtw_p}}(\sigma,\tau) \in R_\varepsilon$. By definition the function g of Definition 10 outputs 1 with σ, τ and $r/(1+\varepsilon)$. Thus $\mathrm{dtw_p}(\sigma,\tau) \leq (1+\varepsilon)r/(1+\varepsilon) = r$. As the algorithm outputs 0 for σ, τ and r it follows that $\mathrm{dtw_p}(\sigma,\tau) > r/(1+\varepsilon)$ implying the claim. ◀

Moreover from the definition of $\widetilde{\mathrm{dtw_p}}$, we conclude that g serves as a membership predicate for balls defined by $\widetilde{\mathrm{dtw_p}}$.

▶ **Lemma 12.** *Let $\varepsilon \in (0,1]$, $\tau \in \mathbb{X}^d_\ell$ and $r \in R_\varepsilon$. For any $\sigma \in \mathbb{X}^d_m$ the output of the function g of Definition 10 with σ, τ and r corresponds to the decision whether the curve σ is in the r-ball $\{x \in \mathbb{X}^d_m \mid \widetilde{\mathrm{dtw_p}}(x,\tau) \leq r\}$ centered at τ.*

Proof. Let $r' = \widetilde{\mathrm{dtw_p}}(\sigma,\tau) \in R_\varepsilon$. Assume $r' \leq r$ which by Lemma 11 implies that $\mathrm{dtw_p}(\sigma,\tau) \leq r$. Then the function g of Definition 10 with σ, τ and r outputs 1. Now let $r < r' \in R_\varepsilon$. In this case however g with σ, τ and r will by definition of $\widetilde{\mathrm{dtw_p}}$ output 0. Thus membership to a ball range corresponds to the output of the function g of Definition 10. ◀

We conclude with the main result of this section, namely an upper bound on the VC dimension of the range space that approximates the p-DTW range space.

▶ **Theorem 13.** *Let $\varepsilon \in (0,1]$ and $\widetilde{\mathcal{R}}^p_{m,\ell} = \{\{x \in \mathbb{X}^d_m \mid \widetilde{\mathrm{dtw_p}}(x,\tau) \leq r\} \subset \mathbb{X}^d_m \mid \tau \in \mathbb{X}^d_\ell, r > 0\}$ be the range set consisting of all balls centered at elements of $\mathbb{X}^d_\ell$ under $\widetilde{\mathrm{dtw_p}}$ in $\mathbb{X}^d_m$. The VC dimension of $(\mathbb{X}^d_m, \widetilde{\mathcal{R}}^p_{m,\ell})$ is at most*

$$2(d+1)\ell \log_2(12\ell m\lfloor (m+\ell)^{1/p}\varepsilon^{-1} + 1\rfloor + 12m + 12\ell) = O(d\ell\log(\ell m\varepsilon^{-1})).$$

Proof. This follows from Theorem 7, Lemma 9 and Lemma 12, and the fact that any ball of radius $r > 0$ under $\widetilde{\mathrm{dtw_p}}$ coincides with some ball with radius $\widetilde{r} \in R_\varepsilon$ under $\widetilde{\mathrm{dtw_p}}$. Finally, the statement is implied by the injectivity of the functions π_m and π_ℓ. ◀

In this section, we defined a distance function $\widetilde{\mathrm{dtw_p}}$ between curves in $\mathbb{X}^d_m$ and those in $\mathbb{X}^d_\ell$ that $(1+\varepsilon)$-approximates $\mathrm{dtw_p}$ and an upper bound on the VC dimension of the range space induced by balls of $\widetilde{\mathrm{dtw_p}}$, thereby producing an approximation of the p-DTW range space that we make use of below. We emphasize that the sole purpose of $\widetilde{\mathrm{dtw_p}}$ is to obtain bounds on the size of a sample constituting a coreset through the knowledge of the VC dimension. At no point do we compute $\widetilde{\mathrm{dtw_p}}$.

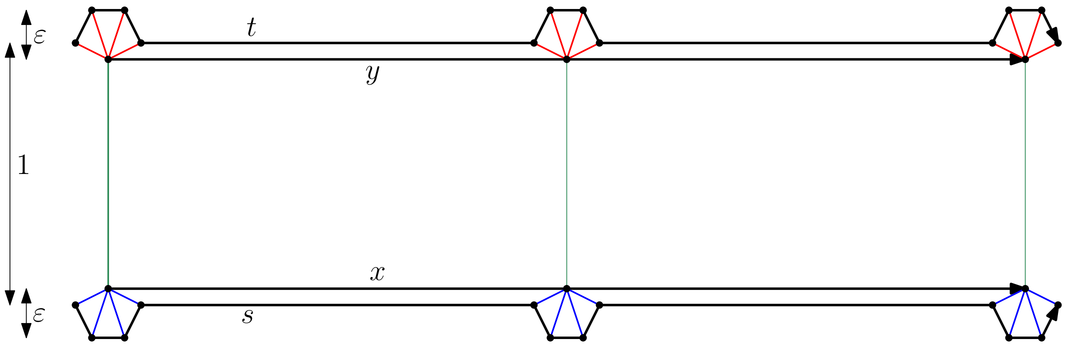

Figure 3 Violated triangle inequality as $\mathrm{dtw}(s,t) \approx 12$, but $\mathrm{dtw}(s,x) \approx 0$ (matching in blue), $\mathrm{dtw}(y,t) \approx 0$ (red matching) and $\mathrm{dtw}(x,y) \approx 3$ (green matching).

4 Sensitivity bounds and coresets for DTW

The proofs in this section are deferred to the full version. To make use of the sensitivity sampling framework for coresets by Feldman and Langberg [22], we recast the input set $T \subset \mathbb{X}^d_m$ as a set of functions. Consider for any $y \in \mathbb{X}^d_m$ the real valued function f_y defined on (finite) subsets of $\mathbb{X}^d_\ell$ by $f_y(C) = \min_{c \in C} \mathrm{dtw}_\mathrm{p}(y,c)$ for $C \subset \mathbb{X}^d_\ell$, transforming T into $F_T = \{f_\tau \mid \tau \in T\}$. To construct a coreset, one draws elements from T according to a fixed probability distribution over T, and reweighs each drawn element. Both the weight and sampling probability are expressed in terms of the *sensitivity* of the drawn element t, which describes the maximum possible relative contribution of t to the cost of any query evaluation. In our case, as we restrict a solution to a size of k, it turns out that it suffices to analyze the sensitivity with respect to inputs of size k.

▶ **Definition 14** (sensitivity). *Let F be a finite set of functions from $\mathcal{P}(\mathbb{X}^d_\ell) \setminus \{\emptyset\}$ to $\mathbb{R}$. For any $f \in F$ define the sensitivity*

$$\mathfrak{s}(f,F) = \sup_{C=\{c_1,\dots,c_k\} \subset Z : \sum_{g \in F} g(C) > 0} \frac{f(C)}{\sum_{g \in F} g(C)}.$$

The total sensitivity $\mathfrak{S}(F)$ of F is defined as $\sum_{f \in F} \mathfrak{s}(f,F)$.

A crucial step in our approach is to show that any (α,β)-approximation for (k,ℓ)-median under dtw_p can be used to obtain a bound on the total sensitivity associated to approximate distances. This is facilitated by the following lemma, that is a weaker version of the triangle inequality, as in general dtw_p is not a metric (see Figure 3).

▶ **Lemma 15** (weak triangle inequality [29]). *For two curves x and z of complexity $m > 0$ and any curve y of complexity $\ell > 0$ it holds that* $\mathrm{dtw}_\mathrm{p}(x,z) \le m^{1/p}(\mathrm{dtw}_\mathrm{p}(x,y) + \mathrm{dtw}_\mathrm{p}(y,z))$.

Note that the distance function in question is not dtw_p, but the $(1+\varepsilon)$-approximation $\widetilde{\mathrm{dtw}}_\mathrm{p}$ of DTW from before. For any $y \in \mathbb{X}^d_m$ and $\varepsilon > 0$, let $\widetilde{f}_y : \mathcal{P}(\mathbb{X}^d_\ell) \setminus \{\emptyset\} \to \mathbb{R}$ with $\widetilde{f}_y(C) = \min_{c \in C} \widetilde{\mathrm{dtw}}_\mathrm{p}(y,c)$. Similarly, let $\widetilde{F}_T$ be the set $\{\widetilde{f}_\tau \mid \tau \in T\}$ for any $T \subset \mathbb{X}^d_m$.

▶ **Lemma 16.** *Let $0 < \varepsilon \le 1$ and let $T \subset \mathbb{X}^d_m$ be the input of size n for (k,ℓ)-median and let $\hat{C} = \{\hat{c}_1,\dots,\hat{c}_{\hat{k}}\} \subset \mathbb{X}^d_\ell$ be an (α,β)-approximation to the (k,ℓ)-median problem on T with cost $\hat{\Delta} = \sum_{\tau \in T} \min_{\hat{c} \in \hat{C}} \mathrm{dtw}_\mathrm{p}(\tau,\hat{c})$, of size $\hat{k} \le \beta k$. For any $i \in [\hat{k}]$ let $\hat{V}_i = \{\tau \in T \mid$*

$\mathrm{dtw_p}(\tau, \hat{c}_i) = \min_{\hat{c}\in\hat{C}} \mathrm{dtw_p}(\tau, \hat{c})\}$ *be the Voronoi region of* $\hat{c}_i$*, the set of which (breaking ties arbitrarily) partitions* T*. Let* $\hat{\Delta}_i = \sum_{\tau\in\hat{V}_i} \mathrm{dtw_p}(\tau, \hat{c}_i)$ *be the cost of* $\hat{V}_i$*. For all* $\tau \in \hat{V}_i$ *let*

$$\gamma(\widetilde{f}_\tau) := (m\ell)^{1/p}\left(\frac{2\alpha\,\mathrm{dtw_p}(\tau,\hat{c}_i)}{\hat{\Delta}} + \frac{4}{|\hat{V}_i|} + \frac{8\alpha\hat{\Delta}_i}{\hat{\Delta}|\hat{V}_i|}\right).$$

Then $\mathfrak{s}(\widetilde{f}_\tau, \widetilde{F}_T) \le \gamma(\widetilde{f}_\tau)$ *for any* $\tau \in T$*, and* $\mathfrak{S}(\widetilde{F}_T) \le \sum_{\tau\in T}\gamma(\widetilde{f}_\tau) \le (m\ell)^{1/p}(4\hat{k} + 10\alpha)$*.*

▶ **Theorem 17.** *For* $\widetilde{f} \in \widetilde{F}$*, let* $\lambda(\widetilde{f}) = 2^{\lceil \log(\gamma(\widetilde{f}))\rceil}$*, with* $\gamma(\widetilde{f})$ *the sensitivity bound of Lemma 16, associated to an* α*-approximation consisting of* $\hat{k}$ *curves, for* (k,ℓ)*-median for curves in* $\mathbb{X}^d_m$ *under* $\mathrm{dtw_p}$*,* $\Lambda = \sum_{\widetilde{f}\in\widetilde{F}}\lambda(\widetilde{f})$*,* $\psi(\widetilde{f}) = \frac{\lambda(\widetilde{f})}{\Lambda}$ *and* $\delta, \varepsilon \in (0,1)$*. A sample* S *of*

$$\Theta\Big(\varepsilon^{-2}\alpha\hat{k}(m\ell)^{1/p}\left((d\ell\log(\ell m\varepsilon^{-1}))k\log(k)\log(\alpha n)\log(\alpha\hat{k}(m\ell)^{1/p}) + \log(1/\delta)\right)\Big)$$

elements $\tau_i \in T$*, drawn independently with replacement with probability* $\psi(\widetilde{f}_i)$ *and weighted by* $w(\widetilde{f}_i) = \frac{\Lambda}{|S|\lambda(\widetilde{f}_i)}$ *is a weighted* ε*-coreset for* (k,ℓ)*-median clustering of* T *under* $\mathrm{dtw_p}$ *with probability at least* $1-\delta$*.*

We remark that in the limit $p \to \infty$, the constructed coreset has a very similar size as a recent construction for coresets for the Fréchet distance [17].

5 Linear time $(O((m\ell)^{1/p}), 1)$-approximation algorithm for (k,ℓ)-median

In this section, we develop approximation algorithms for (k,ℓ)-median for a set $T \subset \mathbb{X}^d_m$ of n curves. For this, we approximate DTW on T by a metric using a new inequality for DTW (Lemma 18). This allows the use of any approximation algorithm for k-median in metric spaces, leading to a first approximation algorithm of the original problem. However, computing the whole metric space would take $O(n^3)$ time. We circumvent this by in turn using the DTW distance to approximate the metric space. Combined with a k-median algorithm in metric spaces [24], we obtain a linear time $(O((m\ell)^{1/p}), 1)$-approximation algorithm.

5.1 Dynamic time warping approximating metric

We begin with the following more general triangle inequality for $\mathrm{dtw_p}$, which motivates analysing the metric closure of the input set. While $\mathrm{dtw_p}$ does not satisfy the triangle inequality (see Figure 3), the inequality shows it is never "too far off". Remarkably, the inequality does not depend on the complexity of the curves "visited". The missing proofs in this section are deferred to the full version. Lemma 18 is illustrated in Figure 4.

▶ **Lemma 18** (Iterated triangle inequality)**.** *Let* $s \in \mathbb{X}^d_\ell$*,* $t \in \mathbb{X}^d_{\ell'}$ *and* $X = (x_1, \dots, x_r)$ *be an arbitrary ordered set of curves in* $\mathbb{X}^d_m$*. Then*

$$\mathrm{dtw_p}(s,t) \le (\ell+\ell')^{1/p}\left(\mathrm{dtw_p}(s,x_1) + \sum_{i<r}\mathrm{dtw_p}(x_i,x_{i+1}) + \mathrm{dtw_p}(x_r,t)\right).$$

▶ **Definition 19** (metric closure)**.** *Let* (X,ϕ) *be a finite set endowed with a distance function* $\phi : X \times X \to \mathbb{R}$*. The metric closure* $\overline{\phi}$ *of* ϕ *is the function*

$$\overline{\phi} : X \times X \to \mathbb{R}, (s,t) \mapsto \min_{\substack{r\ge 2, \{\tau_1,\dots,\tau_r\}\subset X \\ s=\tau_1, t=\tau_r}} \sum_{i<r}\phi(\tau_i,\tau_{i+1}).$$

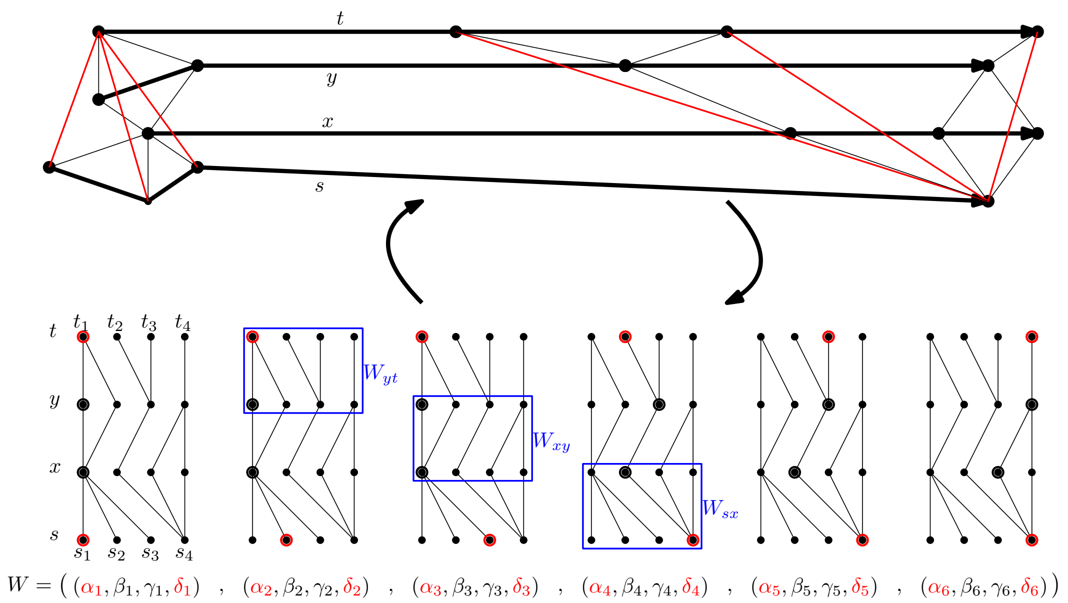

Figure 4 Illustration of how the optimal traversals W_{sx}, W_{xy} and W_{yt} of visited curves can be "composed" to yield a set W that induces a traversal $\widetilde{W}$ (in red) of s and t. Any single matched pair of vertices in W_{sx}, W_{xy} or W_{yt} is at most $|W| \leq \ell + \ell'$ times a part of W.

The metric closure of any distance function is a semimetric and can be extended to a metric by removing duplicates or small (symbolic) perturbations. Note that the metric closure of $\mathrm{dtw_p}$ can be strictly smaller than $\mathrm{dtw_p}$ because DTW may violate the triangle inequality (see Figure 3).

▶ **Observation 20.** *Let X be a finite set with distance function ϕ. Let $Y \subset X$. Then for any $\sigma, \tau \in Y$ it holds that $\overline{\phi}(\sigma, \tau) \leq \overline{\phi|_Y}(\sigma, \tau) \leq \phi(\sigma, \tau)$.*

By Lemma 18 and Observation 20, $\mathrm{dtw_p}$ on any finite set of curves in $\mathbb{X}^d_m$ is approximated by its metric closure, with approximation constant depending on m.

▶ **Lemma 21.** *For any set of curves X and two curves $\sigma, \tau \in X$ of complexity at most m it holds that $\mathrm{dtw_p}(\sigma, \tau) \leq (2m)^{1/p}\overline{\mathrm{dtw_p}\,|_X}(\sigma, \tau) \leq (2m)^{1/p}\,\mathrm{dtw_p}(\sigma, \tau)$.*

▶ **Lemma 22.** *Let $X \subset \mathbb{X}^d_m$ be a set of n curves and k and ℓ be given. Let $X^* = \{\tau^* \mid \tau \in X\}$, where τ^* is a $(1+\varepsilon)$-approximate ℓ-simplification of τ. Let $C \subset X^*$ be an (α, β)-approximation of the k-median problem of X^* in the metric space $(X^*, \overline{\mathrm{dtw_p}\,|_{X^*}})$. Then C is a $\left((4m\ell)^{1/p}\left((4+2\varepsilon)\alpha + 1 + \varepsilon\right), \beta\right)$-approximation of the (k, ℓ)-median problem on X.*

The idea of Lemma 22 is to consider a Voronoi decomposition of X induced by an optimal solution, together with standard arguments using the triangle inequality in the metric closure to bound distances, as illustrated in Figure 6.

▶ **Lemma 23.** *Let $X \subset \mathbb{X}^d_\ell$ be a set of n curves. The metric closure $\overline{\mathrm{dtw_p}\,|_X}$ for all pairs of curves in X can be computed in $O(n^2\ell^2 d + n^3)$ time.*

▶ **Theorem 24** ([19])**.** *Given a set P of n points in a metric space, for $0 < \varepsilon < 1$, one can compute a $(10+\varepsilon)$-approximate k-median clustering of P in $O(nk + k^7\varepsilon^{-5}\log^5 n)$ time, with constant probability of success.*

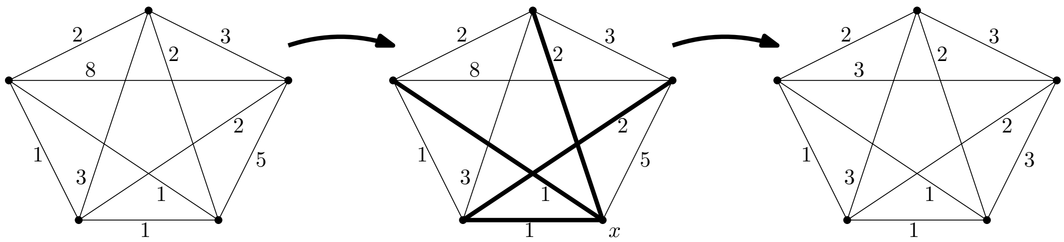

Figure 5 Illustration of the metric closure. On the left a distance function on five points represented as a graph. In the middle the shortest path tree rooted at x inducing all values of the metric closure of the distance function from some element to x. On the right the metric closure.

▶ **Theorem 25.** *Let X be a set of curves of complexity at most m. Let k and ℓ be given. Let $X^* = \{\tau^* \mid \tau \in X\}$ be a set of $(1+\varepsilon)$-approximate optimal ℓ-simplifications. There is an algorithm with input X^*, which computes a $(10+\varepsilon, 1)$-approximation to the k-median problem of X^* in $(X^*, \overline{\mathrm{dtw_p}|_{X^*}})$ in $O(n^2\ell^2 d + n^3 + nk + k^7\varepsilon^{-5}\log^5 n)$ time.*

Proof. This is a direct consequence of Lemma 23 and Theorem 24. ◀

We next show how to combine our ideas with Indyk's sampling technique for bicriteria k-median approximation [24] to achieve linear dependence on n.

5.2 Linear time algorithm

With Theorem 25 we have ran into the following predicament: We would like to apply linear time algorithms to the metric closure of $\mathrm{dtw_p}$. However, constructing the metric closure takes cubic time. We circumvent this by applying the following algorithm, which reduces a k-median instance with n points to two k-median instances with $O(\sqrt{n})$ points, simply by sampling. More precisely, we will apply this technique twice, so that we will compute the metric closure only on sampled subsets of size $O(n^{1/4})$. In this section we want to analyse the problem of computing a k-median of a set X in the metric space $(X, \overline{\phi})$, where ϕ is a distance function on X with the guarantee that there is a constant ζ such that for any $x, y \in X$ it holds that $\phi(x,y) \leq \zeta\overline{\phi}(x,y)$, with a linear running time, and more precisely, only a linear number of calls to the distance function ϕ, and no calls to $\overline{\phi}$. By Lemma 21 the results in this section translate directly to $\phi = \mathrm{dtw_p}|_X$ with $\zeta = (m+\ell)^{1/p}$.

Observe, that similar to Theorem 25, the following lemma holds.

▶ **Lemma 26.** *Let X be a set of n points, equipped with a distance function ϕ that can be computed in time T_ϕ. There is a $(10+\varepsilon, 1)$-approximate algorithm for k-median of X in $(X, \overline{\phi})$ that has constant probability of success and has running time $O(n^2 T_\phi + n^3 + nk + k^7\varepsilon^{-5}\log^5 n)$.*

▶ **Lemma 27.** *Let X be a set of n points, equipped with a distance function ϕ, such that $\phi \leq \zeta\overline{\phi}$ for some $\zeta > 0$, and $Y \subset X$. A (α, β)-approximation for the k-median problem for Y in $(Y, \overline{\phi|_Y})$ is a $(\alpha\zeta, \beta)$-approximation for the k-median problem for Y in $(Y, \overline{\phi}|_Y)$.*

Lemma 27 is a straightforward consequence of Observation 20, as thereby any good solution in the metric closure of the restriction of ϕ onto Y incurs at most a multiplicative loss of ζ with respect to the actual metric closure of ϕ.

▶ **Theorem 28** ([24])**.** *Let $\mathcal{A}$ be a (α, β)-approximate algorithm for k-median in metric spaces with constant success probability. Then for any $\varepsilon > 0$ the k-Routine in Algorithm 1 provided with $\mathcal{A}$ is a $(3(1+\varepsilon)(2+\alpha), 2\beta)$-approximate algorithm for k-median in metric spaces with constant success probability.*

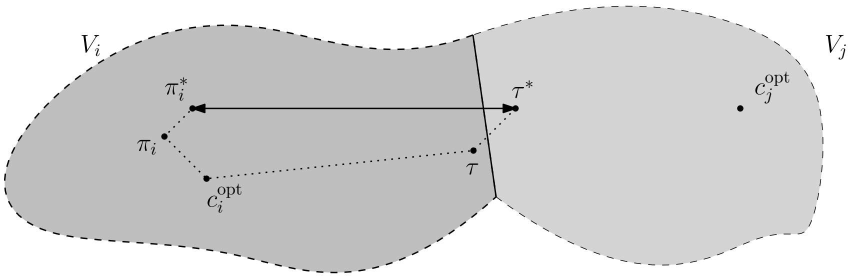

Figure 6 Illustration to Proof of Lemma 22: Assigning τ^* (the $(1+\varepsilon)$-simplification of τ which lies inside the Voronoi cell V_i of c_i^{opt}) to π_i^* (the $(1+\varepsilon)$-simplification of the closest element π_i in V_i to c_i^{opt}) under $\overline{\text{dtw}_p}|_{X^*}$ is at most $4+2\varepsilon$ times as bad as assigning τ to c_i^{opt} under dtw_p.

Algorithm 1 k-median framework.

---

procedure k-ROUTINE$((X,\phi),\varepsilon,\mathcal{A})$ ▷ $\mathcal{A}$ is (α,β)-approximate metric k-median
  $a \leftarrow \Theta(\varepsilon^{-1}\sqrt{\log(\varepsilon^{-1})})$, $b \leftarrow \Theta(a^2)$ ▷ Determine the success probabilities.
  $s \leftarrow a\sqrt{kn\log k}$
  Choose a set S of s points sampled without replacement from X
  $C' \leftarrow \mathcal{A}((S,\overline{\phi|_S}))$
  Select the set M of points x with the $b\frac{kn\log k}{s}$ largest values of $\min_{c'\in C'}\overline{\phi}(x,c')$
  return $C = C' \cup \mathcal{A}((M,\overline{\phi|_M}))$
end procedure
procedure k-MEDIAN$((X,\phi),\varepsilon,\mathcal{A})$ ▷ $\mathcal{A}$ is (α,β)-approximate metric k-median
  $a \leftarrow \Theta(\varepsilon^{-1}\sqrt{\log(\varepsilon^{-1})})$, $b \leftarrow \Theta(a^2)$ ▷ Determine the success probabilities.
  $s \leftarrow a\sqrt{kn\log k}$
  Choose a set S of s points sampled without replacement from X
  $C' \leftarrow k$-ROUTINE$((S,\phi|_S),\varepsilon,\mathcal{A})$
  Select the set M of points x with the $b\frac{kn\log k}{s}$ largest values of $\min_{c'\in C'}\phi(x,c')$
  return $C = C' \cup k$-ROUTINE$((M,\phi|_M),\varepsilon,\mathcal{A})$
end procedure

▶ **Lemma 29.** *Let X be a set of n points, and let ϕ be a distance function that can be computed in T_ϕ time for any $x,y \in X$. Let $T_{\mathcal{A}}(n)$ be the running time of the (α,β)-approximate algorithm for k-median on n elements. Then k-ROUTINE has a running time of $O(n^2T_\phi + T_{\mathcal{A}}(\min(n,\varepsilon^{-1}\sqrt{kn\log(k)\log(\varepsilon^{-1})})))$.*

▶ **Lemma 30.** *Let X be a set of n points, and let ϕ be a distance function on X, which can be computed in time T_ϕ, and further there is a constant ζ such that $\phi \le \zeta\overline{\phi}$. Let $Y \subset X$. Let $\varepsilon > 0$ and let $\mathcal{A}$ be the $(10+\varepsilon,1)$-approximation for metric k-median of Lemma 26. Then k-ROUTINE returns a $(3(1+\varepsilon)\zeta(12+\varepsilon),2)$-approximation of k-median in the metric space $(Y,\overline{\phi}|_Y)$ in time $O(|Y|^2T_\phi + |Y|^2k\log(k)\varepsilon^{-2}\log(\varepsilon^{-1}) + k^7\varepsilon^{-5}\log^5(|Y|))$.*

Proof. The running time bound follows by Lemma 29 and Lemma 26, together with the fact, that $\min(|Y|,\varepsilon^{-1}\sqrt{k|Y|\log(k)\log(\varepsilon^{-1})})^3 \le |Y|^2k\log(k)\varepsilon^{-2}\log(\varepsilon^{-1})$. The approximation guarantee follows by Theorem 28, Lemma 27 and Lemma 26. ◀

By combining the presented subroutines, we obtain our two main results of the section. The first is Theorem 31, which provides a linear time approximation algorithm for k-median in metric closures, assuming the underlying distance is reasonably well approximated by its metric closure. The second is Corollary 33, combining Theorem 31 with Lemma 22 to yield an approximation algorithm for p-DTW with an unoptimized approximation guarantee.

▶ **Theorem 31.** *Let X be a set of points and let ϕ be a distance function on X with $\phi \leq \zeta\overline{\phi}$. Let $\varepsilon > 0$ and let $\mathcal{A}$ be the $(10+\varepsilon, 1)$-approximation for metric k-median of Theorem 25. Then k-MEDIAN returns a $(11\zeta^2(1+\varepsilon)^2(12+\varepsilon), 4)$-approximation of k-median of X in the metric space $(X, \overline{\phi})$ in time $O(nk\log(k)T_\phi + nk^2\log^2 k + k^7\varepsilon^{-5}\log^5(n))$.*

We briefly discuss simplification schemes for curves under p-DTW (for more details refer to the full version). We reduce the problem of finding an $(1+\varepsilon)$-approximate simplification to finding a $(1+\varepsilon)$-approximation of a center point for a set of $\leq m$ points, where the objective is to minimize the sum of the individual distances to the center point raised to the pth power. Note that for $p = \infty$, the problem is that of finding a minimum enclosing ball, and for $p = 2$, the problem can be reduced to that of finding the center of gravity of the set of discrete points, which can both be solved exactly. Furthermore, we show (Proposition 32) that for all $\mathrm{dtw_p}$, there is a deterministic 2-approximation that is a crucial ingredient for our approximation algorithms of (k, ℓ)-median under p-DTW.

▶ **Proposition 32.** *For $\sigma = (\sigma_1, \ldots, \sigma_m) \in \mathbb{X}^d_m$ and integer $\ell > 0$, one can compute in $O(m^2(d + \ell + m))$ time a curve $\sigma^* \in \mathbb{X}^d_\ell$ such that*

$$\inf_{\sigma_\ell \in \mathbb{X}^d_\ell} \mathrm{dtw_p}(\sigma_\ell, \sigma) \leq \mathrm{dtw_p}(\sigma^*, \sigma) \leq 2 \inf_{\sigma_\ell \in \mathbb{X}^d_\ell} \mathrm{dtw_p}(\sigma_\ell, \sigma).$$

▶ **Corollary 33.** *For any $\varepsilon > 0$ the procedure k-MEDIAN from Algorithm 1 can be used to compute a $(72(1+\varepsilon)^2(12+\varepsilon)(16m\ell^3)^{1/p}, 4)$-approximation for (k, ℓ)-median for an input set X of n curves of complexity m under $\mathrm{dtw_p}$ in time $O(nm^3d + nk\log(k)\ell^2 d + nk^2\log^2(k)\varepsilon^{-4}\log^2(\varepsilon^{-1}) + k^7\varepsilon^{-5}\log^5(n))$.*

Proof. Let $X^* = \{\tau^* \mid \tau \in X\}$ be a set of 2-approximate optimal ℓ-simplifications of X under $\mathrm{dtw_p}$. By Proposition 32, X^* can be computed in $O(nm^3d)$ time. We now apply Theorem 31 and Lemma 21 to obtain a $(12(2\ell)^{2/p}(1+\varepsilon)^2(12+\varepsilon), 4)$-approximation of k-median of X^* in $(X^*, \overline{\mathrm{dtw_p}|_{X^*}})$ in time $O(nk\log(k)\ell^2 d + nk^2\log^2(k)\varepsilon^{-4}\log^2(\varepsilon^{-1}) + k^7\varepsilon^{-5}\log^5(n))$. By Lemma 22, the computed set is a $(6(4m\ell)^{1/p}12(2\ell)^{2/p}(1+\varepsilon)^2(12+\varepsilon), 4)$-approximation for (k, ℓ)-median for X under $\mathrm{dtw_p}$. ◀

6 Coreset Application

The theoretical derivations of the previous sections culminate in an approximation algorithm (Theorem 35) to (k, ℓ)-median that is particularly useful in the big data setting, where $n \gg m$. Our strategy is to first compute an efficient but not very accurate approximation(Corollary 33) of (k, ℓ)-median. Subsequently, we use the approximation to construct a coreset. The coreset is then investigated using its metric closure, where by virtue of the size reduction we can greatly reduce the running time of slower more accurate algorithms metric approximation algorithms, yielding a better approximation.

▶ **Theorem 34** ([4, 19])**.** *Given a set X of n points in a metric space, one can compute a $(5+\varepsilon)$-approximate k-median clustering of X in $O(\varepsilon^{-1}n^2k^3\log n)$ time. If P is a weighted point set, with total weight W, then the time required is in $O(\varepsilon^{-1}n^2k^3\log W)$.*

▶ **Theorem 35.** *Let $0 < \varepsilon \leq 1$. The algorithm (k, ℓ)-MEDIAN in Algorithm 2 is a $((32+\varepsilon)(4m\ell)^{1/p}, 1)$-approximate algorithm of constant success probability for (k, ℓ)-median on curves under $\mathrm{dtw_p}$ with a running time of $\widetilde{O}\Big(n(m^3d + k^2 + k\ell^2 d) + \varepsilon^{-6}d^3\ell^3k^7\sqrt[p]{m^6\ell^{12}}\Big)$, where $\widetilde{O}$ hides polylogarithmic factors in n, m, ℓ, k and ε^{-1}.*

Algorithm 2 $((32+\varepsilon)(4m\ell)^{1/p})$-approximate (k,ℓ)-median.

procedure (k,ℓ)-MEDIAN$(X \subset \mathbb{X}^d_m, p, \varepsilon)$
  $\varepsilon' \leftarrow \varepsilon/46$
  Compute $(O((16m\ell^3)^{1/p}), 4)$-approximation C' (Corollary 33)
  Compute bound of sensitivity for each curve $x \in X$ from C' (Lemma 16)
  Compute sample size $s \leftarrow O(\varepsilon^{-2} d\ell k^2 (m^2\ell^4)^{1/p} \log^3(m\ell)\log^2(k)\log(\varepsilon^{-1})\log(n))$
  Sample and weigh ε'-coreset S of X of size s (Theorem 17)
  Compute a 2-simplification for every $s \in S$ resulting in the set S^* (Proposition 32)
  Compute metric closure values $\overline{\phi} = \overline{\mathrm{dtw_p}|_{S^*}}$ (Lemma 23)
  Return $(5+\varepsilon', 1)$-approximation of weighted k-median in $(S^*, \overline{\phi})$ (Theorem 34)
end procedure

Combining the computed ε-coreset with the (k,ℓ)-median algorithm from [16, Theorem 35] instead, we achieve a matching approximation guarantee and improve the dependency on n. The improved approximation guarantee from Corollary 36 compared to Theorem 35 comes at the cost of an exponential dependency in k, as is also present in their results.

▶ **Corollary 36.** *Let $0 < \varepsilon \le 1$ and $0 < \delta \le 1$. There is an $((8+\varepsilon)(m\ell)^{1/p}, 1)$-approximation for (k,ℓ)-median with $\Theta(1-\delta)$ success probability and running time in*

$$\widetilde{O}\left(n(m^3 d + k^2 + k\ell^2) + k^7 + \left(32k^2\varepsilon^{-1}\log(1/\delta)\right)^{k+2} md\left(m^3 + \varepsilon^{-2} d\ell k^2 \sqrt[p]{m^2\ell^4}\right)\right),$$

where $\widetilde{O}$ hides polylogarithmic factors in n, m, ℓ, k and ε^{-1}.

Finally, combining Theorem 35 with Theorem 17 yields the following result.

▶ **Corollary 37.** *The algorithm (k,ℓ)-MEDIAN in Algorithm 2 can be used to construct an ε-coreset for (k,ℓ)-median in time $\widetilde{O}\left(n(m^3 d + k^2 + k\ell^2 d) + \varepsilon^{-6} d^3\ell^3 k^7 \sqrt[p]{m^6\ell^{12}}\right)$ of size*

$$O(\varepsilon^{-2} d\ell k^2 (m^2\ell^2)^{1/p} \log^3(m\ell)\log^2(k)\log(\varepsilon^{-1})\log(n)).$$

7 Conclusion

Our first contribution involves investigating the VC dimension of range spaces characterized by arbitrarily small perturbations of DTW distances. While our results hold for a relaxed variant of the range spaces in question, they establish a robust link between numerous sampling results dependent on the VC dimension and DTW distances. Indeed, our first algorithmic contribution is the construction of coresets for (k,ℓ)-median through the sensitivity sampling framework by Feldman and Langberg [22]. Apart from the VC dimension, the crux of adapting the sensitivity sampling framework to our (non-metric) setting was to use an already known weak version of the triangle inequality satisfied by DTW. This inequality prompted us to further explore approximation algorithms by approximating DTW with a metric. By reducing to the metric case and plugging in our coresets, we designed an algorithm for the (k,ℓ)-median problem, with running time linear in the number of the input sequences, and an approximation factor predominantly determined by the weak triangle inequality.

Although our primary motivation lies in constructing coresets, there are additional direct consequences through sampling bounds that establish a connection between the sample size and the VC dimension. For instance, suppose that we have a large set of time series, following some unknown distribution, and we want to estimate the probability that a new time series falls within a given DTW ball b. Suppose that we also allow for small perturbations of the distances, i.e., we only want to guarantee that the estimated probability is realized by *some* small perturbations of the distances. This probability can be approximated within a constant additive error, by considering a random sample of size depending solely on the VC dimension and the probability of success (over the random sampling) and measuring its intersection with b. Such an estimation can be used for example in anomaly detection, where one aims to detect time series with a small chance of occurring, or in time series segmentation, where diverse patterns may emerge throughout the series.

References

1 Waleed H. Abdulla, David Chow, and Gary Sin. Cross-words reference template for dtw-based speech recognition systems. In *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*, volume 4, pages 1576–1579 Vol.4, 2003.

2 Marcel R. Ackermann, Johannes Blömer, and Christian Sohler. Clustering for metric and nonmetric distance measures. *ACM Transactions on Algorithms*, 6(4):59:1–59:26, 2010.

3 Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999. doi:10.1017/CBO9780511624216.

4 Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local Search Heuristics for k-Median and Facility Location Problems. *SIAM Journal on Computing*, 33(3):544–562, 2004.

5 Milutin Brankovic, Kevin Buchin, Koen Klaren, André Nusser, Aleksandr Popov, and Sampson Wong. (k, l)-Medians Clustering of Trajectories Using Continuous Dynamic Time Warping. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, volume 1, pages 99–110, New York, NY, USA, November 2020. ACM. doi:10.1145/3397536.3422245.

6 Vladimir Braverman, Vincent Cohen-Addad, Shaofeng H.-C. Jiang, Robert Krauthgamer, Chris Schwiegelshohn, Mads Bech Toftrup, and Xuan Wu. The power of uniform sampling for coresets. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 462–473. IEEE, 2022.

7 Vladimir Braverman, Vincent Cohen-Addad, Shaofeng H.-C. Jiang, Robert Krauthgamer, Chris Schwiegelshohn, Mads Bech Toftrup, and Xuan Wu. The power of uniform sampling for coresets. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 462–473. IEEE, 2022. doi:10.1109/FOCS54457.2022.00051.

8 Markus Brill, Till Fluschnik, Vincent Froese, Brijnesh J. Jain, Rolf Niedermeier, and David Schultz. Exact mean computation in dynamic time warping spaces. *Data Min. Knowl. Discov.*, 33(1):252–291, 2019.

9 Kevin Buchin, Anne Driemel, Joachim Gudmundsson, Michael Horton, Irina Kostitsyna, Maarten Löffler, and Martijn Struijs. Approximating (k, ℓ)-center clustering for curves. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 2922–2938, San Diego, California, USA, January 2019. SIAM.

10 Kevin Buchin, Anne Driemel, and Martijn Struijs. On the hardness of computing an average curve. In *17th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT 2020, June 22-24, 2020, Tórshavn, Faroe Islands*, pages 19:1–19:19, 2020.

11 Kevin Buchin, Anne Driemel, and Martijn Struijs. On the Hardness of Computing an Average Curve. In Susanne Albers, editor, *17th Scandinavian Symposium and Workshops on Algorithm Theory*, volume 162 of *LIPIcs*, pages 19:1–19:19, Tórshavn, Faroe Islands, June 2020. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

12 Kevin Buchin, Anne Driemel, Natasja van de L'Isle, and André Nusser. klcluster: Center-based Clustering of Trajectories. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 496–499, 2019.

13 Maike Buchin, Anne Driemel, and Dennis Rohde. Approximating (k, ℓ)-median clustering for polygonal curves. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 2697–2717. SIAM, 2021.

14 Maike Buchin, Anne Driemel, and Dennis Rohde. Approximating (k, ℓ)-Median Clustering for Polygonal Curves. In Dániel Marx, editor, *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 2697–2717, Virtual Conference, January 2021. SIAM.

15 Maike Buchin, Anne Driemel, and Dennis Rohde. Approximating (k, ℓ)-median clustering for polygonal curves. *ACM Trans. Algorithms*, 19(1):4:1–4:32, 2023.

16 Maike Buchin, Anne Driemel, Koen van Greevenbroek, Ioannis Psarros, and Dennis Rohde. Approximating length-restricted means under dynamic time warping. In Parinya Chalermsook and Bundit Laekhanukit, editors, *Approximation and Online Algorithms - 20th International Workshop, WAOA 2022, Potsdam, Germany, September 8-9, 2022, Proceedings*, volume 13538 of *Lecture Notes in Computer Science*, pages 225–253. Springer, 2022.

17 Maike Buchin and Dennis Rohde. Coresets for (k, ℓ)-Median Clustering Under the Fréchet Distance. In Niranjan Balachandran and R. Inkulu, editors, *Algorithms and Discrete Applied Mathematics - 8th International Conference, CALDAM, Puducherry, India, February 10-12, Proceedings*, volume 13179 of *Lecture Notes in Computer Science*, pages 167–180. Springer, 2022.

18 Laurent Bulteau, Vincent Froese, and Rolf Niedermeier. Tight hardness results for consensus problems on circular strings and time series. *SIAM J. Discret. Math.*, 34(3):1854–1883, 2020.

19 Ke Chen. On Coresets for k-Median and k-Means Clustering in Metric and Euclidean Spaces and Their Applications. *SIAM Journal on Computing*, 39(3):923–947, 2009.

20 Siu-Wing Cheng and Haoqiang Huang. Curve simplification and clustering under fréchet distance. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 1414–1432. SIAM, 2023.

21 Anne Driemel, Amer Krivosija, and Christian Sohler. Clustering time series under the Fréchet distance. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 766–785, 2016.

22 Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing*, pages 569–578. ACM, 2011.

23 Ville Hautamäki, Pekka Nykanen, and Pasi Franti. Time-series clustering by approximate prototypes. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, 2008.

24 Piotr Indyk. Sublinear time algorithms for metric space problems. In Jeffrey Scott Vitter, Lawrence L. Larmore, and Frank Thomson Leighton, editors, *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing, May 1-4, 1999, Atlanta, Georgia, USA*, pages 428–434. ACM, 1999. doi:10.1145/301250.301366.

25 Youngseon Jeong, Myong Kee Jeong, and Olufemi A. Omitaomu. Weighted dynamic time warping for time series classification. *Pattern Recognit.*, 44(9):2231–2240, 2011.

26 Rohit J. Kate. Using dynamic time warping distances as features for improved time series classification. *Data Min. Knowl. Discov.*, 30(2):283–312, 2016.

27 Amit Kumar, Yogish Sabharwal, and Sandeep Sen. A Simple Linear Time $(1+\varepsilon)$-Approximation Algorithm for k-Means Clustering in Any Dimensions. In *45th Symposium on Foundations of Computer Science (FOCS), 17-19 October, Rome, Italy, Proceedings*, pages 454–462. IEEE Computer Society, 2004.

28 Michael Langberg and Leonard J. Schulman. Universal ε-approximators for Integrals. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 598–607, 2010.

29 Daniel Lemire. Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recognition*, 42(9):2169–2180, 2009.

30 François Petitjean, Germain Forestier, Geoffrey I. Webb, Ann E. Nicholson, Yanping Chen, and Eamonn J. Keogh. Dynamic time warping averaging of time series allows faster and more accurate classification. In Ravi Kumar, Hannu Toivonen, Jian Pei, Joshua Zhexue Huang, and Xindong Wu, editors, *2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, December 14-17, 2014*, pages 470–479. IEEE Computer Society, 2014.

31 François Petitjean, Germain Forestier, Geoffrey I. Webb, Ann E. Nicholson, Yanping Chen, and Eamonn J. Keogh. Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm. *Knowl. Inf. Syst.*, 47(1):1–26, 2016.

32 François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognit.*, 44(3):678–693, 2011.

33 Lawrence Rabiner and Jay Wilpon. Considerations in applying clustering techniques to speaker independent word recognition. In *ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 578–581, 1979.

34 Thanawin Rakthanmanon, Bilson J. L. Campana, Abdullah Mueen, Gustavo E. A. P. A. Batista, M. Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn J. Keogh. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Trans. Knowl. Discov. Data*, 7(3):10:1–10:31, 2013.

35 Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory Series A*, 13:145–147, 1972.

36 Nathan Schaar, Vincent Froese, and Rolf Niedermeier. Faster binary mean computation under dynamic time warping. In *31st Annual Symposium on Combinatorial Pattern Matching, CPM 2020, June 17-19, 2020, Copenhagen, Denmark*, pages 28:1–28:13, 2020.

37 Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1), 1972.

38 Tuan Minh Tran, Xuan-May Thi Le, Hien T. Nguyen, and Van-Nam Huynh. A novel non-parametric method for time series classification based on k-nearest neighbors and dynamic time warping barycenter averaging. *Eng. Appl. Artif. Intell.*, 78:173–185, 2019.

39 Vladimir Vapnik and Alexey Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.