Simple Sublinear-Time Edit Distance Approximation

Tomasz Kociumaka

MAX PLANCK INSTITUTE FOR INFORMATICS

SIC Saarland Informatics Campus

↓

INSAIT

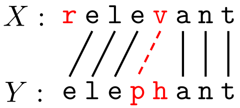
Based on a FOCS'22 paper
with Elazar Goldenberg, Robert Krauthgamer, and Barna Saha

ADFOCS 2024, August 29th, 2024

Edit Distance

Edit distance $\mathsf{ED}(X, Y)$

Minimum number of character insertions, deletions, and substitutions to transform X to Y.



$\mathsf{ED}(X, Y) = 3$

Computing Edit Distance

Exact algorithms

Reference	Time	Remarks
Vintsyuk'68, Needleman & Wunsch'70, ...	$\mathcal{O}(n^2)$	
Backurs, Indyk; STOC'15	$\Omega(n^{2-o(1)})$	conditioned on SETH

Computing Edit Distance

Exact algorithms

Reference	Time	Remarks
Vintsyuk'68, Needleman & Wunsch'70, ...	$\mathcal{O}(n^2)$	
Backurs, Indyk; STOC'15	$\Omega(n^{2-o(1)})$	conditioned on SETH

Approximation algorithms (selected results, all randomized)

Reference	Time	Approx.
Andoni, Onak; STOC'09	$n^{1+o(1)}$	$n^{o(1)}$
Chakraborty, Das, Goldenberg, Koucky, Saks; FOCS'18	$\widetilde{\mathcal{O}}(n^{12/7})$	$\mathcal{O}(1)$
Goldenberg, Rubinstein, Saha; STOC'20	$n^{1.6+o(1)}$	$3+o(1)$
Andoni, Nosatzki; FOCS'20	$\mathcal{O}(n^{1+\varepsilon})$	$\mathcal{O}_\varepsilon(1)$

Computing Edit Distance

Exact algorithms

Reference	Time	Remarks
Vintsyuk'68, Needleman & Wunsch'70, ...	$\mathcal{O}(n^2)$	
Backurs, Indyk; STOC'15	$\Omega(n^{2-o(1)})$	conditioned on SETH

Approximation algorithms (selected results, all randomized)

Reference	Time	Approx.
Andoni, Onak; STOC'09	$\hat{\mathcal{O}}(n)$	$\hat{\mathcal{O}}(1)$
Chakraborty, Das, Goldenberg, Koucky, Saks; FOCS'18	$\widetilde{\mathcal{O}}(n^{12/7})$	$\mathcal{O}(1)$
Goldenberg, Rubinstein, Saha; STOC'20	$\hat{\mathcal{O}}(n^{1.6})$	$3 + o(1)$
Andoni, Nosatzki; FOCS'20	$\mathcal{O}(n^{1+\varepsilon})$	$\mathcal{O}_{\varepsilon}(1)$

Sublinear-Time Edit Distance Approximation

Model

- $\mathcal{O}(1)$-time random access to X and Y
- Monte-Carlo randomization

Adaptive algorithms: may issue access queries one by one.

Non-adaptive algorithms: issue a *single batch* of access queries.

Sublinear-Time Edit Distance Approximation

Model

- $\mathcal{O}(1)$-time random access to X and Y
- Monte-Carlo randomization

Adaptive algorithms: may issue access queries one by one.
Non-adaptive algorithms: issue a *single batch* of access queries.

Bad news: Testing $\mathrm{ED}(X, Y) = 0$ requires $\Omega(n)$ queries!

Sublinear-Time Edit Distance Approximation

Model

- $\mathcal{O}(1)$-time random access to X and Y
- Monte-Carlo randomization

Adaptive algorithms: may issue access queries one by one.
Non-adaptive algorithms: issue a *single batch* of access queries.

Bad news: Testing $\mathrm{ED}(X,Y)=0$ requires $\Omega(n)$ queries!

(k,K)-Gap Edit Distance Problem [BEK+03,AO09,GKS19,KS20,GKKS22,BCFN22,BCFK24]

Given random access to strings $X,Y \in \Sigma^n$, and integer parameters $0 \le k \le K \le n$, return:

- **CLOSE** if $\mathrm{ED}(X,Y) \le k$,
- **FAR** if $\mathrm{ED}(X,Y) > K$,
- an arbitrary answer otherwise.

Gap Edit Distance Problem: Selected Results

Theorem — Goldenberg, K., Krauthgamer, Saha; FOCS'22

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

- The first algorithm to achieve sublinear runtime for the whole spectrum of parameters.

Gap Edit Distance Problem: Selected Results

Theorem — Goldenberg, K., Krauthgamer, Saha; FOCS'22

There is a non-adaptive algorithm that solves the (k,K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

- The first algorithm to achieve sublinear runtime for the whole spectrum of parameters.

Theorem — Goldenberg, K., Krauthgamer, Saha; FOCS'22

There is a non-adaptive algorithm that solves the (k,K)-Gap Edit distance problem using $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K})$ queries and in $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K} + \frac{nk^2}{K^2})$ time provided that $K > k \cdot n^{\Omega(1)}$.

- Up to $\widetilde{\mathcal{O}}(\cdot)$, the query complexity is unconditionally optimal for non-adaptive algorithms.

Gap Edit Distance Problem: Selected Results

Theorem — Goldenberg, K., Krauthgamer, Saha; FOCS'22

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

- The first algorithm to achieve sublinear runtime for the whole spectrum of parameters.

Theorem — Goldenberg, K., Krauthgamer, Saha; FOCS'22

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit distance problem using $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K})$ queries and in $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K} + \frac{nk^2}{K^2})$ time provided that $K > k \cdot n^{\Omega(1)}$.

- Up to $\widetilde{\mathcal{O}}(\cdot)$, the query complexity is unconditionally optimal for non-adaptive algorithms.

Theorem — Bringmann, Cassis, Fischer, K.; SODA'24

There is an adaptive algorithm that solves the (k, K)-Gap Edit distance problem in $\widetilde{\mathcal{O}}(\frac{n}{K} + \sqrt{nk} + k^2)$ time provided that $K > k \cdot n^{\Omega(1)}$.

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \leq k \cdot \hat{\mathcal{O}}(1)$.

1. Partition $X = X_1 \cdots X_m$ and $Y = Y_1 \cdots Y_m$ into m blocks of equal length.
2. For each $i \in [1\,..\,m]$:
   - With probability ρ, use the oracle to distinguish $\mathsf{ED}(X_i, Y_i) \leq k$ from $\mathsf{ED}(X_i, Y_i) > \hat{k}$.
3. Return CLOSE if and only if all oracle calls returned CLOSE.

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. Partition $X = X_1 \cdots X_m$ and $Y = Y_1 \cdots Y_m$ into m blocks of equal length.
2. For each $i \in [1 \mathinner{.\,.} m]$:
   - With probability ρ, use the oracle to distinguish $\mathsf{ED}(X_i, Y_i) \le k$ from $\mathsf{ED}(X_i, Y_i) > \hat{k}$.
3. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \le k$, then $\mathsf{ED}(X_i, Y_i) \le \mathsf{ED}(X, Y) \le k$, so all oracle calls return CLOSE.

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \leq k \cdot \hat{\mathcal{O}}(1)$.

1. Partition $X = X_1 \cdots X_m$ and $Y = Y_1 \cdots Y_m$ into m blocks of equal length.
2. For each $i \in [1 \mathinner{.\,.} m]$:
   - With probability ρ, use the oracle to distinguish $\mathsf{ED}(X_i, Y_i) \leq k$ from $\mathsf{ED}(X_i, Y_i) > \hat{k}$.
3. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \leq k$, then $\mathsf{ED}(X_i, Y_i) \leq \mathsf{ED}(X, Y) \leq k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{i : \mathsf{ED}(X_i, Y_i) > \hat{k}\}| \lesssim \frac{1}{\rho}$.

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. Partition $X = X_1 \cdots X_m$ and $Y = Y_1 \cdots Y_m$ into m blocks of equal length.
2. For each $i \in [1\,.\,.\,m]$:
   - With probability ρ, use the oracle to distinguish $\mathsf{ED}(X_i, Y_i) \le k$ from $\mathsf{ED}(X_i, Y_i) > \hat{k}$.
3. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \le k$, then $\mathsf{ED}(X_i, Y_i) \le \mathsf{ED}(X, Y) \le k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{i : \mathsf{ED}(X_i, Y_i) > \hat{k}\}| \lesssim \frac{1}{\rho}$. Thus,

$$\mathsf{ED}(X, Y) \le \sum_{i=1}^{m} \mathsf{ED}(X_i, Y_i) \lesssim m\hat{k} + \frac{1}{\rho} \cdot \frac{n}{m}$$

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. Partition $X = X_1 \cdots X_m$ and $Y = Y_1 \cdots Y_m$ into $m \approx \frac{K}{k}$ blocks of equal length.
2. For each $i \in [1\,..\,m]$:
   - With probability $\rho \approx \frac{n}{mK}$, use the oracle to distinguish $\mathsf{ED}(X_i, Y_i) \le k$ from $\mathsf{ED}(X_i, Y_i) > \hat{k}$.
3. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \le k$, then $\mathsf{ED}(X_i, Y_i) \le \mathsf{ED}(X, Y) \le k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{i : \mathsf{ED}(X_i, Y_i) > \hat{k}\}| \lesssim \frac{1}{\rho}$. Thus,

$$\mathsf{ED}(X, Y) \le \sum_{i=1}^{m} \mathsf{ED}(X_i, Y_i) \lesssim m\hat{k} + \frac{1}{\rho} \cdot \frac{n}{m} \lesssim K.$$

Algorithm of Andoni and Onak [STOC'09]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. Partition $X = X_1 \cdots X_m$ and $Y = Y_1 \cdots Y_m$ into $m \approx \frac{K}{k}$ blocks of equal length.
2. For each $i \in [1\,.\,.\,m]$:
   - With probability $\rho \approx \frac{n}{mK}$, use the oracle to distinguish $\mathsf{ED}(X_i, Y_i) \le k$ from $\mathsf{ED}(X_i, Y_i) > \hat{k}$.
3. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \le k$, then $\mathsf{ED}(X_i, Y_i) \le \mathsf{ED}(X, Y) \le k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{i : \mathsf{ED}(X_i, Y_i) > \hat{k}\}| \lesssim \frac{1}{\rho}$. Thus,

$$\mathsf{ED}(X, Y) \le \sum_{i=1}^{m} \mathsf{ED}(X_i, Y_i) \lesssim m\hat{k} + \frac{1}{\rho} \cdot \frac{n}{m} \lesssim K.$$

Expected running time: $\hat{\mathcal{O}}(\rho \cdot m \cdot \frac{n}{m}) = \hat{\mathcal{O}}(\rho \cdot n) = \hat{\mathcal{O}}(\frac{n^2}{mK}) = \hat{\mathcal{O}}(\frac{n^2 k}{K^2})$.

Simple Sublinear-Time Edit Distance Approximation [GKKS, FOCS'22]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \leq k \cdot \hat{\mathcal{O}}(1)$.

1. For $q = 0$ to $\log n$:
   a. Partition $X = X_{q,1} \cdots X_{q,2^q}$ and $Y = Y_{q,1} \cdots Y_{q,2^q}$ into 2^q blocks of equal length.
   b. For each $i \in [1 \mathinner{.\,.} 2^q]$:
      - With probability $\rho \approx \frac{k}{K}$, use the oracle to distinguish $\mathsf{ED}(X_{q,i}, Y_{q,i}) \leq k$ from $\mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}$.
2. Return CLOSE if and only if all oracle calls returned CLOSE.

Simple Sublinear-Time Edit Distance Approximation [GKKS, FOCS'22]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. For $q = 0$ to $\log n$:
   a. Partition $X = X_{q,1} \cdots X_{q,2^q}$ and $Y = Y_{q,1} \cdots Y_{q,2^q}$ into 2^q blocks of equal length.
   b. For each $i \in [1 \,..\, 2^q]$:
      - With probability $\rho \approx \frac{k}{K}$, use the oracle to distinguish $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le k$ from $\mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}$.
2. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \le k$, then $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le \mathsf{ED}(X, Y) \le k$, so all oracle calls return CLOSE.

Simple Sublinear-Time Edit Distance Approximation [GKKS, FOCS'22]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. For $q = 0$ to $\log n$:
   a. Partition $X = X_{q,1} \cdots X_{q,2^q}$ and $Y = Y_{q,1} \cdots Y_{q,2^q}$ into 2^q blocks of equal length.
   b. For each $i \in [1\,..\,2^q]$:
      - With probability $\rho \approx \frac{k}{K}$, use the oracle to distinguish $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le k$ from $\mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}$.
2. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X,Y) \le k$, then $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le \mathsf{ED}(X,Y) \le k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{(q,i) : \mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}\}| \lesssim \frac{1}{\rho}$.

Key Lemma

Lemma

If there are at most $1/\rho$ far blocks $X_{q,i}$ in total and the last level $q = \log n$ contains only close blocks, then X can be partitioned into $2/\rho$ close blocks.

<table>
<tr><td colspan="16">$X_{0,1}$</td></tr>
<tr><td colspan="8">$X_{1,1}$</td><td colspan="8">$X_{1,2}$</td></tr>
<tr><td colspan="4">$X_{2,1}$</td><td colspan="4">$X_{2,2}$</td><td colspan="4">$X_{2,3}$</td><td colspan="4">$X_{2,4}$</td></tr>
<tr><td colspan="2">$X_{3,1}$</td><td colspan="2">$X_{3,2}$</td><td colspan="2">$X_{3,3}$</td><td colspan="2">$X_{3,4}$</td><td colspan="2">$X_{3,5}$</td><td colspan="2">$X_{3,6}$</td><td colspan="2">$X_{3,7}$</td><td colspan="2">$X_{3,8}$</td></tr>
<tr><td>$X_{4,1}$</td><td>$X_{4,2}$</td><td>$X_{4,3}$</td><td>$X_{4,4}$</td><td>$X_{4,5}$</td><td>$X_{4,6}$</td><td>$X_{4,7}$</td><td>$X_{4,8}$</td><td>$X_{4,9}$</td><td>$X_{4,10}$</td><td>$X_{4,11}$</td><td>$X_{4,12}$</td><td>$X_{4,13}$</td><td>$X_{4,14}$</td><td>$X_{4,15}$</td><td>$X_{4,16}$</td></tr>
</table>

Key Lemma

Lemma

If there are at most $1/\rho$ far blocks $X_{q,i}$ in total and the last level $q = \log n$ contains only close blocks, then X can be partitioned into $2/\rho$ close blocks.

$X_{0,1}$															
$X_{1,1}$								$X_{1,2}$							
$X_{2,1}$				$X_{2,2}$				$X_{2,3}$				$X_{2,4}$			
$X_{3,1}$		$X_{3,2}$		$X_{3,3}$		$X_{3,4}$		$X_{3,5}$		$X_{3,6}$		$X_{3,7}$		$X_{3,8}$	
$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$	$X_{4,11}$	$X_{4,12}$	$X_{4,13}$	$X_{4,14}$	$X_{4,15}$	$X_{4,16}$

Key Lemma

Lemma

If there are at most $1/\rho$ far blocks $X_{q,i}$ in total and the last level $q = \log n$ contains only close blocks, then X can be partitioned into $2/\rho$ close blocks.

<table>
<tr><td colspan="16">$X_{0,1}$</td></tr>
<tr><td colspan="8">$X_{1,1}$</td><td colspan="8">$X_{1,2}$</td></tr>
<tr><td colspan="4">$X_{2,1}$</td><td colspan="4">$X_{2,2}$</td><td colspan="4">$X_{2,3}$</td><td colspan="4">$X_{2,4}$</td></tr>
<tr><td colspan="2">$X_{3,1}$</td><td colspan="2">$X_{3,2}$</td><td colspan="2">$X_{3,3}$</td><td colspan="2">$X_{3,4}$</td><td colspan="2">$X_{3,5}$</td><td colspan="2">$X_{3,6}$</td><td colspan="2">$X_{3,7}$</td><td colspan="2">$X_{3,8}$</td></tr>
<tr><td>$X_{4,1}$</td><td>$X_{4,2}$</td><td>$X_{4,3}$</td><td>$X_{4,4}$</td><td>$X_{4,5}$</td><td>$X_{4,6}$</td><td>$X_{4,7}$</td><td>$X_{4,8}$</td><td>$X_{4,9}$</td><td>$X_{4,10}$</td><td>$X_{4,11}$</td><td>$X_{4,12}$</td><td>$X_{4,13}$</td><td>$X_{4,14}$</td><td>$X_{4,15}$</td><td>$X_{4,16}$</td></tr>
</table>

Key Lemma

Lemma

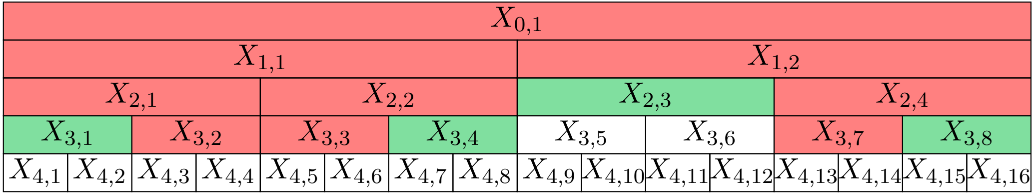
If there are at most $1/\rho$ far blocks $X_{q,i}$ in total and the last level $q = \log n$ contains only close blocks, then X can be partitioned into $2/\rho$ close blocks.

<table>
<tr><td colspan="16">$X_{0,1}$</td></tr>
<tr><td colspan="8">$X_{1,1}$</td><td colspan="8">$X_{1,2}$</td></tr>
<tr><td colspan="4">$X_{2,1}$</td><td colspan="4">$X_{2,2}$</td><td colspan="4">$X_{2,3}$</td><td colspan="4">$X_{2,4}$</td></tr>
<tr><td colspan="2">$X_{3,1}$</td><td colspan="2">$X_{3,2}$</td><td colspan="2">$X_{3,3}$</td><td colspan="2">$X_{3,4}$</td><td colspan="2">$X_{3,5}$</td><td colspan="2">$X_{3,6}$</td><td colspan="2">$X_{3,7}$</td><td colspan="2">$X_{3,8}$</td></tr>
<tr><td>$X_{4,1}$</td><td>$X_{4,2}$</td><td>$X_{4,3}$</td><td>$X_{4,4}$</td><td>$X_{4,5}$</td><td>$X_{4,6}$</td><td>$X_{4,7}$</td><td>$X_{4,8}$</td><td>$X_{4,9}$</td><td>$X_{4,10}$</td><td>$X_{4,11}$</td><td>$X_{4,12}$</td><td>$X_{4,13}$</td><td>$X_{4,14}$</td><td>$X_{4,15}$</td><td>$X_{4,16}$</td></tr>
</table>

Key Lemma

Lemma

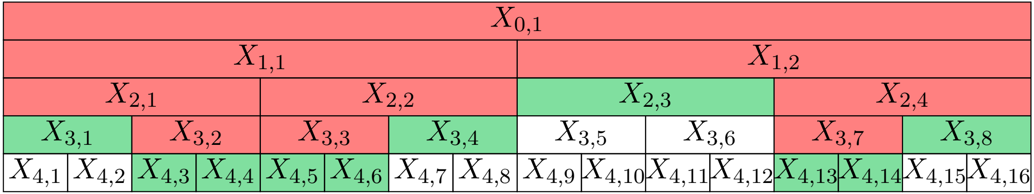
If there are at most $1/\rho$ far blocks $X_{q,i}$ in total and the last level $q = \log n$ contains only close blocks, then X can be partitioned into $2/\rho$ close blocks.



Key Lemma

Lemma

If there are at most $1/\rho$ far blocks $X_{q,i}$ in total and the last level $q = \log n$ contains only close blocks, then X can be partitioned into $2/\rho$ close blocks.



Simple Sublinear-Time Edit Distance Approximation [GKKS, FOCS'22]

Oracle (almost linear-time approximation of edit distance)

The $(k, \hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \leq k \cdot \hat{\mathcal{O}}(1)$.

1. For $q = 0$ to $\log n$:
   a. Partition $X = X_{q,1} \cdots X_{q,2^q}$ and $Y = Y_{q,1} \cdots Y_{q,2^q}$ into 2^q blocks of equal length.
   b. For each $i \in [1\,..\,2^q]$:
      - With probability $\rho \approx \frac{k}{K}$, use the oracle to distinguish $\mathsf{ED}(X_{q,i}, Y_{q,i}) \leq k$ from $\mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}$.
2. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X, Y) \leq k$, then $\mathsf{ED}(X_{q,i}, Y_{q,i}) \leq \mathsf{ED}(X, Y) \leq k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{(q, i) : \mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}\}| \lesssim \frac{1}{\rho}$. By the lemma, X can be decomposed into $\approx \frac{2}{\rho}$ blocks $X_{q,i}$ satisfying $\mathsf{ED}(X_{q,i}, Y_{q,i}) \leq \hat{k}$:

$$\mathsf{ED}(X, Y) \lesssim \frac{2}{\rho} \cdot \hat{k} \lesssim \frac{K}{k} \cdot k = K.$$

Simple Sublinear-Time Edit Distance Approximation [GKKS, FOCS'22]

Oracle (almost linear-time approximation of edit distance)

The $(k,\hat{k})$-Gap Edit Distance problem can be solved in $\hat{\mathcal{O}}(n)$ time for some $\hat{k} \le k \cdot \hat{\mathcal{O}}(1)$.

1. For $q = 0$ to $\log n$:
   a. Partition $X = X_{q,1} \cdots X_{q,2^q}$ and $Y = Y_{q,1} \cdots Y_{q,2^q}$ into 2^q blocks of equal length.
   b. For each $i \in [1\,..\,2^q]$:
      - With probability $\rho \approx \frac{k}{K}$, use the oracle to distinguish $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le k$ from $\mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}$.
2. Return CLOSE if and only if all oracle calls returned CLOSE.

Correctness:

- If $\mathsf{ED}(X,Y) \le k$, then $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le \mathsf{ED}(X,Y) \le k$, so all oracle calls return CLOSE.
- If all oracle calls return CLOSE, then $|\{(q,i) : \mathsf{ED}(X_{q,i}, Y_{q,i}) > \hat{k}\}| \lesssim \frac{1}{\rho}$. By the lemma, X can be decomposed into $\approx \frac{2}{\rho}$ blocks $X_{q,i}$ satisfying $\mathsf{ED}(X_{q,i}, Y_{q,i}) \le \hat{k}$:

$$\mathsf{ED}(X,Y) \lesssim \frac{2}{\rho} \cdot \hat{k} \lesssim \frac{K}{k} \cdot k = K.$$

Expected running time: $\hat{\mathcal{O}}(\sum_q \rho \cdot 2^q \cdot \frac{n}{2^q}) = \hat{\mathcal{O}}(\rho \cdot n) = \hat{\mathcal{O}}(\frac{nk}{K})$.

Conclusions & Open Problems

Simple Algorithm

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

Conclusions & Open Problems

Simple Algorithm

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

Optimal Non-Adaptive Algorithm

There is a non-adaptive algorithm that solves (k, K)-Gap Edit distance problem with $K > k \cdot n^{\Omega(1)}$ using $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K})$ queries and in $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K} + \frac{nk^2}{K^2})$ time.

Any non-adaptive algorithm for (k, K)-Gap Edit distance with $K < \frac{n}{6}$ needs $\Omega(\frac{n\sqrt{k}}{K})$ queries.

Conclusions & Open Problems

Simple Algorithm

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

Optimal Non-Adaptive Algorithm

There is a non-adaptive algorithm that solves (k, K)-Gap Edit distance problem with $K > k \cdot n^{\Omega(1)}$ using $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K})$ queries and in $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K} + \frac{nk^2}{K^2})$ time.

Any non-adaptive algorithm for (k, K)-Gap Edit distance with $K < \frac{n}{6}$ needs $\Omega(\frac{n\sqrt{k}}{K})$ queries.

Open problems:

1. Bridge the gap between time and query complexity. Conditional lower bounds?
2. Use adaptive algorithms to circumvent the lower bound also for $k > \sqrt{n}$.

Conclusions & Open Problems

Simple Algorithm

There is a non-adaptive algorithm that solves the (k, K)-Gap Edit Distance problem in $\hat{\mathcal{O}}(\frac{nk}{K})$ time provided that $K > k \cdot n^{\Omega(1)}$.

Optimal Non-Adaptive Algorithm

There is a non-adaptive algorithm that solves (k, K)-Gap Edit distance problem with $K > k \cdot n^{\Omega(1)}$ using $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K})$ queries and in $\widetilde{\mathcal{O}}(\frac{n\sqrt{k}}{K} + \frac{nk^2}{K^2})$ time.

Any non-adaptive algorithm for (k, K)-Gap Edit distance with $K < \frac{n}{6}$ needs $\Omega(\frac{n\sqrt{k}}{K})$ queries.

Open problems:

1. Bridge the gap between time and query complexity. Conditional lower bounds?
2. Use adaptive algorithms to circumvent the lower bound also for $k > \sqrt{n}$.

Thank you for your attention!