Wiktor Zuba

MIMUW Colloquium

13.03.2025

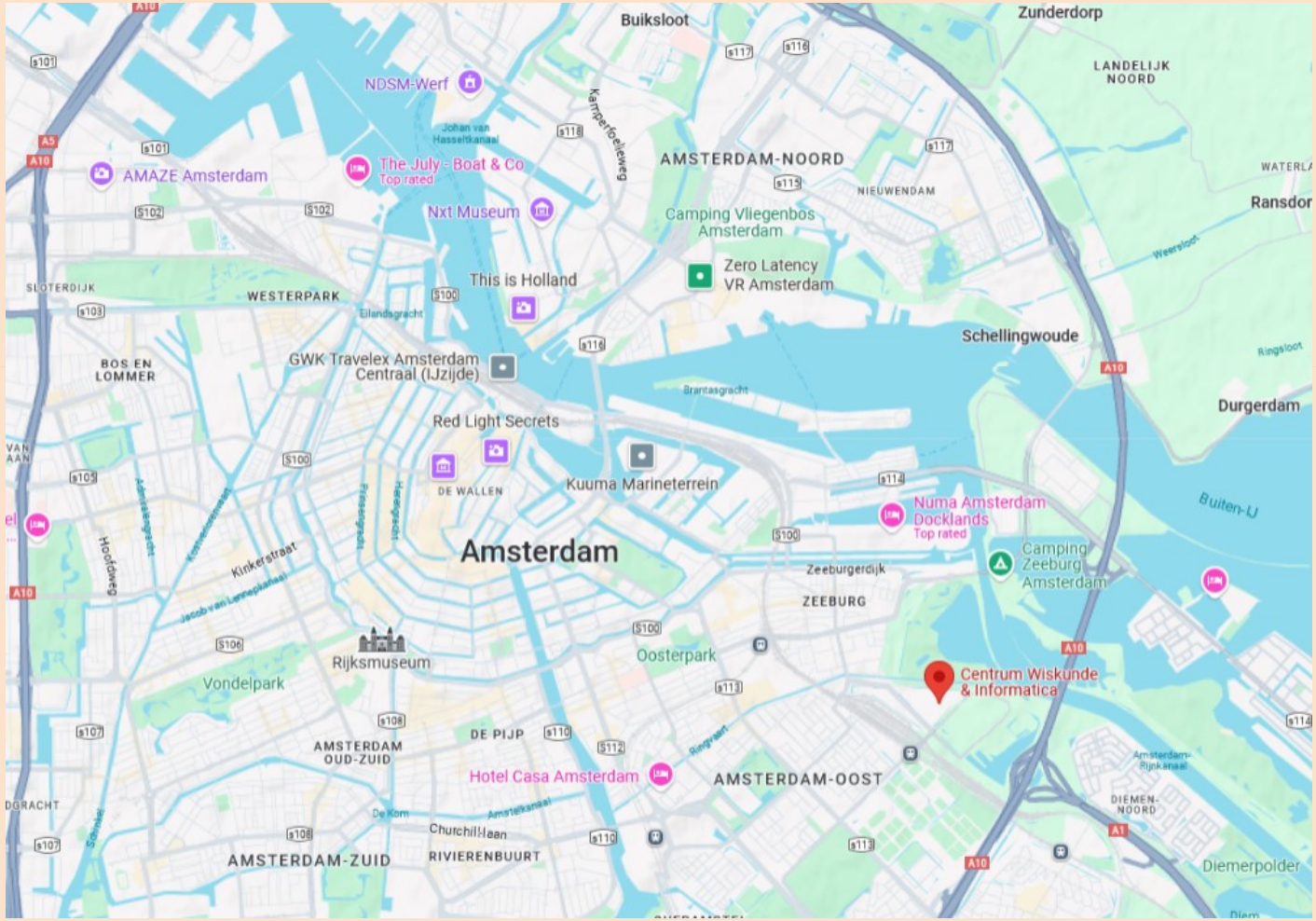
Background

- 2011-2021: Studies at MIMUW
- 2022-2024: PostDoc at CWI
- 2024-present: Assistant Professor at MIMUW

CWI


Centrum Wiskunde & Informatica
Netherlands
Amsterdam
Alkmaar
Zwolle
Groningen
Aurich
Emden
Oldenburg
Bremerhaven
Naturpark Wildeshauser Geest
Osnabrück
Bielefeld
Münster
Paderborn
Dortmund
Essen
Düsseldorf
Cologne
Bonn
The Hague
Rotterdam
Eindhoven
Antwerp
Bruges
Ghent
Dunkirk
Brussels
A7
A28
A32
A6
A50
A1
A2
A73
A76
E19
E34
E17
E313
E403
A25

CWI


Buiksloot
Zunderdorp
LANDELIJK NOORD
NDSM-Werf
Johan van Hasseltkanaal
Kamperfoelieweg
AMSTERDAM-NOORD
NIEUWENDAM
AMAZE Amsterdam
The July - Boat & Co
Top rated
Nxt Museum
Camping Vliegenbos Amsterdam
Zero Latency VR Amsterdam
Ransdorp
Weersloot
SLOTERDIJK
WESTERPARK
Eilandsgracht
This is Holland
Schellingwoude
Ringsloot
BOS EN LOMMER
GWK Travelex Amsterdam Centraal (IJzijde)
Brantasgracht
Durgerdam
Red Light Secrets
DE WALLEN
Kuuma Marineterrein
Numa Amsterdam Docklands
Top rated
Buiten-IJ
Amsterdam
Kinkerstraat
Hoofdweg
Zeeburgerdijk
Camping Zeeburg Amsterdam
ZEEBURG
Rijksmuseum
Oosterpark
Vondelpark
Centrum Wiskunde & Informatica
AMSTERDAM OUD-ZUID
DE PIJP
Hotel Casa Amsterdam
AMSTERDAM-OOST
Amsterdam-Rijnkanaal
DIEMEN-NOORD
DORACHT
De Kom
Amstelkanaal
Churchilllaan
AMSTERDAM-ZUID
RIVIERENBUURT
Diemerpolder

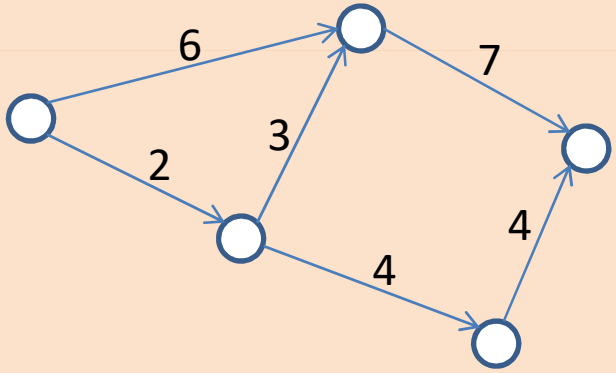
CWI


Flevopark
Valentijnkade
Natuurspeeltuin Jeugdland Amsterdam
Waterkeringpad
Tokyo Nail Art
Bedrijven Te
edrijf verkopen
Het grootste netwerk
AMOLF
Centrum Wiskunde & Informatica
Netherlands eScience Center
Science Park
hannieskeramiek
Carolina Maccill avylaan
Universiteit van Amsterdam
Right2Education
Lab42
Science Park campus
Equinix AM3
Hoogspannings
Watergraafsme
Amsterdam-Waterg
boerderij De Werf
Carolina Maccillavylaan
erdam Science Park
Ringslangpad
Science Park

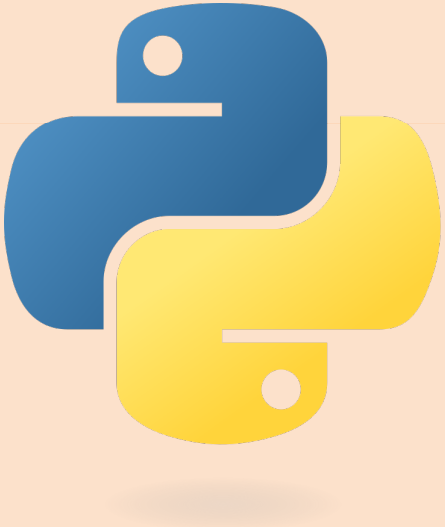
CWI

CWI

Dijkstra Algorithm 1956



Python 1980s



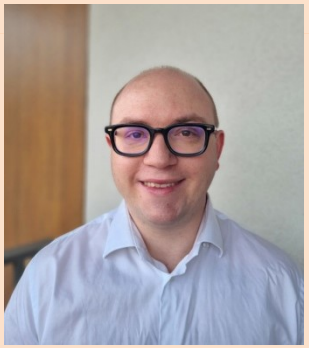
Now - 

Text Algorithms Team

Now -

Text Algorithms Team

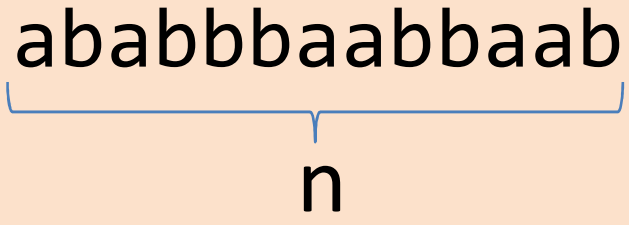
Wojciech Rytter



Jakub Radoszewski

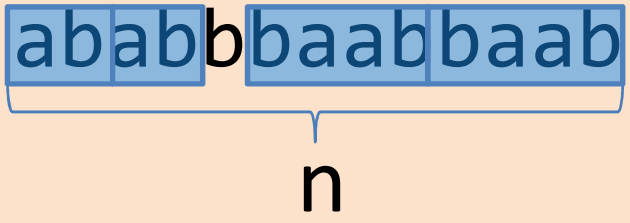
Tomasz Waleń

Text Algorithms



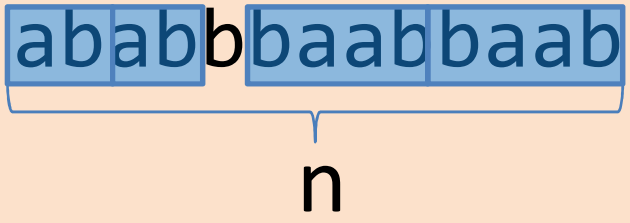
(Check / find all / count all) regular parts

Text Algorithms



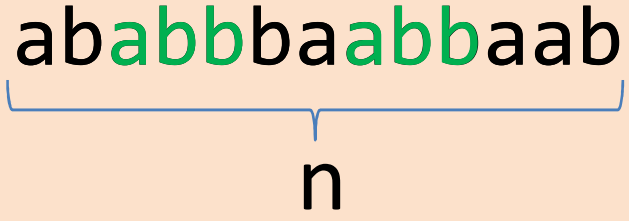
(Check / find all / count all) regular parts

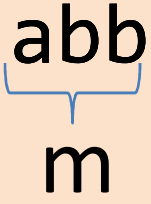
Text Algorithms



(Check / find all / count all) regular parts
(in a substring)

Text Algorithms

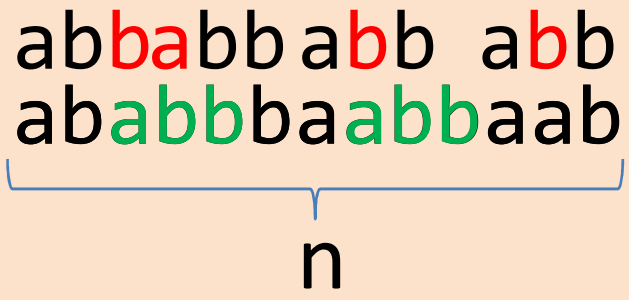


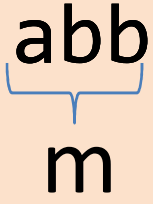


(Check / find all / count all) regular parts
(in a substring)

Pattern matching

Text Algorithms

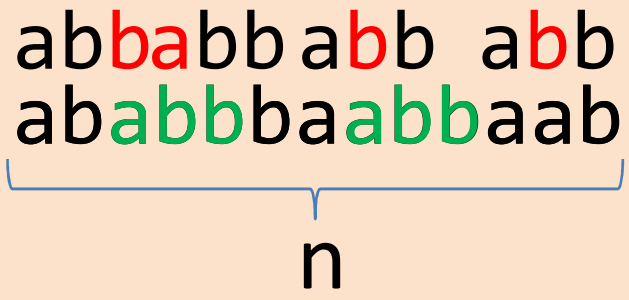


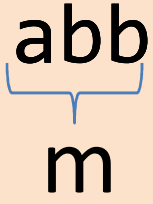


(Check / find all / count all) regular parts (in a substring)

(Approximate = up to k errors) Pattern matching

Text Algorithms





(Check / find all / count all) regular parts (in a substring)

(Approximate = up to k errors) Pattern matching

String similarities (common substrings, ...)

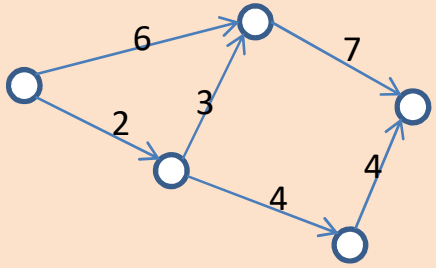
Applications of regularities and similarities of strings

- Bioinformatics – tandem repeats are associated to genetic diseases,

  similarities between DNA sequences~ close relationship between organisms
- Compression – high regularity = better compression rate
- Data analisys – plagiarism detection

Aims

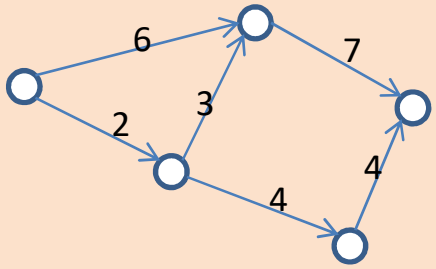
Graph algorithms



brute force:
usually exponential

Aims

Graph algorithms



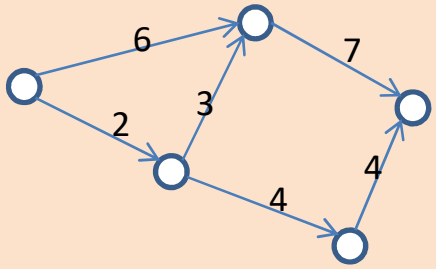
brute force:
usually exponential

Is there an $n^{O(1)}$-time algorithm?

NP-complete

Aims

Graph algorithms



brute force:
usually exponential

Is there an $n^{O(1)}$-time algorithm?

NP-complete

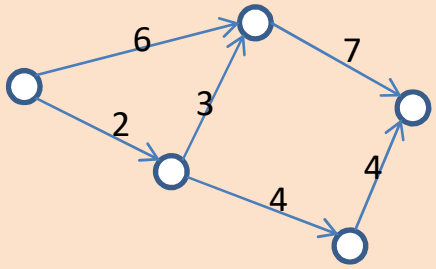
Smaller exponent?

Approximation?

FPT?

Aims

Graph algorithms



brute force:
usually exponential

Is there an $n^{O(1)}$-time algorithm?

NP-complete

Smaller exponent?

Approximation?

FPT?

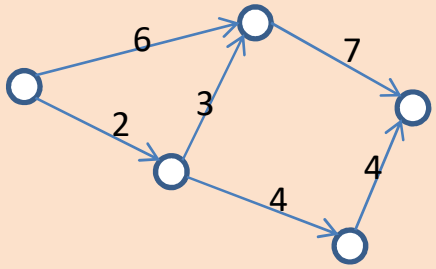
Text algorithms

ababbbaabbaab

brute force:
usually $O(n^3)$ or $O(n^2)$ time.

Aims

Graph algorithms



brute force:
usually exponential

Is there an $n^{O(1)}$-time algorithm?

NP-complete

Smaller exponent?

Approximation?

FPT?

Text algorithms

ababbbaabbbaab

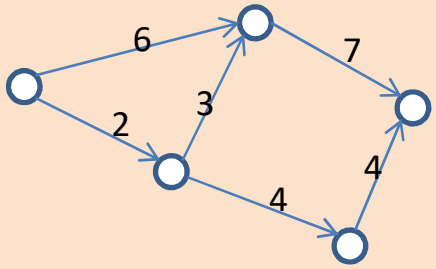
brute force:
usually $O(n^3)$ or $O(n^2)$ time.

Try to reach O(n) or at least O(n polylog n).

No $O(n^{2-\varepsilon})$ time algorithm
(by e.g. SETH)

Aims

Graph algorithms



brute force:
usually exponential

Is there an $n^{O(1)}$-time algorithm?

NP-complete

Smaller exponent?

Approximation?

FPT?

Text algorithms

ababbbaabbbaab

brute force:
usually $O(n^3)$ or $O(n^2)$ time.

Try to reach O(n) or at least O(n polylog n).

Sometimes even O(n/log n).

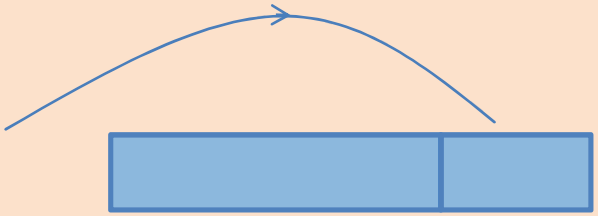
No $O(n^{2-\varepsilon})$ time algorithm
(by e.g. SETH)

Different models

- Circular
- Approximate
- Abelian
- Order preserving
- 2D

- Labelled graphs

Different models



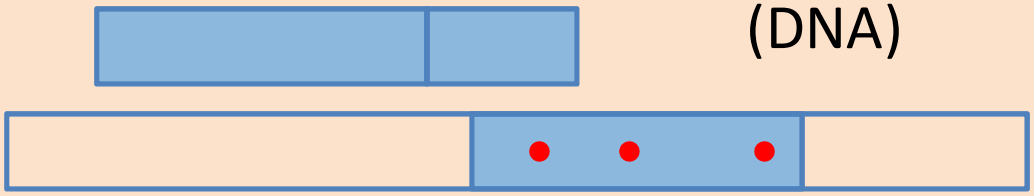
(DNA)

- Circular
- Approximate
- Abelian
- Order preserving
- 2D

- Labelled graphs

Different models

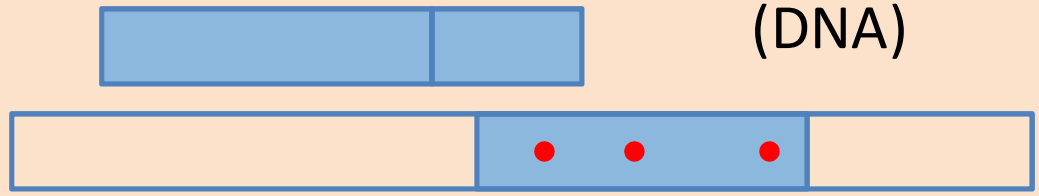
- Circular
- Approximate
- Abelian
- Order preserving
- 2D



- Labelled graphs

Different models

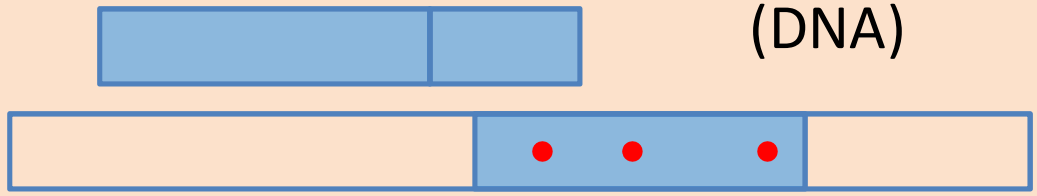
- Circular
- Approximate
- Abelian ababbbaabbaab~ aaaaaabbbbbbb
- Order preserving
- 2D



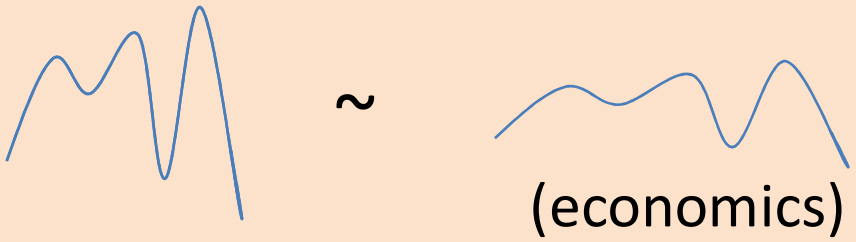
- Labelled graphs

Different models

- Circular
- Approximate
- Abelian
- Order preserving
- 2D



ababbbaabbaab~ aaaaaabbbbbbb



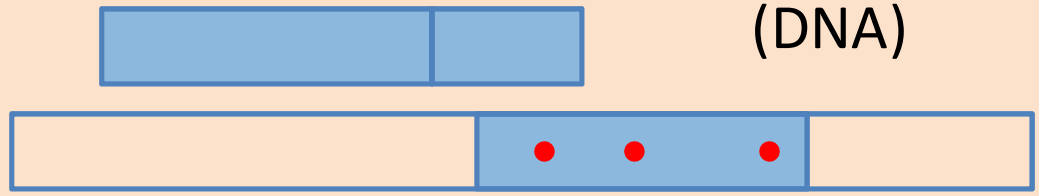
- Labelled graphs

Different models

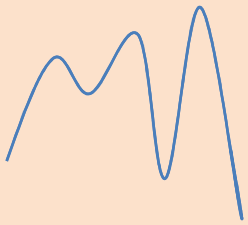
- Circular
- Approximate
- Abelian ababbbaabbaab~ aaaaaabbbbbbb
- Order preserving
- 2D

abcab
bcabc
cabca

- Labelled graphs

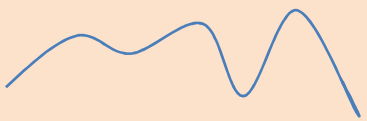


(DNA)



(image processing)

~



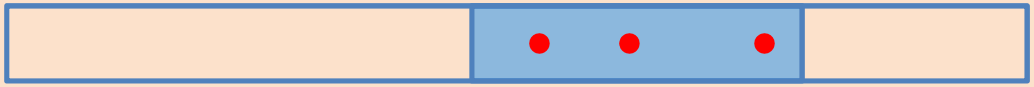
(economics)

Different models

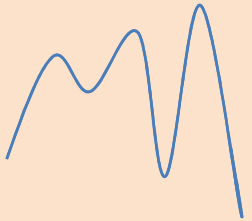
- Circular

(DNA)

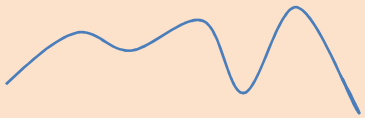
- Approximate



- Abelian ababbbaabbbaab~ aaaaaabbbbbbb
- Order preserving



~



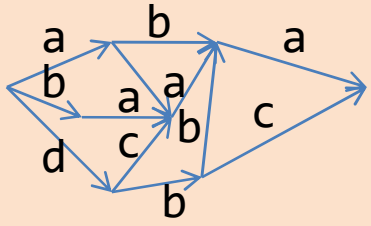
(economics)

- 2D

abcab
bcabc
cabca

(image processing)

- Labelled graphs



(representation of pangenomes)

RECENT EXAMPLES

(IN MY PUBLICATIONS)

Space-Efficient Indexes for Uncertain Strings (ICDE 2024)

X	1	2	3	4	5	6
A	1	1/2	3/4	4/5	1/2	1/4
B	0	1/2	1/4	1/5	1/2	3/4

Space-Efficient Indexes for Uncertain Strings (ICDE 2024)

X	1	2	3	4	5	6
A	1	1/2	3/4	4/5	1/2	1/4
B	0	1/2	1/4	1/5	1/2	3/4

P(X[3..5] = ABA)= 3/40

Space-Efficient Indexes for Uncertain Strings (ICDE 2024)

X	1	2	3	4	5	6
A	1	1/2	3/4	4/5	1/2	1/4
B	0	1/2	1/4	1/5	1/2	3/4

P(X[3..5] = ABA)= 3/40

Occurrence if probablity ≥ 1/z

O(nz) size indexes are too large to be useful.

Our result: O(nz/l) size indexes for patterns of length ≥ l.

Space-Efficient Indexes for Uncertain Strings (ICDE 2024)

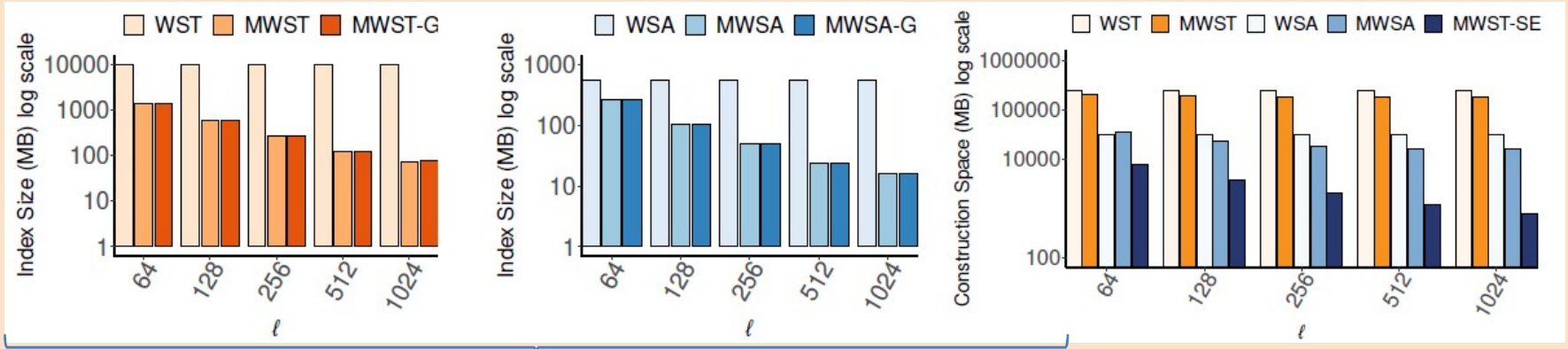
X	1	2	3	4	5	6
A	1	1/2	3/4	4/5	1/2	1/4
B	0	1/2	1/4	1/5	1/2	3/4

P(X[3..5] = ABA)= 3/40

Occurrence if probablity ≥ 1/z

O(nz) size indexes are too large to be useful.

Our result: O(nz/l) size indexes for patterns of length ≥ l.



SARS-Cov-2

EFM

Scalable Order-Preserving Pattern Mining (ICDM 2024)

Pattern mining in time series (IEEE Trans. Cybern. 2023)

Their algorithm finds frequent patterns, but it takes $\Omega(n^3)$ time.

Scalable Order-Preserving Pattern Mining (ICDM 2024)

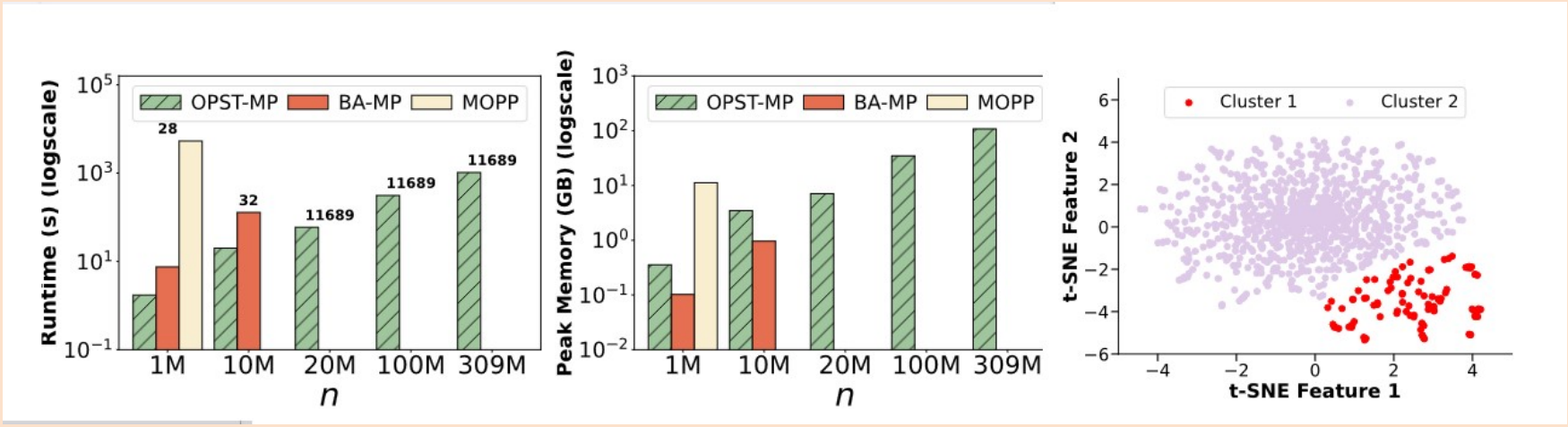
Pattern mining in time series (IEEE Trans. Cybern. 2023)
Their algorithm finds frequent patterns, but it takes $\Omega(n^3)$ time.

Our result: Structure that mines frequent patterns in O(n) time after O(n log σ)-time construction, where σ is the alphabet size.

Scalable Order-Preserving Pattern Mining (ICDM 2024)

Pattern mining in time series (IEEE Trans. Cybern. 2023)
Their algorithm finds frequent patterns, but it takes $\Omega(n^3)$ time.

Our result: Structure that mines frequent patterns in O(n) time after O(n log σ)-time construction, where σ is the alphabet size.



Whale Signals

EEG of 2 groups of patients

Thank you for your attention!