HUAWEI
Software/Hardware Codesign and Performance Optimization Summit 2023
12-14 October 2023, Nizhny Novgorod, Russia

High-performance subsequence anomaly discovery in long time series

明智是了解事件的人
Wise is the person who understands events.
Chinese proverb

SUSU 1943

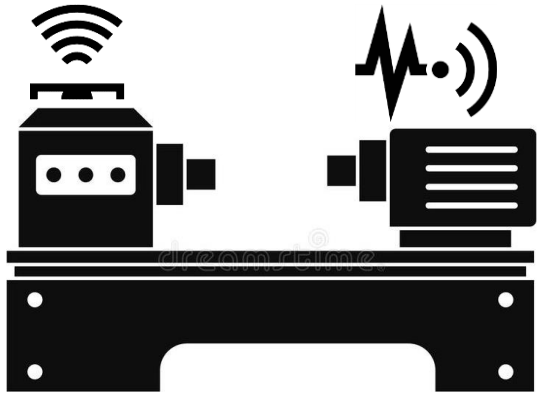
Mikhail Zymbler, Yana Kraeva

{mzym, kraevaya}@susu.ru

Big Data and Machine Learning Lab, South Ural State University, Chelyabinsk, Russia

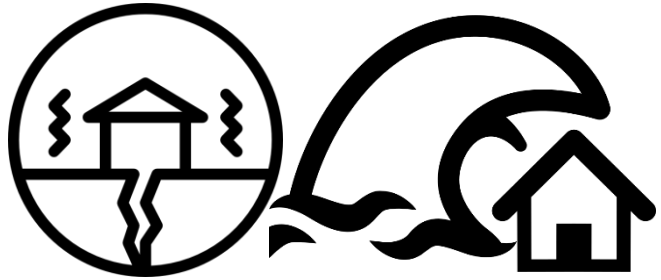
This work is financially supported by the Russian Science Foundation (grant No. 23-21-00465)

Time series are ubiquitous

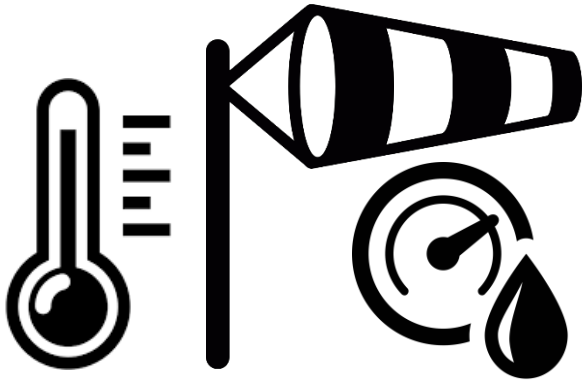


Smart manufacturing,
Predictive maintenance

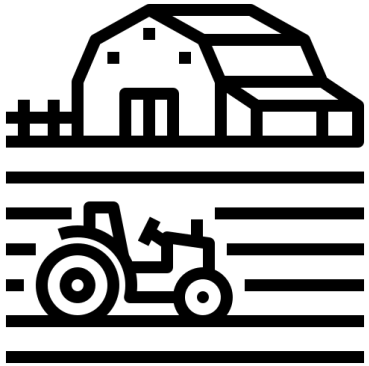
Internet
of Things



Prediction
of natural disasters



Weather forecasting,
Climate modelling

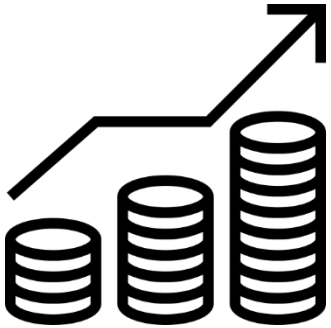


Agriculture
and farming

Personal
healthcare

Chemo-
and bioinformatics

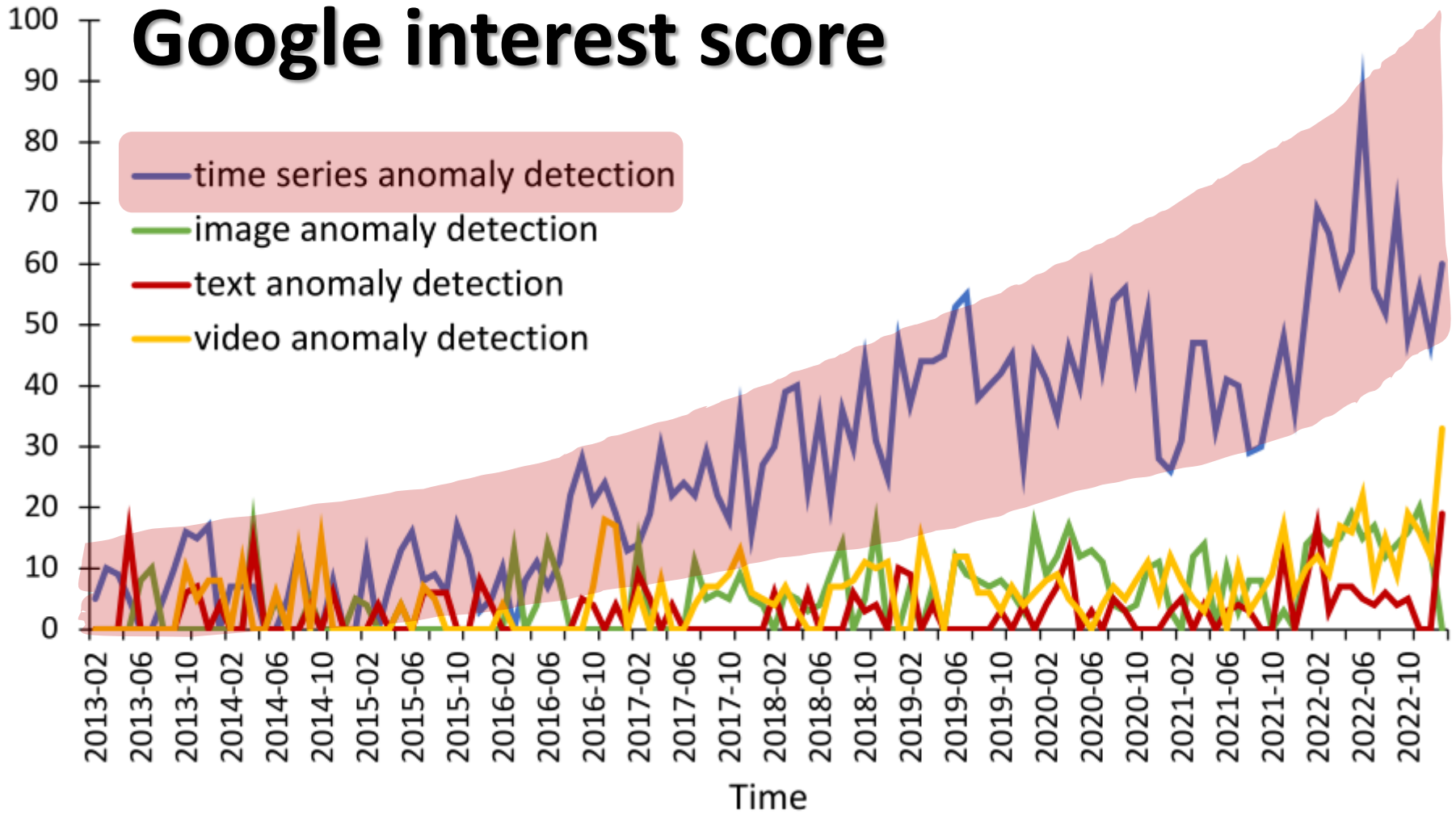
Fighting
crime



Business
and economics

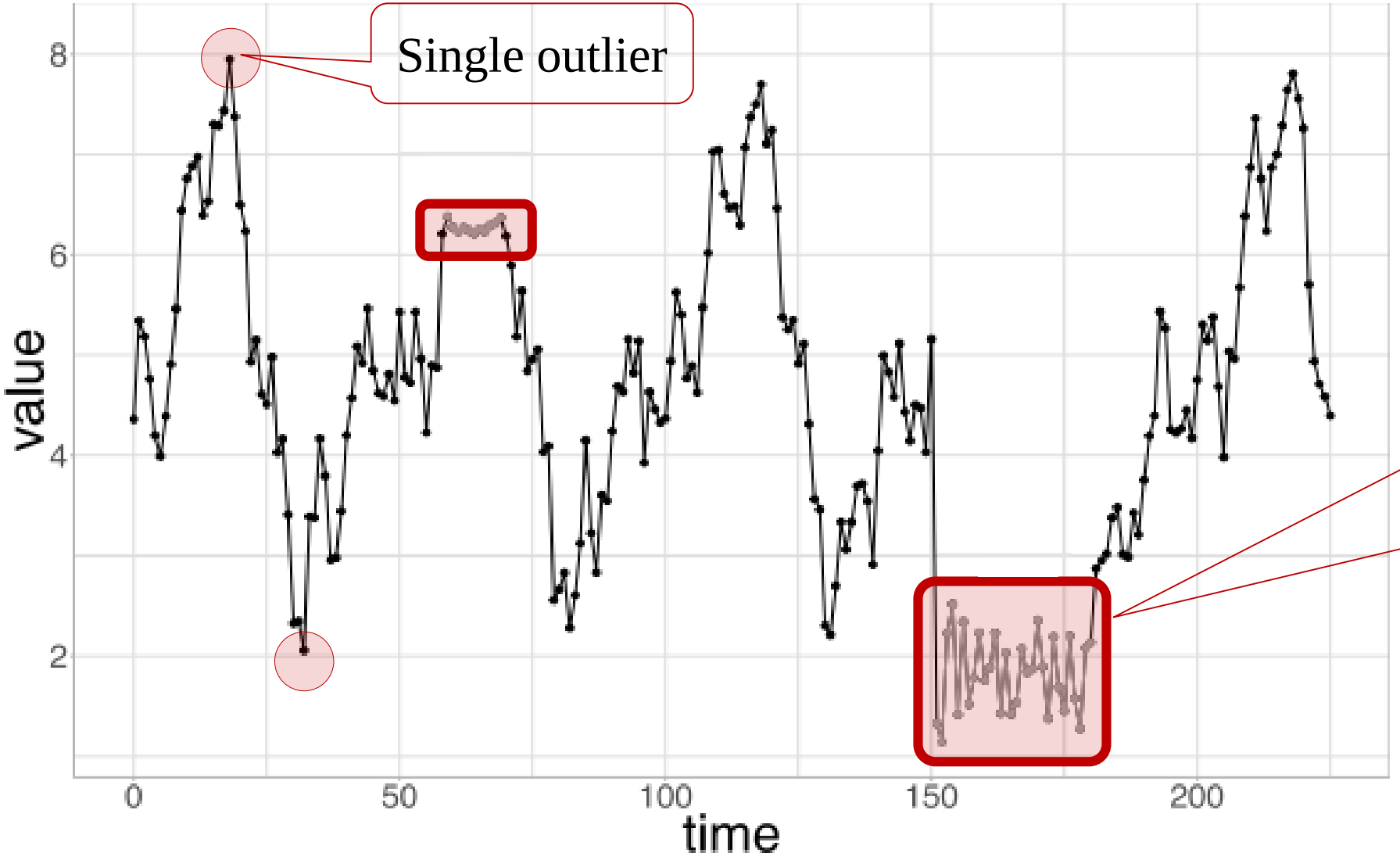
Electronic
education

In time series, people are most interested in **anomalies**



* Boniol P. *et al.* New trends in time-series anomaly detection. EDBT'2023. pp. 847-850. DOI: 10.48786/edbt.2023.80

Subsequence anomaly is the challenge



Collective behavior
of points
is anomalous,
but a point
may not be an outlier

Anomalies are informal and domain-dependent


Homer


Marge


Bart

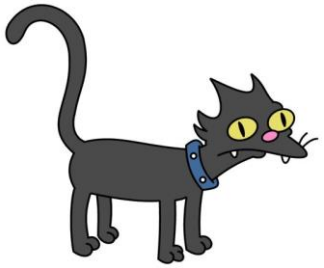

Selma


Patty


Barney


Quasimodo


Itchy

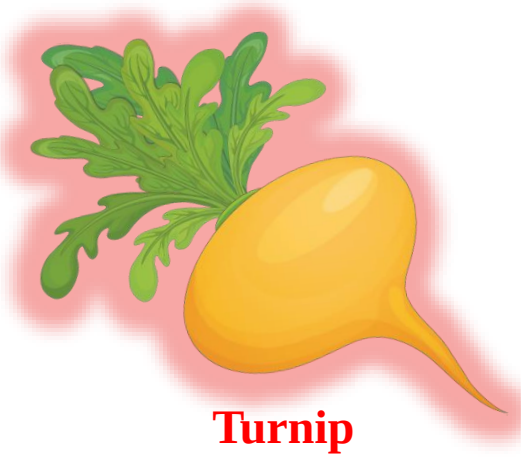

Scratchy


Santa's Little Helper


Lisa

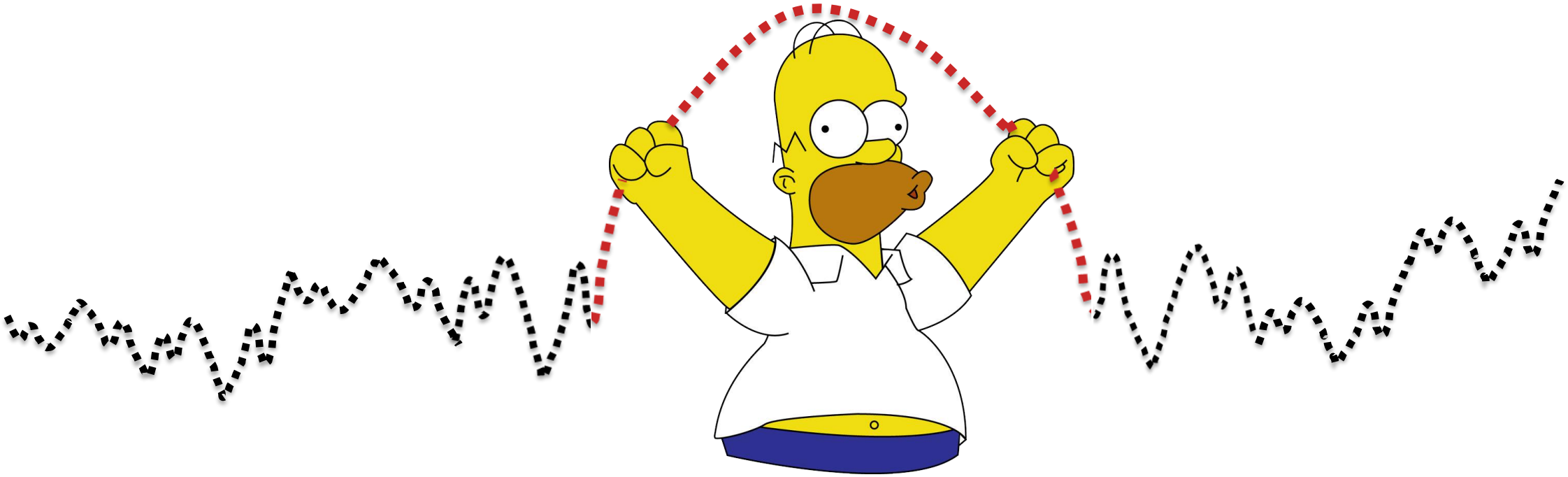

Mona


Abe


Turnip

Discord formalizes subsequence anomaly

Discord* is a subsequence of the given length whose distance to its nearest neighbor is greatest



* Keogh E. *et al.* HOT SAX: Efficiently finding the most unusual time series subsequence. ICDM 2005. pp. 226-233. DOI: 10.1109/ICDM.2005.79

Discord concept

Distance matrix:
the close neighbors,
the similar they are

Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0						
	0					
		0				
			0			
				0		
					0	
						0

Discord concept

Distance matrix
with calculated distances
to neighbors

Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

Discord concept

Distance matrix
with
distances to their nearest neighbors
(i.e. column-wise minima)

 Homer	 Marge	 Bart	 Selma	 Patty	 Barney	 Quasimodo
0	5	**2**	4	4	6	8
5	0	2.5	3	3	6	10
2	**2.5**	0	4	4	6	9
4	3	4	0	**0.5**	**5**	8
4	3	4	**0.5**	0	**5**	8
6	6	6	5	5	0	**7**
8	10	9	8	8	7	0

Discord concept

Distance matrix
with the
farthest distance
to the nearest neighbor
(i.e. maximum
among
column-wise minima)

Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

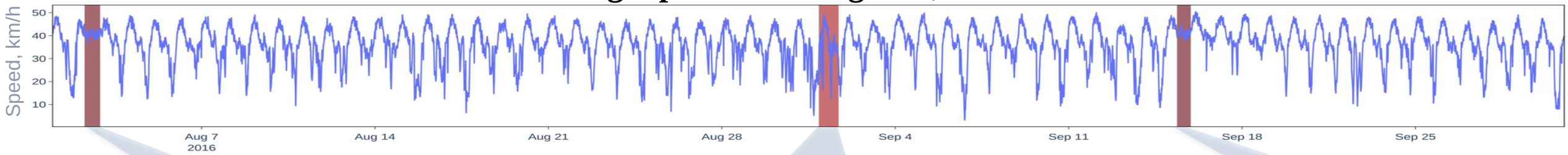
Discord concept

Discord is an object with the **farthest nearest neighbor** (i.e. argument of the maximum among column-wise minima)

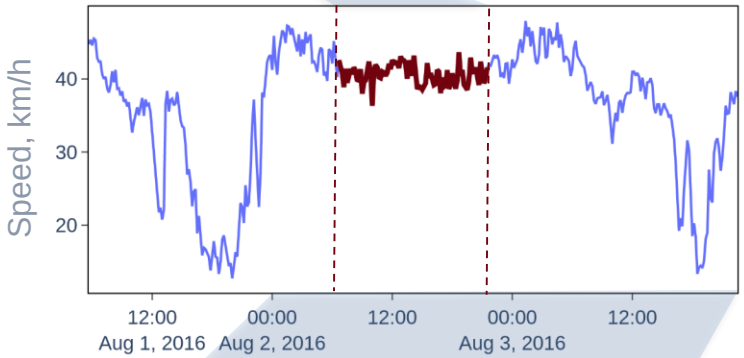
Homer	Marge	Bart	Selma	Patty	Barney	**Quasimodo**
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	**7**
8	10	9	8	8	7	0

Discords reflect anomalies in real life

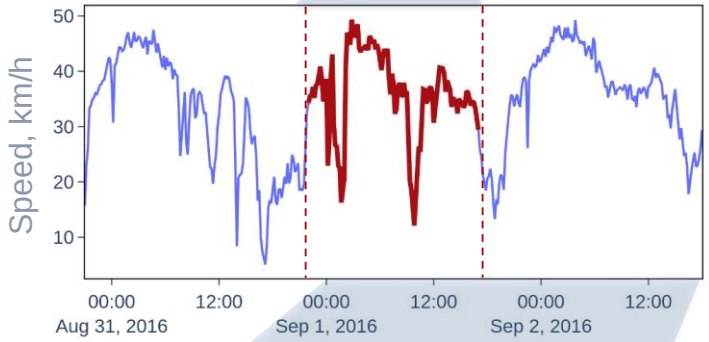
Average speed in Guangzhou, China*



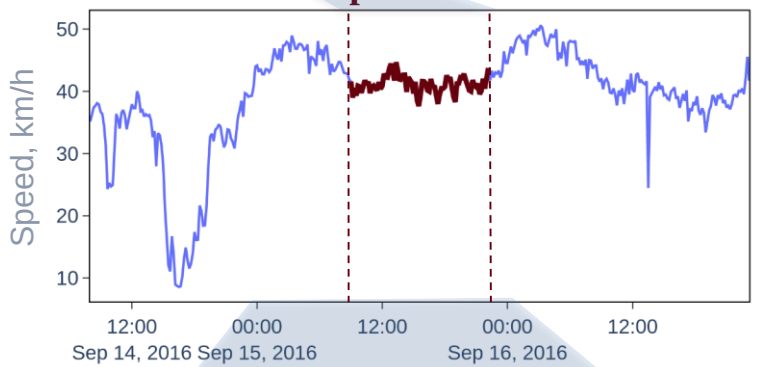
Top-2 discord



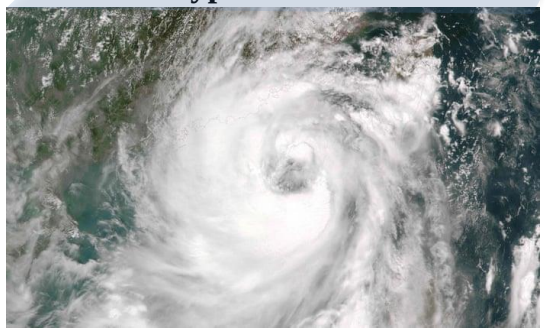
Top-3 discord



Top-1 discord



Typhoon Nida

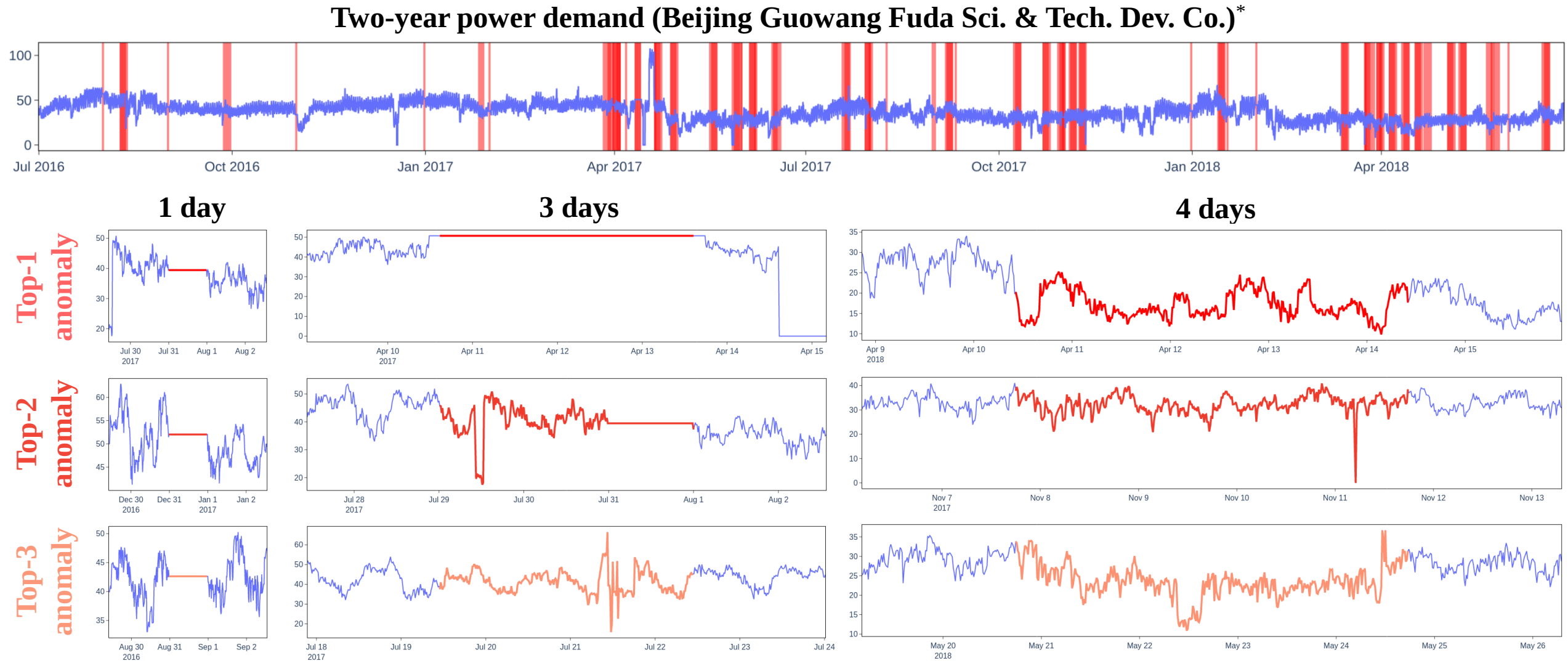


Day of Victory over Japan

Mid-Autumn Festival

* Chen X, Chen Y, He Z. Urban traffic speed dataset of Guangzhou, China. 2018. DOI: 10.5281/zenodo.1205229.

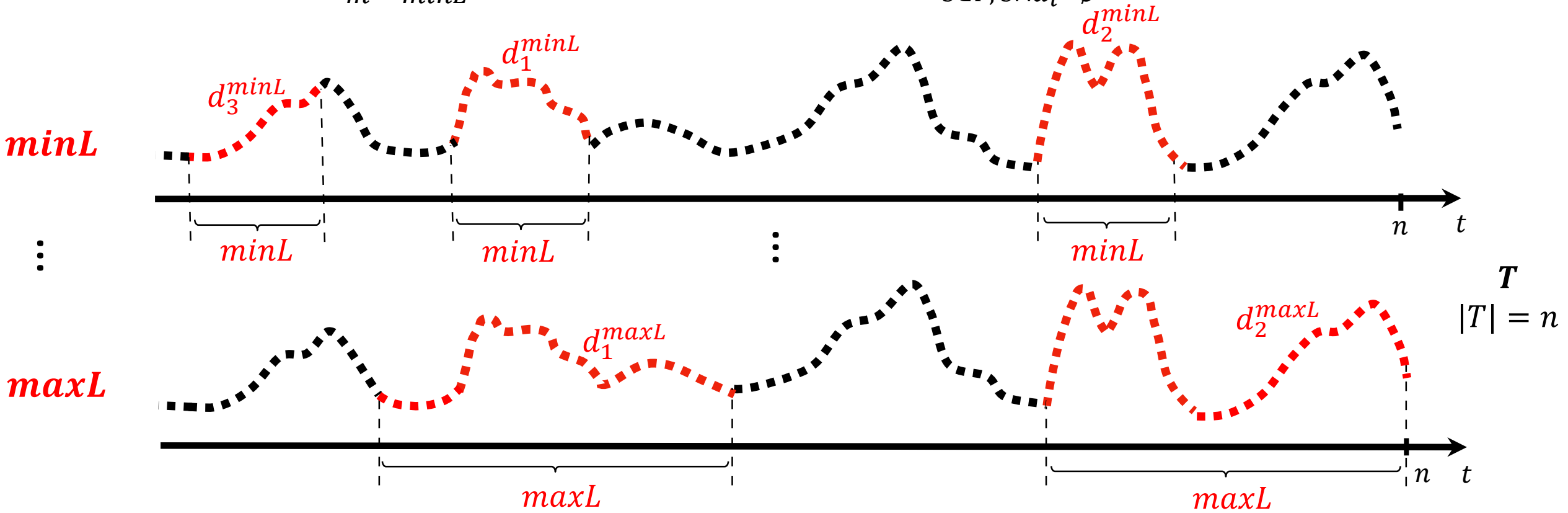
For long series, we need **all-length** subsequence anomalies!



* Zhou H. *et al.* Informer: beyond efficient transformer for long sequence time-series forecasting. AAAI 2021. pp. 11106-11115. DOI: 10.1609/aaai.v35i12.17325.

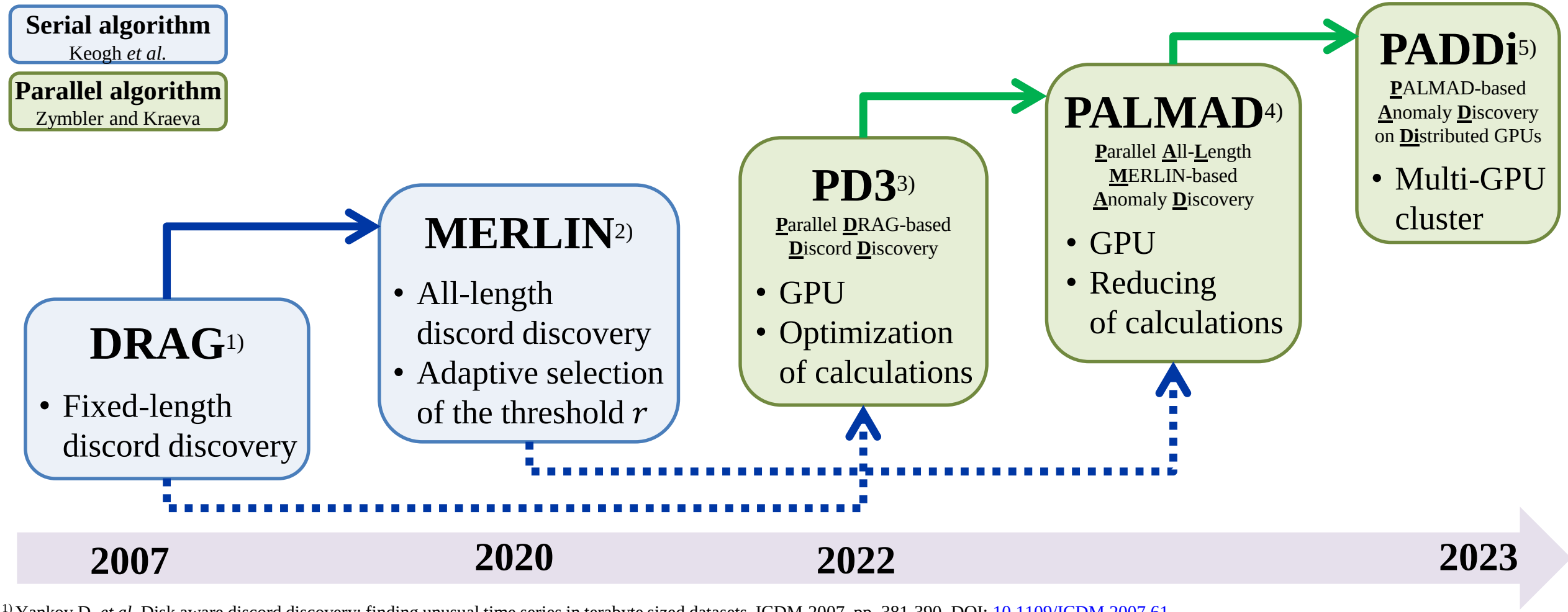
For all-length anomaly subsequence discovery, we need **range discords**

- **Range discord*** is a discord that is at least a **threshold** r away from its nearest neighbor
- **We are given:** time series T, **range of discord length** $minL \dots maxL$
- **We are to find:** $\mathcal{D} = \bigcup_{m=minL}^{maxL} D_m,\ D_m = \{d_1^m, d_2^m, \dots\}$, where $\min_{s \in T,\ s \cap d_i = \emptyset} \text{dist}(d_i^m, s) \geq r$



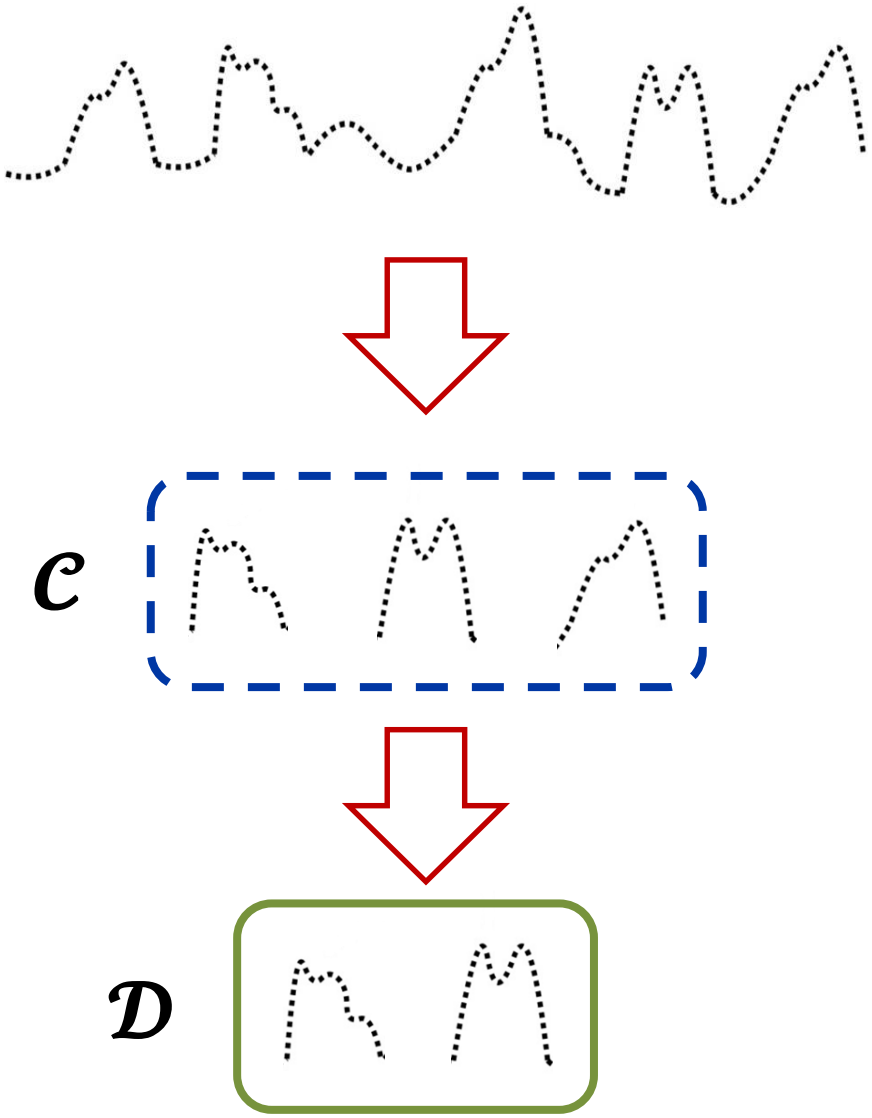
* Yankov D. *et al.* Disk aware discord discovery: finding unusual time series in terabyte sized datasets. ICDM 2007. pp. 381-390. DOI: 10.1109/ICDM.2007.61.

Discord discovery algorithms



[1)] Yankov D. *et al.* Disk aware discord discovery: finding unusual time series in terabyte sized datasets. ICDM 2007. pp. 381-390. DOI: 10.1109/ICDM.2007.61.
[2)] Nakamura T. *et al.* MERLIN: parameter-free discovery of arbitrary length anomalies in massive time series archives. ICDM 2020. pp. 1190-1195. DOI: 10.1109/ICDM50108.2020.00147.
[3)] Kraeva Y., Zymbler M. A parallel discord discovery algorithm for a graphics processor. PRIA. 2023. 33(2). pp. 101-112. DOI: 10.1134/S1054661823020062.
[4)] Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14). art. 3193. DOI: 10.3390/math11143193.
[5)] Kraeva Y., Zymbler M. Anomaly detection in long time series on high-performance cluster with GPUs. Numerical Methods and Programming. 2023. 24(3). pp. 291-304. DOI: 10.26089/NumMet.v24r320.

Discovery of fixed-length discords (DRAG algorithm)



1. Selection

Through one full scan of the time series, create a **set of candidates** to discords

2. Refinement

Through one full scan of the time series, **prune false positive** candidates

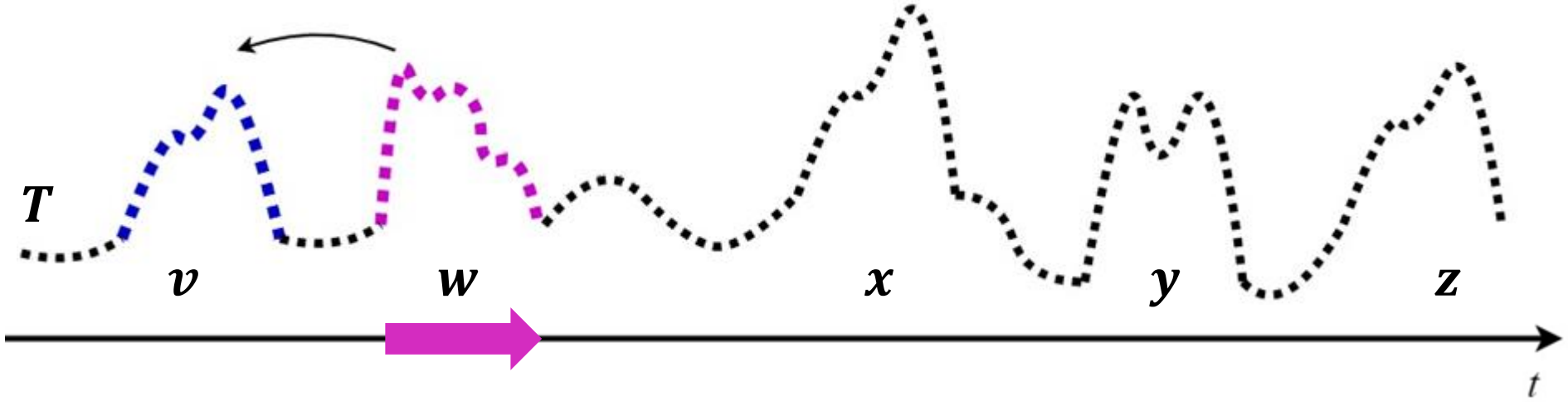
Candidate selection

$\mathcal{C} := \{T_{1,m}\}$
while not end of $\boldsymbol{T}$
    get next subsequence $\boldsymbol{s}$
    $isCandidate :=$ TRUE
    for each $\boldsymbol{c_i} \in \mathcal{C}$ **and** $\boldsymbol{s} \cap \boldsymbol{c_i} = \emptyset$
        if $\text{dist}(\boldsymbol{s}, \boldsymbol{c_i}) < r$ **then**
            $\mathcal{C} := \mathcal{C} \setminus \boldsymbol{c_i}$; $isCandidate :=$ FALSE
    if $isCandidate =$ TRUE **then** $\mathcal{C} := \mathcal{C} \cup \boldsymbol{s}$

$\mathcal{C} = \{\boldsymbol{v}\}$

$\text{dist}(\boldsymbol{w}, \boldsymbol{v}) \geq r$

$\mathcal{C} = \{\boldsymbol{v}, \boldsymbol{w}\}$



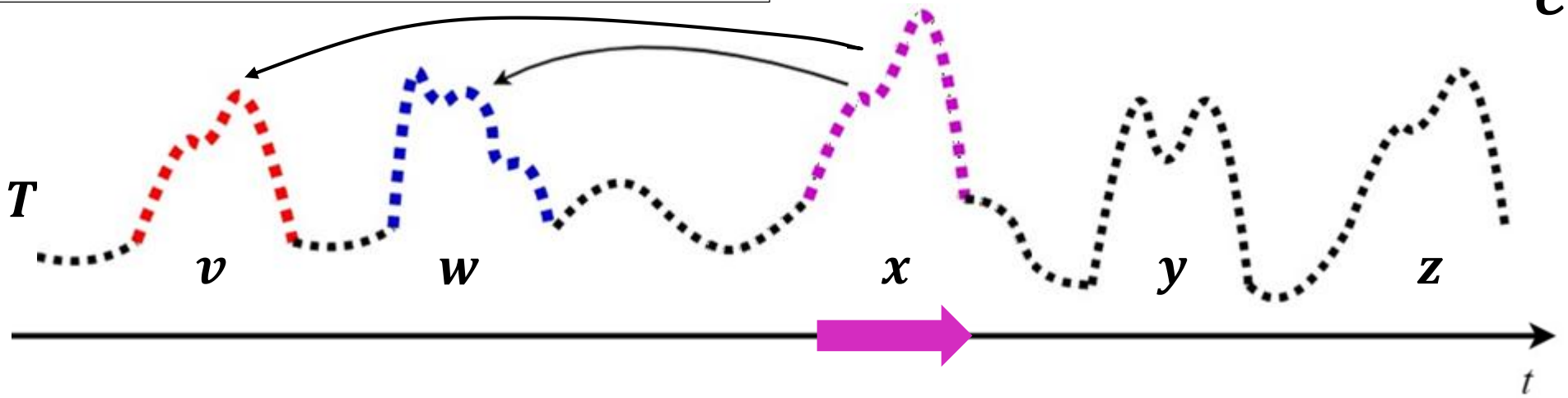
Candidate selection

$\mathcal{C} := \{T_{1,m}\}$
while not end of $\boldsymbol{T}$
    get next subsequence $\boldsymbol{s}$
    $isCandidate :=$ TRUE
    for each $\boldsymbol{c_i} \in \mathcal{C}$ **and** $\boldsymbol{s} \cap \boldsymbol{c_i} = \emptyset$
        if $\text{dist}(\boldsymbol{s}, \boldsymbol{c_i}) < r$ **then**
            $\mathcal{C} := \mathcal{C} \setminus \boldsymbol{c_i}$; $isCandidate :=$ FALSE
    if $isCandidate =$ TRUE **then** $\mathcal{C} := \mathcal{C} \cup \boldsymbol{s}$

$\mathcal{C} = \{\boldsymbol{v}, \boldsymbol{w}\}$

$\text{dist}(\boldsymbol{x}, \boldsymbol{v}) < r$
$\text{dist}(\boldsymbol{x}, \boldsymbol{w}) \geq r$

$\mathcal{C} = \{\boldsymbol{w}\}$



Candidate selection

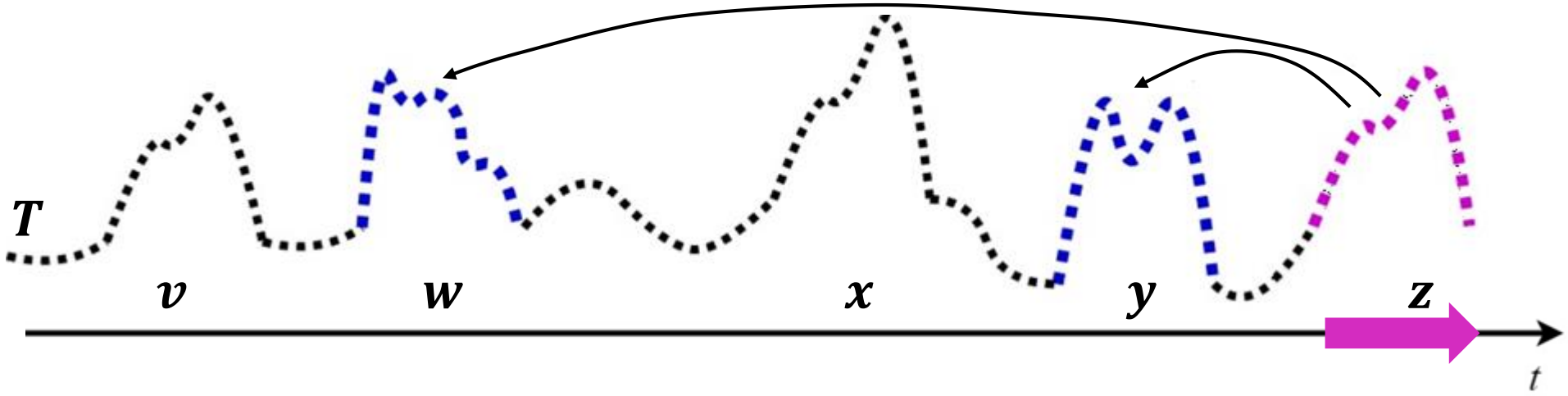
$\mathcal{C} := \{T_{1,m}\}$
while not end of $\boldsymbol{T}$
    get next subsequence $\boldsymbol{s}$
    $isCandidate :=$ TRUE
    for each $\boldsymbol{c_i} \in \mathcal{C}$ **and** $\boldsymbol{s} \cap \boldsymbol{c_i} = \emptyset$
        if $\text{dist}(\boldsymbol{s}, \boldsymbol{c_i}) < r$ **then**
            $\mathcal{C} := \mathcal{C} \setminus \boldsymbol{c_i}$; $isCandidate :=$ FALSE
    if $isCandidate =$ TRUE **then** $\mathcal{C} := \mathcal{C} \cup \boldsymbol{s}$

$\mathcal{C} = \{\boldsymbol{w}, \boldsymbol{y}\}$

$\text{dist}(\boldsymbol{z}, \boldsymbol{w}) \geq r$

$\text{dist}(\boldsymbol{z}, \boldsymbol{y}) \geq r$

$\mathcal{C} = \{\boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}\}$



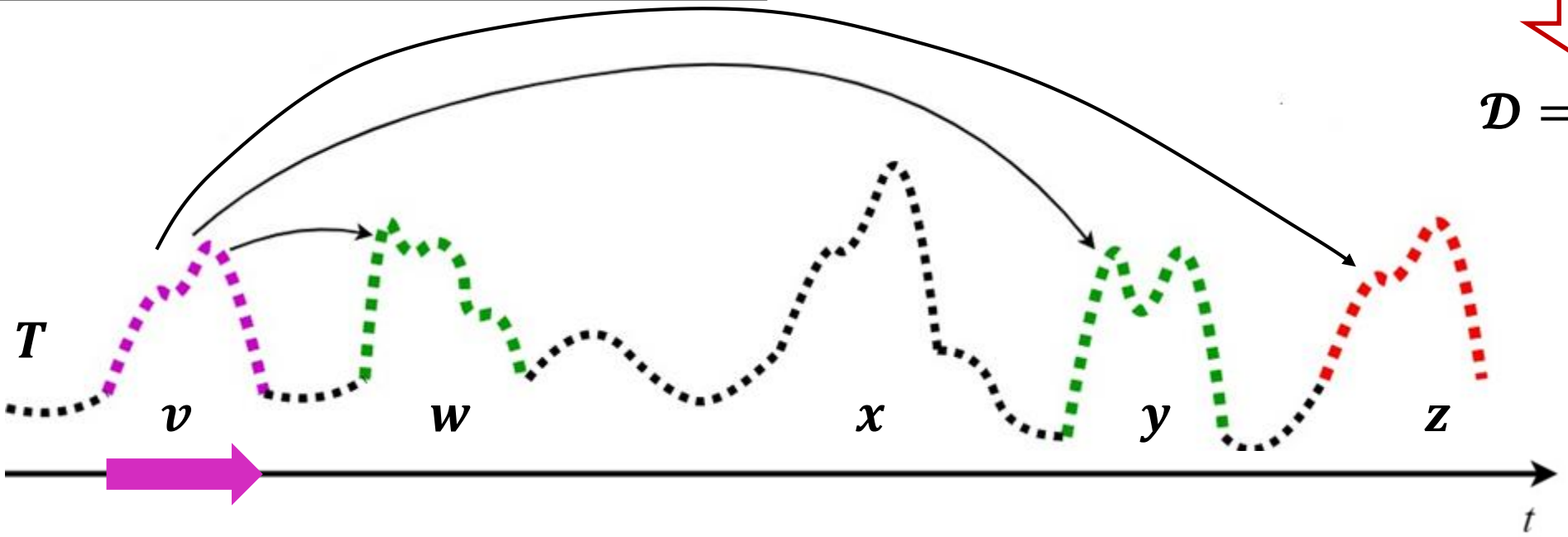
Discord refinement

$\mathcal{D} := \mathcal{C}$
while not end of T
  get next subsequence s
  for each $d_i \in \mathcal{D}$ **and** $s \cap d_i = \emptyset$
    if $\text{dist}(s, d_i) < r$ **then**
      $\mathcal{D} := \mathcal{D} \setminus d_i$

$\mathcal{D} = \{w, y, z\}$

$\text{dist}(v, w) \geq r$
$\text{dist}(v, y) \geq r$
$\text{dist}(v, z) < r$

$\mathcal{D} = \{w, y\}$



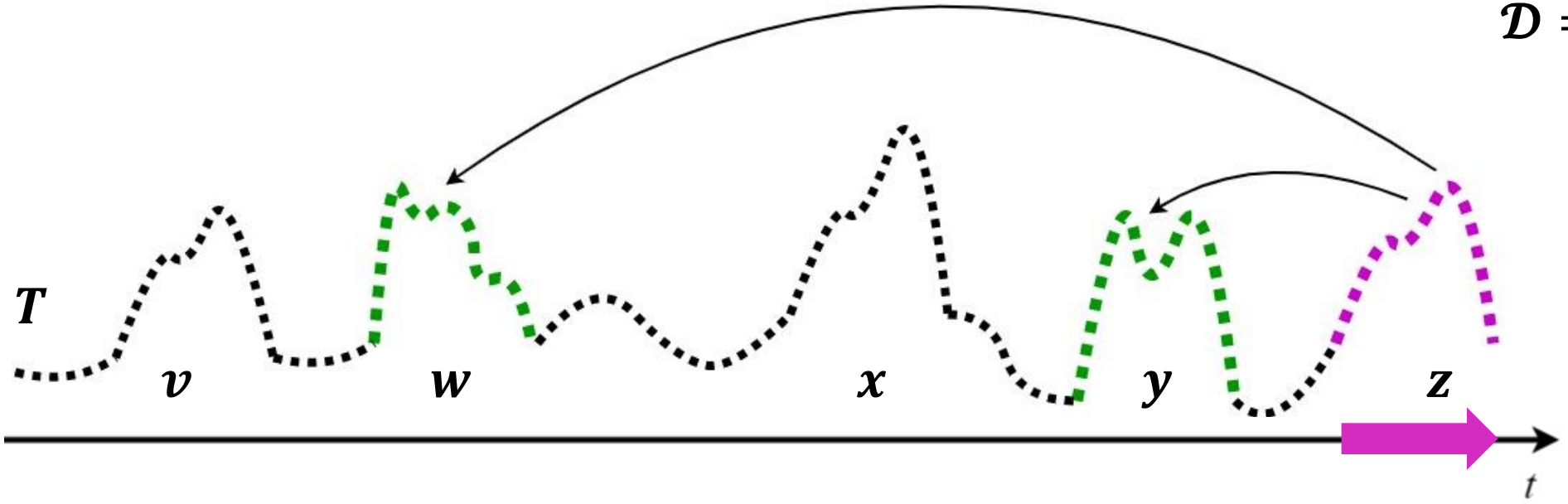
Discord refinement

$\mathcal{D} := \mathcal{C}$
while not end of T
    get next subsequence s
    for each $d_i \in \mathcal{D}$ **and** $s \cap d_i = \emptyset$
        if $\text{dist}(s, d_i) < r$ **then**
            $\mathcal{D} := \mathcal{D} \setminus d_i$

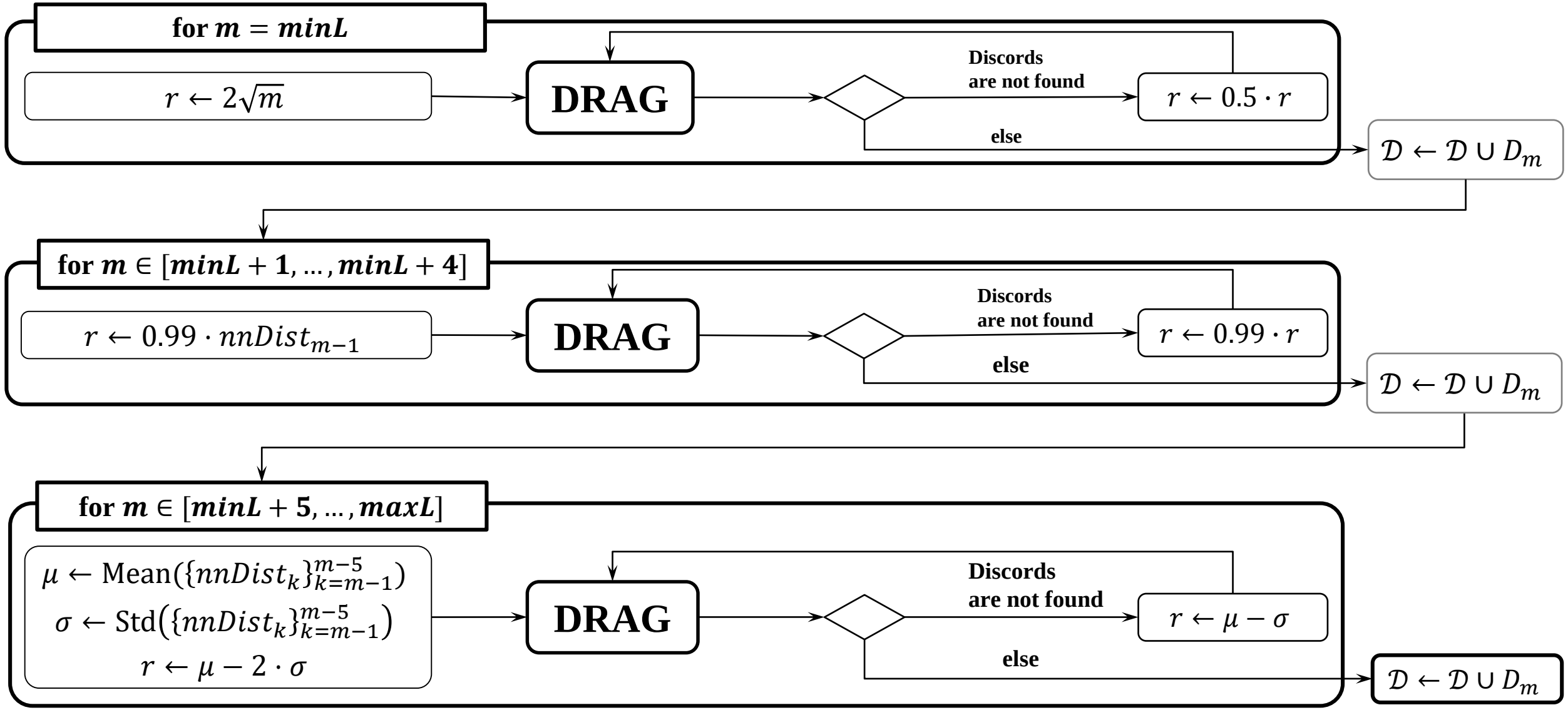
$\mathcal{D} = \{w, y\}$

$\text{dist}(z, w) \geq r$
$\text{dist}(z, y) \geq r$

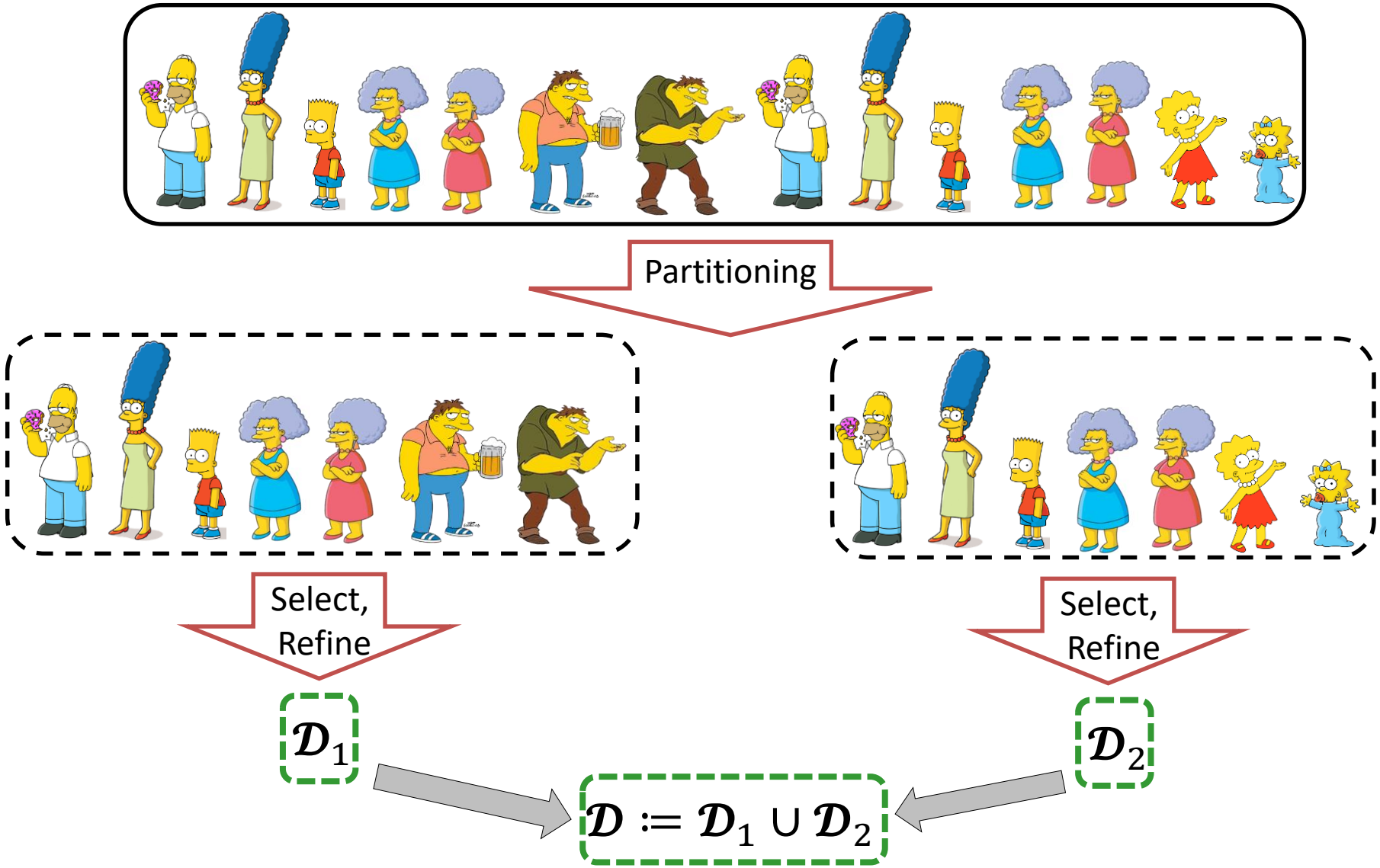
$\mathcal{D} = \{w, y\}$



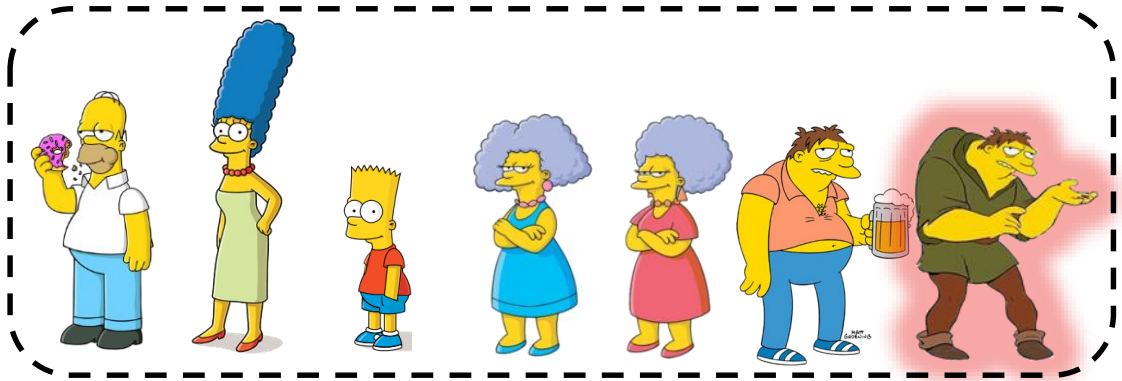
Discovery of all-length discords (MERLIN algorithm)



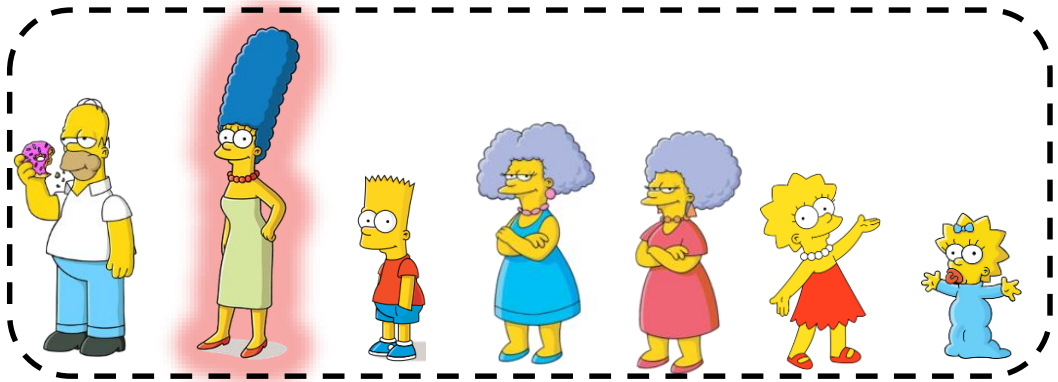
Can we discover discords in parallel?



Naïve parallel discord discovery does not work

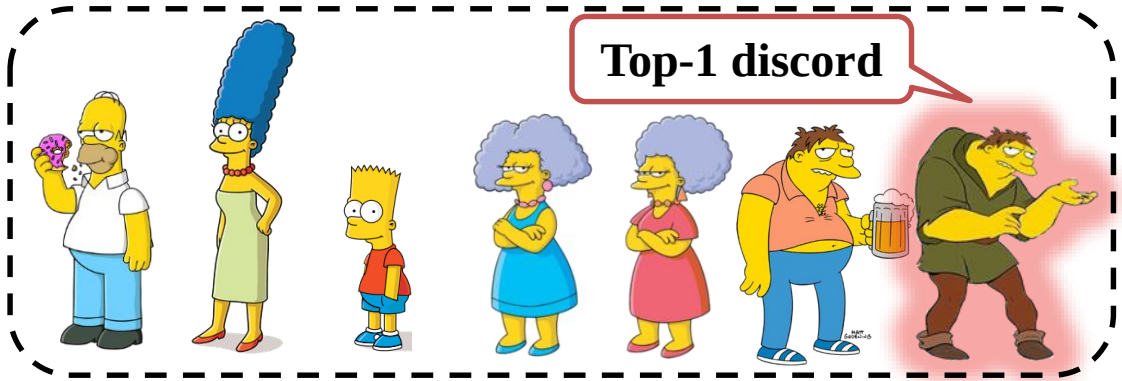


0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	**7**
8	10	9	8	8	7	0

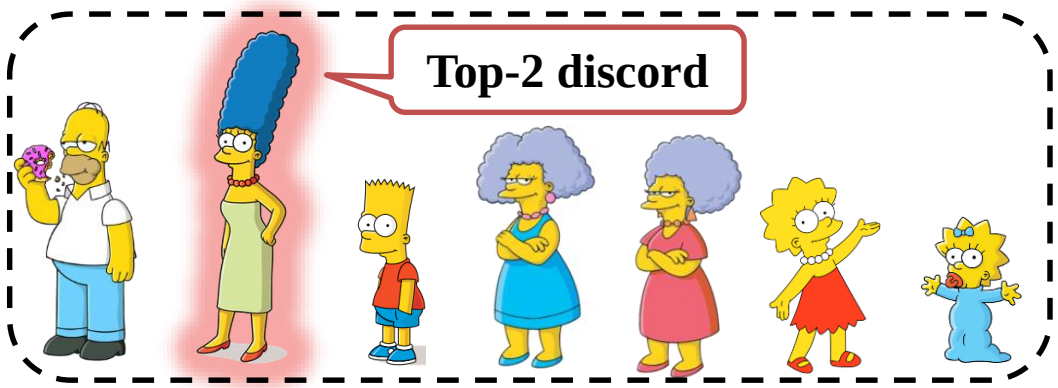


0	5	2	4	4	2.5	2.5
5	0	2.5	3	3	2	2
2	**2. 5**	0	4	4	0.5	0.5
4	3	4	0	0.5	4	4
4	3	4	0.5	0	4	4
2.5	2	0.5	4	4	0	0.5
2.5	2	0.5	4	4	0.5	0

Naïve parallel discord discovery does not work

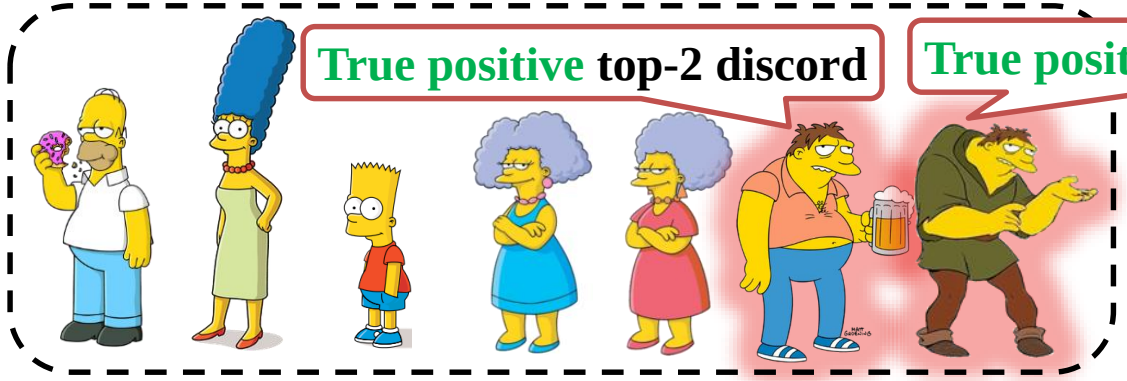


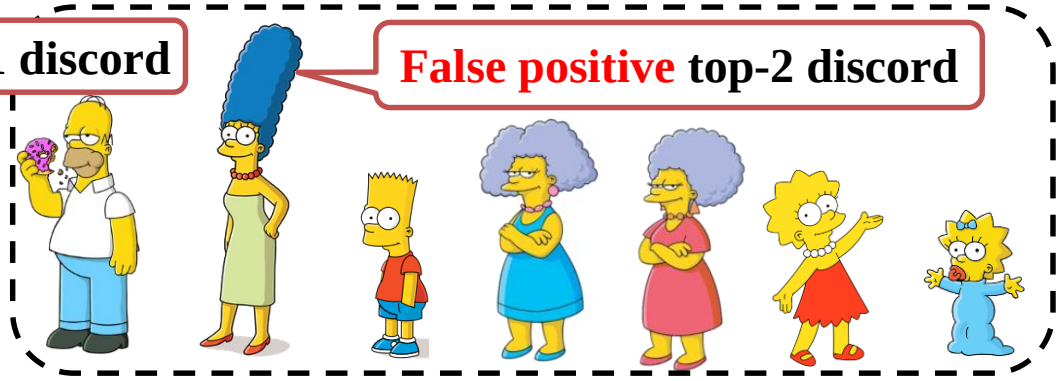
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	**7**
8	10	9	8	8	7	0



0	5	2	4	4	2.5	2.5
5	0	2.5	3	3	2	2
2	**2. 5**	0	4	4	0.5	0.5
4	3	4	0	0.5	4	4
4	3	4	0.5	0	4	4
2.5	2	0.5	4	4	0	0.5
2.5	2	0.5	4	4	0.5	0

Naïve parallel discord discovery does not work

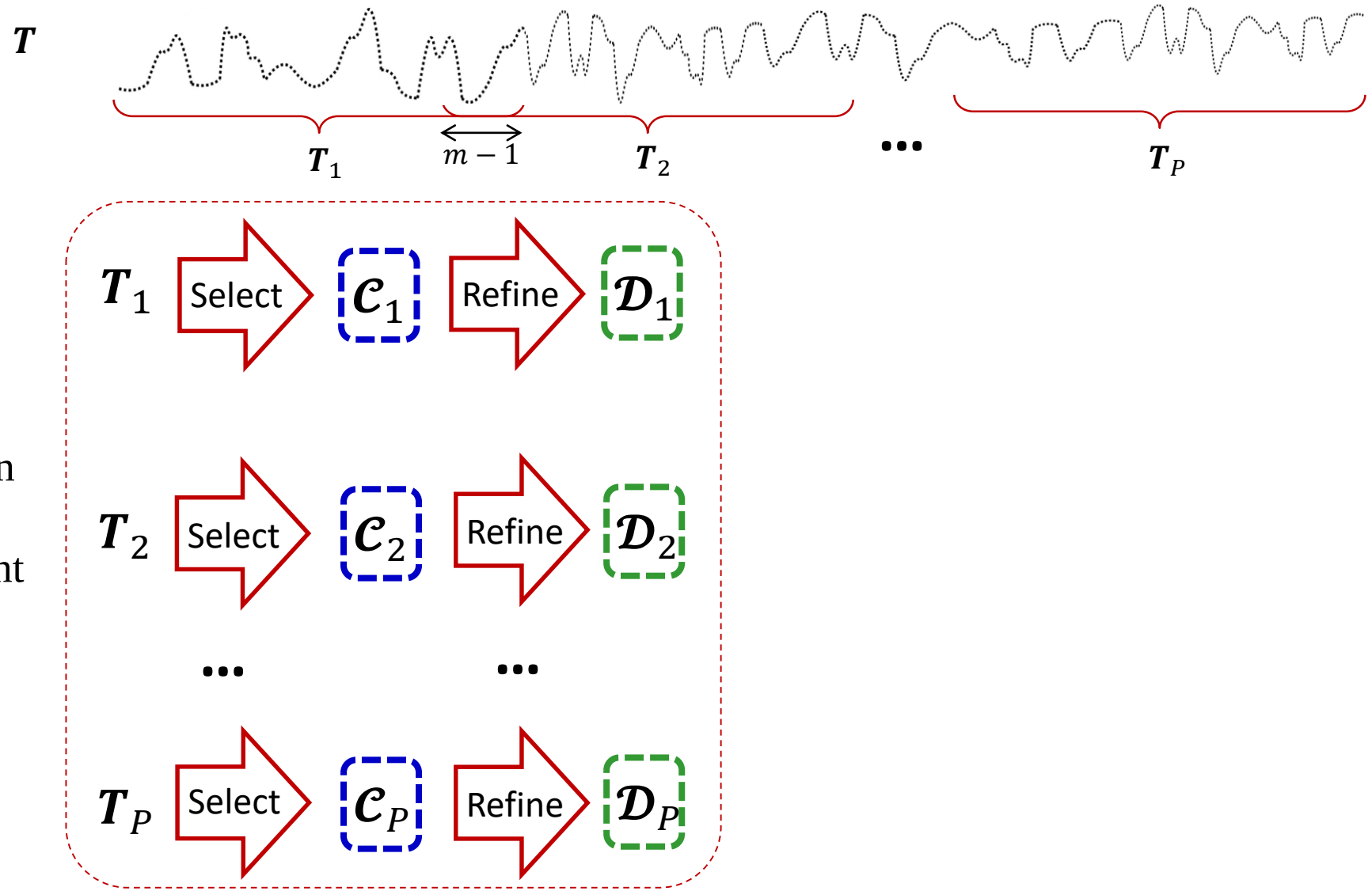




0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	**5**	8
4	3	4	0.5	0	**5**	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

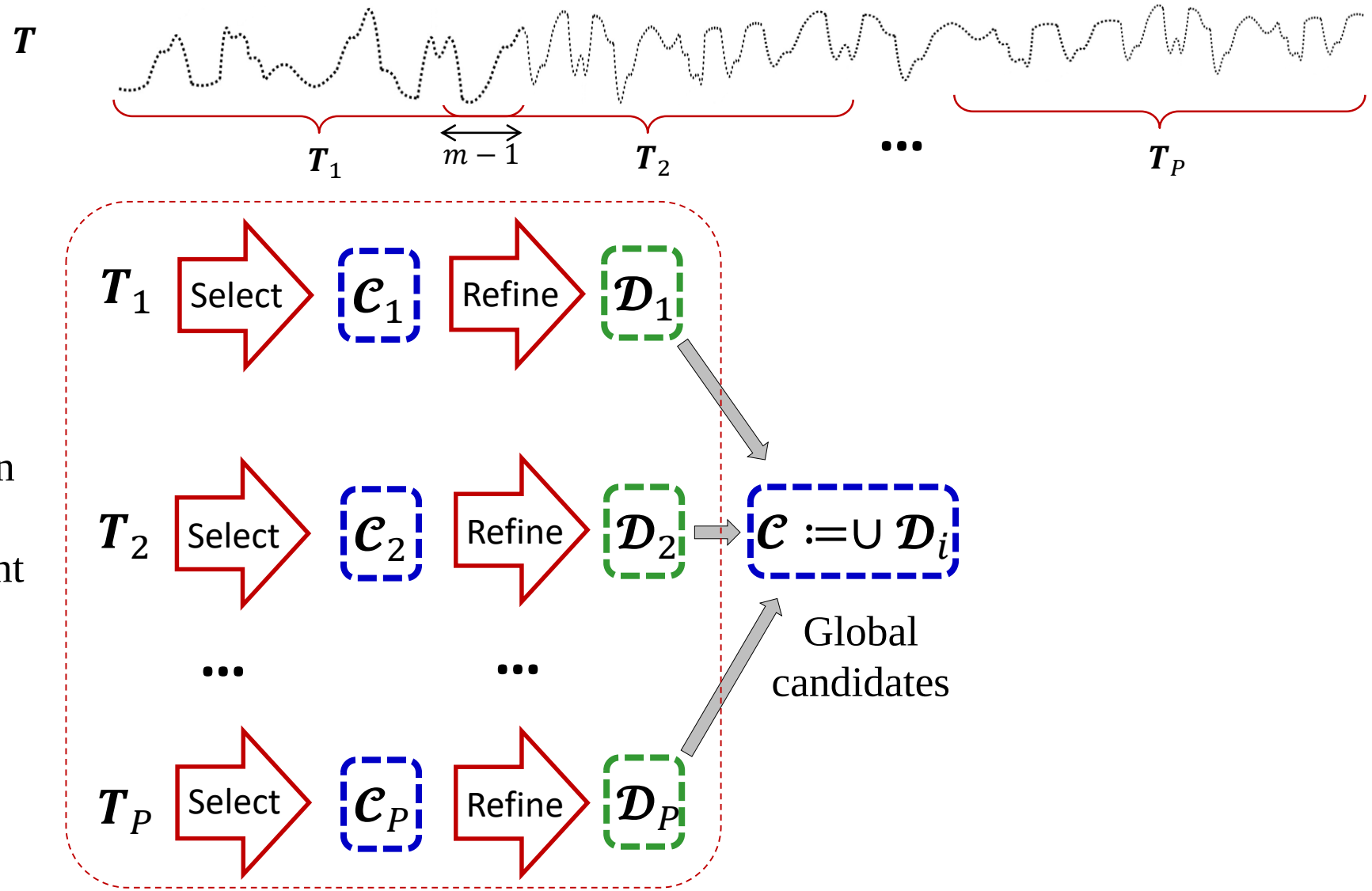
0	5	2	4	4	2.5	2.5
5	0	2.5	3	3	2	2
2	**2. 5**	0	4	4	0.5	0.5
4	3	4	0	0.5	4	4
4	3	4	0.5	0	4	4
2.5	2	0.5	4	4	0	0.5
2.5	2	0.5	4	4	0.5	0

Parallel discovery: Local discords must be globally refined



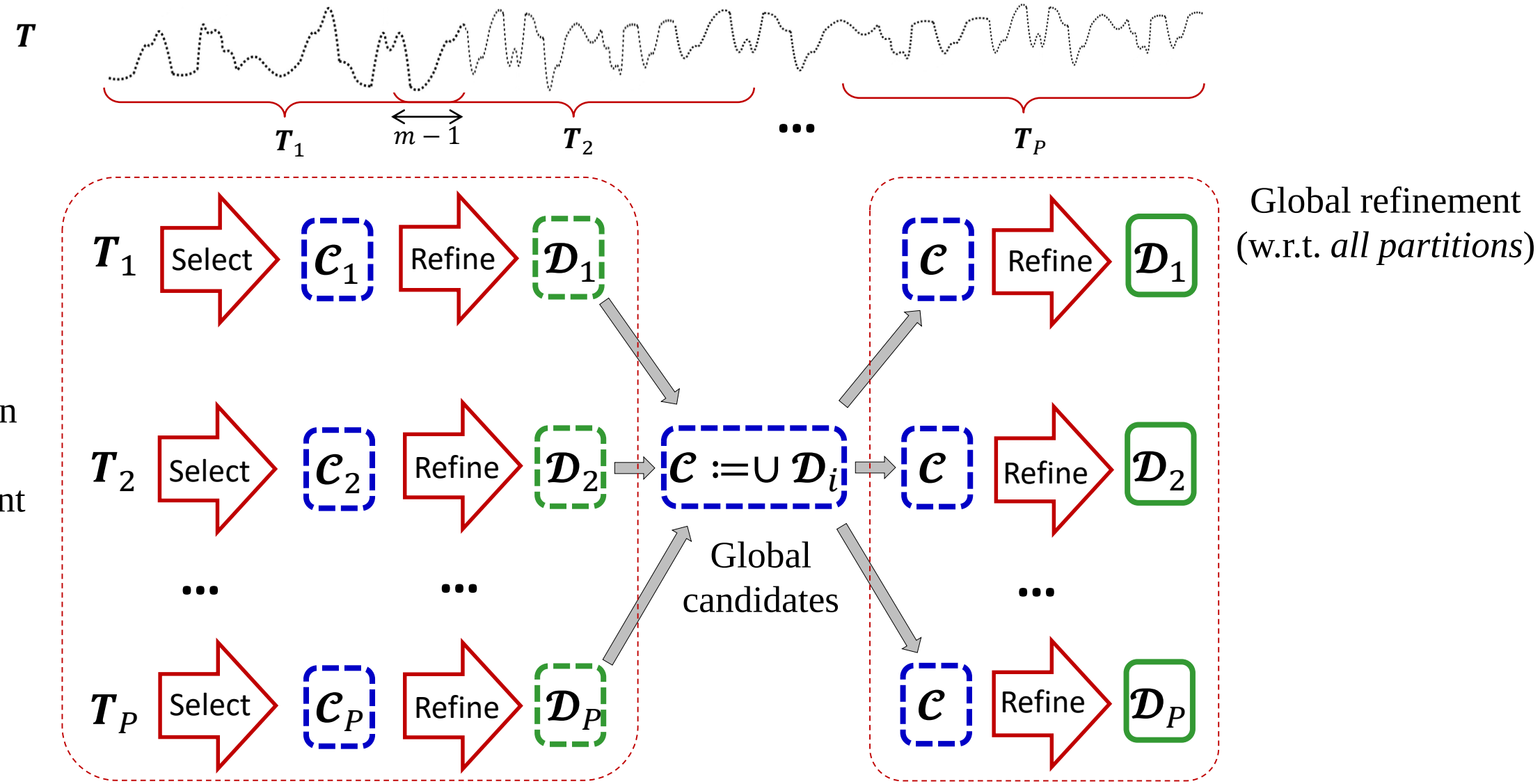
Local selection and local refinement (w.r.t. *each partition*)

Parallel discovery: Local discords must be globally refined

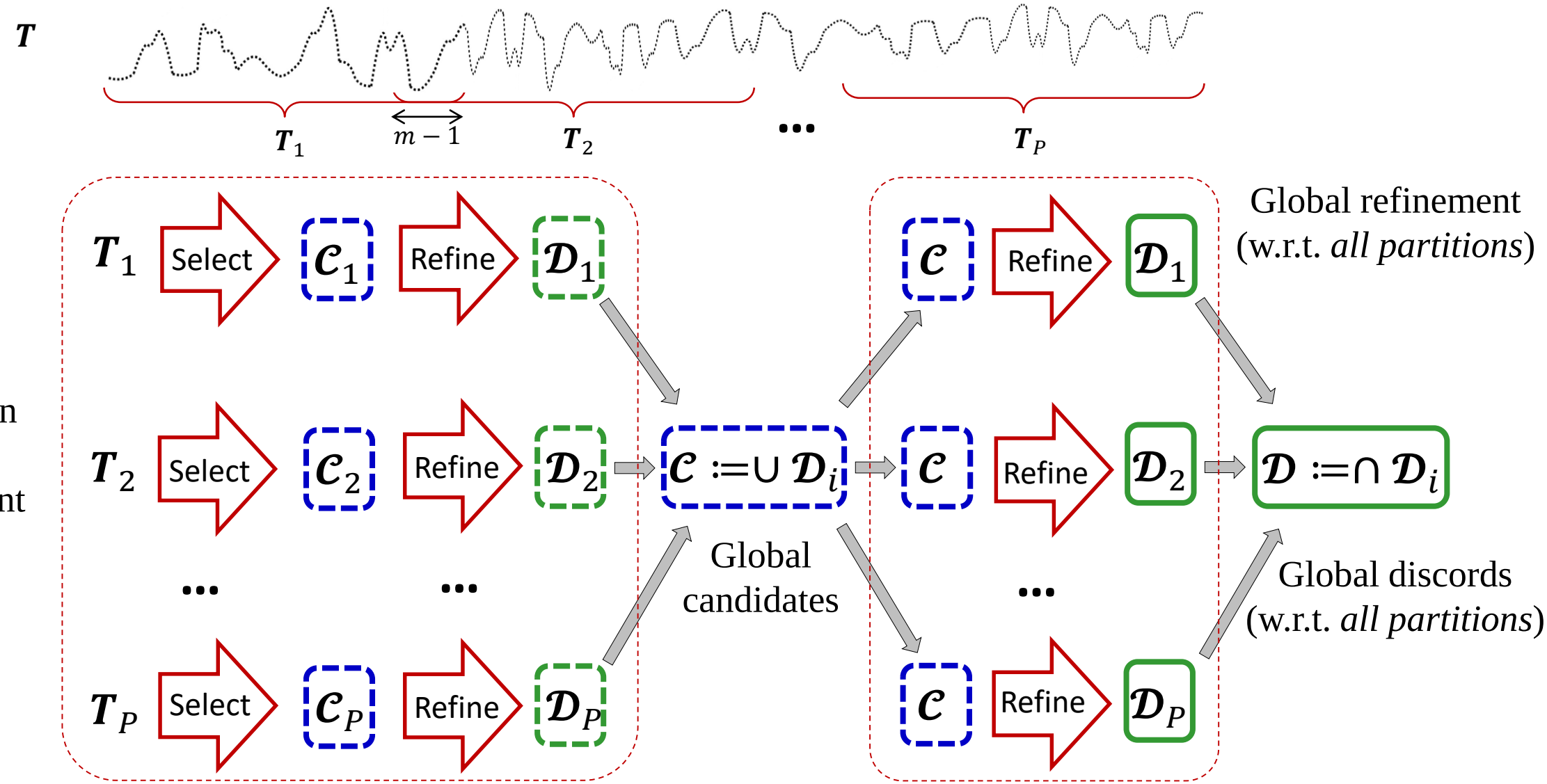


Local selection and local refinement (w.r.t. *each partition*)

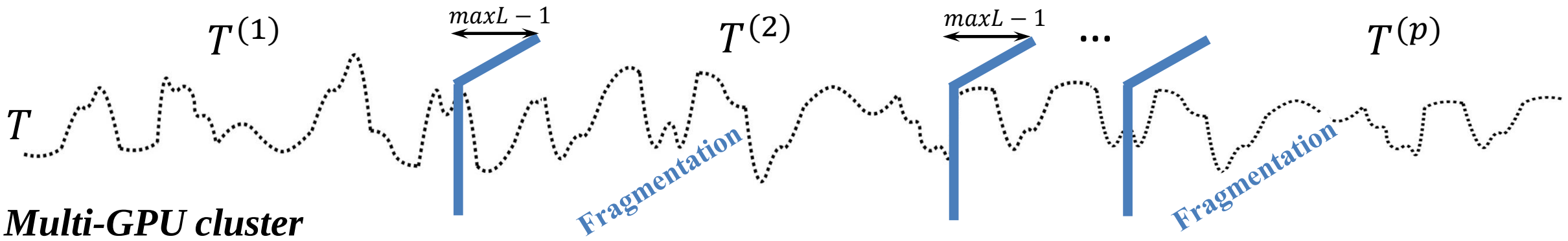
Parallel discovery: Local discords must be globally refined



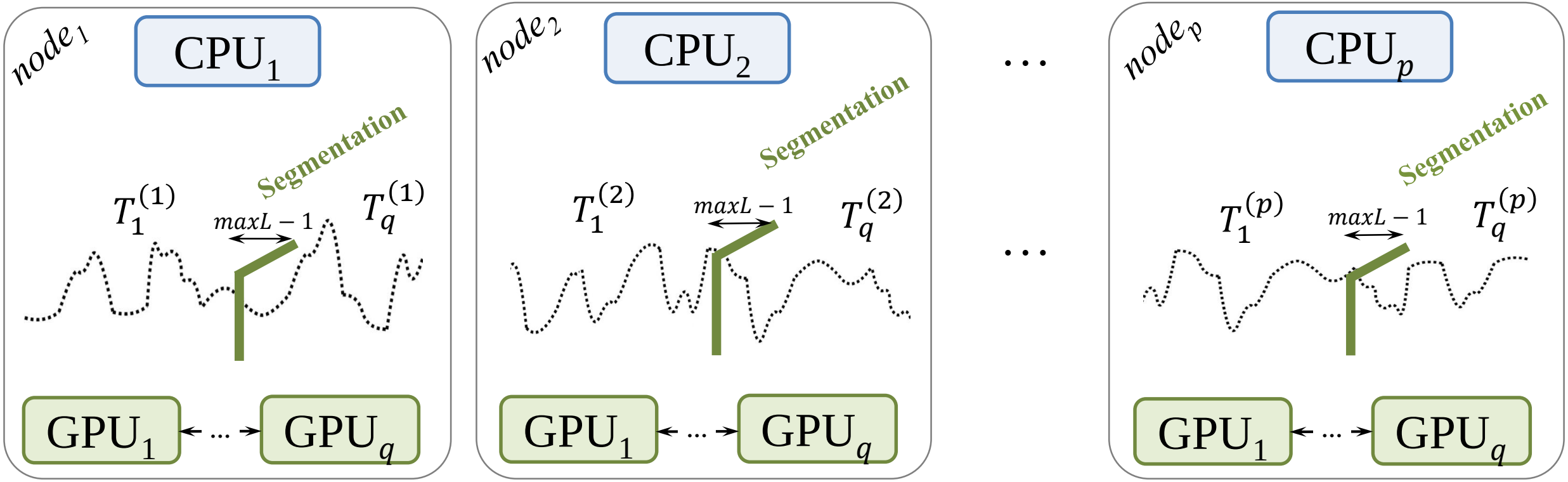
Parallel discovery: Local discords must be globally refined



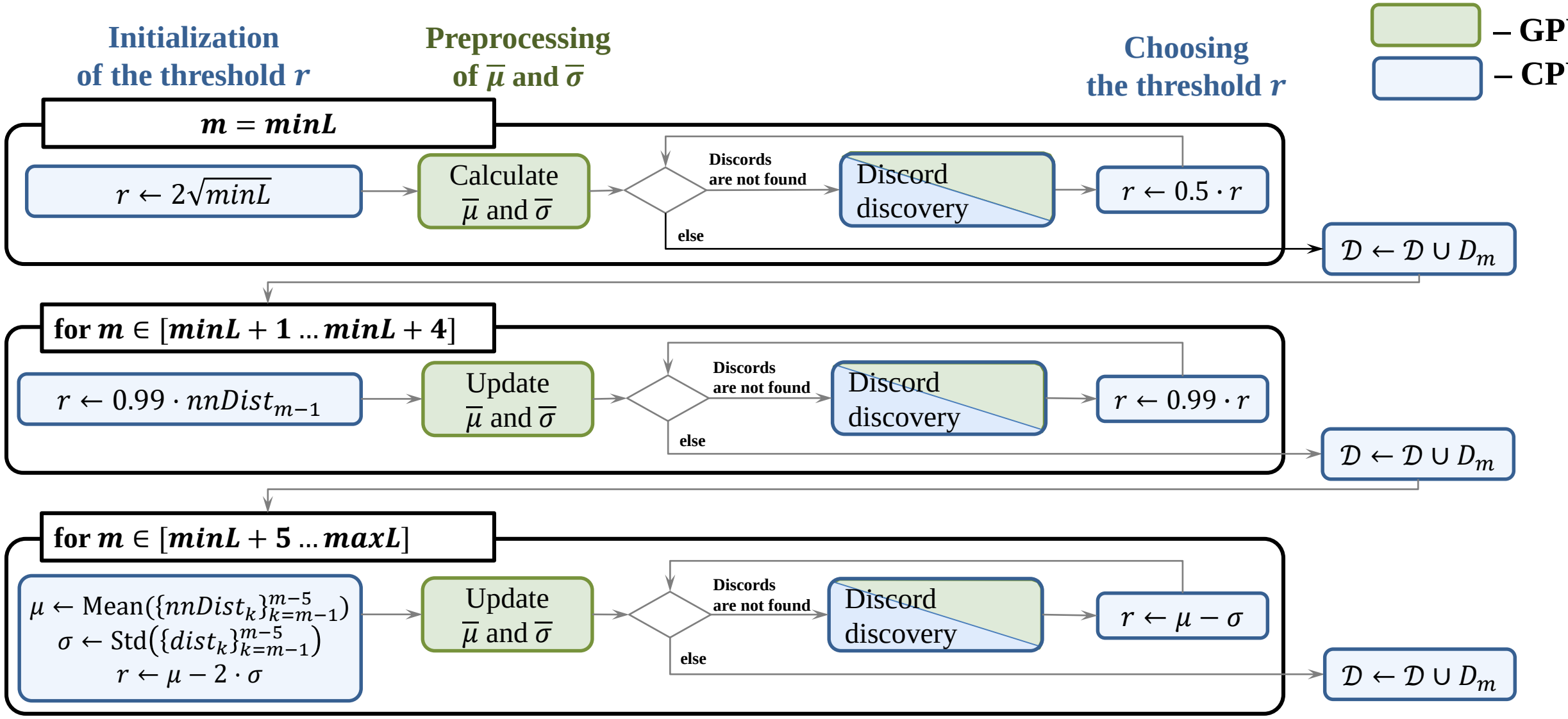
Two-level data parallelism (PADDi algorithm)



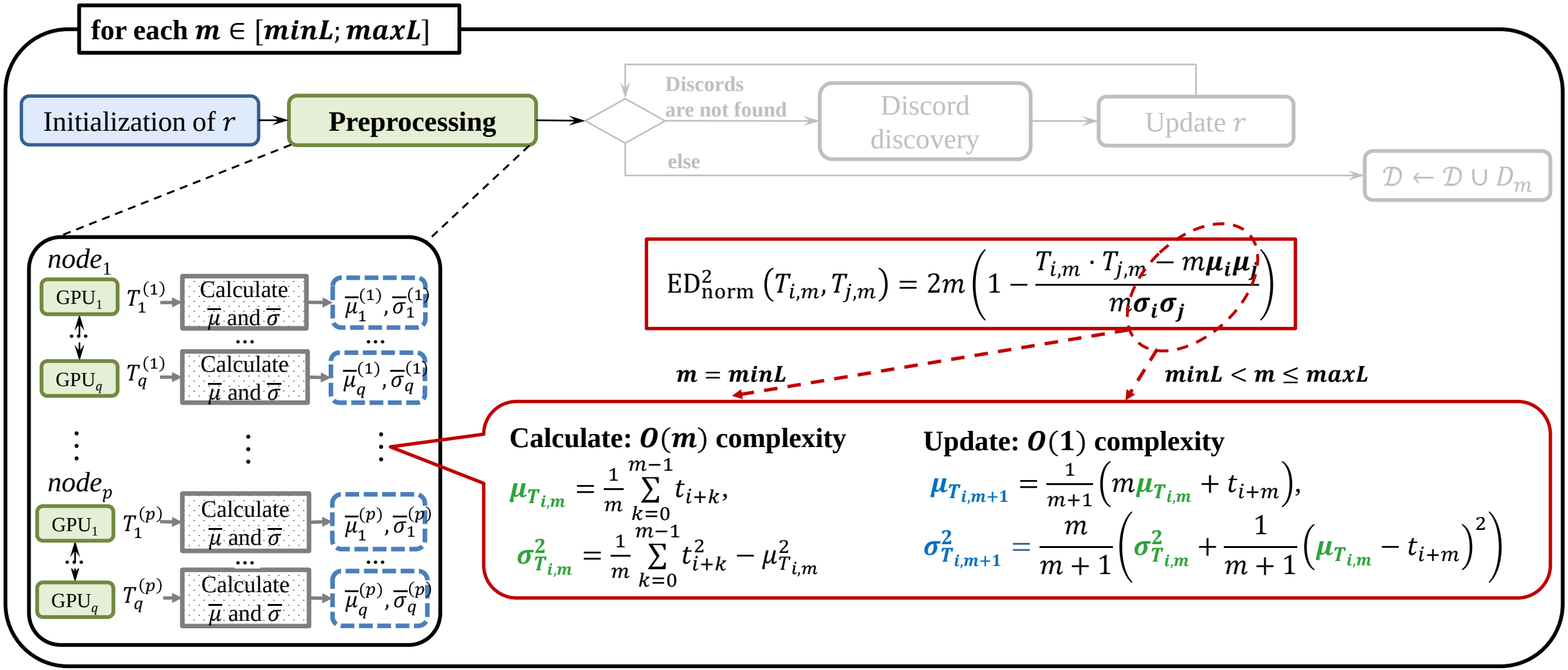
Multi-GPU cluster



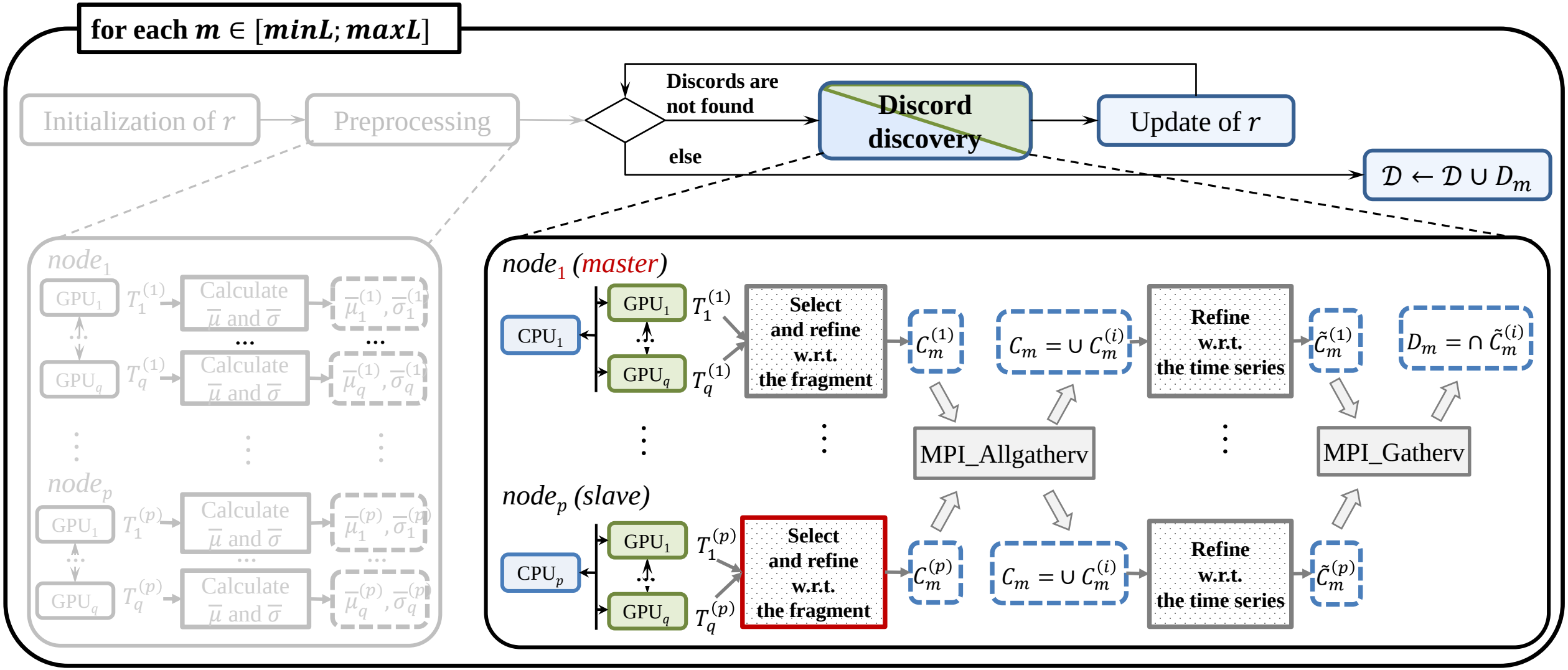
General scheme of calculations



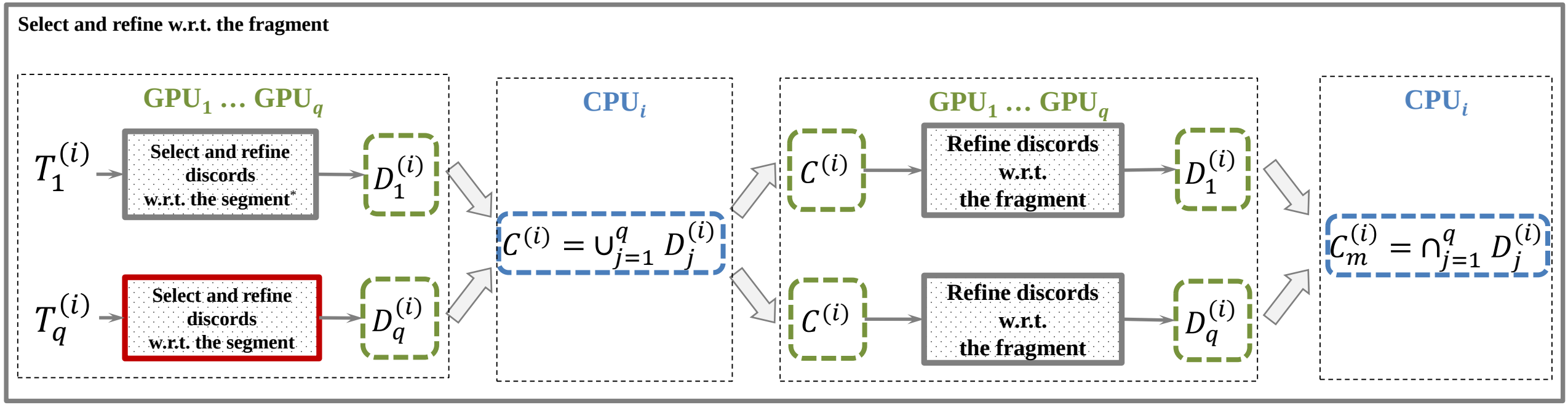
Preprocessing



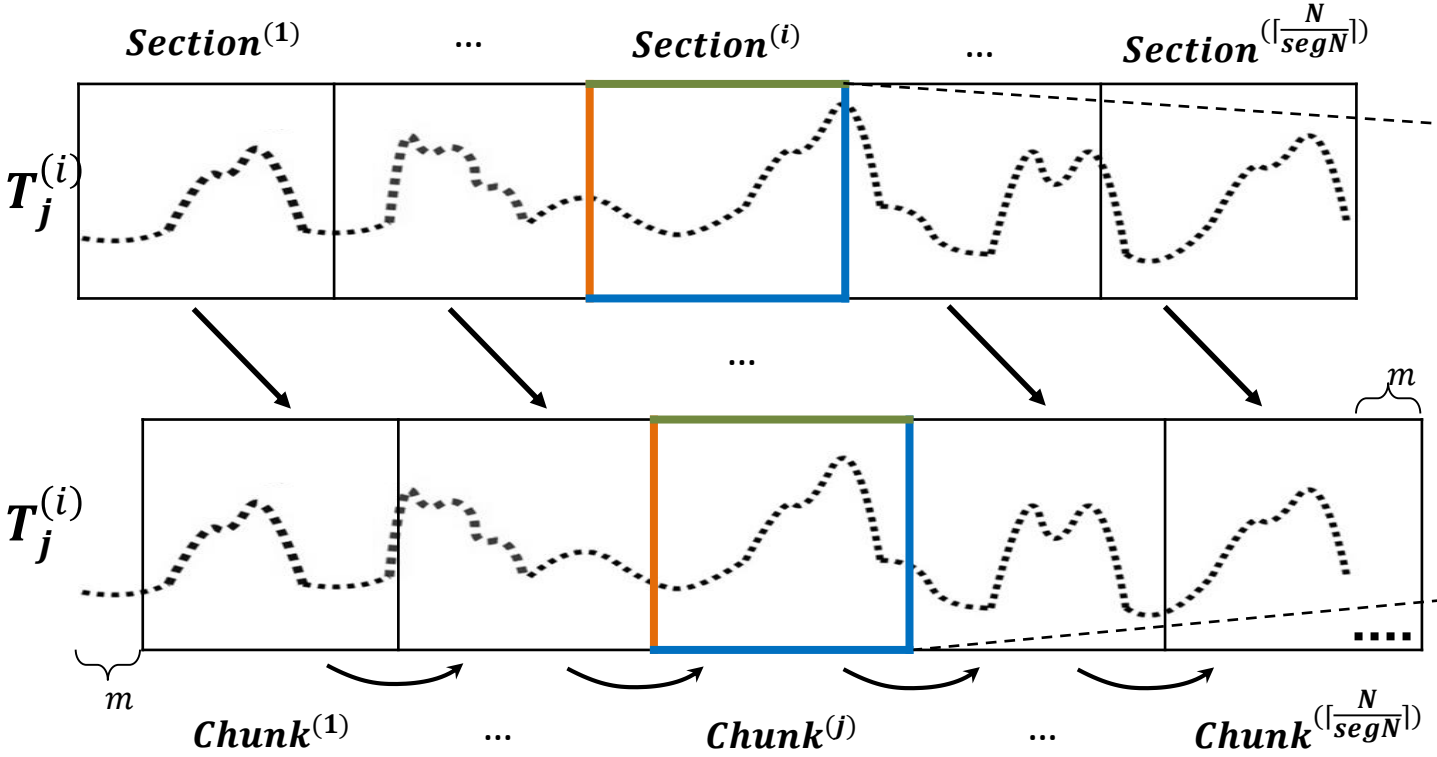
Discord discovery

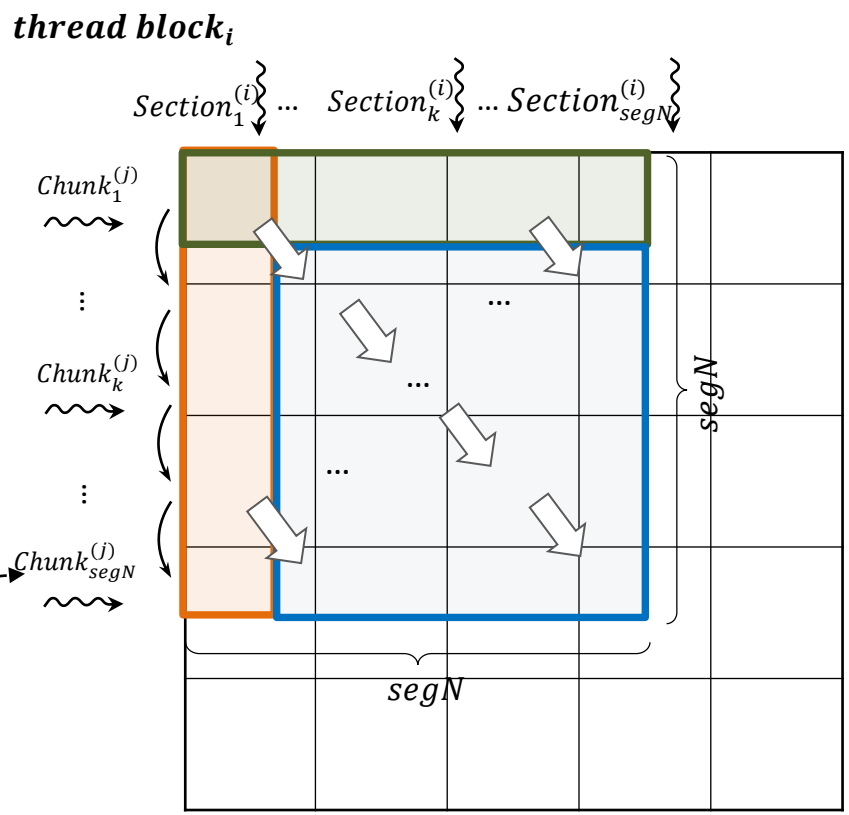


Candidate selection and discord refinement w.r.t. the fragment



Candidate selection and discord refinement w.r.t. the segment





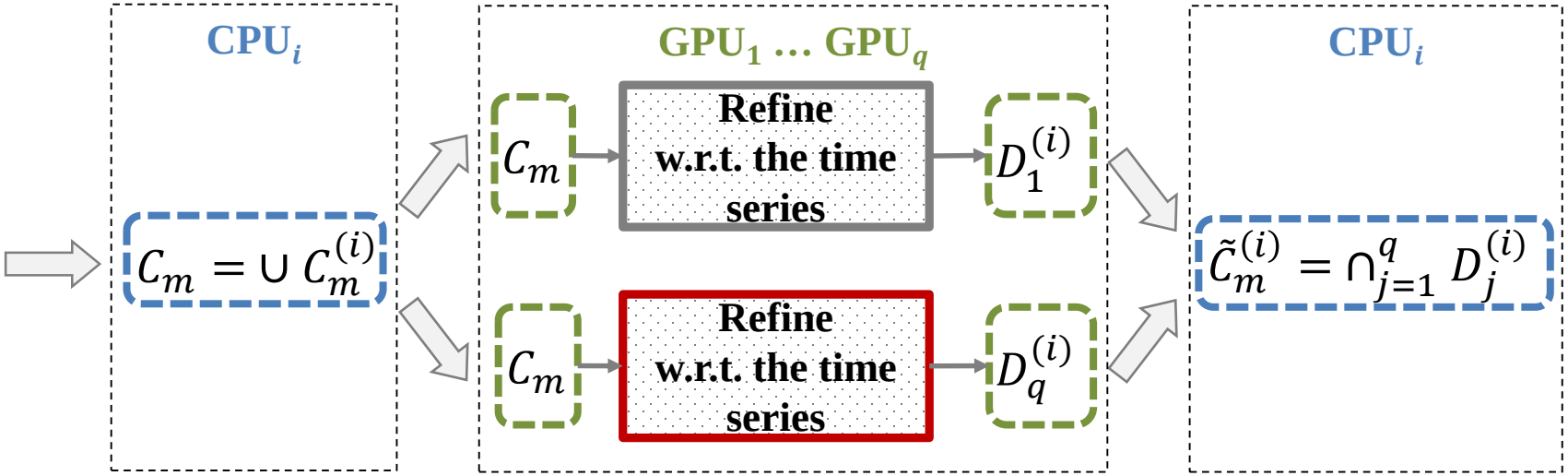
$$\mathrm{ED}^2_{\mathrm{norm}}\left(T_{i,m}, T_{j,m}\right) = 2m\left(1 - \frac{T_{i,m} \cdot T_{j,m} - m\mu_i\mu_j}{m\sigma_i\sigma_j}\right)$$

$$\boldsymbol{QTrow^{(i)}(tid)} = \sum_{k=1}^{m} T_{tid}^{(i)}(k) \cdot Chunk_1^{(j)}(k) \qquad \boldsymbol{QTcol^{(i)}(tid)} = \sum_{k=1}^{m} T_1^{(i)}(k) \cdot Chunk_{tid}^{(j)}(k)$$

$O(1)$ instead of $O(m)$!

$$\boldsymbol{QTrow^{(i)}(tid)} = QTrow^{(i)}(tid-1) - T_{tid-1}^{(i)}(1) \cdot Chunk_{tid-1}^{(j)}(1) + T_{tid}^{(i)}(m) \cdot Chunk_{tid}^{(j)}(m)$$

Discord refinement w.r.t. the time series

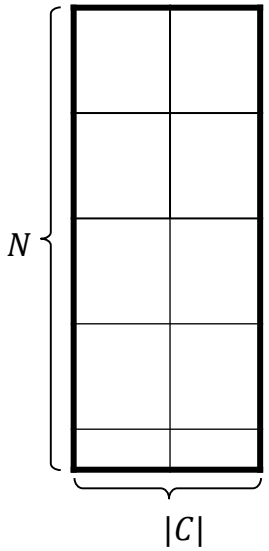


Discord refinement

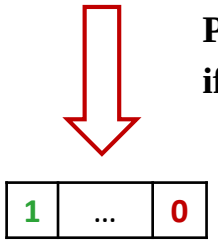
Distance matrix

$Dist \in \mathbb{R}^{N \times |C|}$

$$Dist_{i,j} = 2m\left(1 - \frac{T_{i,m} \cdot T_{j,m} - m\mu_i\mu_j}{m\sigma_i\sigma_j}\right)$$



Discord
Bitmap



Prune the candidate C_j,
if $\min_{1<i<N}(Dist_{i,j}) < r^2$

Segment $T_j^{(i)}$

Segment subsequences

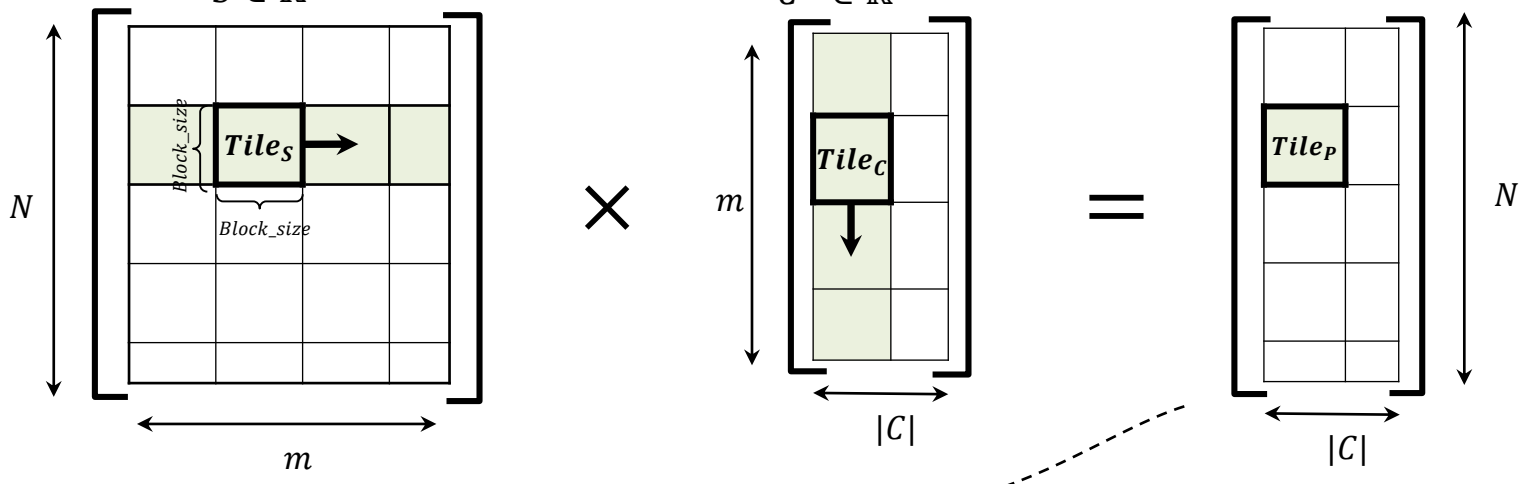
$S \in \mathbb{R}^{N \times m}$

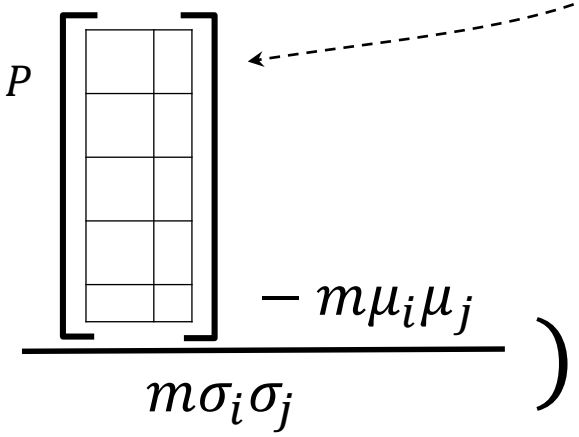
Candidates

$C^T \in \mathbb{R}^{|C| \times m}$

Matrix of scalar products

$P \in \mathbb{R}^{N \times |C|}$





Distance matrix

$Dist \in \mathbb{R}^{N \times |C|}$

Experimental evaluation

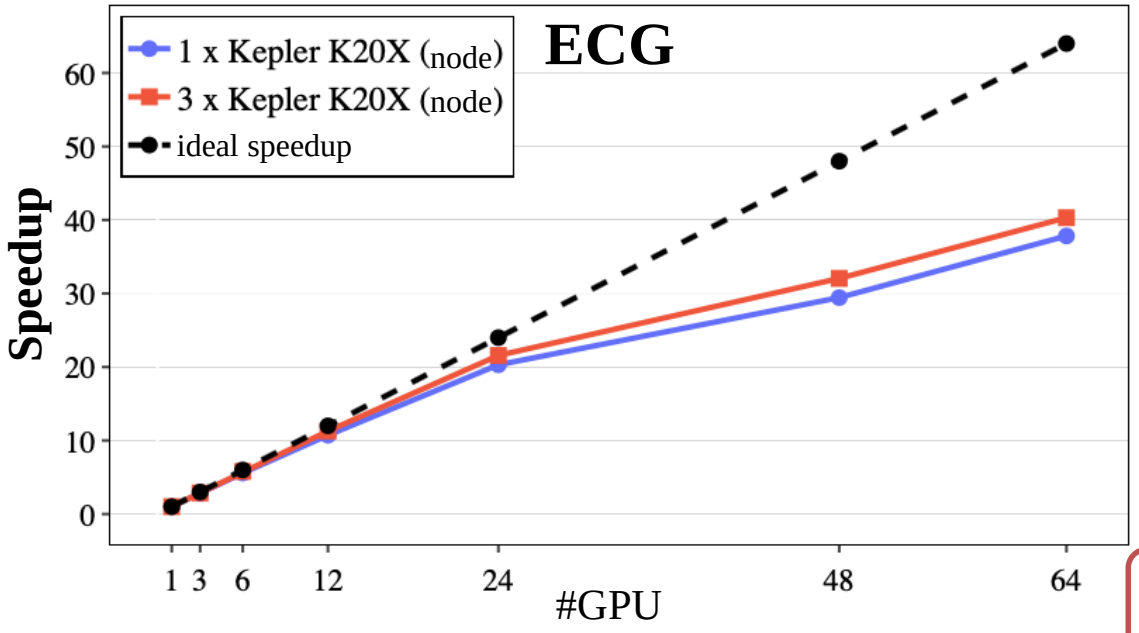
- **HPC cluster Lobachevsky (University of Nizhny Novgorod)**
  64 × (2×Sandy Bridge E5-2660 + **3×NVIDIA Kepler K20X** (2 688 cores @0.732 GHz, 1.3 TFLOPS))
- **Data**

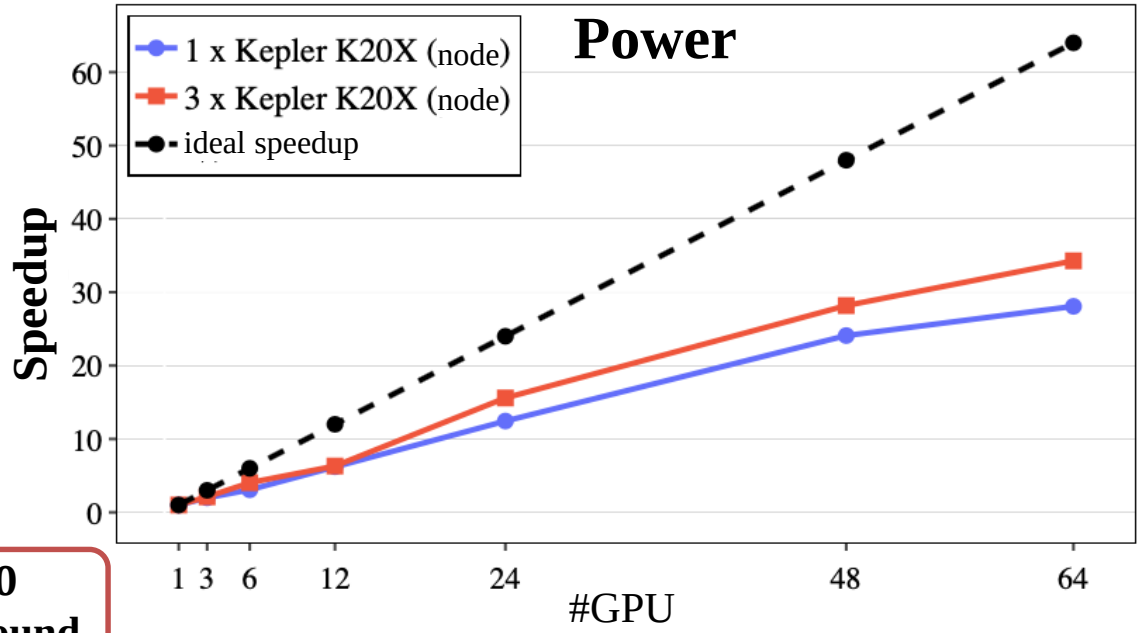
Time series	Length, n	Discord length range, $minL..maxL$	Subject domain
ECG[1)]	$2 \cdot 10^6$	$64..128$	ECG of an adult patient
Power[2)]			Annual household power consumption

- **Metrics**
  - ***Performance***: average running time over 10 launches (without I/O)
  - ***Speedup***: $s(p) = {}^{t_1}/_{t_p}$, where t_1 and t_p are performance on one and p GPUs, respectively

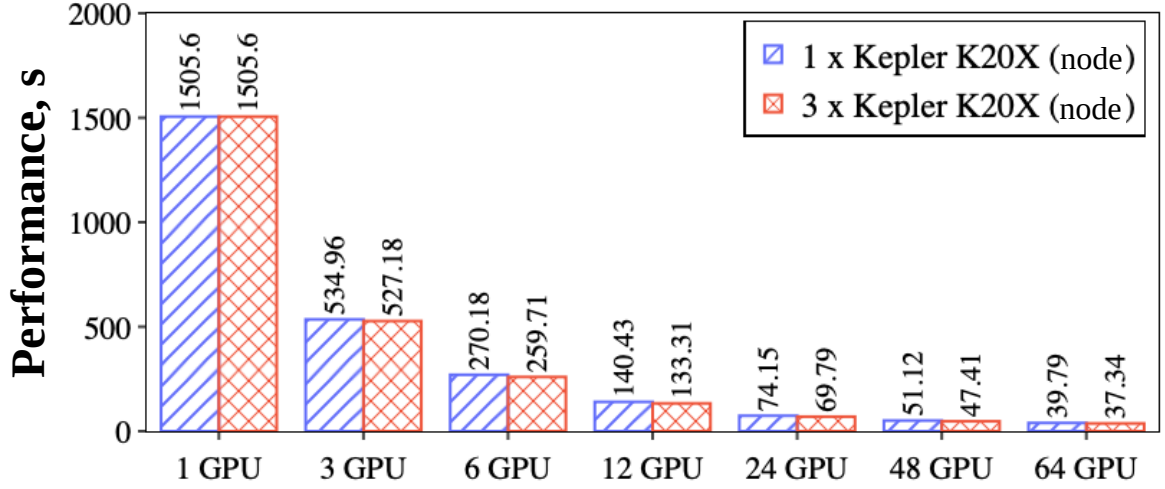
[1)] Goldberger A. *et al*. PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals. Circulation. 2000. 101(23). pp. 215-220. DOI: 10.1161/01.CIR.101.23.e215.
[2)] Individual household electric power consumption. URL: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption/.

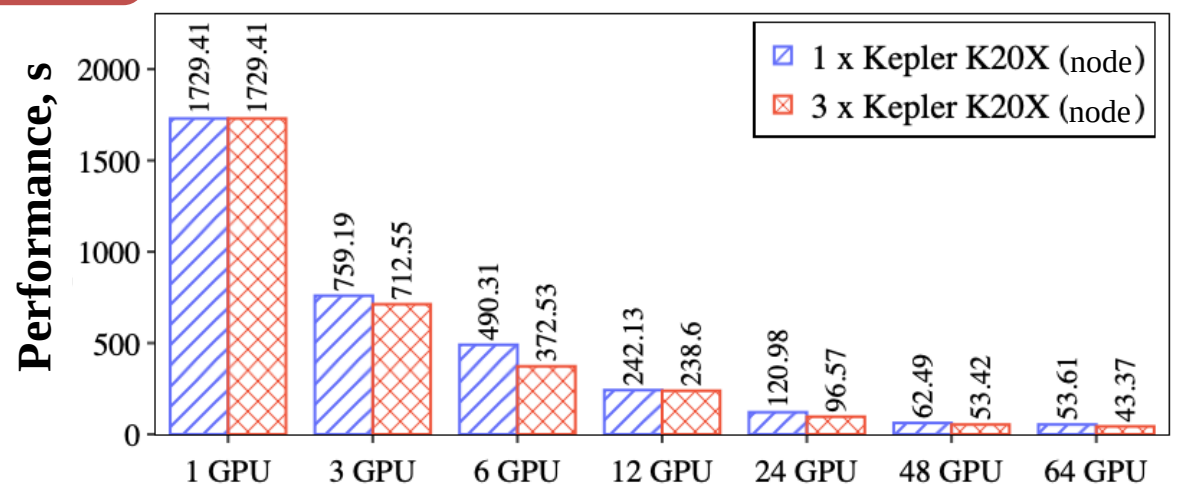
Speedup and performance





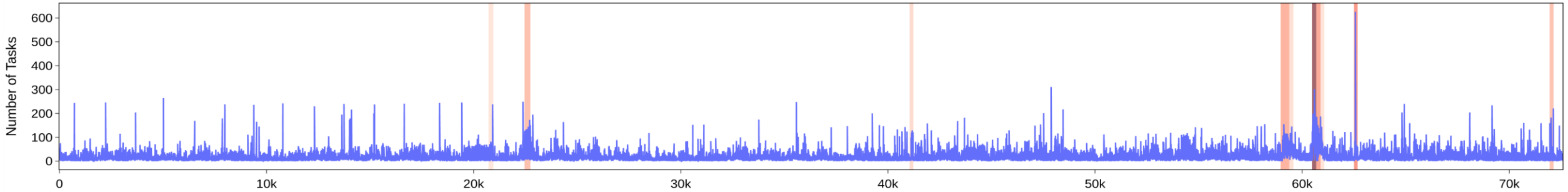
≈ 1500 discords found





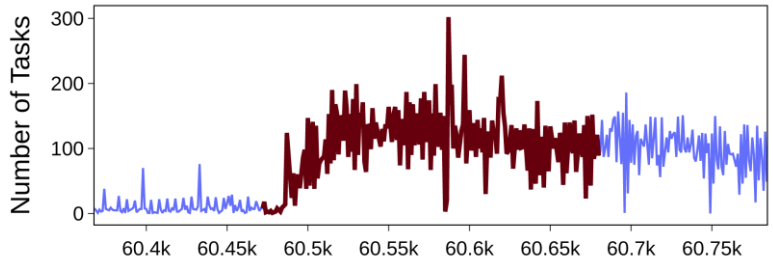
Case study: anomalies in Alibaba Cloud HPC cluster workload

Tasks running in Alibaba PAI (Platform for Artificial Intelligence), July-August 2020*

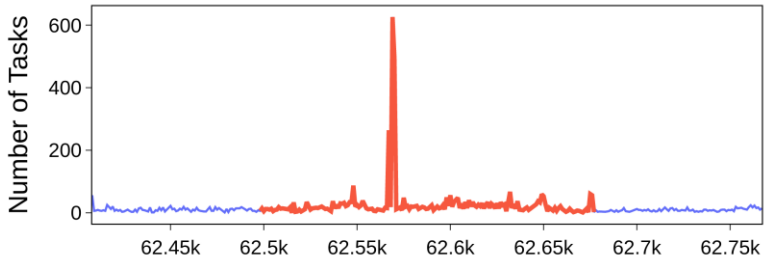


Top discords (range is 3-7 min)

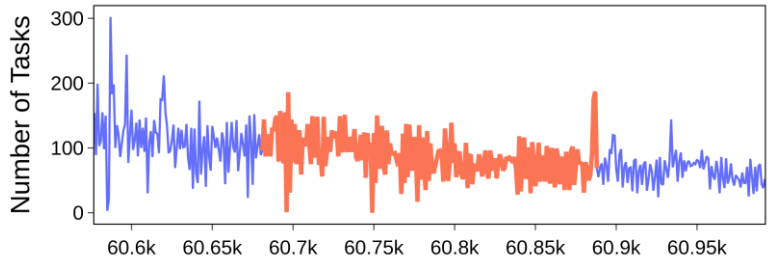
Top-1: $m = 209$ (3 min 29 sec)



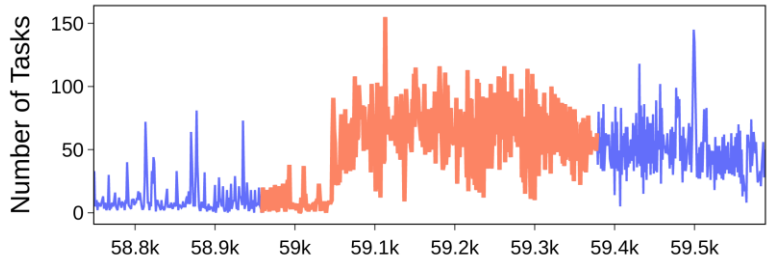
Top-2: $m = 180$ (3 min)



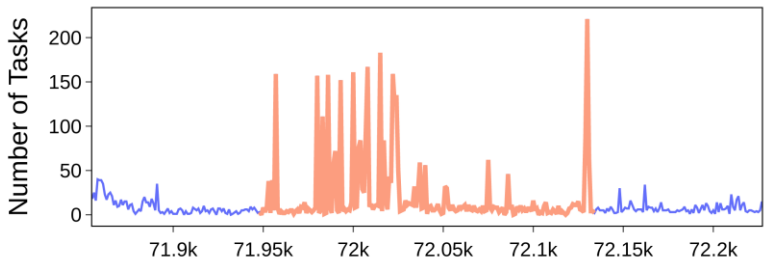
Top-3: $m = 208$ (3 min 28 sec)



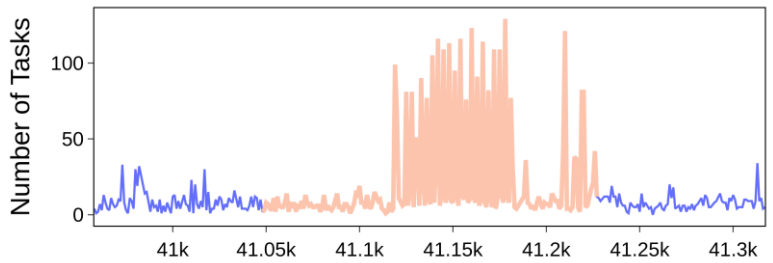
Top-4: $m = 420$ (7 min)



Top-5: $m = 283$ (4 min 43 sec)

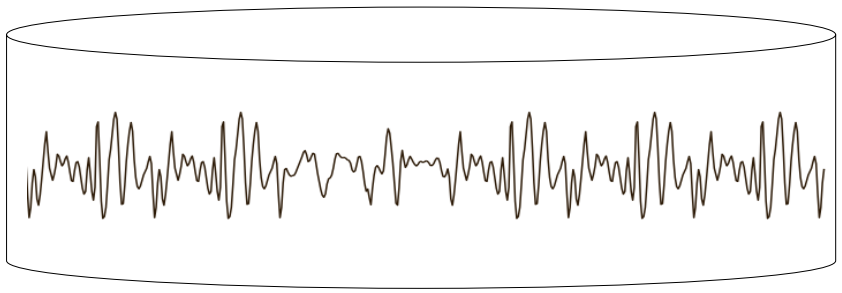


Top-6: $m = 187$ (3 min 7 sec)



* Weng Q. *et al.* MLaaS in the wild: workload analysis and scheduling in large-scale heterogeneous GPU clusters. NSDI 2022. pp. 945-960. https://www.usenix.org/conference/nsdi22/presentation/weng

Q: Can we discover anomalies **online**? **A:** DiSSiD[1)]



Representative fragment
of the time series

PSF[2)]:
Parallel
Snippet-
Finder

Label
activities

PADDi

Discover
discords

Training
set

Learn
the model

Sensor

…

subsequence

DiSSiD:
Discord, Snippet, and Siamese Net-based
Detector of anomalies

Norm

OR

Anomaly

[1)] Kraeva Ya.A. Detection of time series anomalies based on data mining and neural network technologies. Bulletin of SUSU, CMSE. 2023. 12(3). pp. 50-71. DOI: 10.14529/cmse230304.
[2)] Zymbler M., Goglachev A. Fast summarization of long time series with graphics processor. Mathematics. 2022. 10(10). 1781. DOI: 10.3390/math10101781.

We're ready to discover anomalies in your time series ASAP!

- **PADDi (recently completed research)**
  The first, only, and scalable algorithm
  for all-length discord discovery
  in long time series on multi-GPU clusters
- **DiSSiD (ongoing research)**
  Deep learning method based on PADDi
  for *online anomaly discovery in time series*
- **Further research**
  Extension of the proposed solutions
  to *multidimensional time series*

Mikhail Zymbler
Dr.Sci.,
Assoc. Prof.

mzym@susu.ru
https://mzym.susu.ru

Yana Kraeva
PhD student,
RF President stipend holder

kraevaya@susu.ru