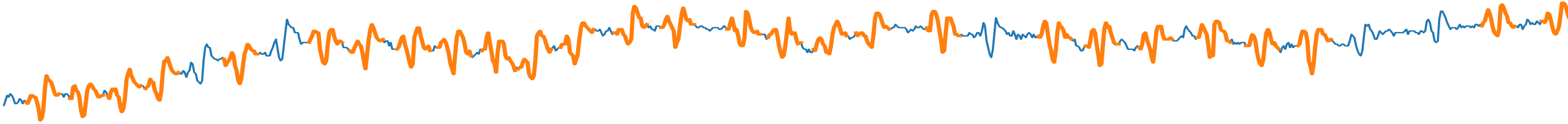
Motiflets –
Simple and Accurate Detection of Motifs in Time Series



Patrick Schäfer, Ulf Leser

Humboldt-University of Berlin, Germany
patrick.schaefer@hu-berlin.de

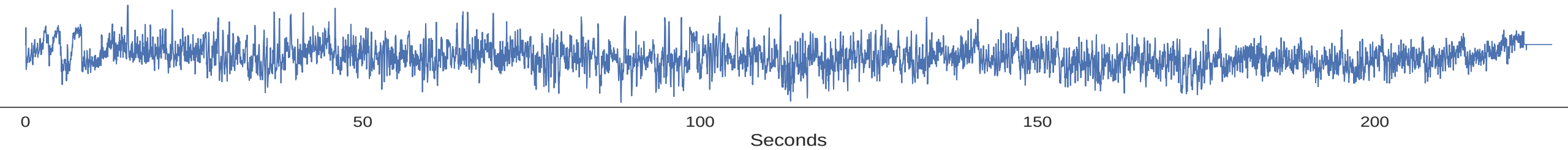
VLDB, Vancouver, Canada - August 28 to September 1, 2023

Source Codes

Time Series

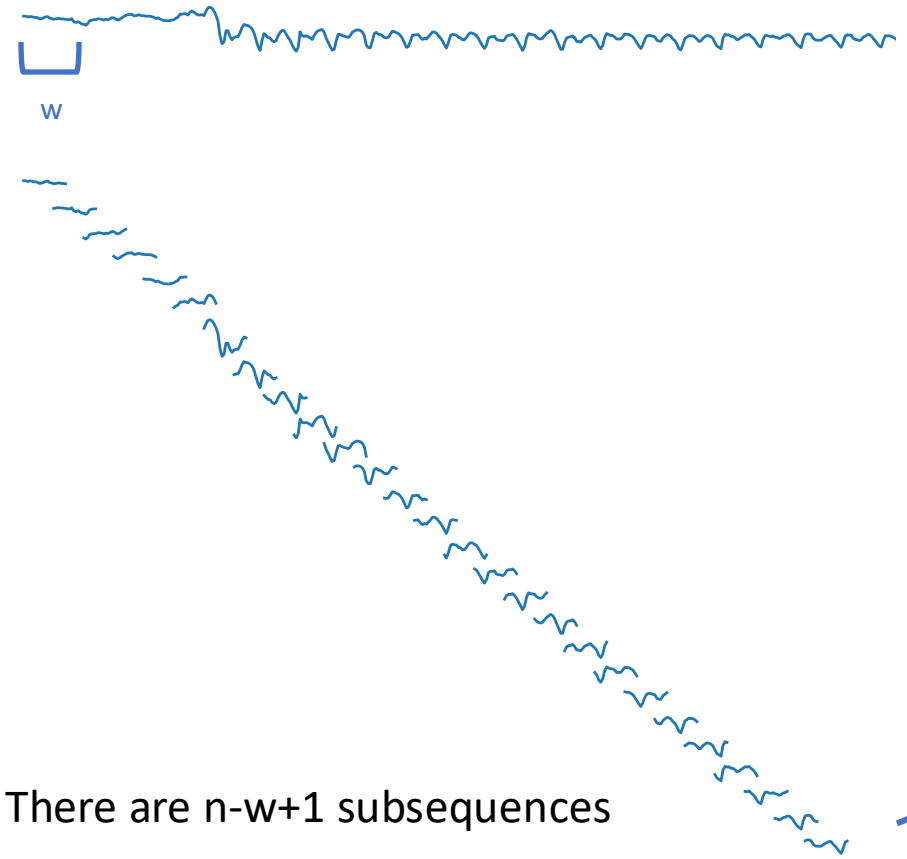
- Time series are **ubiquitous!**

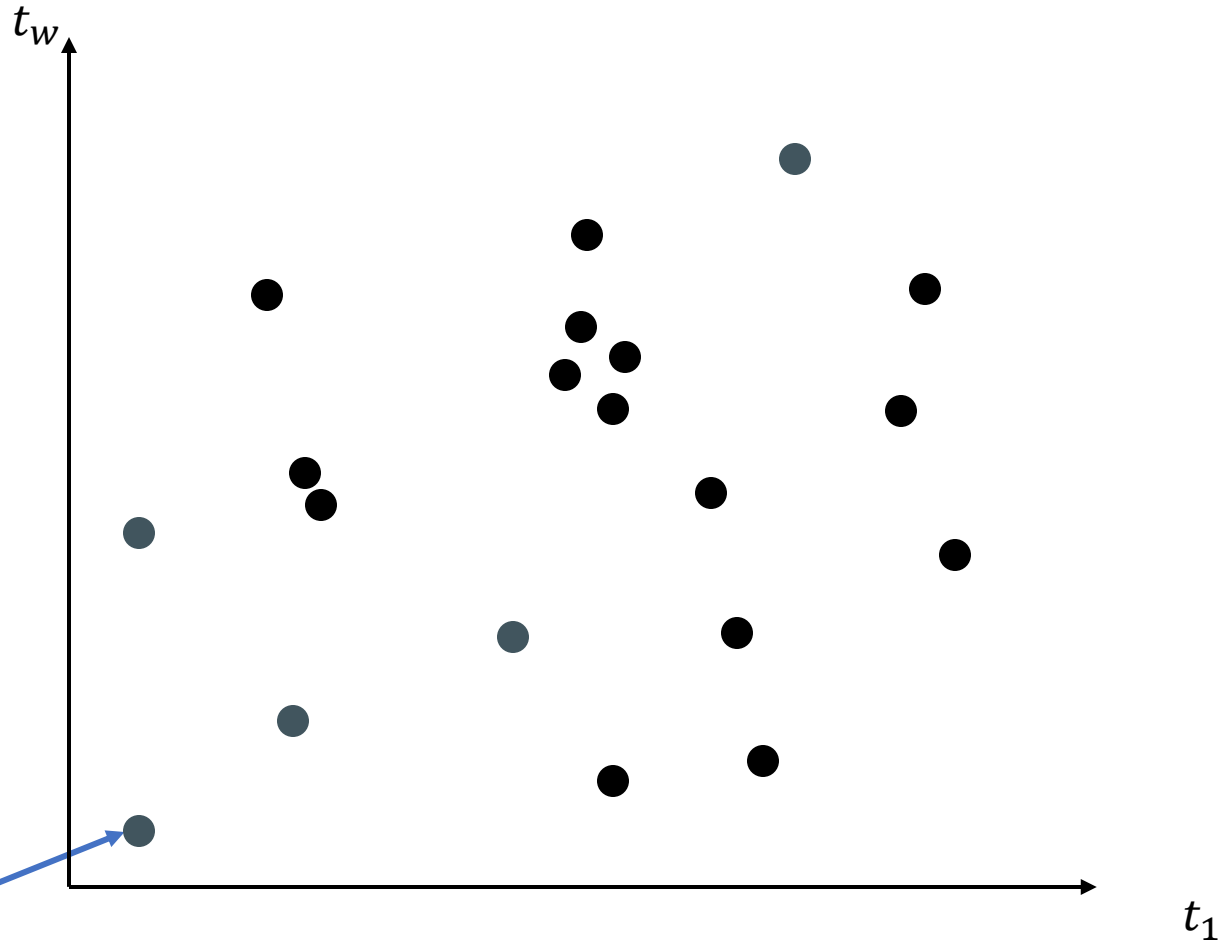
- We record data over time:
  - Sensors data, Medical (ECG and EEG), Financial, Meteorological, Biological Processes, Motion Data and Video, Music



Geometric Representation

We think of w-length subsequences of a time series as points in w-dim space

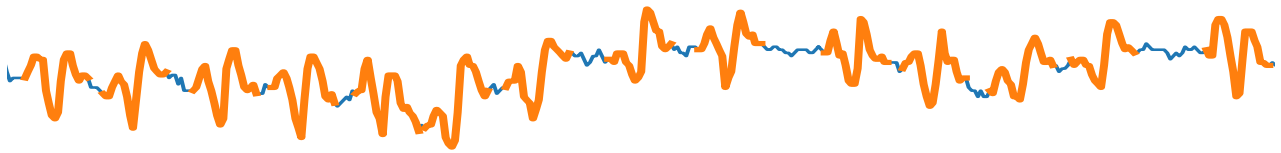


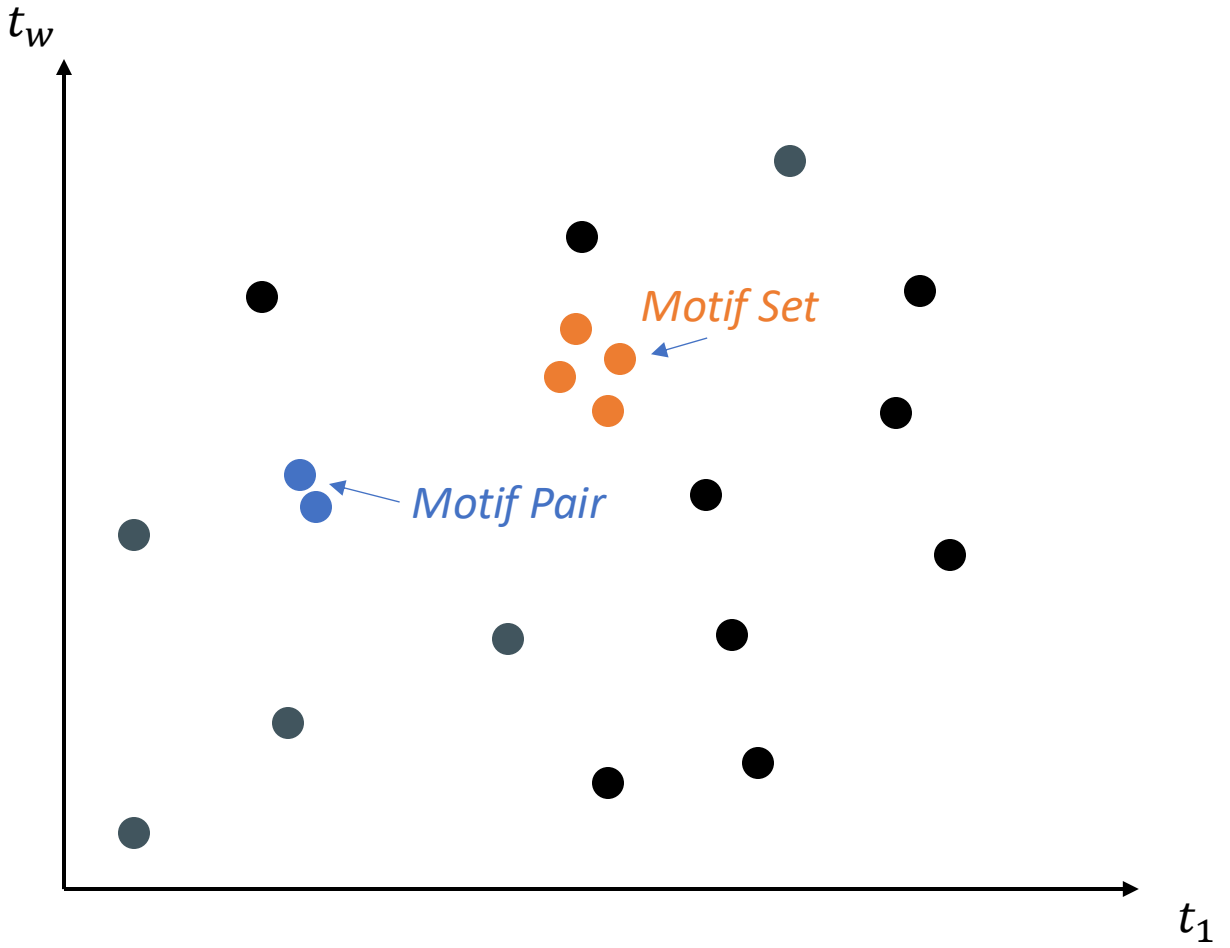


There are n-w+1 subsequences

Motifs

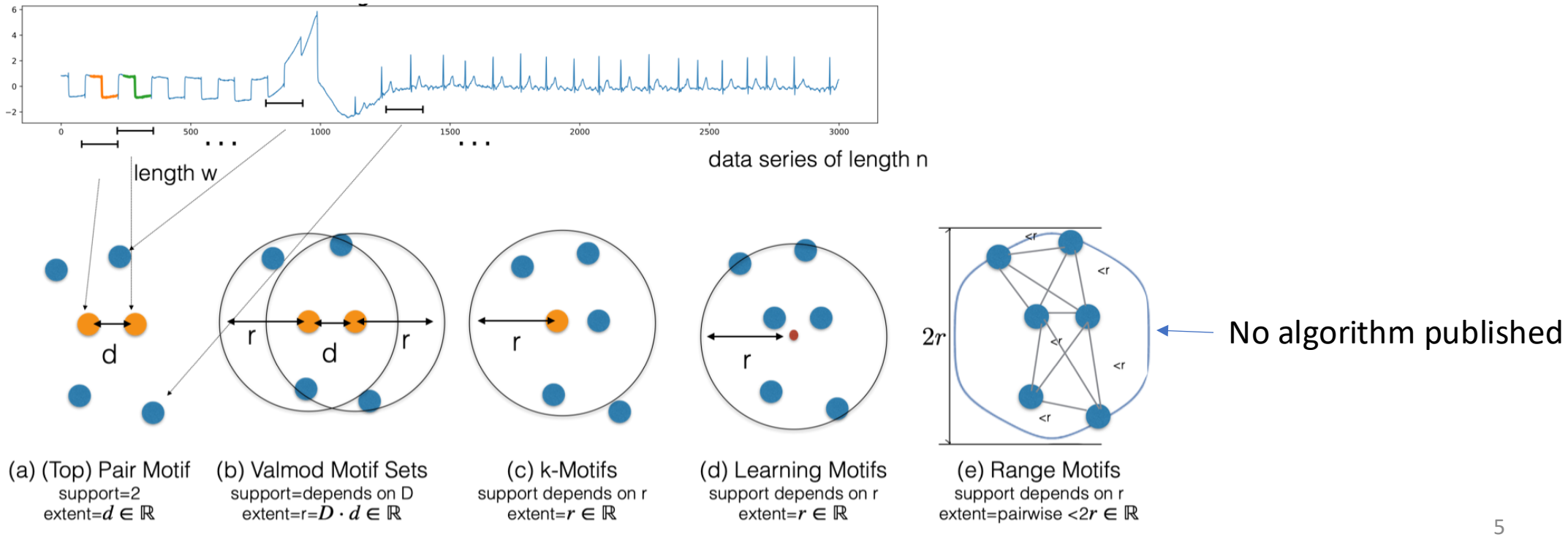
- **Pair Motif:** Two subsequences closest to each other – (smallest distance)
- But interesting motifs seldom occur in mere pairs
- **Motif Sets**: Frequent approximately repeated patterns of a time series





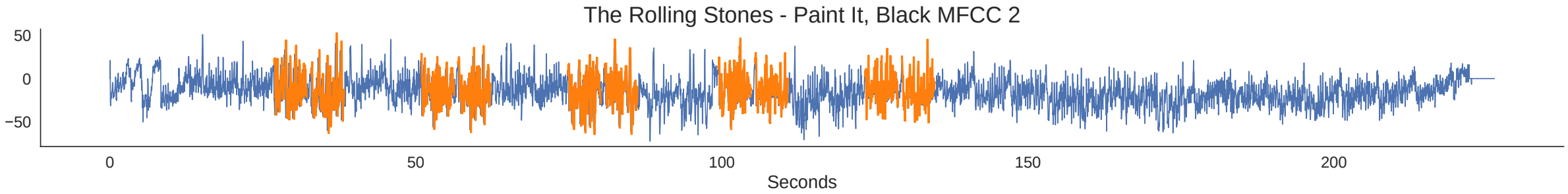
SotA in Motif Set Discovery

- Each definition can be mapped to a geometric shape (hypersphere)
- **The distance threshold r is the crucial input parameter**



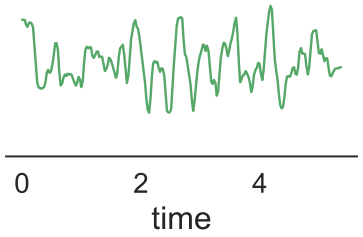
Rolling Stones – Paint it black
(Motivating Example)

Can you spot, what is frequent?



MFCC representation of the song: Rollings Stones, Paint it Black

I see the girls walk by,
dressed in their summer clothes



Rhythmical section (10x)

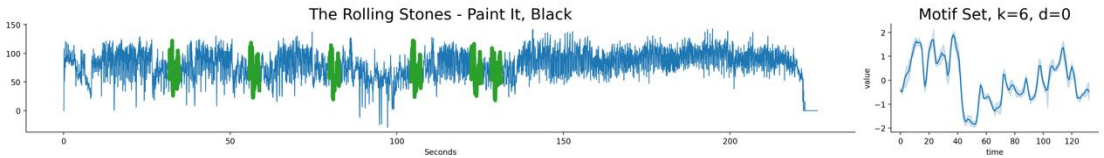
Are we confident, we found all occurrences?

1.) SotA: Motif set found for small r

Motif Set with 6 occurrences of 3 sec

1. I have to turn my head...
2. Like a newborn baby...
3. It's not easy facing up...
4. My love will laugh with me...
5. I see the girls walk by...
6. I have to turn my head...

(similar rhythmical section)

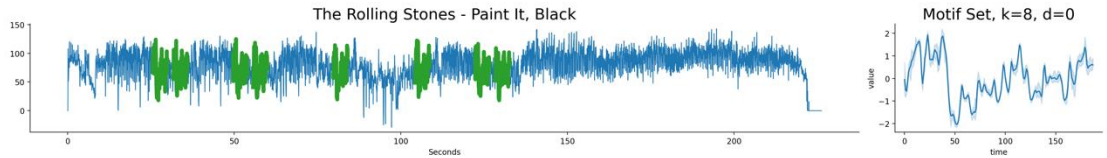


2.) SotA: Motif set found for larger r

Motif Set with 8 occurrences of 4.3 sec

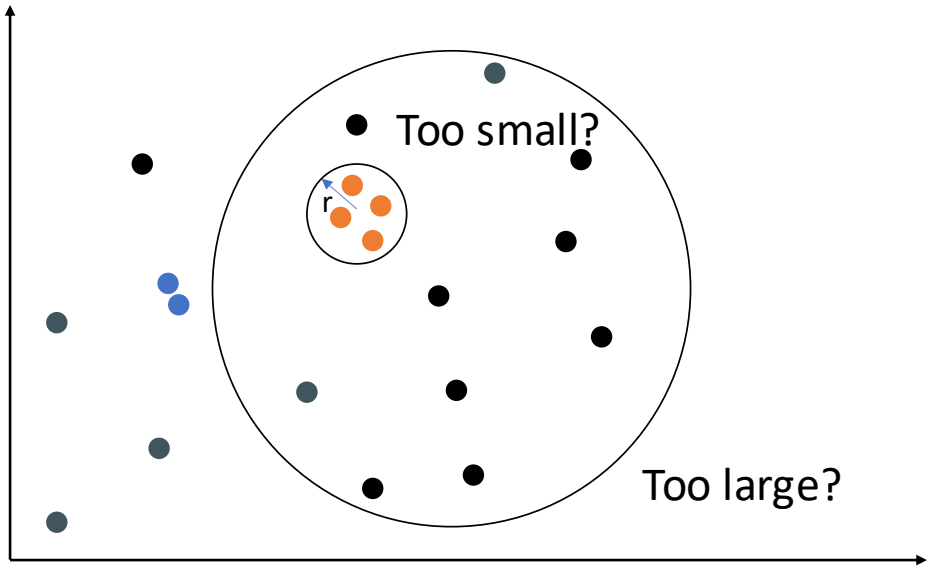
1. I see the girls walk by, dressed in their ...
2. I have to turn my head until my...
3. I see people turn their heads and quickly...
4. Like a newborn baby, it just happens...
5. It's not easy facing up when your whole...
6. My love will laugh with me before the...
7. I see the girls walk by, dressed in their...
8. I have to turn my head until my...

(similar rhythmical section)



Distance 1/2

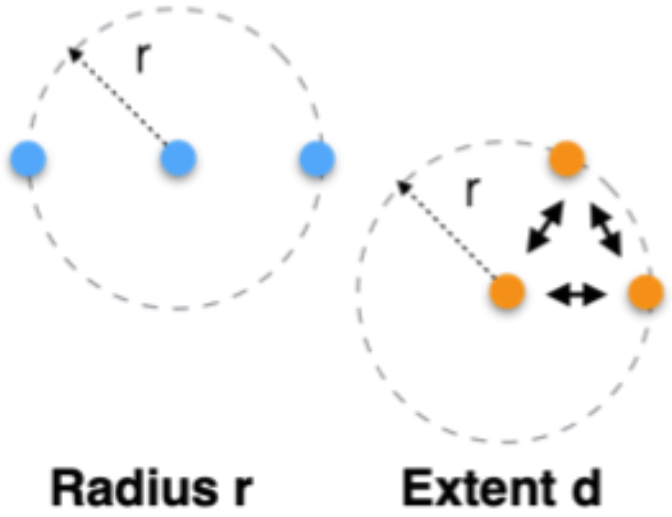
- Problem 1: The distance is a **real value**, and it is hard to estimate by a user
  - **If r set too small, we miss occurrences**
  - **If r set too large, we find everything**
- Similar problems exist with the **length of the motif**



Distance 2/2

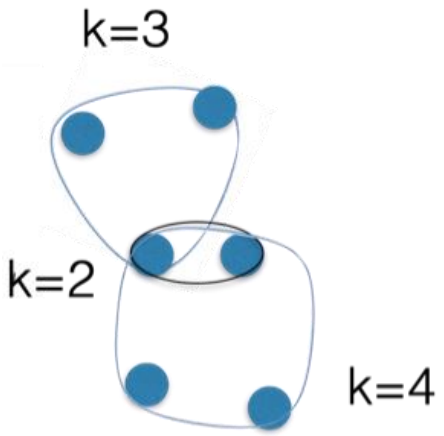
- Problem 2: The hypersphere defined by SotA **overestimates the actual hypersphere** of the TOP motif set

- Radius is equal for both motifs of size 3
- But the extent (maximum over pairwise distances) is smaller for the orange motif set



k-Motiflets

- k-Motiflets is a motif set discovery algorithm
- It finds **k-frequent motifs of smallest pairwise distance**
  - For k=2 it is equal to pair motif discovery
- It has two input parameters:
  - **k: size of desired motif set**
  - w: length of subsequences
  - **... and no radius/distance as input**



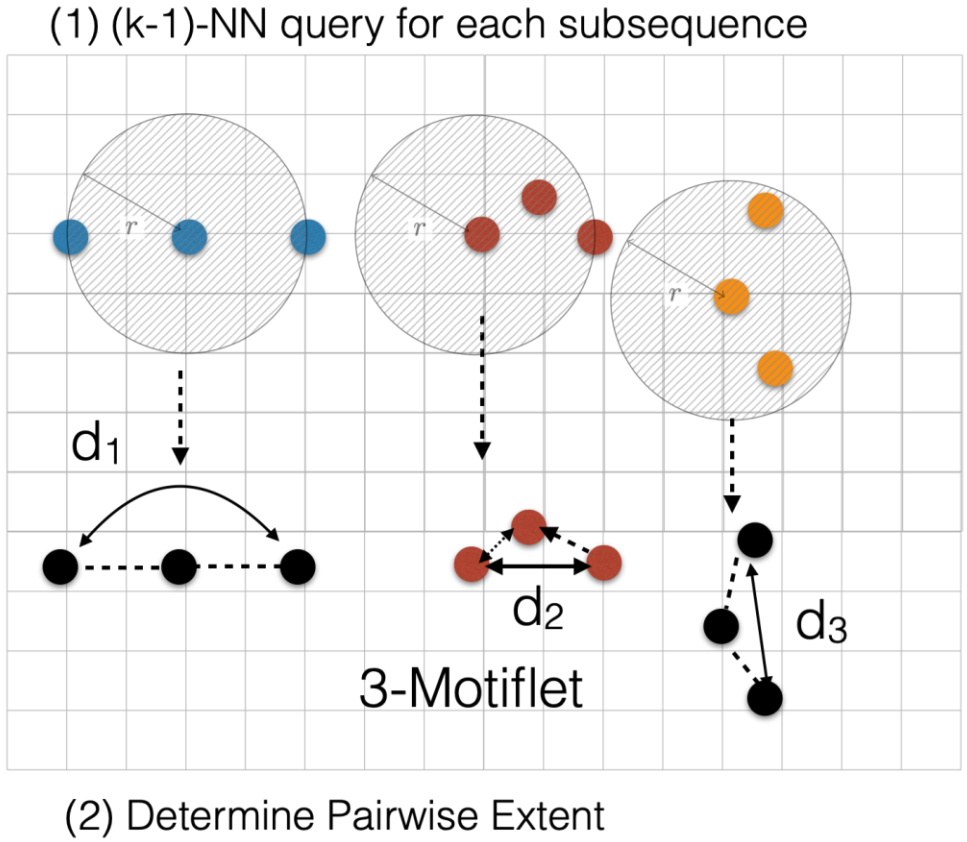
(f) k-Motiflets
support=$k \in \mathbb{N}$
extent depends on k

Geometric Shape: Reuleaux polygon

Motiflets - Algorithm

- To find a k-Motiflet:
  1. Input k as parameter
  2. Perform a k-NN search around each of $O(n)$ subsequences of the TS
  3. For each candidate, compute maximum over pairwise distances (*"extent"*)
  4. Minimize over all candidate sets

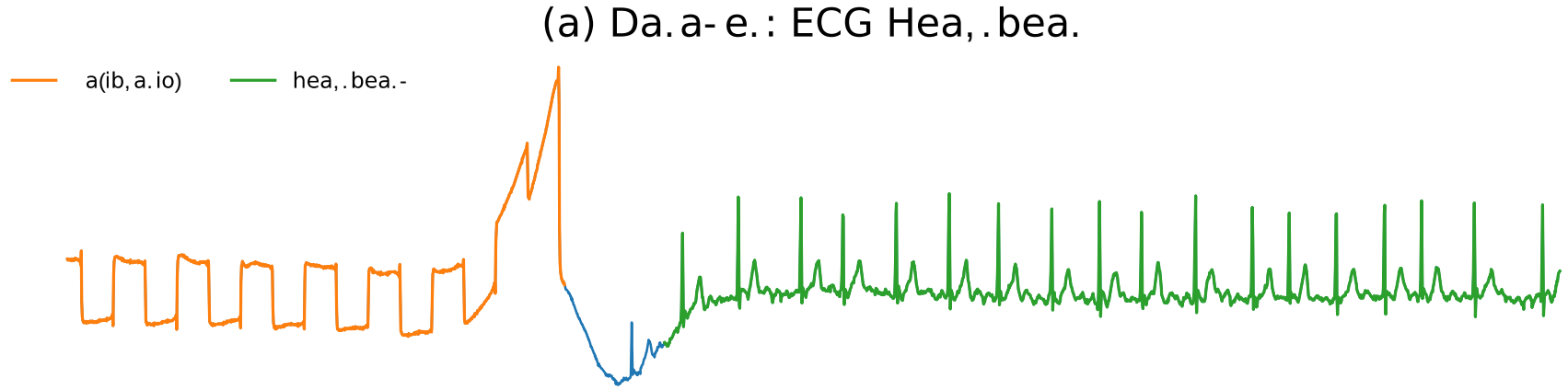
- Complexity: $O(n^2k + k^2n)$

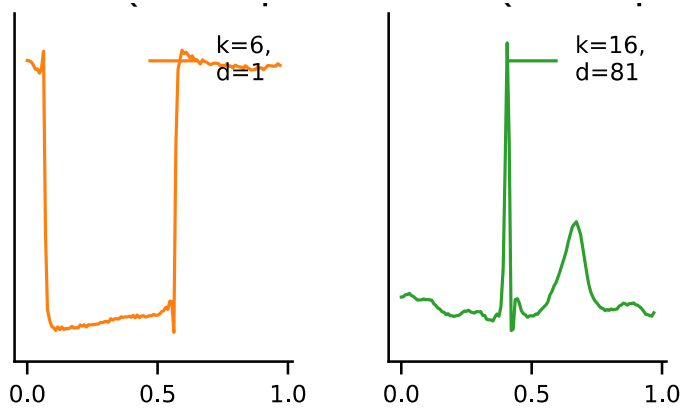


d_2 has smallest extent

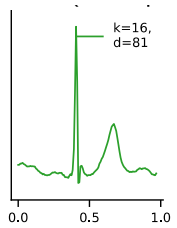
Finding representative k's

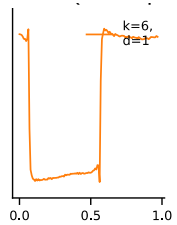
- Consider this ECG trace

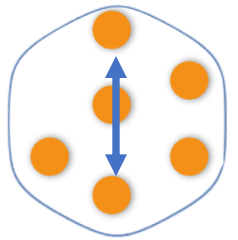




- It contains two motifs: 6 calibration signals and ~16 full ECG waves
- **How to find the largest set of repeats of each motif?**



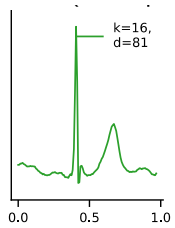


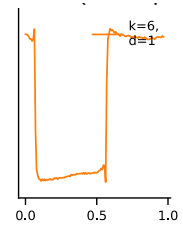


Up to k=6, extent increases linearly/slightly

6-Motiflet with smallest extent

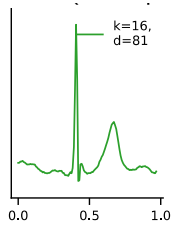
For k=7,
extent increases
sharply

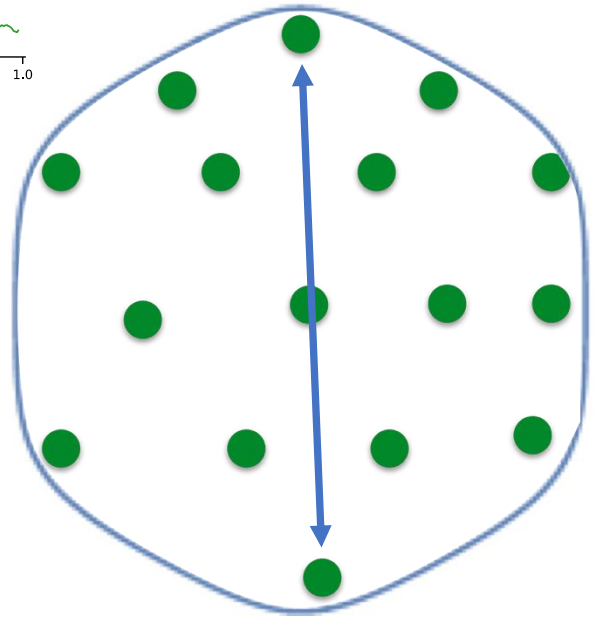


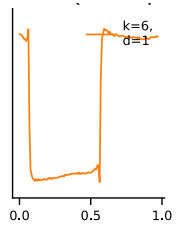


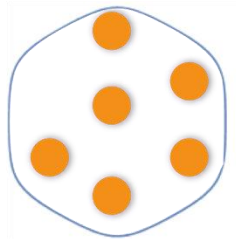
7-Motiflet with
smallest extent

Up to k=16, extent increases linearly/slightly



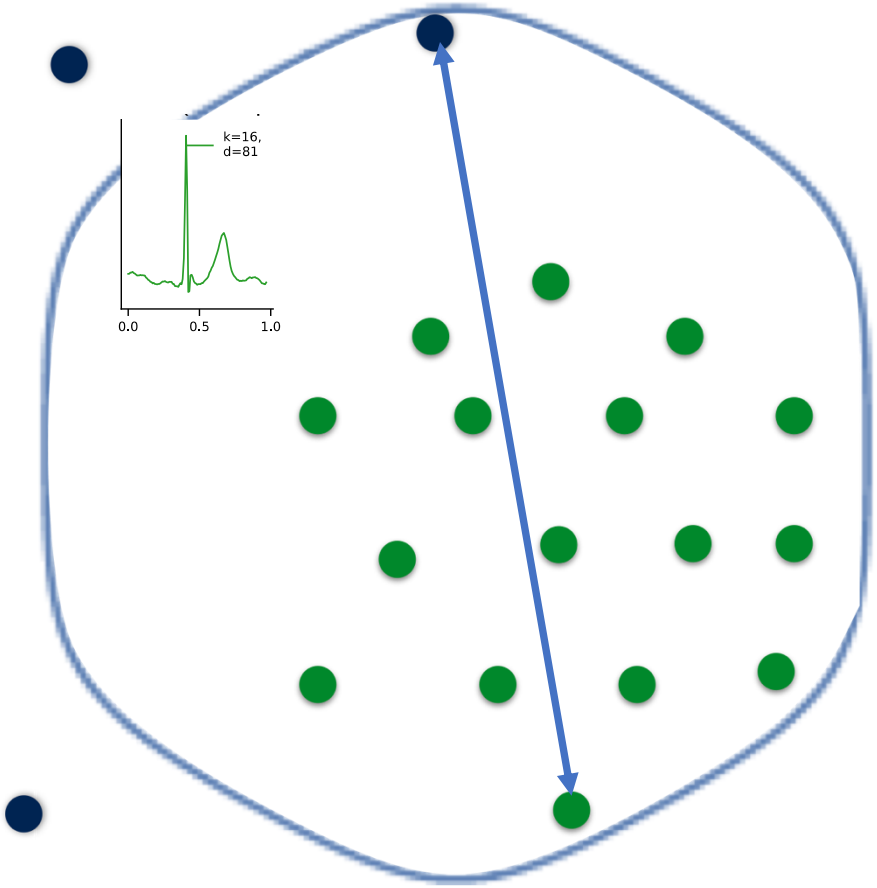


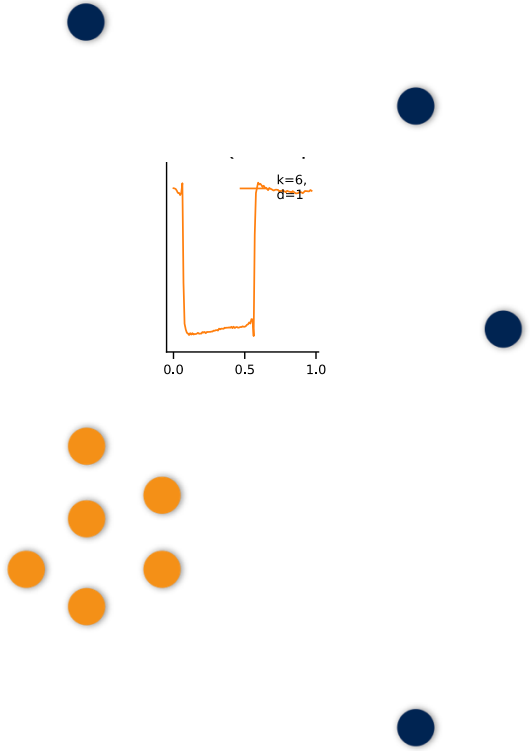




16-Motiflet with smallest extent

Once we increase k>16, extent increases sharply

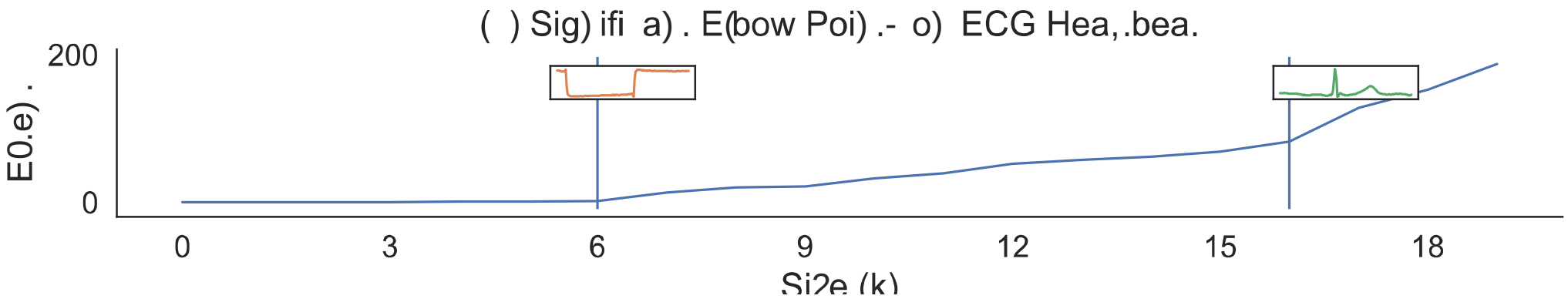




17-Motiflet with smallest extent

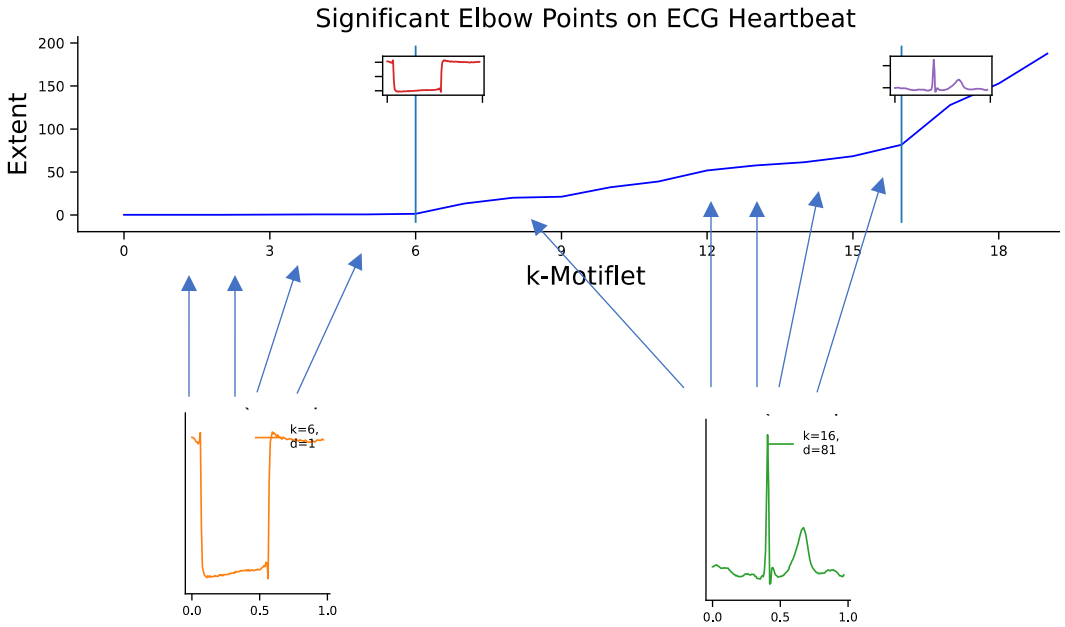
Elbow Plots

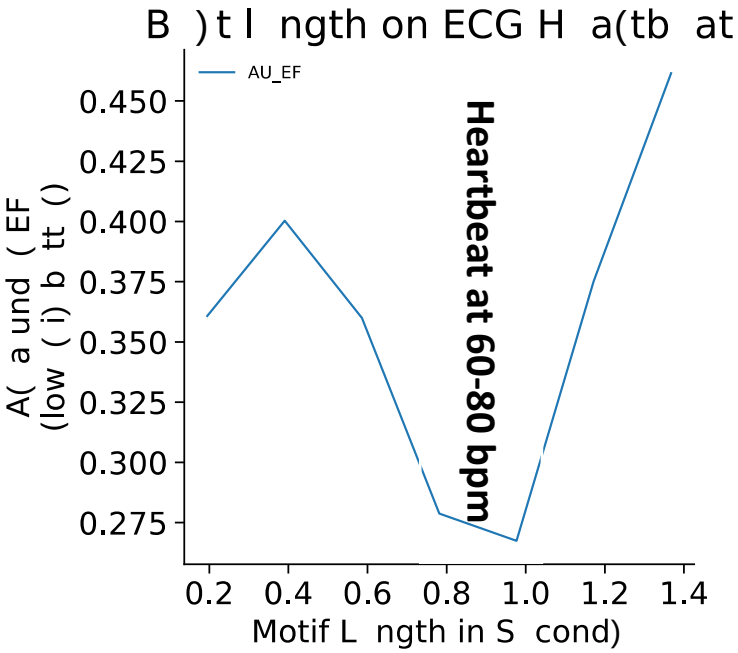
- Plot the size k of the k-Motiflet against its extent



- **Elbows** represent the largest meaningful motif set
- A sharp increase in extent indicates adding a dissimilar subsequence

Finding Motif Lengths

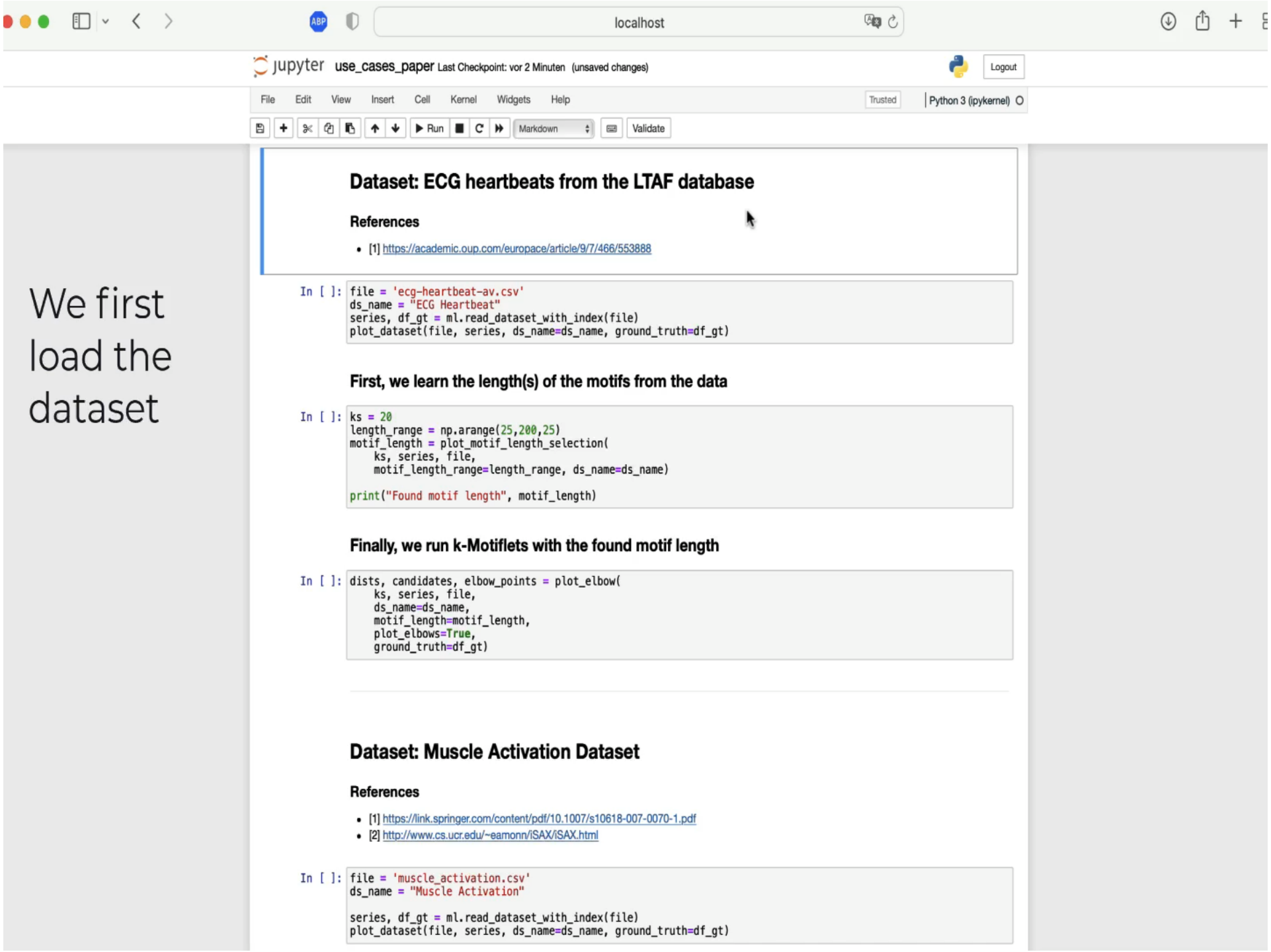




- **Long flat stretches**: correspond to high numbers of repeats of the same motif
- Measured by **area under the elbow plot**

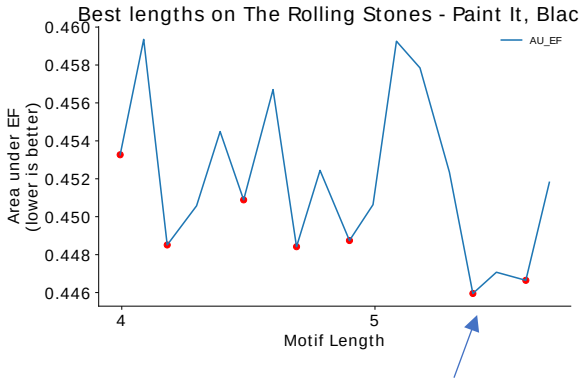
- For each length, we compute its area
- Minima correspond to long stretches

We first load the dataset

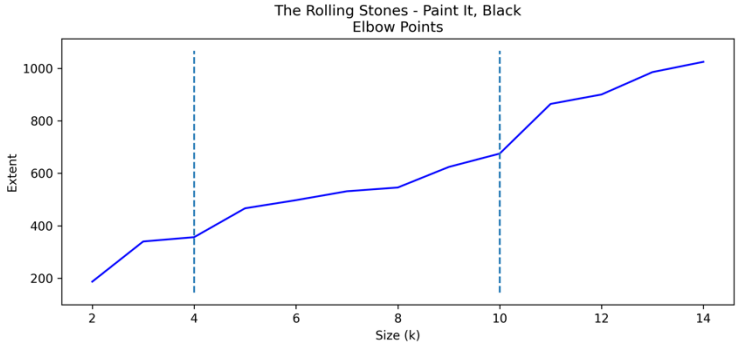


Rolling Stones – Paint it black
(Motivating Example)

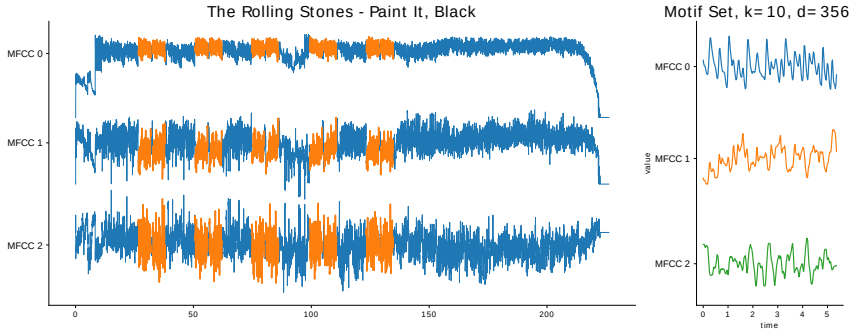
- Human voice is represented by low frequencies
  - we use the first MFCCs from the wave file
  - and search for k-Motiflets
- We found all 10 repeats of length 5.4 seconds



Best length ~5.4 sec



Found elbows at
k=4 and k=10

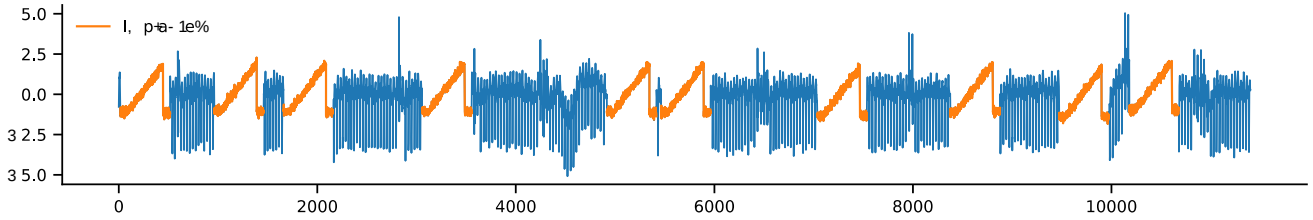


Found 10-Motiflet:

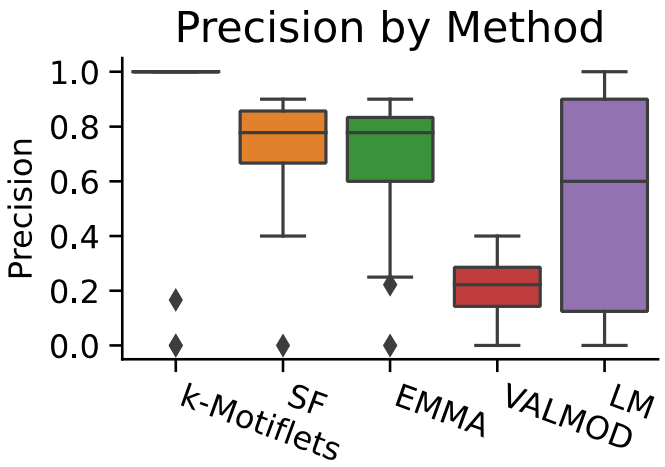
I see the girls walk by
Dressed in their summer clothes
... (10x)

Motiflets find better Motifs

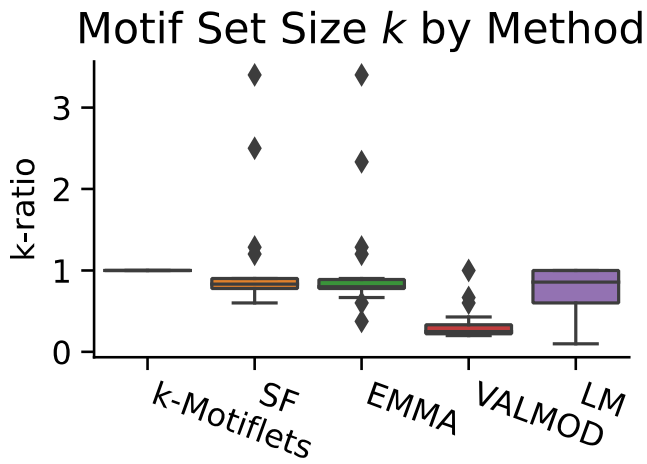
- We created a semi-synthetic benchmark with 25 ts with implanted motif sets
  - Use gold standard parameters as input to each method
  - (i.e. k for k-Motiflets or the radius for SoTA)



Recall: hypersphere defined by SotA **overestimates the actual hypersphere**

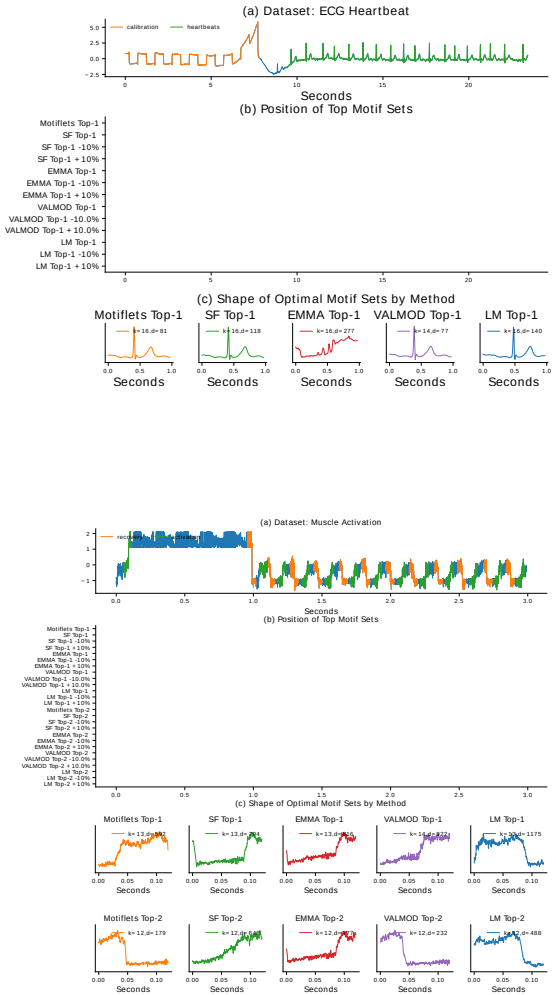


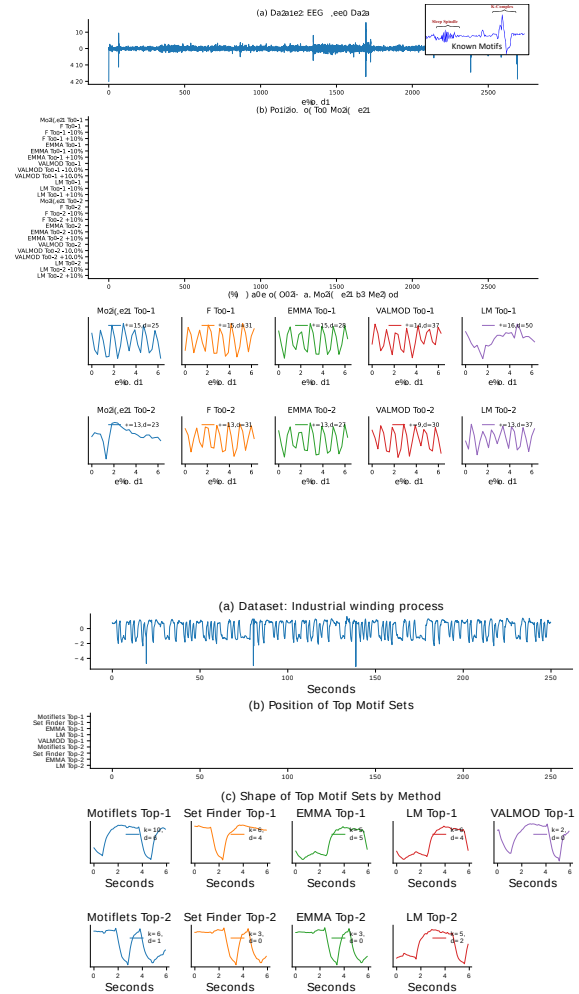
Motiflets outperform SoTA by a large margin

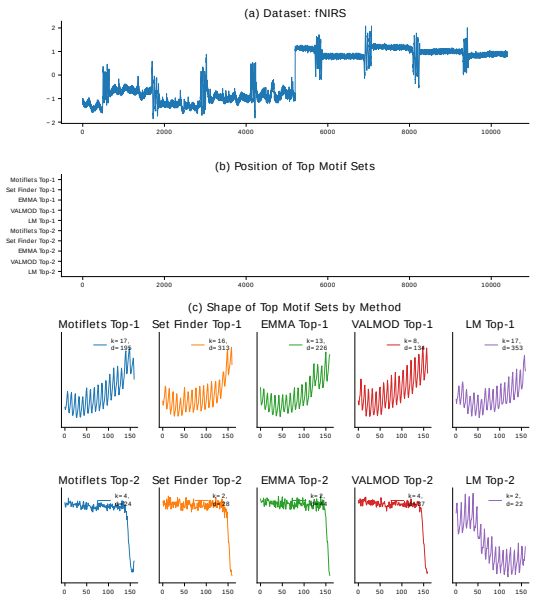


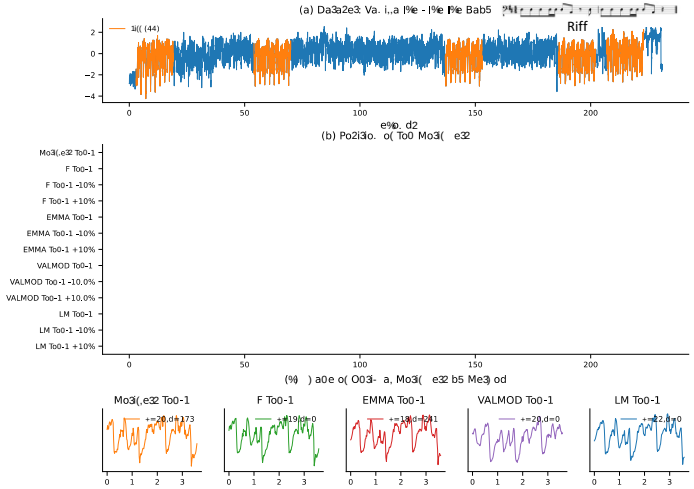
SoTA often finds too small or too large motif sets

Many Other Real-World Use-Cases









Please see
the paper and supporting website

Conclusion

Source Codes

- Many different motif set discovery algorithms exist
  - As central parameters they all take **length w and distance r**

- We propose k-Motiflets as the set of exactly k occurrences of a motif of length w
  - We argue that the value of k is much easier to set than the distance r
  - We introduce tools to automatically determine meaningful input values
  - Motiflets produce better motifs than all competitors at lower runtimes

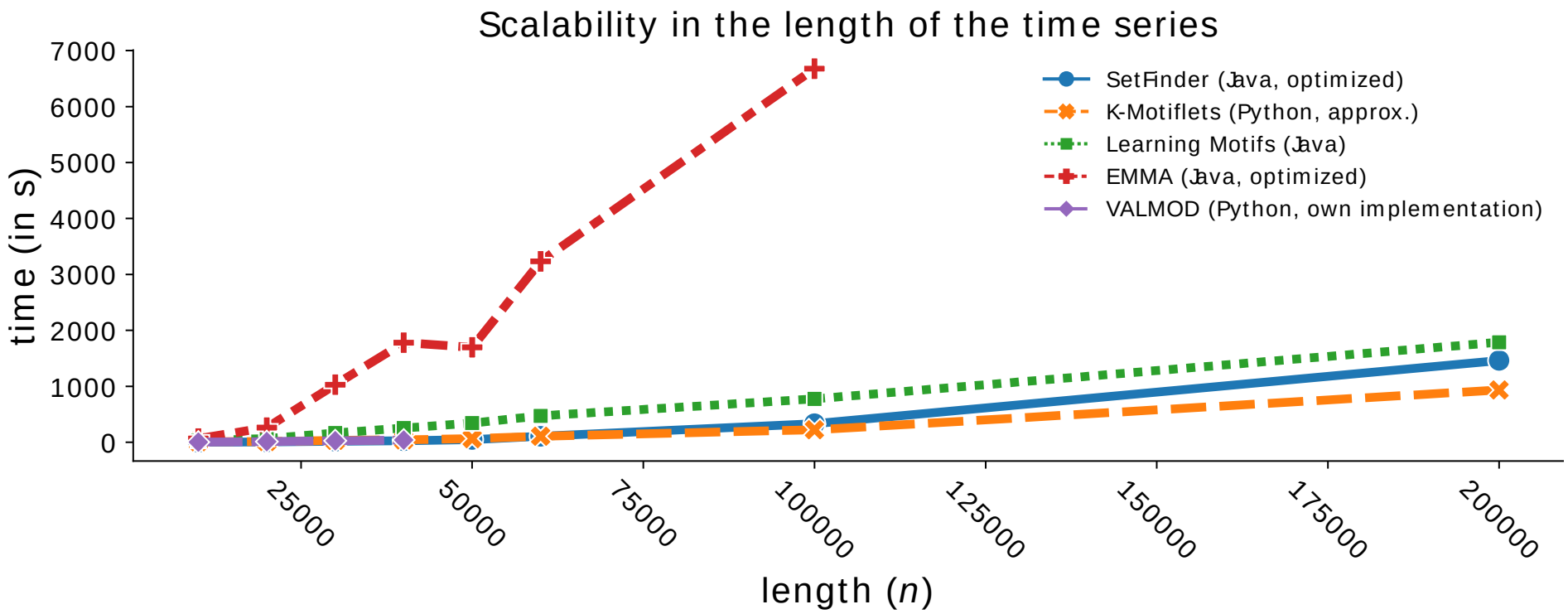
Source Codes

Thank you!

Questions?

Backup Slides

Backup – Scalability



Backup – Rolling Stones – Paint it black

B1

- I see a red door
  And I want it painted black
  No colors anymore
  I want them to turn black

C1

- I see the girls walk by (1)
  Dressed in their summer clothes
  I have to turn my head (2)
  Until my darkness goes

B2

- I see a line of cars
  And they're all painted black
  With flowers and my love
  Both never to come back

C2

- I've seen people turn their heads (3)
  And quickly look away
  Like a newborn baby (4)
  It just happens everyday

B3

- I look inside myself
  And see my heart is black
  I see my red door
  I must have it painted black

C3

- Maybe then, I'll fade away (5)
  And not have to face the facts
  It's not easy facing up (6)
  When your whole world is black

D1

- No more will my green sea
  Go turn a deeper blue
  I could not foresee this thing
  Happening to you

C4

- If I look hard enough (7)
  Into the setting sun
  My love will laugh with me (8)
  Before the morning comes

B1

- I see a red door
  And I want it painted black
  No colors anymore
  I want them to turn black

C1

- I see the girls walk by (9)
  Dressed in their summer clothes
  I have to turn my head (10)
  Until my darkness goes

F

- I wanna see it painted
  Painted black
  Black as night
  Black as coal
  I wanna see the sun
  Blotted out from the sky
  I wanna see it painted, painted, painted
  Painted black, yeah

The instruments and the style are all played in the same manner for the B and C parts.

In the verse D1, the instrumentation changes slightly with all the other instruments dropping out except for the sitar.

Backup – 10-Motiflet Found

B1
- I see a red door
  And I want it painted black
  No colors anymore
  I want them to turn black

C1
- I see the girls walk by (1)
  Dressed in their summer clothes
  I have to turn my head (2)
  Until my darkness goes

B2
- I see a line of cars
  And they're all painted black
  With flowers and my love
  Both never to come back

C2
- I've seen people turn their heads (3)
  And quickly look away
  Like a newborn baby (4)
  It just happens everyday

B3
- I look inside myself
  And see my heart is black
  I see my red door
  I must have it painted black

C3
- Maybe then, I'll fade away (5)
  And not have to face the facts
  It's not easy facing up (6)
  When your whole world is black

D1
- No more will my green sea
  Go turn a deeper blue
  I could not foresee this thing
  Happening to you

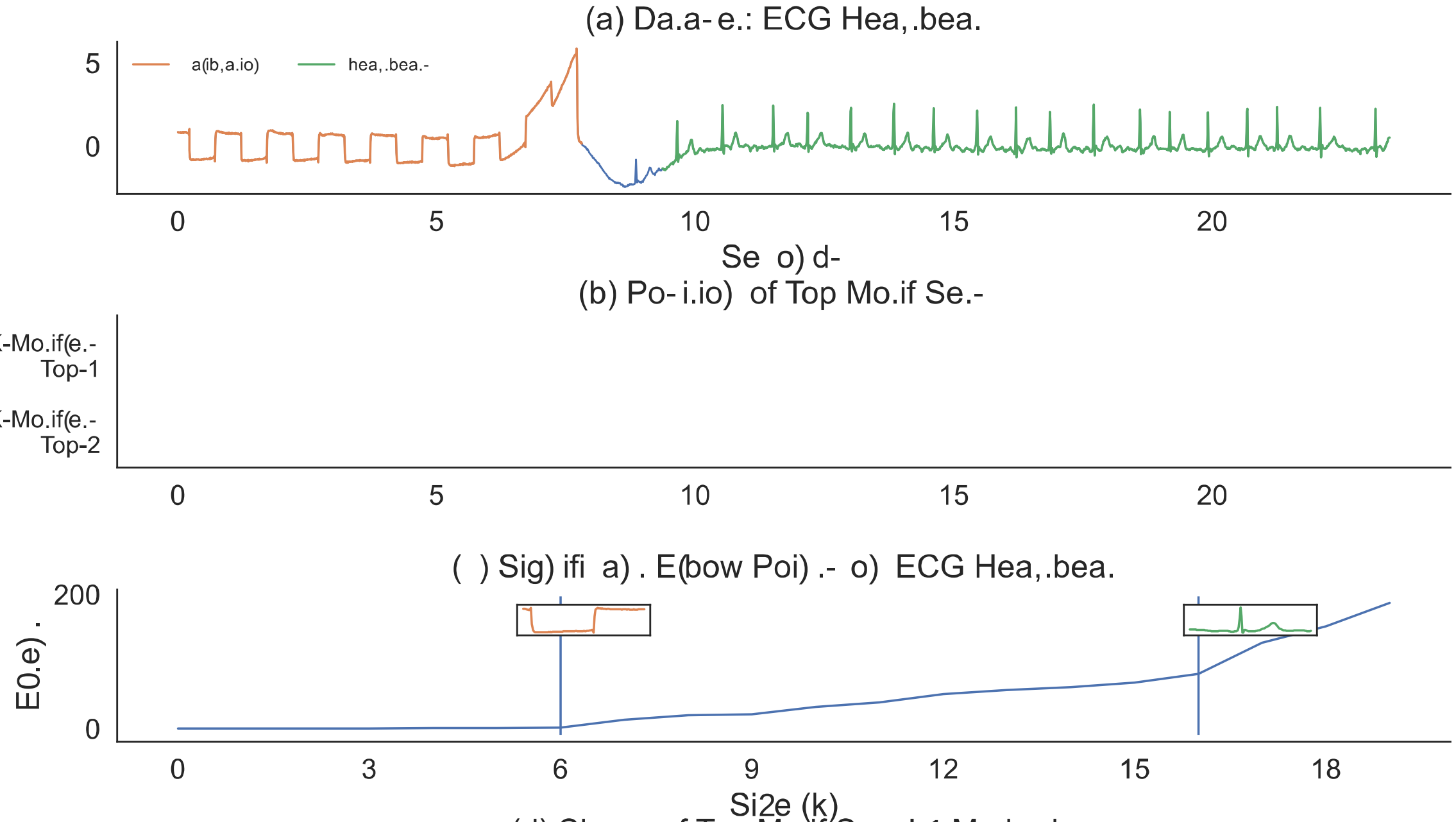
C4
- If I look hard enough (7)
  Into the setting sun
  My love will laugh with me (8)
  Before the morning comes

B1
- I see a red door
  And I want it painted black
  No colors anymore
  I want them to turn black

C1
- I see the girls walk by (9)
  Dressed in their summer clothes
  I have to turn my head (10)
  Until my darkness goes

F
- I wanna see it painted
  Painted black
  Black as night
  Black as coal
  I wanna see the sun
  Blotted out from the sky
  I wanna see it painted, painted, painted
  Painted black, yeah

The instruments and the style are all played in the same manner for the B and C parts.
In the verse D1, the instrumentation changes slightly with all the other instruments dropping out except for the sitar.


(a) Da.a- e.: ECG Hea,.bea.
5
0
a(ib,a.io)
hea,.bea.-
0
5
10
15
20
Se o) d-
(b) Po- i.io) of Top Mo.if Se.-
K-Mo.if(e.-
Top-1
K-Mo.if(e.-
Top-2
0
5
10
15
20
( ) Sig) ifi a) . E(bow Poi) .- o) ECG Hea,.bea.
200
E0.e) .
0
0
3
6
9
12
15
18
Si2e (k)