THE WEB CONFERENCE ACM

Dynamic Multi-Network Mining of Tensor Time Series

Kohei Obata, Koki Kawabata, Yasuko Matsubara, Yasushi Sakurai

SANKEN, Osaka University

Outline

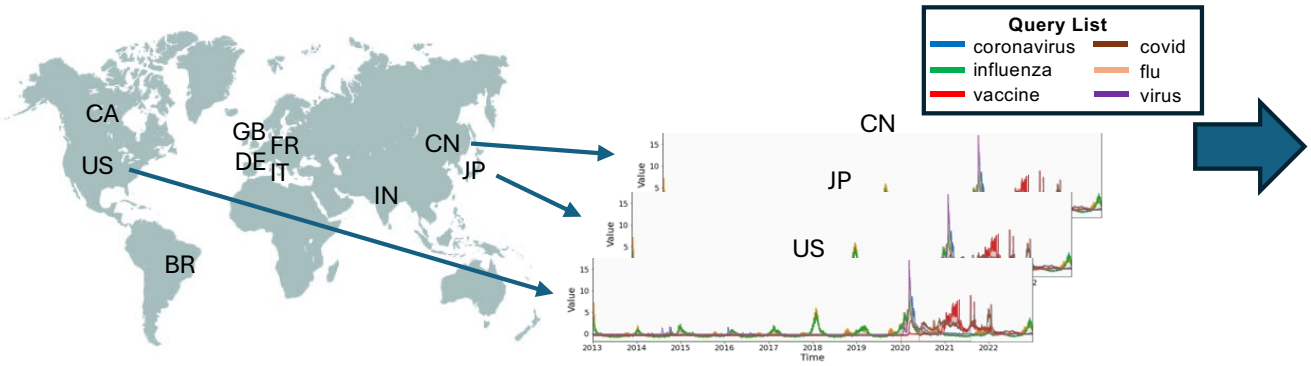
- Introduction
- Preliminaries
- Methodology
- Experiments
- Conclusion

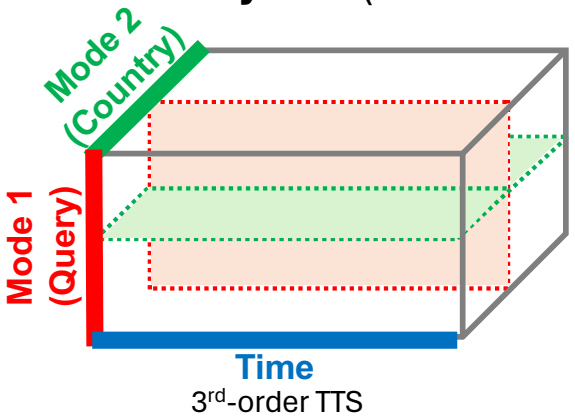
Ubiquity of Tensor Time Series (TTS)

- IoT has facilitated the collection of TTS.
- TTS consists of multiple modes including Time.
  - e.g., Online activity data {Query, Country, Timestamp},
    Air pollutant data {Pollutant, Site, Timestamp},
    Automobile data {Sensor, Lap, Driver, Meter}.



Example of Online activity data:
The search amount of **6 Queries** related to Covid19 taken from **10 Countries** for **10 years** (2013 ~ 2023).



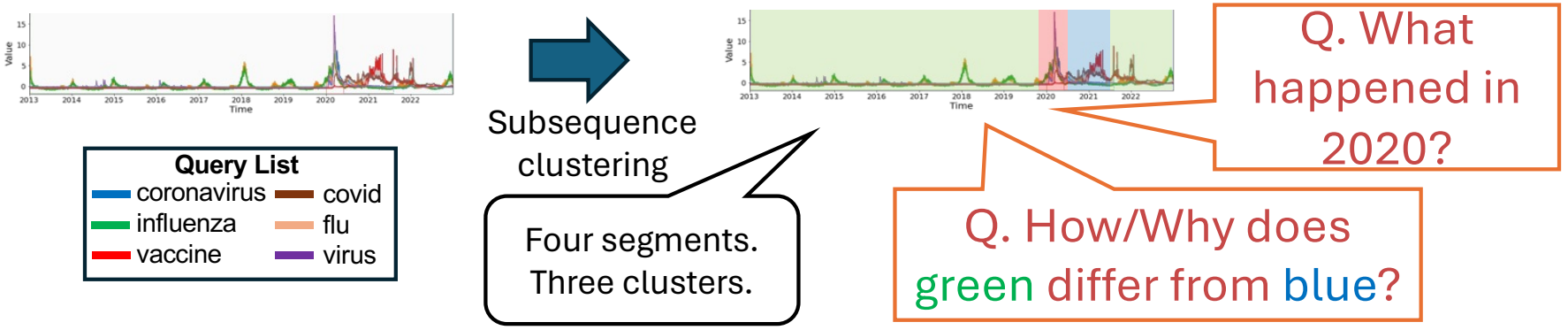


3rd-order TTS

Subsequence Clustering

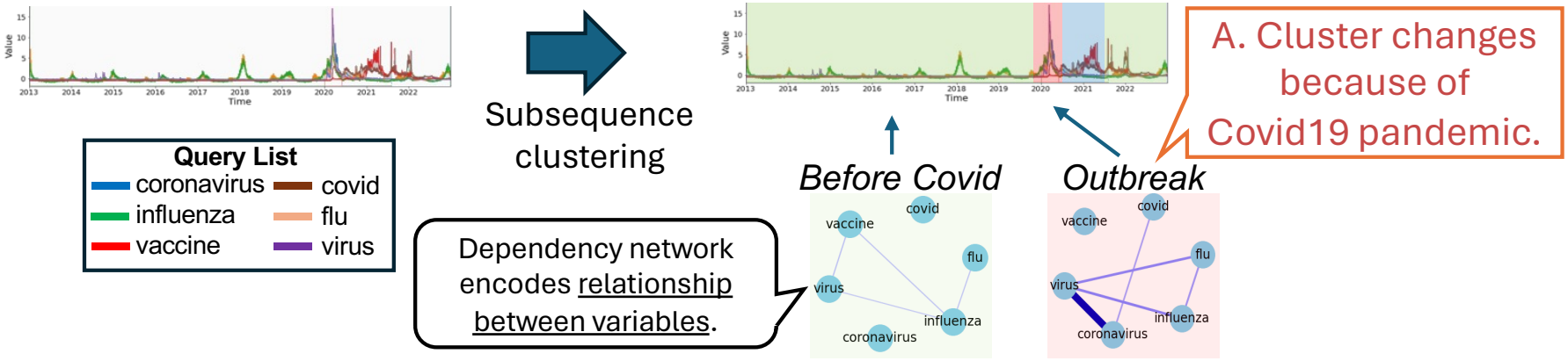
- Important task for data mining.
  - uncover interesting patterns, useful for downstream tasks
- **Interpretability** of resulting clusters is also essencial.

Example of Online activity data:
The search amount of **6 Queries** related to Covid19 for **10 years** (2013 ~ 2023).



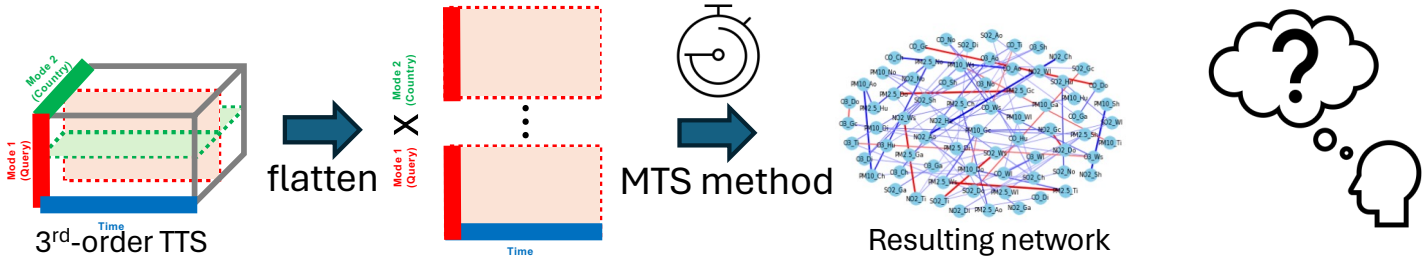
Subsequence Clustering

- Important task for data mining.
  - uncover interesting patterns, useful for downstream tasks
- **Interpretability** of resulting clusters is also essencial.
- **Dependency network** gives a clear explanation.
  - why a particular cluster distinguishes itself from another?
  - what happened during a period belonging to the cluster?



Existing Works of Subsequence Clustering

- Univariate time series clustering (e.g., K-means, DTW).
  - Focus on matching raw values.
  - Do not consider relationship between variables.
- Multivariate time series clustering (e.g., TICC[1], TAGM[2]).
  - Characterize cluster with network based on graphical lasso.
  - Discover clusters that other traditional methods cannot find.
- TTS clustering is more challenging.
  - It has an intricate dependency and huge data size due to the modes.
  - Applying MTS methods to TTS hinders interpretability and requires time.



3rd-order TTS

[1] David Hallac, Sagar Vare, Stephen P. Boyd, and Jure Leskovec. 2017. Toeplitz Inverse Covariance-Based Clustering of Multivariate Time Series Data. In KDD. 215–223.
[2] Veronica Tozzo, Federico Ciech, Davide Garbarino, and Alessandro Verri. 2021. Statistical Models Coupling Allows for Complex Local Multivariate Time Series Analysis.InKDD.1593–1603.

Outline

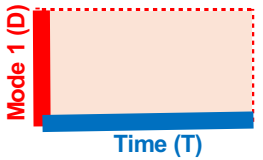
- Introduction
- Preliminaries
- Methodology
- Experiments
- Conclusion

Network & Graphical Lasso

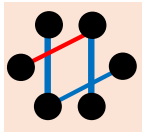
- Inverse covariance matrix (i.e. network) of Gaussian distribution encodes dependency between features.
  - If there is an edge, then they are dependent given the rest of the variables.
- Graphical lasso infers a sparse network, which helps us understand the important relationship.

Given: Multivariate time series: $X \in \mathbb{R}^{D \times T}$

Output: Network (precision matrix): $\theta \in \mathbb{R}^{D \times D}$



Given MTS


Network

minimize $\boxed{-ll(X,\theta)} + \boxed{\lambda\|\theta\|_1}$

Gaussian Log-Likelihood

L1 norm

λ controls the sparsity of network

Problem Formulation

Given: $(N+1)$th-order TTS $\mathcal{X} \in \mathbb{R}^{D_1 \times \cdots \times D_N \times T}$

Find: $\mathcal{M}$ that minimize the cost function $Cost_T(\mathcal{X}; \mathcal{M})$

Cluster parameter $\mathcal{M} = \{\Theta, \mathcal{F}\}$

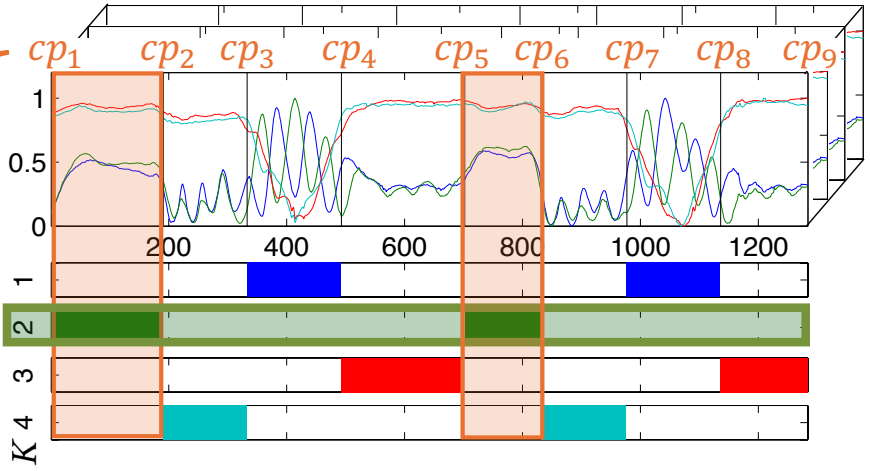
- Cluster assignment set $\mathcal{F} = \{f_k\}_{k=1}^{K}$
  e.g., $f_2 = \{cp_1, \ldots, cp_2 - 1, cp_5, \ldots, cp_6 - 1\}$
- Model parameter set $\Theta = \{\theta_k\}_{k=1}^{K}$

Cut-Point: starting point of a segment

Segment: TTS is separated into m segments

Cluster: set of segment having the same model

# of clusters



Outline

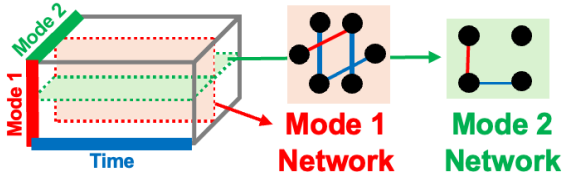
- Introduction
- Preliminaries
- Methodology
- Experiments
- Conclusion

DMM: *Dynamic Multi-Network Mining*

DMM achieves interpretable TTS clustering.

(1) Characterize a cluster with multiple networks.

- Extend graphical lasso to TTS.
- Each mode has a dependency network.



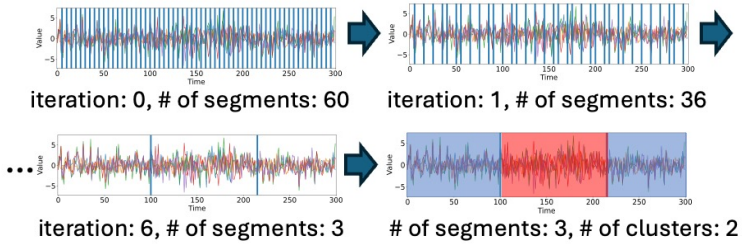
(2) Define the cost function (goodness of clustering).

- Based on the Minimum Description Length (MDL).
- Can determine any hyperparameters (e.g., λ, K).

$$\underbrace{Cost_T(\mathcal{X};\mathcal{M})}_{\text{Total description cost}} = \underbrace{Cost_A(\mathcal{F})}_{\text{Coding length cost}} + \underbrace{Cost_M(\Theta)}_{\text{Model coding cost}} + \underbrace{Cost_C(\mathcal{X}|\mathcal{M})}_{\text{Data coding cost}} + \underbrace{Cost_{\ell_1}(\Theta)}_{\text{L1-norm cost}}.$$

(3) Propose the algorithm that minimize the cost.

- Based on the bottom-up algorithm.
- Scales linearly w.r.t. data size.
- It can be applied to long-range, high-dimensional TTS.



(1) Multimode Graphical Lasso

- We model TTS with *N* networks.
- The model is **interpretable** as each mode has a sparse network.

# of dimensions at mode-N

Timesteps

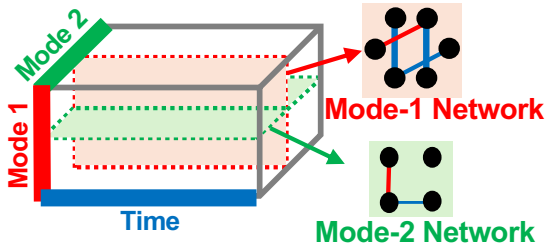
Given: $(N+1)$th-order TTS: $\mathcal{X} \in \mathbb{R}^{D_1 \times \cdots \times D_N \times T}$

Output: N networks: $\theta = \{A^{(1)}, \cdots, A^{(N)}\}, A^{(n)} \in \mathbb{R}^{D_n \times D_n}$

Our model

Mode-n network

$$\text{minimize} \sum_{n=1}^{N} \underbrace{-ll_n\left(\mathcal{X}, A^{(n)}\right)}_{\text{Likelihood}} + \underbrace{\lambda \left\| A^{(n)} \right\|_1}_{\text{L1 norm}}$$



(2) Model Description Cost

- We define the criterion for the goodness of clustering based on Minimum Description Length (MDL).
- It can determine any hyperparameters (e.g., λ, K) by minimizing $Cost_T(\mathcal{X}; \mathcal{M})$.

Description complexity of $\mathcal{M} = \{\Theta, \mathcal{F}\}$

Coding length cost　Model coding cost

$$\underbrace{Cost_T(\mathcal{X}; \mathcal{M})}_{\text{Total description cost}} = Cost_A(\mathcal{F}) + Cost_M(\Theta) + Cost_C(\mathcal{X}|\mathcal{M}) + Cost_{\ell_1}(\Theta).$$

Log likelihood of $\mathcal{X}$ given $\mathcal{M}$

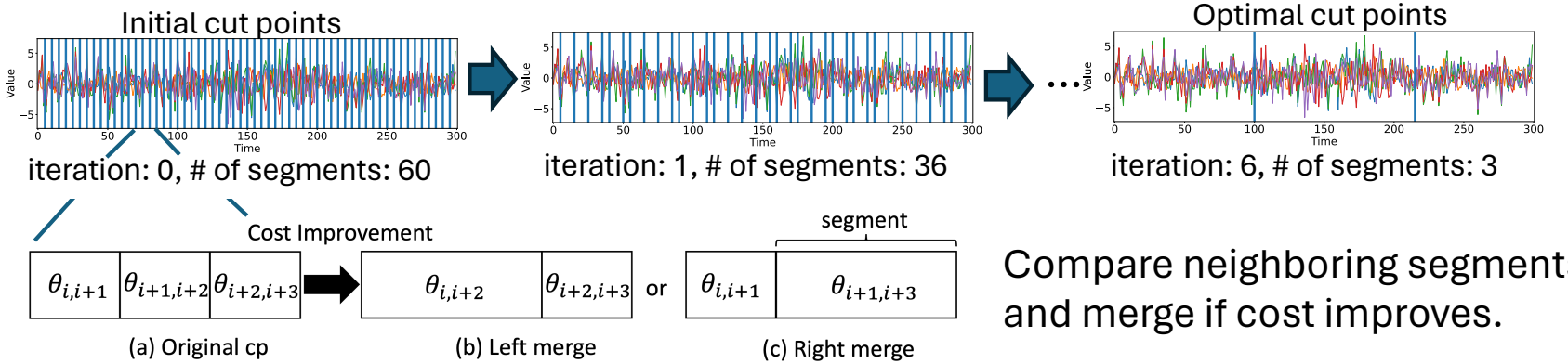
Data coding cost　L1-norm cost

(3) Scalable Algorithm

DMM Input: $\mathcal{X}$ and initial segment size $\mathbf{w}$

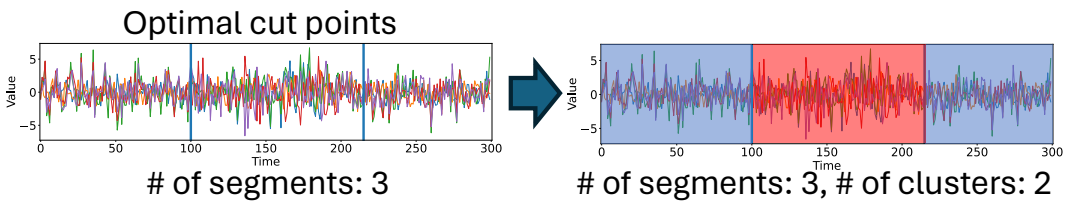
Output: $\mathcal{M} = \{\Theta, \mathcal{F}\}$ that minimizes $Cost_T(\mathcal{X}; \mathcal{M})$

CutPointDetector: Detect optimal cut points via Bottom-up algorithm



Compare neighboring segments and merge if cost improves.

ClusterDetector: Detect # of clusters and cluster assignment



Vary the number of cluster from 1 to m until the cost converges.

Outline

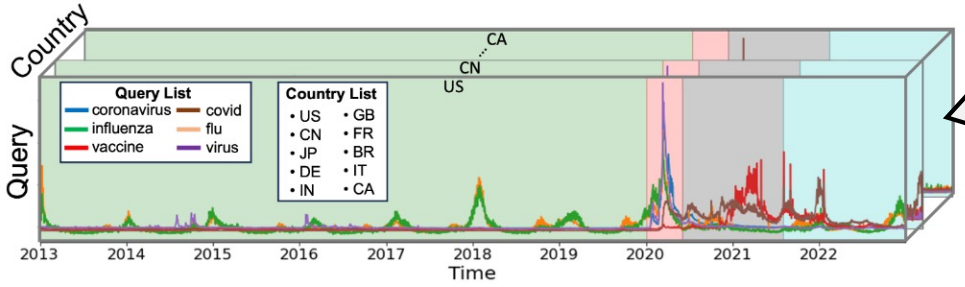
- Introduction
- Preliminaries
- Methodology
- Experiments
- Conclusion

DMM is Interpretable (Covid data)

- DMM helps us understand the real-world data.

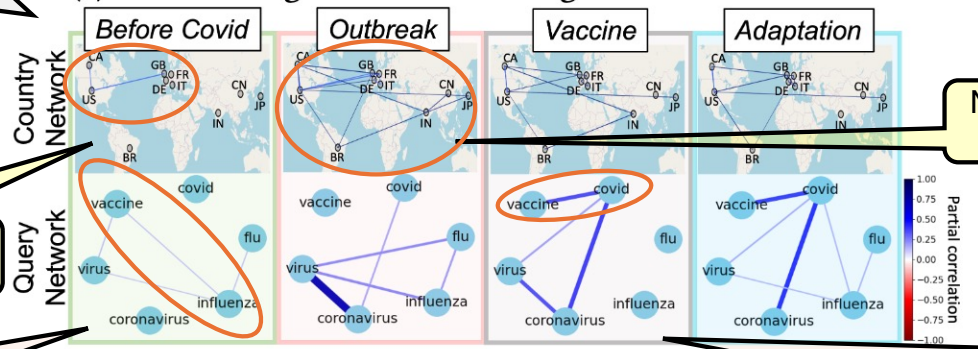
The search amount of **6 Queries** related to Covid19 taken from **10 Countries** for **10 years** (2013 ~ 2023).

Each cluster is characterized with Country and Query Networks.



DMM discovers 4 clusters and **captures the effect against Covid19.**

(a) Cluster assignments on the original tensor time series



(b) Country and query networks change dynamically

Many countries simultaneously react to Covid19.

Edges between English-speaking countries.

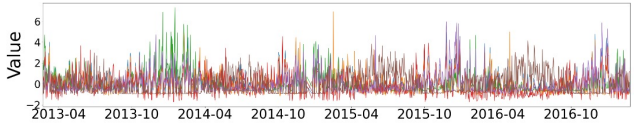
"vaccine" is connected with "influenza".

People started to care about prevention from Covid19, and "vaccine" strongly connects to "covid".

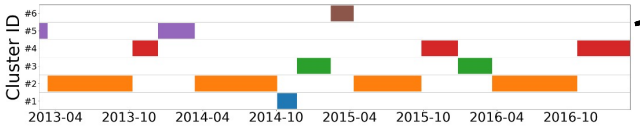
DMM is Interpretable (Air data)

- DMM is more interpretable than existing MTS methods (TAGM and TICC).

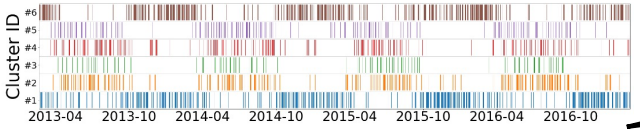
Concentration of **6 Pollutants** taken from **10 Locations** in China for **4 years** (2013 ~ 2017).



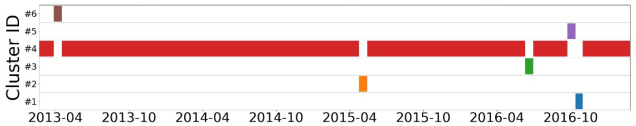
(a) Original sensor data at a site (Aoti Zhongxin)



(b) DMM assigns every Apr. ~ Oct. to cluster #2



(c) TAGM casuses frequent switching



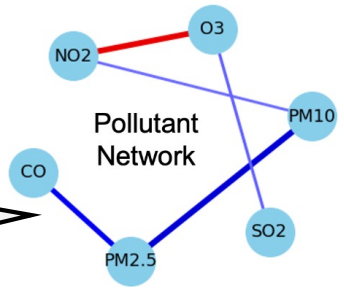
(d) TICC assigns most periods to cluster #4

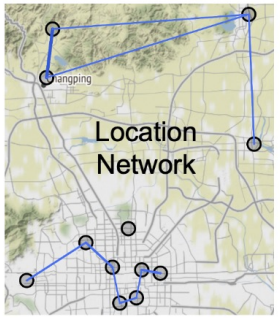
DMM finds **seasonality**.

DMM provides two **simple** networks.

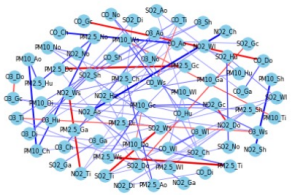
Existing methods **fail** to finds reasonable clusters.

Networks of existing methods have **large # of nodes and edges**.

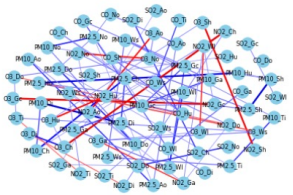




(a) DMM networks of cluster #2



(b) TAGM network of cluster #6



(c) TICC network of cluster #4

DMM is Accurate (Synthetic data)

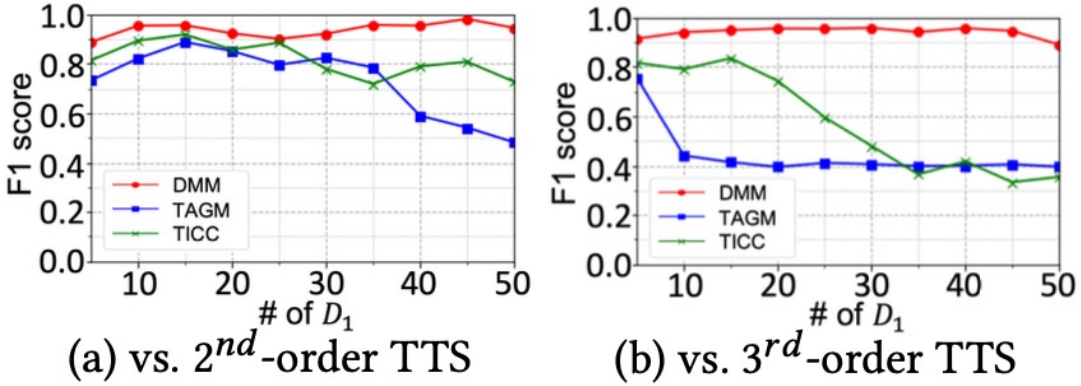
- DMM accurately discovers cluster with different networks even at MTS.

Table 1. Macro-F1 score of Synthetic data.

Data		DMM	TAGM	TAGM †	TICC	TICC †
2nd-order TTS (MTS) (i)	A	0.955	0.915	0.915	**0.997**	**0.997**
	B	**0.926**	0.897	0.756	0.884	0.825
	C	**0.956**	0.770	0.811	0.725	0.756
	D	**0.960**	0.907	0.912	0.857	0.952
3rd-order TTS (ii)	A	**0.961**	0.514	0.514	0.932	0.923
	B	**0.962**	0.462	0.431	0.844	0.770
	C	**0.941**	0.359	0.396	0.704	0.594
	D	**0.980**	0.438	0.432	0.838	0.741

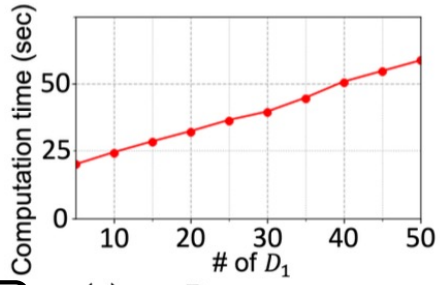
All the hyperparameters, including # of clusters, are determined by the cost function.

- DMM is not affected by the number of variables.



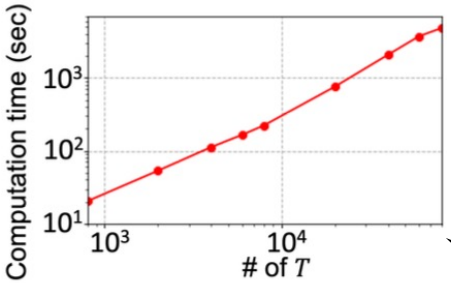
DMM is Scalable

- DMM scales linearly w.r.t. the input data size.



(a) vs. $D_1 = 5 \sim 50$

Linear to Variable



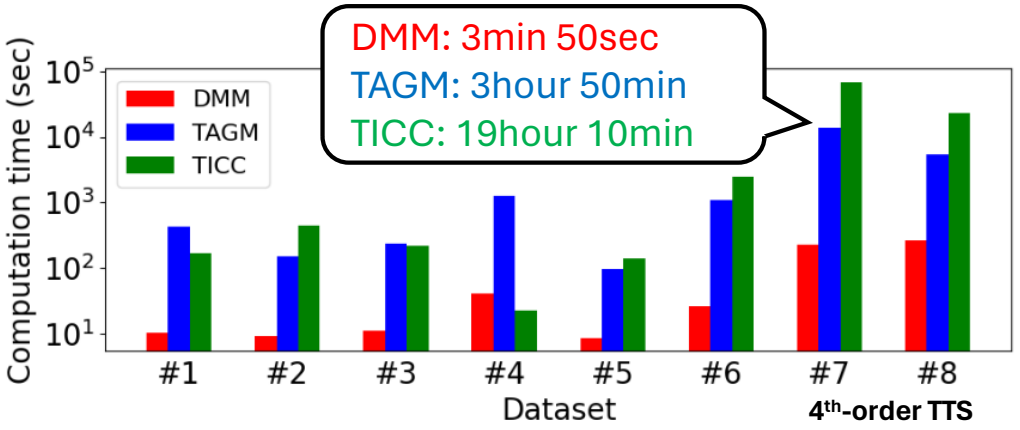
(b) vs. $T = 800 \sim 80000$

Linear to Time

- DMM is up to x300 faster than the existing methods.

Table 3: The data size and attributes for each dataset.

ID	Dataset	Size	Description
#1	E-commerce	(11, 10, 1796)	3rd-order TTS
#2	VoD	(8, 10, 1796)	(query, state, day)
#3	Sweets	(9, 10, 1796)	
#4	Covid	(6, 10, 3652)	(query, country, day)
#5	GAFAM	(5, 10, 1796)	
#6	Air	(6, 12, 1461)	(pollutant, site, day)
#7	Car-A	(6, 10, 4, 3241)	4th-order TTS
#8	Car-H	(6, 10, 4, 4000)	(sensor, lap, driver, meter)



Outline

- Introduction
- Preliminaries
- Methodology
- Experiments
- Conclusion

Conclusion

- DMM is **Interpretable**, **Accurate**, and **Scalable**.
- DMM is a useful tool for **TTS subsequence clustering** that enables multifaceted analysis and understanding of TTS.

- Possible applications could be

Traffic Congestion

Factory

Healthcare

Thank you for listening!

Code & Data are available [1].

[1] https://github.com/KoheiObata/DMM