TENDI: Tensor Disaggregation from Multiple Coarse Views

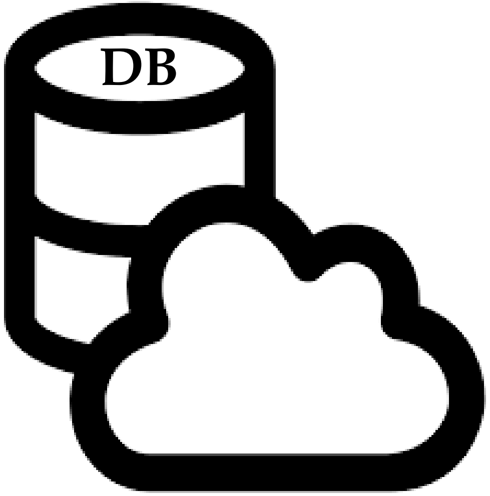
Faisal M. Almutairi[1], Charilaos I. Kanatsoulis[1], and Nicholas D. Sidiropoulos[2]

[1] *ECE Department, University of Minnesota.*
[2] *ECE Department, University of Virginia.*

May 14, 2020

Motivation: data are aggregated

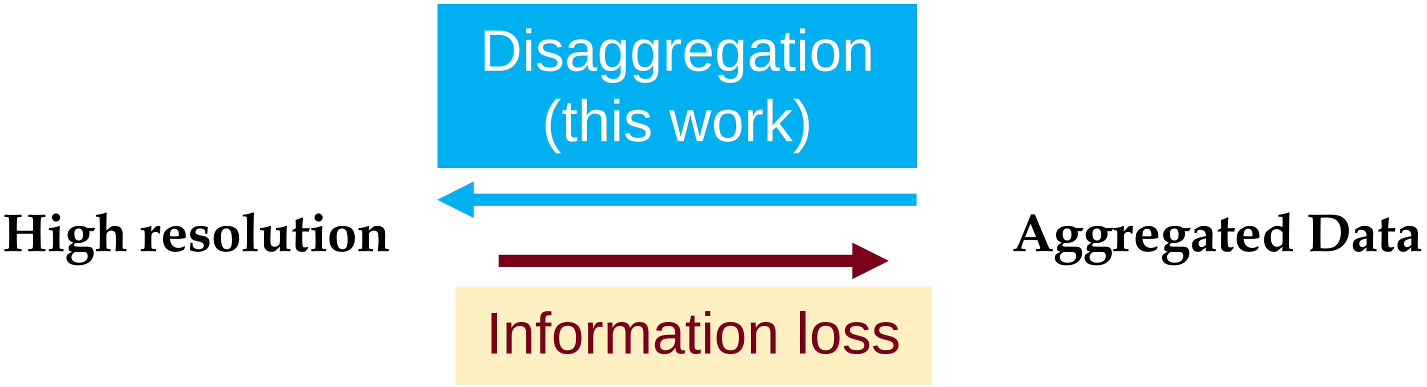
- **Aggregated** → summed (averaged) over multiple data atoms



- **Types:** Temporally, or over other attributes (e.g., geographically)
- **Why:** Scalability, communication cost, or privacy and anonymity

Motivation: data disaggregation

- ML and Data Mining algorithms benefit from data in high-resolution:
  - Epidemiological studies [Panhuis et al., '13]
  - Healthcare (EHR) [Park et sl., '14]
  - Education [Cole, '10]
  - Supply chains [Jin at al., '15]
  - ….



Motivation: retail supply chain application

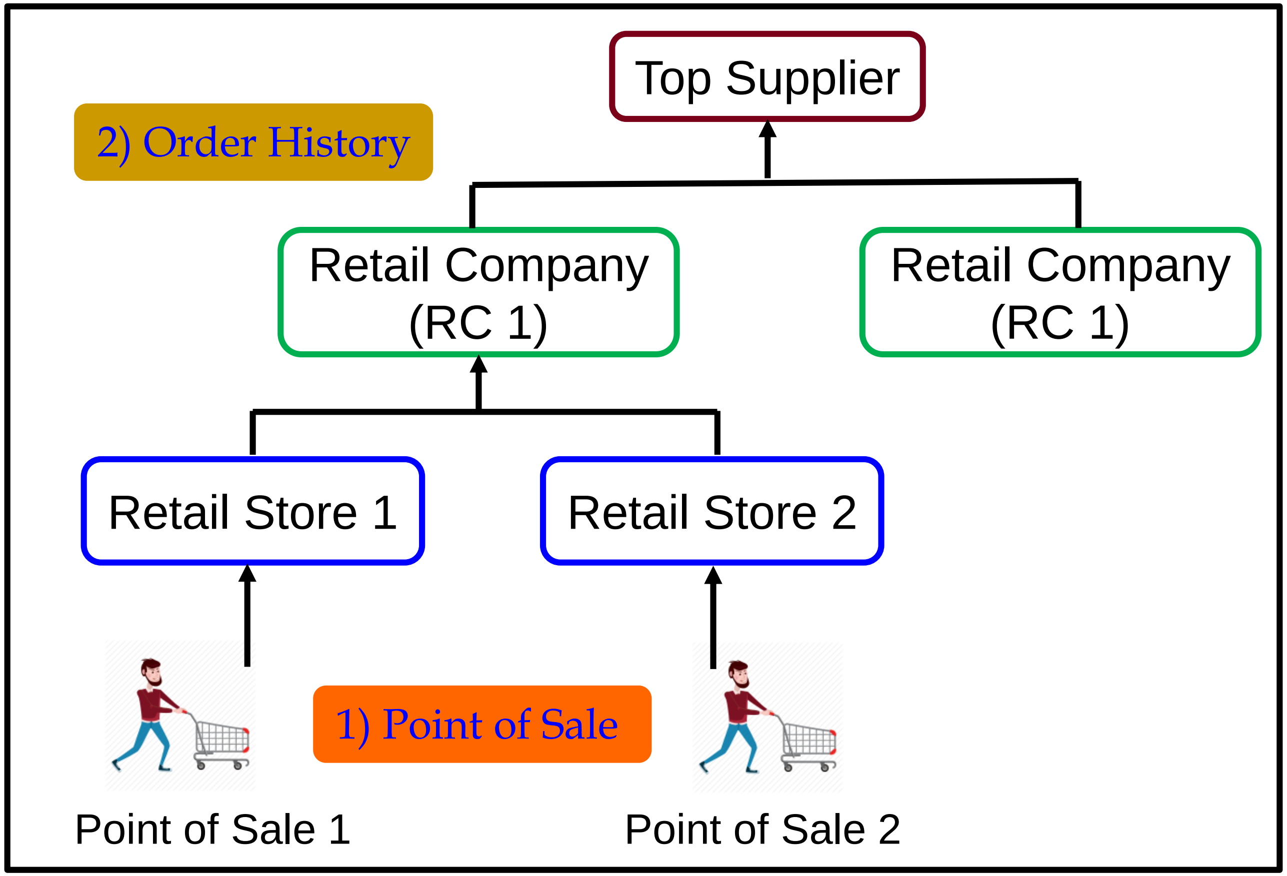
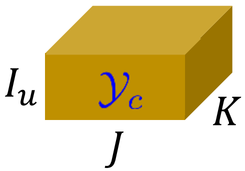


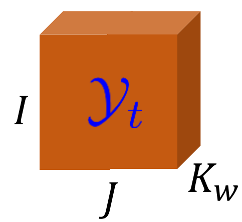
Figure: Retail supply chain model.



Sales (*company*, *item*, *week*)

$I_u < I$

$K_w < K$



Sales (*store*, *item*, *month*)

- Aggregation (in product, location, and time) degrades demand forecasting accuracy in supply chains [1, 2].

[1] G. P. Cachon, T. Randall, and G. M. Schmidt, "*In Search of the Bullwhip Effect*", INFORMS, 2007.
[2] Y. Jin et al., "*Masking the bullwhip effect in retail: the influence of data aggregation*", Int. J. Physical Distribution & Logistics Management, 2015.

Roadmap

- Background
  - (multi-view) data disaggregation
- Tensor Algebra
  - tensor decomposition, mode product
- Proposed Method: TENDI
- TENDIB:
  - *Blind* disaggregation (no knowledge of aggregation pattern)
- Experiments
- Conclusion

Background: data disaggregation

- Data disaggregation: linear inverse problem

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{\mathbf{y}\in\mathbb{R}^M} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{\mathbf{A}\in\mathbb{R}^{M\times N}} \times \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{\mathbf{x}\in\mathbb{R}^N}$$

- Fuse multiple aggregated views
  - E.g., Annual GDP per state (temporal), and quarterly national GDP (contemporaneous)
  - Given: linear (temporal , and contemporaneous) constraints: $\mathbf{y}_t = \mathbf{A}_t\mathbf{x}, \quad \mathbf{y}_c = \mathbf{A}_c\mathbf{x}$

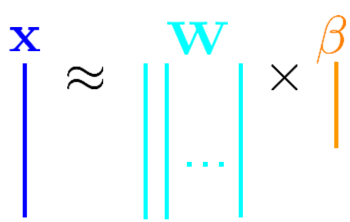
$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_c \end{bmatrix} = \begin{bmatrix} \mathbf{A}_t \\ \mathbf{A}_c \end{bmatrix} \mathbf{x}$$

- Challenge:
  - ill-posed problem (#variables >> #equations)

- Approaches:
  - Impose a prior [1] (increase #equations)

$$\min_{\mathbf{x}} \ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \ + f(\mathbf{x})$$

  - Linear regression [2] (decrease #variables)



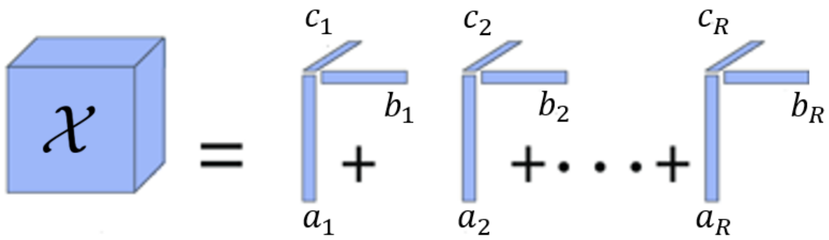
[1] Z. Liu, H. A. Song, V. Zadorozhny, C. Faloutsos, and N. Sidiropoulos, "*H-fuse: Efficient fusion of aggregated historical data*," SDM, 2017
[2] J. Pavía-Miralles, and B. Cabrer-Borrás, "*A survey of methods to interpolate, distribute and extra-polate time series*", J of Service Science and Management, 2010

Canonical Polyadic Decomposition (CPD)

- CPD decomposes a third-order (three mode) tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ into a sum of R rank-one tensors, i.e.,

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

where $\circ$ is a outer products.



$$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

(factor matrices)

$$\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}$$

$$\mathbf{A} = [\mathbf{a}_1 \;\; \mathbf{a}_2 \ldots \mathbf{a}_R]$$ (Similarly for **B** and **C**)

- Property: CPD factors are unique (up to common permutation and scaling)[1].

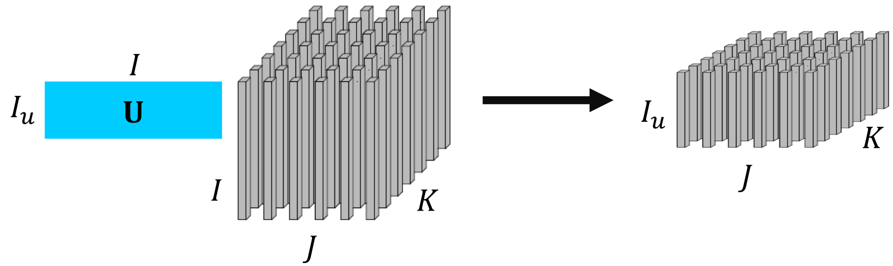
[1] L. Chiantini and G. Ottaviani, "*On generic identifiability of 3-tensors of small rank*," SIAM J. on Matrix Analysis and Applications (SIMAX), 2012.

Tensor Algebra: Mode Product

- Mode product: multiply a *tensor* by a *matrix* in a particular mode, i.e.,

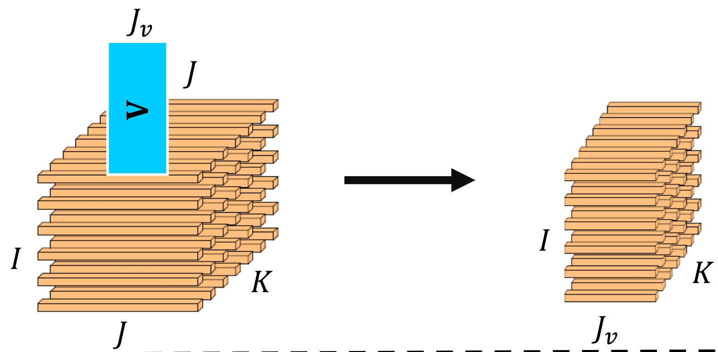
Mode-1 product:

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}$$



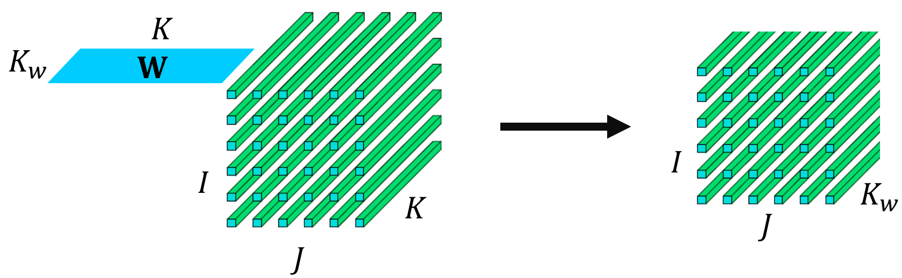
Mode-2 product:

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{V}$$



Mode-3 product:

$$\mathcal{Y} = \mathcal{X} \times_3 \mathbf{W}$$

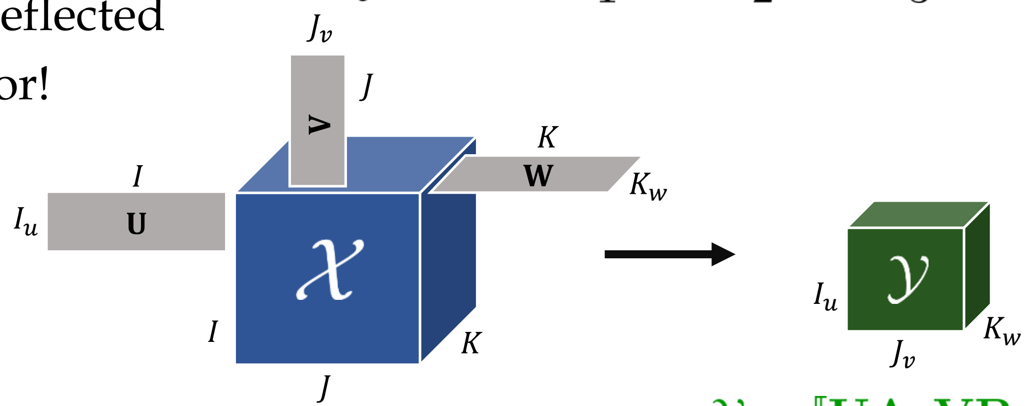


- Tensor size is reduced if multiplied by a "fat" matrix!

Tensor Algebra: Mode Product

- Notes:
  - The mode product is reflected in the CPD of the tensor!

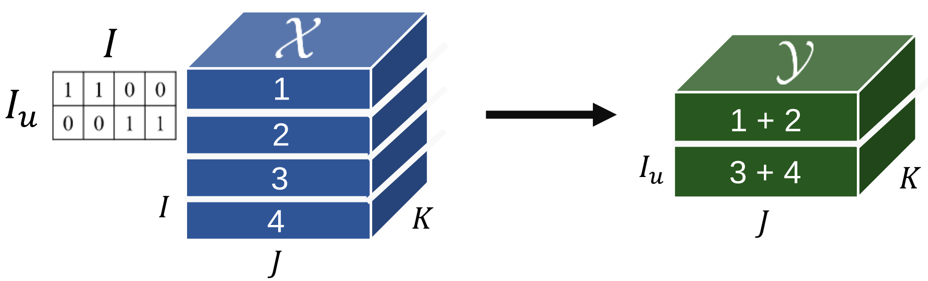
$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$



$$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

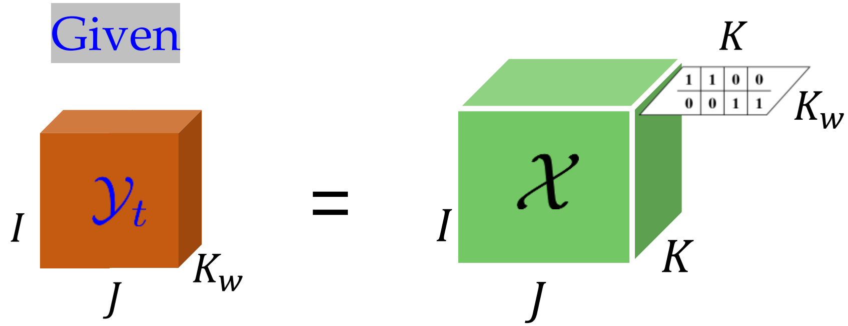
$$\mathcal{Y} = [\![\mathbf{UA}, \mathbf{VB}, \mathbf{WC}]\!]$$

  - Aggregation is a mode product!



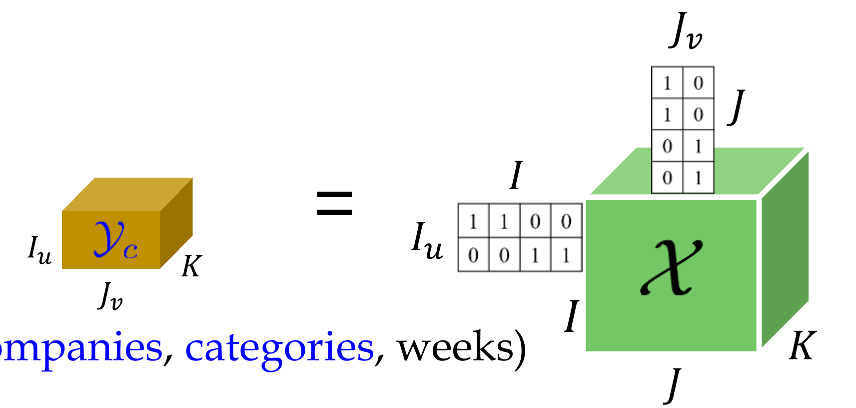
Problem Statement

Given



#items (stores, items, months)

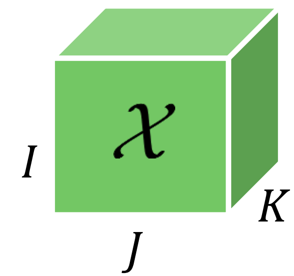
$$\mathcal{Y}_t = \mathcal{X} \times_3 \mathbf{W}$$



#items (companies, categories, weeks)

$$\mathcal{Y}_c = \mathcal{X} \times_1 \mathbf{U} \times_2 \mathbf{V}$$

Recover?



E.g., #items (stores, items, weeks)

TENDI: formulation

- TENDI learns the *factor matrices* of the target tensor by coupled tensor factorization (effective in other applications [1]).

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \quad \mathcal{F} = \|\mathcal{Y}_t - [\![\mathbf{A}, \mathbf{B}, \mathbf{WC}]\!]\|_F^2 + \|\mathcal{Y}_c - [\![\mathbf{UA}, \mathbf{VB}, \mathbf{C}]\!]\|_F^2$$

$$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\mathcal{Y}_t = [\![\mathbf{A}, \mathbf{B}, \mathbf{WC}]\!]$$
$$\mathcal{Y}_c = [\![\mathbf{UA}, \mathbf{VB}, \mathbf{C}]\!]$$

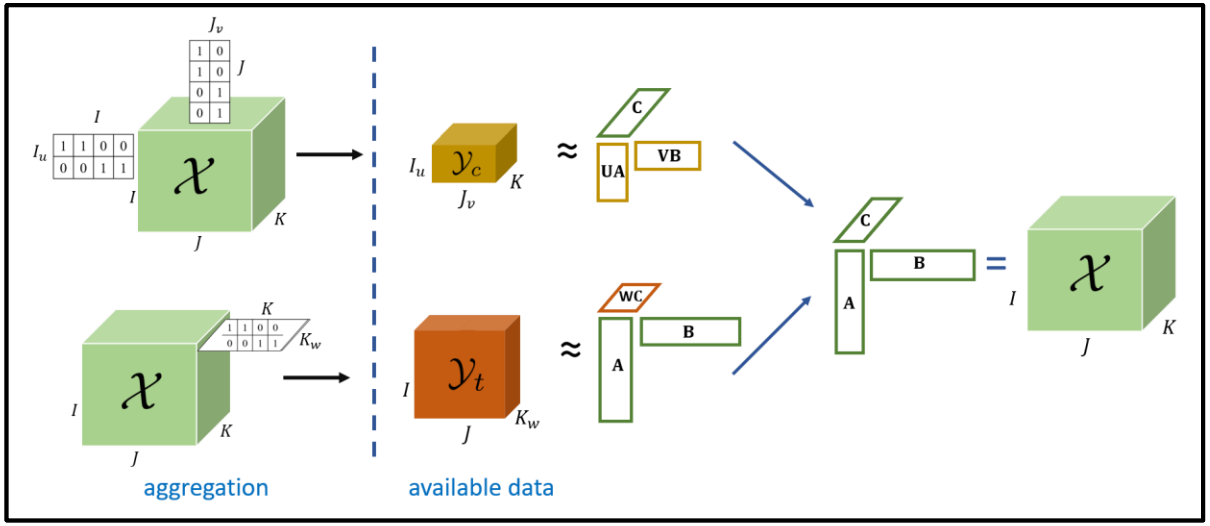


Figure: Overview of TENDI.

- The factors **A, B**, and **C** are identifiable! (under mild conditions).

[1] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos and W. Ma, "*Hyperspectral Super-Resolution: A Coupled Tensor Factorization Approach,*" TSP, 2018

TENDIB: blind disaggregation

- What if we do not know the aggregation matrices?

- TENDIB Blindly disaggregates by treating the "aggregated" factors as separate variables: $\widetilde{\mathbf{A}} = \mathbf{U}\mathbf{A}, \quad \widetilde{\mathbf{C}} = \mathbf{W}\mathbf{C}$

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\widetilde{\mathbf{A}},\widetilde{\mathbf{C}}} \quad \mathcal{F} = \|\mathcal{Y}_t - [\![\mathbf{A}, \mathbf{B}, \widetilde{\mathbf{C}}]\!]\|_F^2 + \|\mathcal{Y}_c - [\![\widetilde{\mathbf{A}}, \mathbf{B}, \mathbf{C}]\!]\|_F^2 + \mu\|\mathbf{1}^T\widetilde{\mathbf{C}} - \mathbf{1}^T\mathbf{C}\|_2^2$$

- Then, factors (**A**, **B**,**C**) are used to reconstruct the target tensor.

- Scaling ambiguity? $\mathcal{Y}_t = [\![\lambda\mathbf{A}, \mathbf{B}, 1/\lambda\widetilde{\mathbf{C}}]\!], \quad \mathcal{Y}_c = [\![\sigma\widetilde{\mathbf{A}}, \mathbf{B}, 1/\sigma\mathbf{C}]\!]$

$$\mathcal{X} = [\![\lambda\mathbf{A}, \mathbf{B}, 1/\sigma\mathbf{C}]\!]$$

- Column sums of **C** = column sums of $(\widetilde{\mathbf{C}} = \mathbf{W}\mathbf{C})$

Experimental Setup (1/2)

- Datasets:
  - Dominick's Finer Foods (DFF): #sold items (stores, items, weeks)
  - Walmart: #sold items (stores, departments, weeks)
  - Crime: #crime incidents (locations, types, months)

Table 1: Summary of datasets and their statistics.

Dataset	Size	Max	Avg	SD	% zeros
Cheese	93×50×393	18176	24.36	84.75	14.10
Fabric Softeners	93×50×397	7168	4.62	17.13	27.48
Mixed Sales	93×500×230	17610	16.10	65.98	17.83
Walmart	45×99×143	6.93e+05	1.05e+04	1.99e+04	19.38
Crime	304×388×221	325	0.26	1.47	91.56

Experimental Setup (2/2)

- Baselines:
  - **Mean**: data atoms contribute equally in their corresponding sum.
  - **LS**: min ||. || solution to the under-determined linear system:

$$\min_{\mathbf{x}} \ \|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}\|_2^2 + \|\mathbf{y}_c - \mathbf{A}_c\mathbf{x}\|_2^2$$

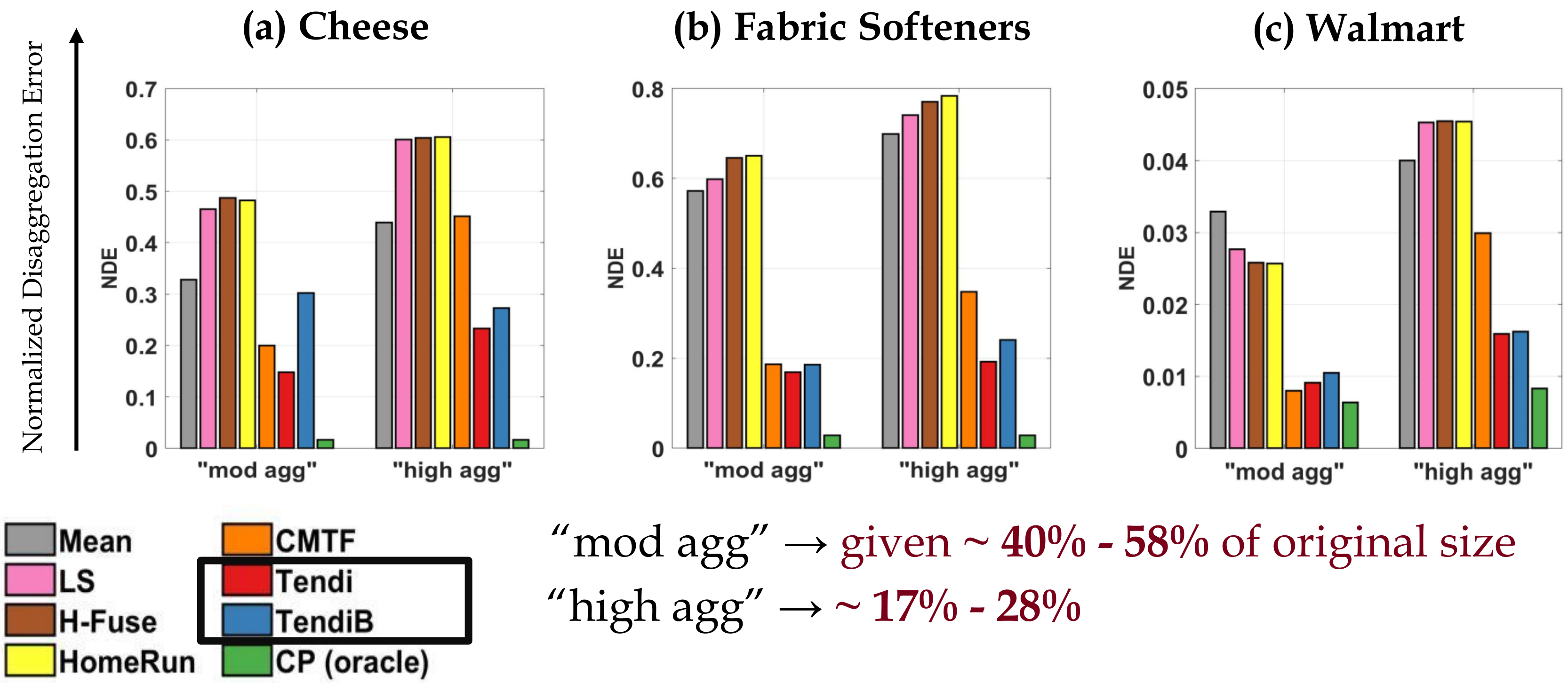
$$\mathbf{x} = \mathrm{vec}(\mathcal{X}), \quad \mathbf{y}_t = \mathrm{vec}(\mathcal{Y}_t), \quad \mathbf{y}_c = \mathrm{vec}(\mathcal{Y}_c)$$

  - **H-Fuse** [Liu et al., SDM '17]: enforces smoothness on the solution **x** to LS above.
  - **HomeRun** [1]: finds the series representation in frequency domain (discrete cosine transform).
  - **CMTF** [2]: coupled metricized tensor factorization.
  - **CP**: tensor decomposition of ground truth (lower bound of disaggregation error).

[1] F. M Almutairi, F. Yang, H. A. Song, C. Faloutsos, N. Sidiropoulos, and V. Zadorozhny, "*HomeRun: Scalable Sparse-Spectrum Reconstruction of Aggregated Historical Data,*" VLDB, 2018.
[2] M. Simoes,J. Bioucas-Dias, L.B. Almeida, J. Chanussot, "*A convex formulation for hyperspectral image superresolution via subspace-based regularization,*" IEEE Transactions on Geoscience and Remote Sensing, 2014

Experimental Results: single aggregation

- Target is (locations × types × time stamps) tensor, we have:
  - Temporal aggregate: (locations × types × agg. time stamps) tensor
  - Contemporaneous aggregate: (agg. locations × types × time stamps) tensor

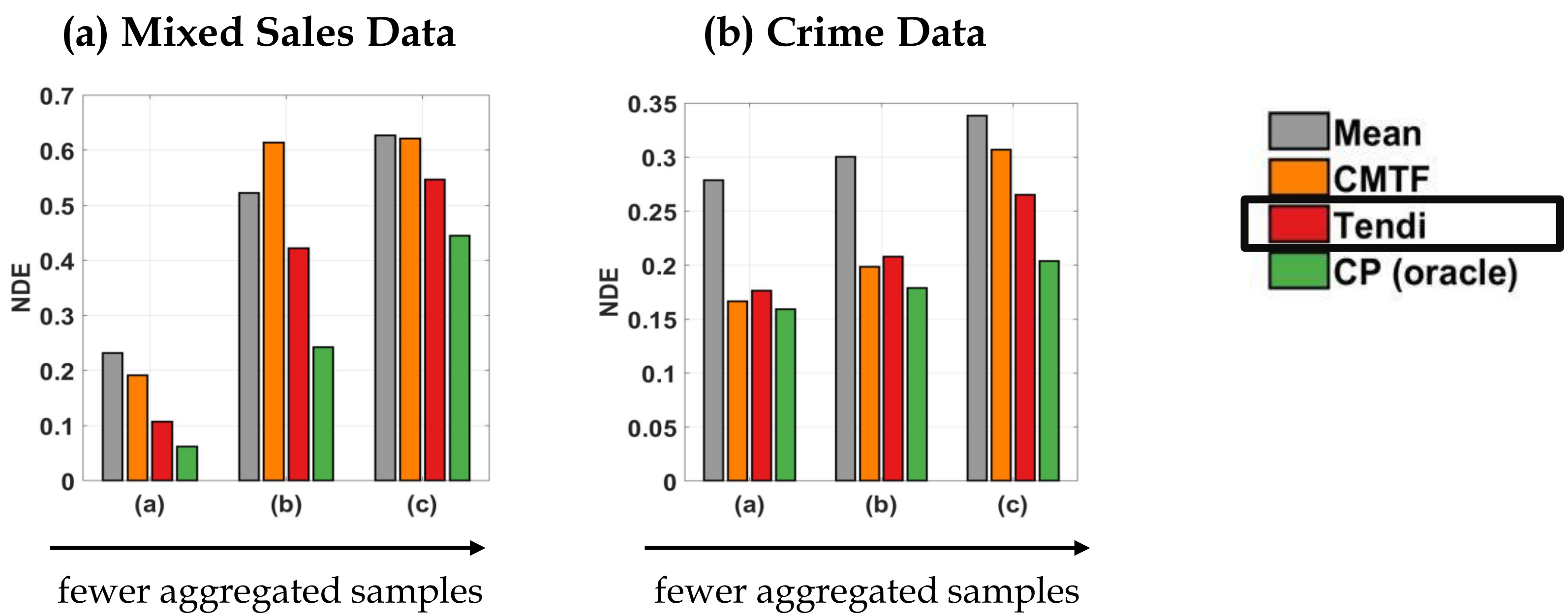


"mod agg" → given ~ **40% - 58%** of original size

"high agg" → ~ **17% - 28%**

Experimental Results: double aggregation

- Target is (locations × types × time stamps) tensor, we have:
  - Temporal aggregate: (locations × types × agg. time stamps) tensor
  - Contemporaneous aggregate: (agg. locations × agg. types × time stamps) tensor



Conclusions & take home points

- Formally defined the problem of multidimensional data disaggregation
- TENDI provably converts an ill-posed problem to an identifiable one
- Significantly improve the disaggregation accuracy of baselines on real data.
- TENDI disaggregates under challenging scenarios: blind disaggregation, double aggregation

THANK YOU!

Questions?

Details of this work:

- TENDI: Tensor Disaggregation from Multiple Coarse Views (PAKDD20202)
- PREMA: Principled Tensor Data Recovery from Multiple Aggregated Views (journal version on arXiv)