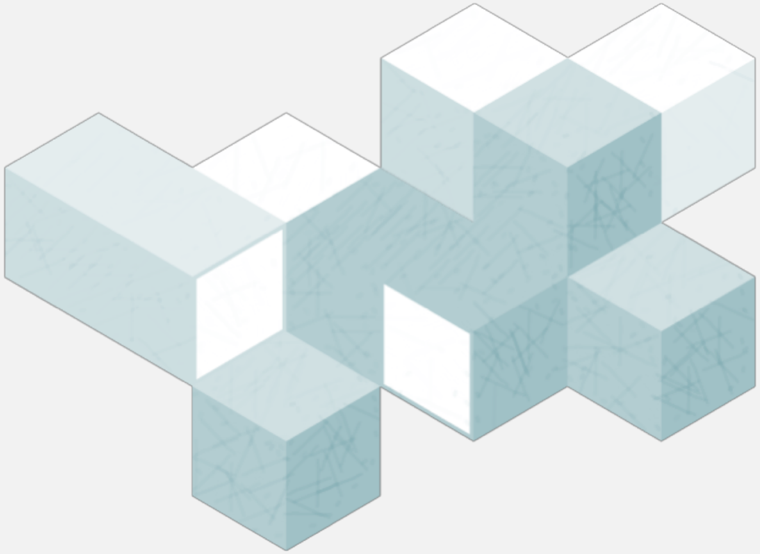
Online CP Decomposition for Sparse Tensors

Shuo Zhou, Sarah Erfani, James Bailey
School of Computing and Information Systems
The University of Melbourne

ICDM'2018, Singapore



Introduction

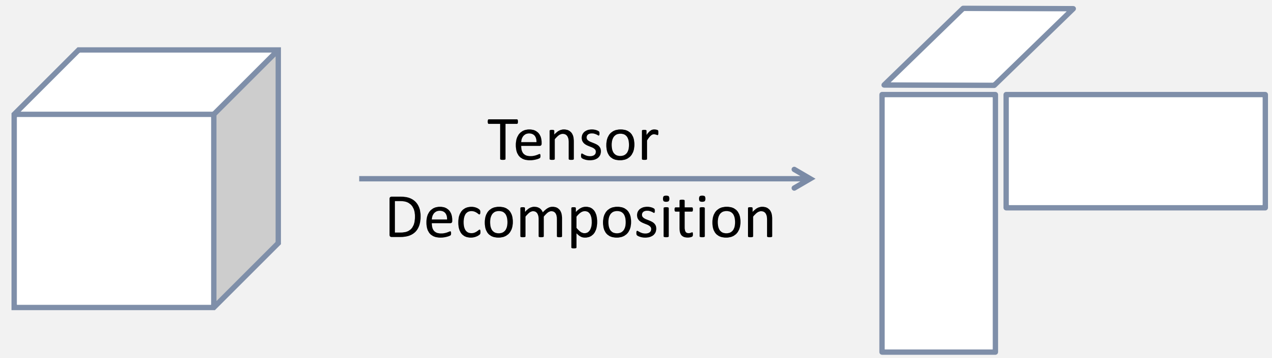
Tensor

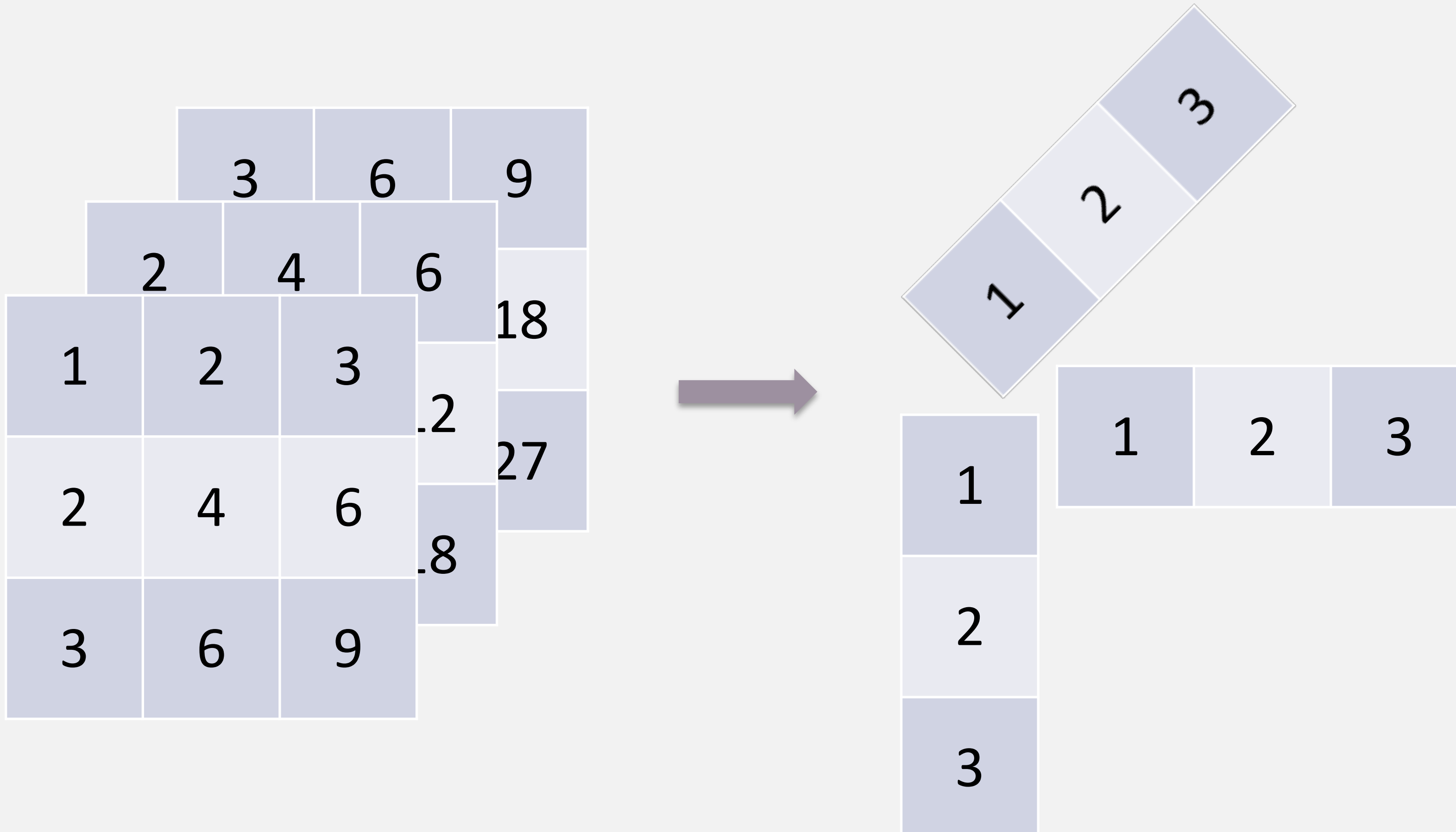
Vector Matrix Tensor

Tensors (multi-way array) are a natural representation for multi-dimensional data, e.g., videos, time-evolving networks

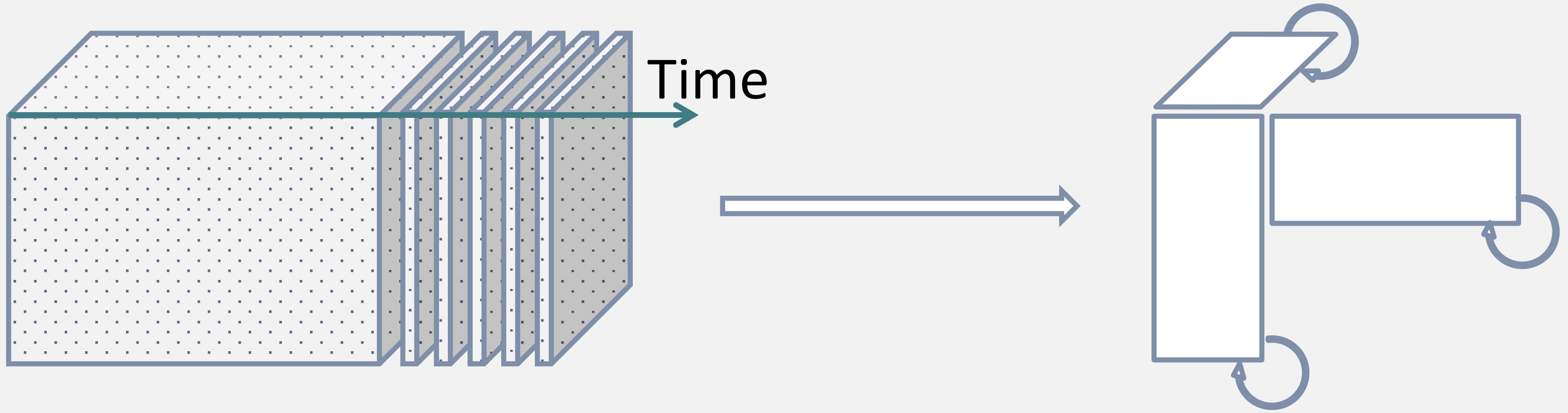

Spotify
NETFLIX
amazon

Tensor Decomposition

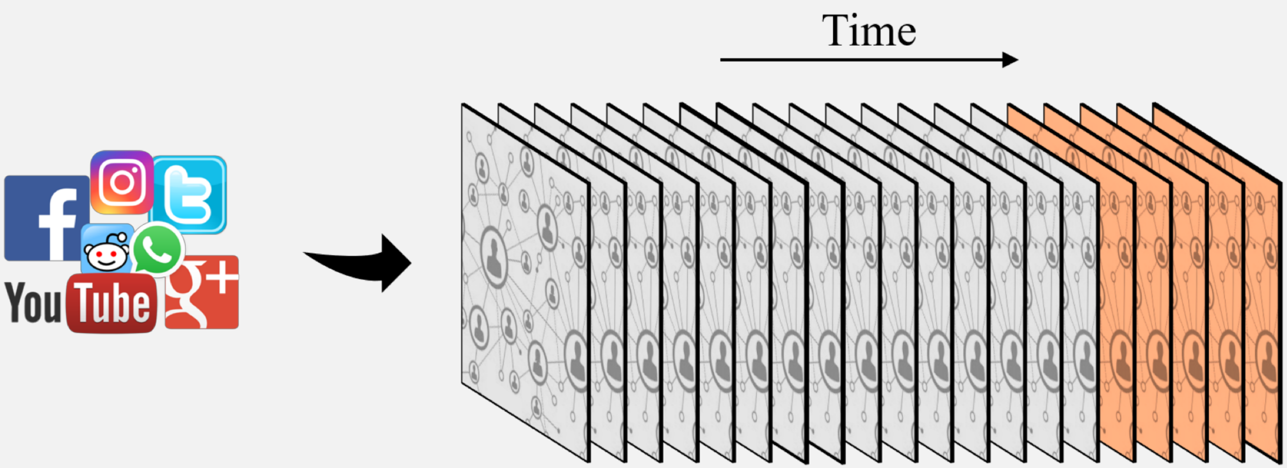



3 6 9
2 4 6
18
27
1 2 3
2 4 6
3 6 9
1 2 3
1 2 3
1 2 3

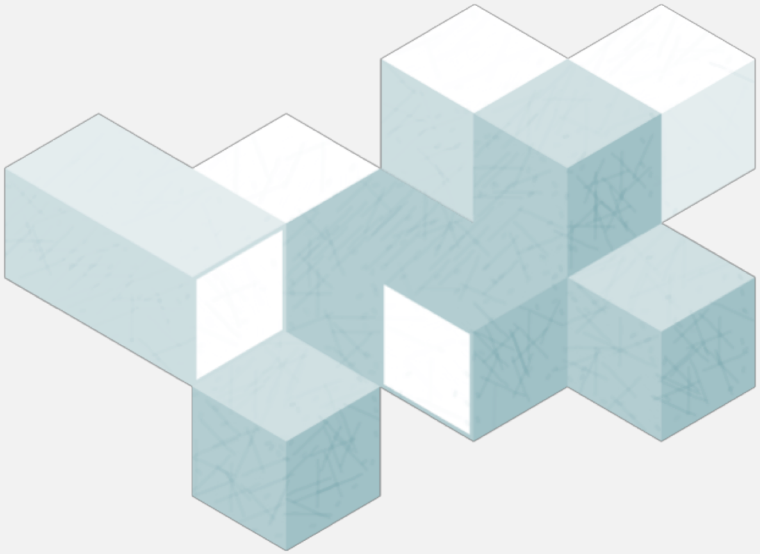
Research Problem



Online Sparse Tensors

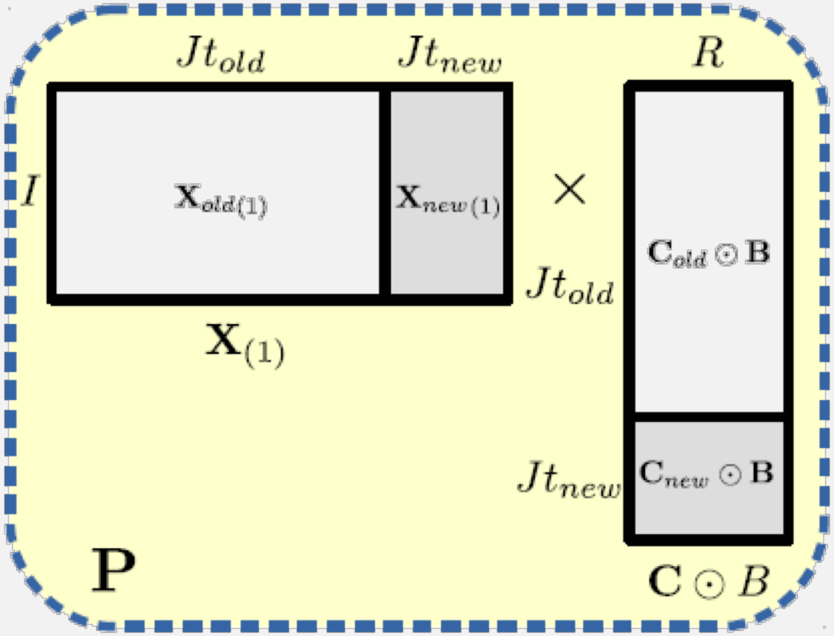


- Online Tensors
  - Snapshots appended to history tensor along the time
  - e.g., time-evolving network, video sequence, fmri
- Sparse Tensors
  - A small set of non-zeros compared to zeros
- Static methods are too expensive
- Online methods are not optimized for sparse data



Methodology

Limitations of OnlineCP on Sparse Tensors

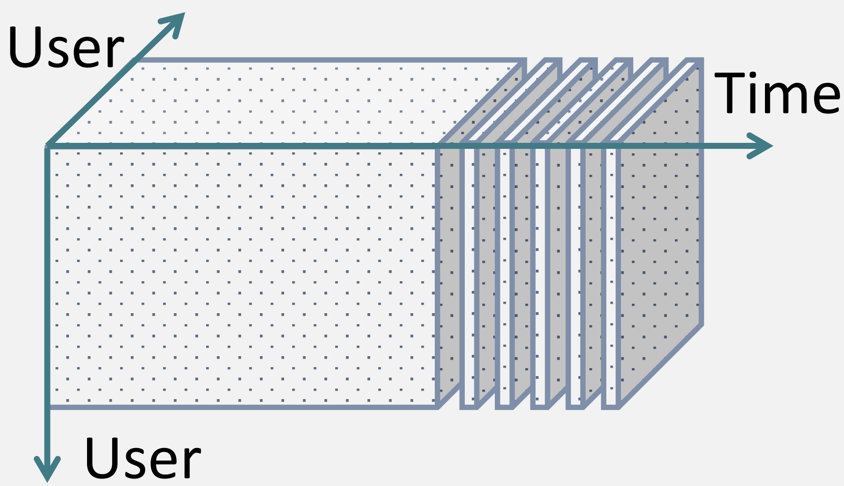


$$\mathbf{P} = \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})$$
$$\leftarrow \mathbf{P} + \mathbf{X}_{new(1)}(\mathbf{C}_{new} \odot \mathbf{B})$$

$\mathcal{O}(IJt_{new})$
$1e6 \times 1e6 \times 4 \approx 3{,}725\ GB$

SpMttkrp
[Bader & Kolda, 2007]

$\mathcal{O}(|\Delta\Omega+|)$
$384\ MB$



1 million users, density: 1×10^{-5}

OnlineSCP Algorithm

Given: history data $\tilde{\mathcal{X}}$, existing decomposition $[\![\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{C}}]\!]$, new data $\Delta\mathcal{X}$.
Find: new decomposition $[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$

OnlineCP

- initialize complementary matrices $\mathbf{P} \in \mathbb{R}^{I\times R}$, $\mathbf{Q} \in \mathbb{R}^{R\times R}$
- project $\Delta\mathcal{X}$ to $\mathbf{C}$ by fixing $\mathbf{A}$, $\mathbf{B}$
- update $\mathbf{A}$ by fixing $\mathbf{B}$, $\mathbf{C}$

$$\mathbf{P} \leftarrow \mathbf{P} + \Delta\mathbf{X}_{(1)}(\Delta\mathbf{C} \odot \mathbf{B})$$
$$\mathbf{Q} \leftarrow \mathbf{Q} + (\Delta\mathbf{C} \odot \mathbf{B})^\top(\Delta\mathbf{C} \odot \mathbf{B})$$
$$\mathbf{A} \leftarrow \mathbf{P}\mathbf{Q}^{-1}$$

OnlineSCP

- initialize complementary matrix $\mathbf{Q} \in \mathbb{R}^{R\times R}$, $\mathbf{Q} = \mathbf{A}^\top\mathbf{A} \circledast \mathbf{B}^\top\mathbf{B} \circledast \mathbf{C}^\top\mathbf{C}$
- project $\Delta\mathcal{X}$ to $\mathbf{C}$ by fixing $\mathbf{A}$, $\mathbf{B}$
- update $\mathbf{A}$ by fixing $\mathbf{B}$, $\mathbf{C}$

$$\mathbf{A} \leftarrow \frac{\tilde{\mathbf{X}}_{(1)}(\tilde{\mathbf{C}} \odot \mathbf{B}) + \Delta\mathbf{X}_{(1)}(\Delta\mathbf{C} \odot \mathbf{B})}{\mathbf{Q} \oslash (\mathbf{A}^\top\mathbf{A})}$$
$$\approx \tilde{\mathbf{A}}\frac{\tilde{\mathbf{Q}} \oslash (\mathbf{A}^\top\mathbf{A})}{\mathbf{Q} \oslash (\mathbf{A}^\top\mathbf{A})} + \frac{\Delta\mathbf{X}_{(1)}(\Delta\mathbf{C} \odot \mathbf{B})}{\mathbf{Q} \oslash (\mathbf{A}^\top\mathbf{A})}$$

$\mathcal{O}(IJt_{new})$ → $\mathcal{O}(|\Delta\Omega+|)$

Experiments

Datasets	Description	Size	nnz*	Density	Batch Size
Facebook-Links	user × user × day	64K × 64K × 886	671K	2×10^{-7}	4
Facebook-Wall	wall owner × poster × day	47K × 47K × 2K	738K	2×10^{-7}	8
MovieLens	user × movie × time	6K × 4K × 1K	1M	4×10^{-5}	5
LastFM	user × artist × time	1K × 1K × 168	3M	2×10^{-2}	1
NIPS	paper × author × year × word	2K × 3K × 17 × 14K	3M	2×10^{-6}	70
Youtube	user × user × day	3M × 3M × 226	18M	8×10^{-9}	1
Enron	sender × receiver × word × day	6K × 6K × 244K × 1K	54M	5×10^{-9}	6
NELL-2	entity × relation × entity	12K × 9K × 30K	77M	2×10^{-5}	144
NELL-1	entity × relation × entity	3M × 2M × 25M	144M	9×10^{-13}	127,476

* number of non-zeros

Results - Effectiveness

Table 1: Mean relative **fitness** to ALS over all batches. The higher the better (boldface means the best results)

Dataset	OnlineCP	SDT	RLST	OnlineSCP
Facebook-Links	n/a*	n/a	n/a	**1.00**
Facebook-Wall	n/a	n/a	n/a	**0.92**
MovieLens	**1.00**	0.31	**1.00**	**1.00**
LastFM	**0.91**	0.22	0.17	**0.91**
NIPS	n/a	n/a	n/a	**0.96**
Youtube	n/a	n/a	n/a	**1.00**
Enron	n/a	n/a	n/a	**0.93**
NELL-2	0.97	n/a	n/a	**0.98**
NELL-1	n/a	n/a	n/a	**0.84**

* n/a means the method is failed since it is not applicable/running out of memory/cannot finish within 12 hours

Results – Time Efficiency

Table 1: Mean relative **speedup** to ALS over all batches. The higher the better (boldface means the best results)

Dataset	OnlineCP	SDT	RLST	OnlineSCP
Facebook-Links	n/a	n/a	n/a	**3.90**
Facebook-Wall	n/a	n/a	n/a	**5.65**
MovieLens	2.03	0.05	0.02	**42.06**
LastFM	70.68	12.41	6.53	**74.83**
NIPS	n/a	n/a	n/a	**102.08**
Youtube	n/a	n/a	n/a	**3.14**
Enron	n/a	n/a	n/a	**160.24**
NELL-2	30.12	n/a	n/a	**258.50**
NELL-1	n/a	n/a	n/a	**27.81**

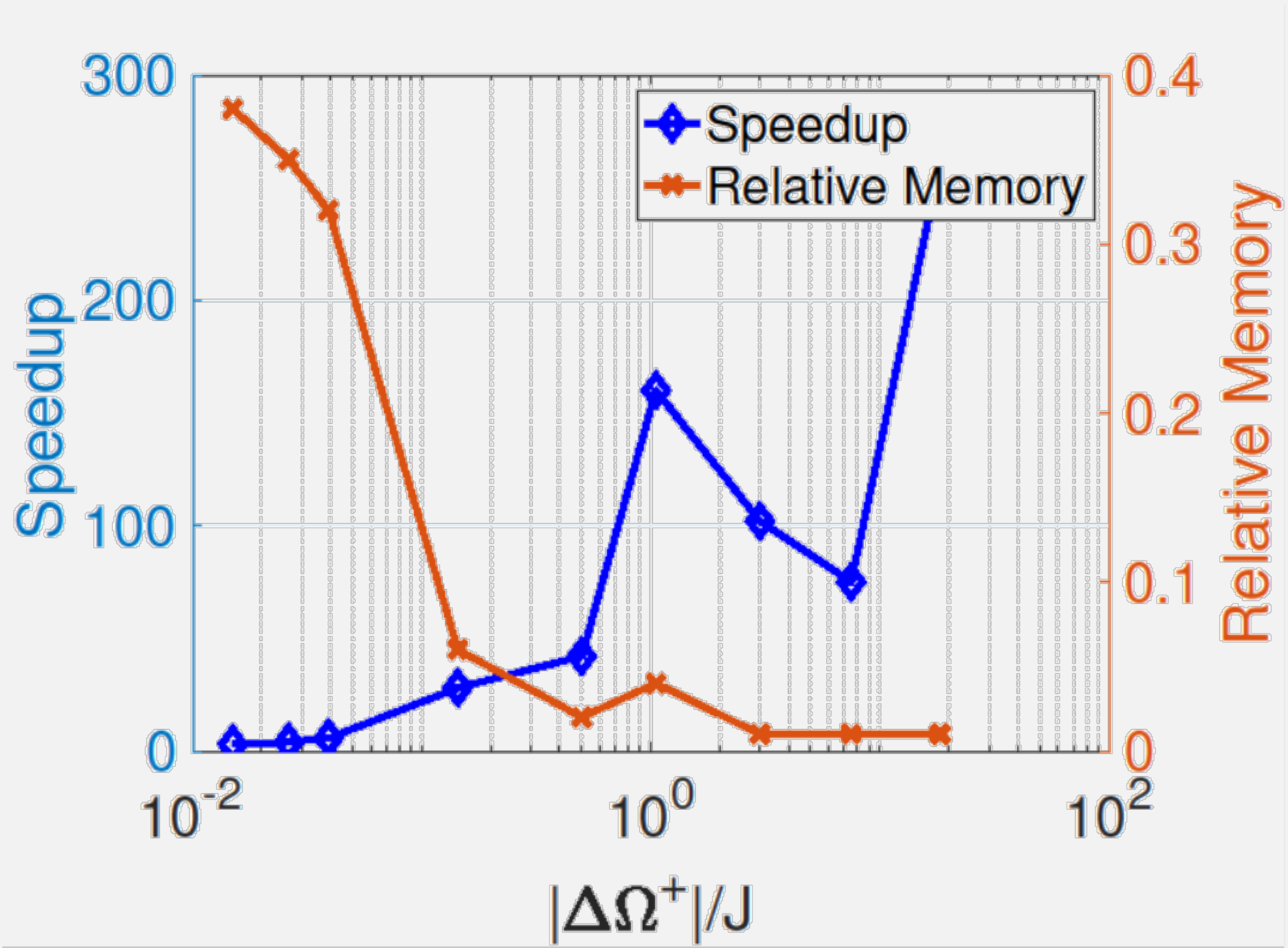
Results – Space Efficiency

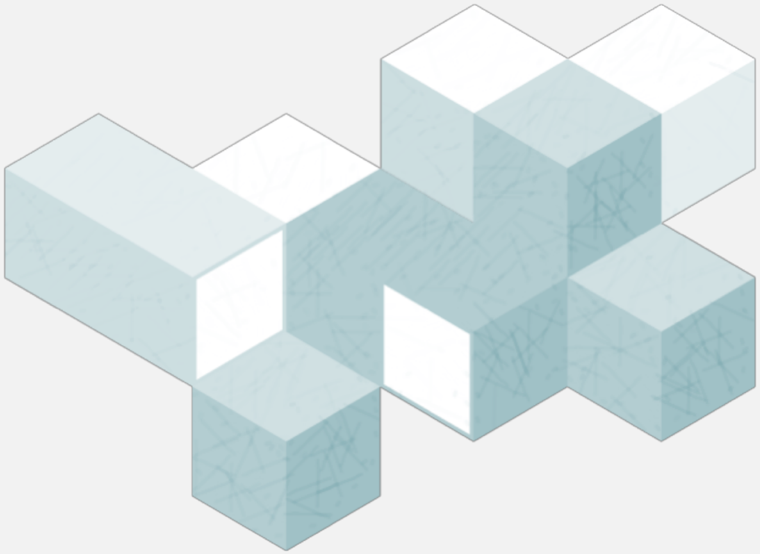
Table 1: Mean relative **memory usage** to ALS over all batches. The lower the better (boldface means the best results)

Dataset	OnlineCP	SDT	RLST	OnlineSCP
Facebook-Links	n/a	n/a	n/a	**0.35**
Facebook-Wall	n/a	n/a	n/a	**0.32**
MovieLens	17.41	60.87	46.38	**0.02**
LastFM	0.36	1.24	0.94	**0.01**
NIPS	n/a	n/a	n/a	**0.01**
Youtube	n/a	n/a	n/a	**0.38**
Enron	n/a	n/a	n/a	**0.04**
NELL-2	1.49	n/a	n/a	**0.01**
NELL-1	n/a	n/a	n/a	**0.06**

Variance on Efficiency

$$\mathcal{O}(JR^2 + |\Delta\Omega^+|NR)$$





Q & A