Constrained Low-rank Approximation

Koby Hayashi
School of Computational Science and Engineering, College of Computing at the Georgia Institute of Technology

Georgia Tech

1. What is Low-rank Approximation?
2. Applications of Low-rank Approximation
3. Randomized Algorithms for Low-rank Approximation

Georgia Tech

Low-rank Matrix Approximation

$$X \approx W \times H^T$$

$$X \in \mathbb{R}^{m \times n}$$
$$W \in \mathbb{R}^{m \times k}$$
$$H \in \mathbb{R}^{n \times k}$$
$$k \ll \min(m, n)$$

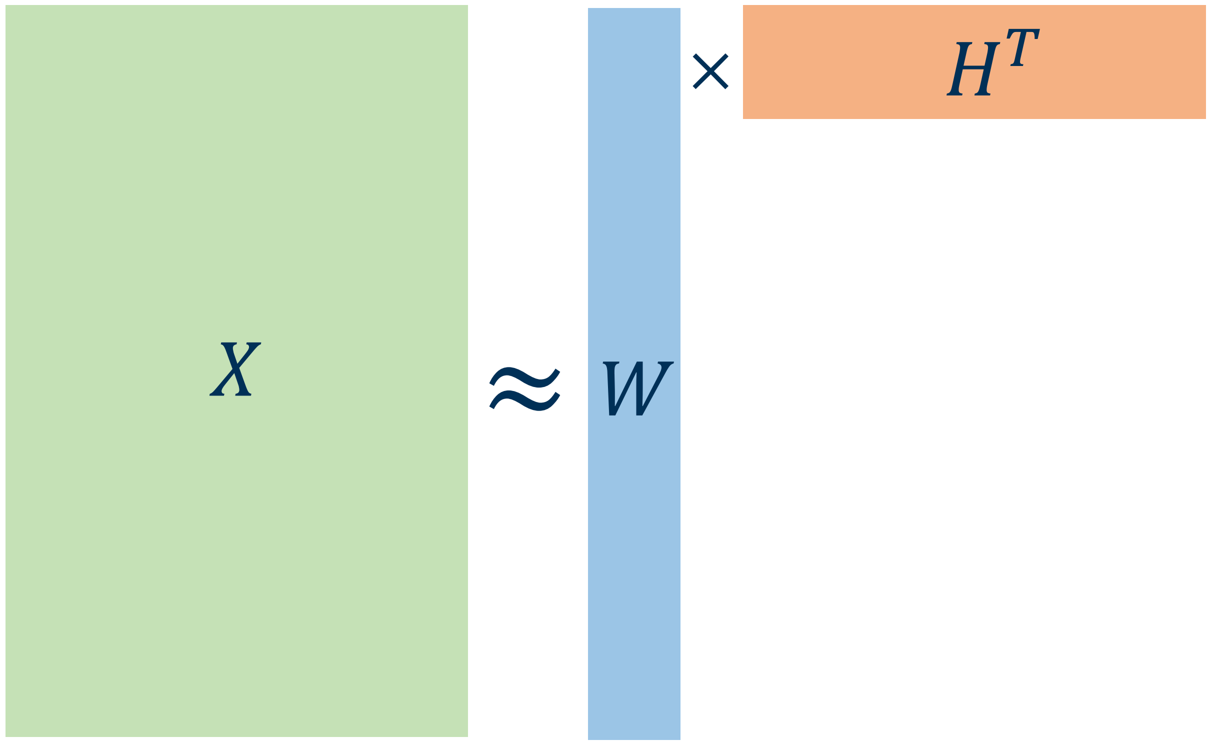
Singular Value Decomposition gives optimal solution

$$\min_{\{W,H\}} ||X - WH^T||_F$$

Georgia Tech

Nonnegative Matrix Factorization (NMF)

$$\min_{\{W,H\}\geq 0} ||X - WH^T||_F$$



Why?

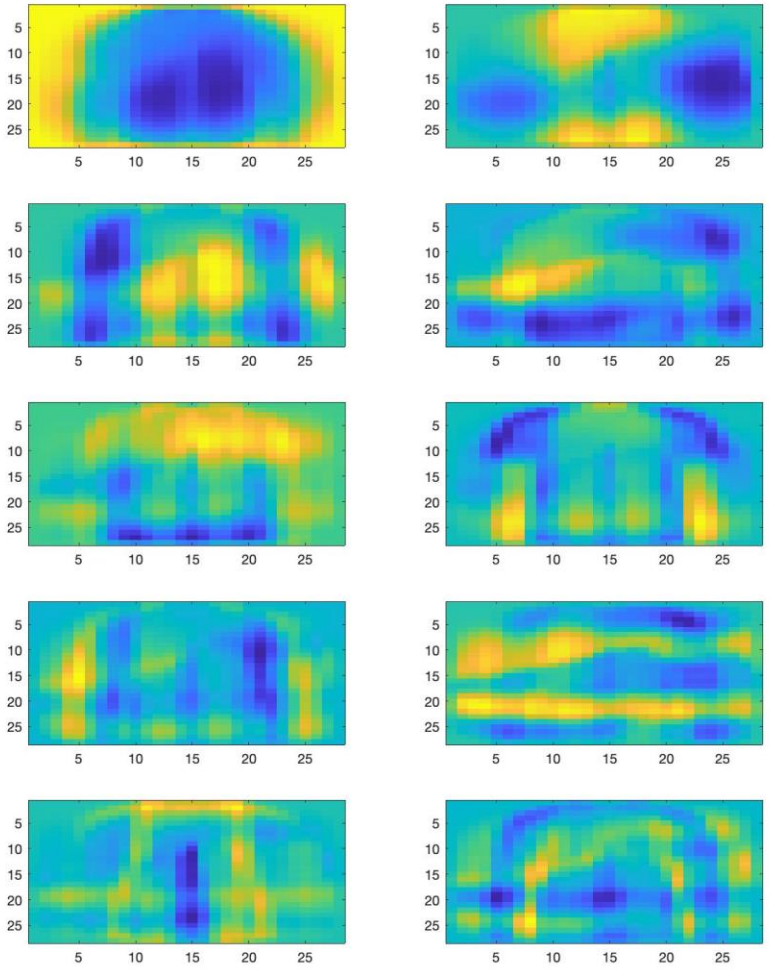
- Images
- Count Data
- (Hyper) Graphs
- Probabilities

NMF Example : Fashion MNIST Data Set

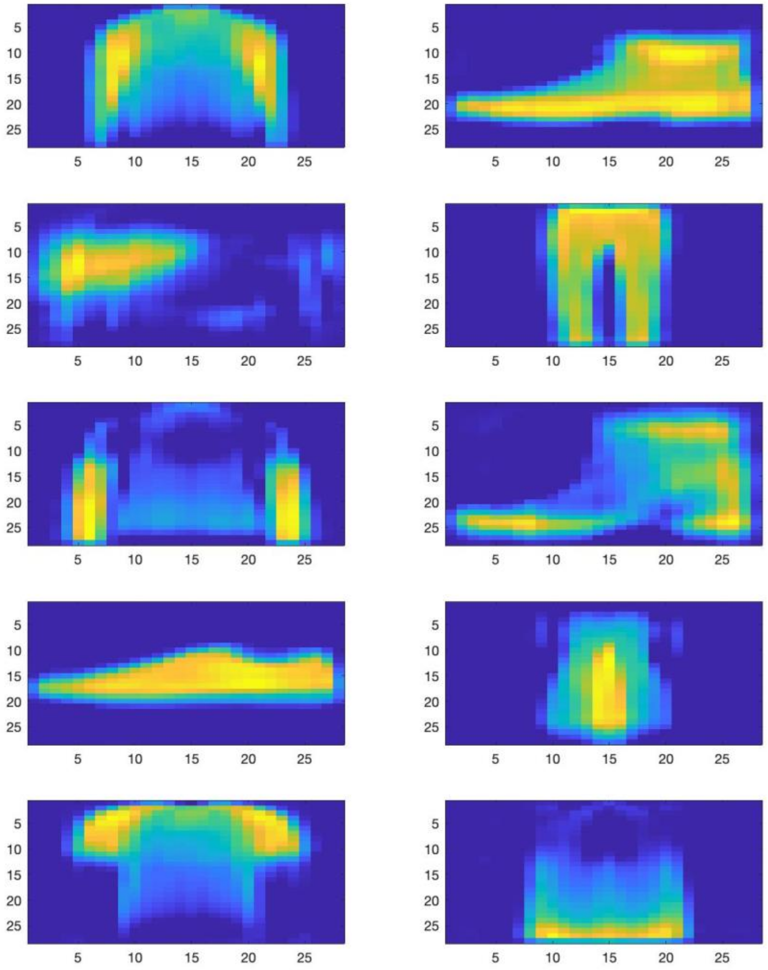
Clothing Categories :
- T-shirt/top
- Trouser
- Pullover
- Dress
- Bag

Ect ...

Interpretability via parts-based representation



SVD



NMF

Netflix Prize

In 2006 Netflix offered $1 million dollars to improve their recommendation algorithm by %10.

In 2009 a team improved the algorithm by %10.6 and won.

Netflix did not use their algorithm but adopted one that improved their method by only %8.43…

Why?

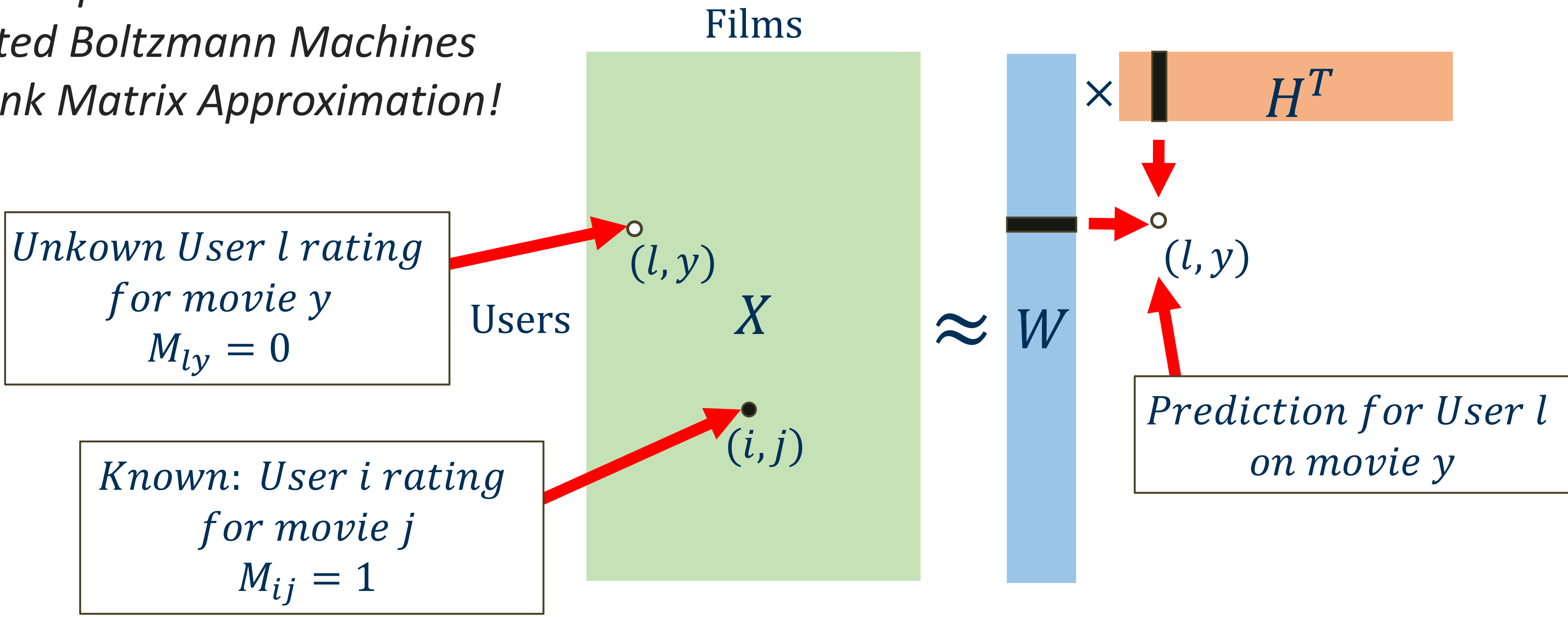
https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429

Georgia Tech

The Adopted Method

2 Main Techniques :

1. *Restricted Boltzmann Machines*
2. *Low-rank Matrix Approximation!*



$$\min_{\{W,H\}\geq 0} ||M * (X - WH^T)||_F$$

Bell, Robert M. et al. “The BellKor solution to the Netflix Prize.” (2007).

Again… Why?

In 2006 Netflix offered $1 million dollars to improve their recommendation algorithm by %10.

In 2009 a team improved the algorithm by %10.6 and won.

Netflix did not use their algorithm but adopted one that improved their method by only %8.43…

Why?

1. Update-ability
2. Scalability

- Training data had 100 million ratings
- Real data had more than 5 billion

https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429

Georgia Tech

How to Scale

Phillip B. Gibbons @CMU

1. Scale up : increase resources on a single node
2. Scale out : use multiple nodes
3. Scale down : reduce the amount of data or resources needed

https://github.com/ramkikannan/planc

S. Eswar, K. Hayashi, G. Ballard, R. Kannan, M. A. Matheson, and H. Park, "Planc: Parallel low-rank approximation with nonnegativity constraints," ACM Trans. Math. Softw.

S. Eswar, K. Hayashi, G. Ballard, R. Kannan, R. Vuduc, and H. Park, "Distributed-memory parallel symmetric nonnegative matrix factorization," in SC20: International Conference for High Performance Computing, Networking, Storage and Analysis, 2020

S. Eswar, B. Cobb, K. Hayashi, R. Kannan, G. Ballard, R. Vuduc, and H. Park. 2023. "Distributed-Memory Parallel JointNMF". In Proceedings of the 37th International Conference on Supercomputing (ICS '23).

Computing a Nonnegative Matrix Factorization

$$\min_{\{W,H\}\geq 0} ||X - WH^T||_F$$

Nonlinear – optimization problem!
Its NP-HARD!

Repeat Until Converged:

1. $W^{new} \leftarrow \min_{\{W\}\geq 0} ||X - WH^T||_F$
2. $H^{new} \leftarrow \min_{\{H\}\geq 0} ||X - W^{new}H^T||_F$

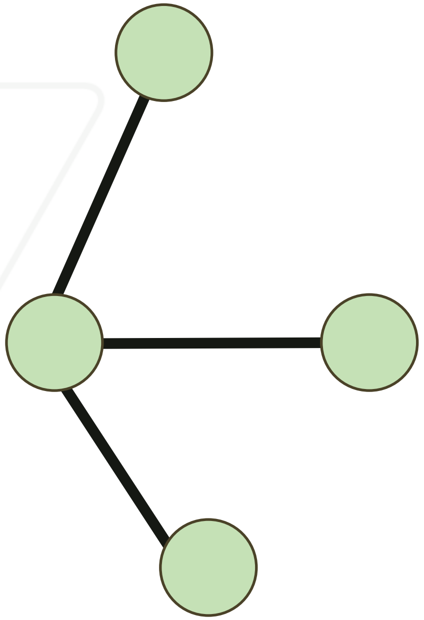
Nonnegative Least Squares Problems

Symmetric NMF

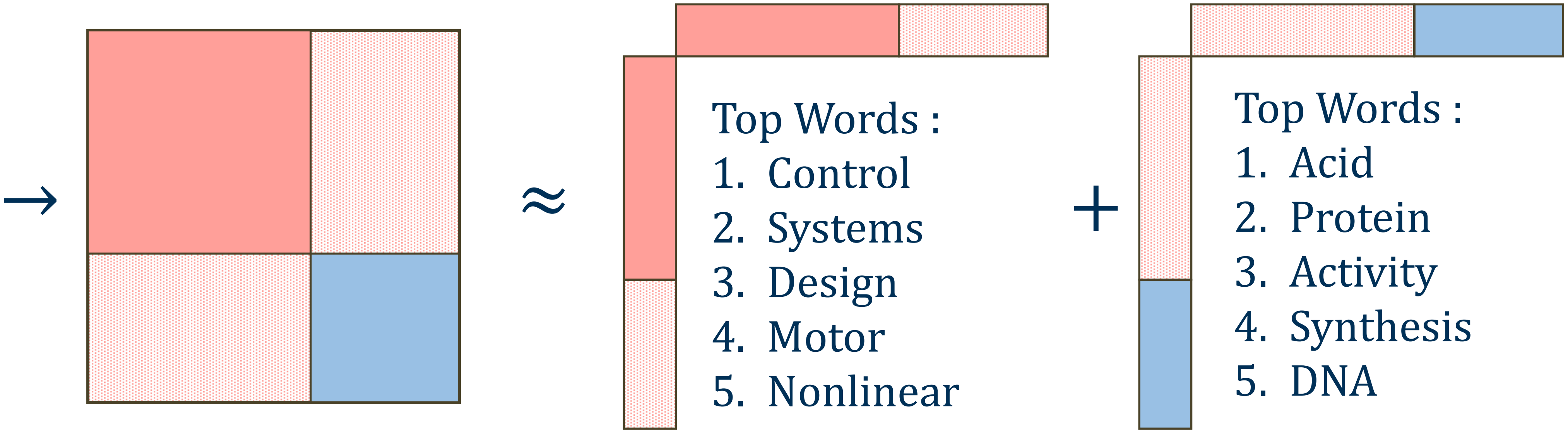
$$\min_{\{H\}\geq 0} ||A - HH^T||_F$$

$$A = A^T$$

Paper



Graph Clustering:

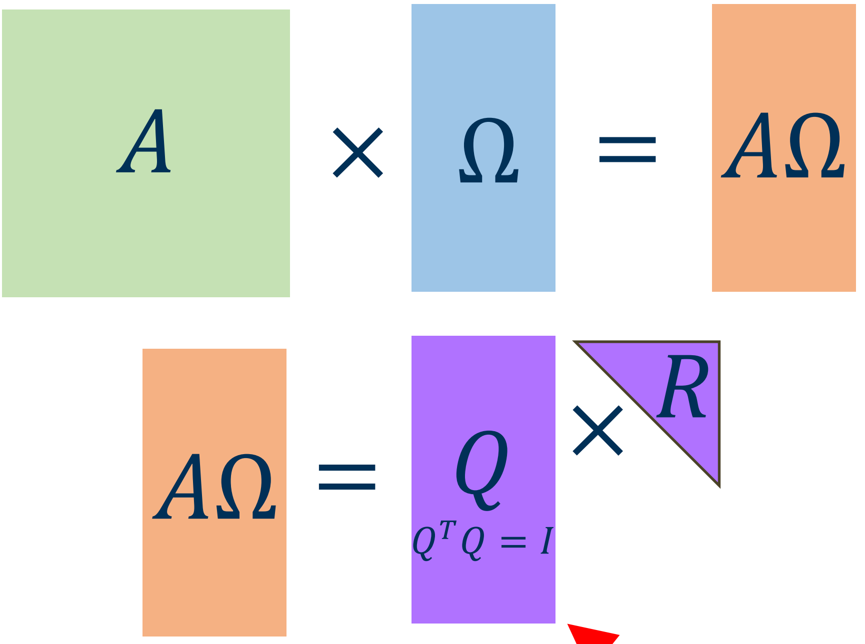


Da Kuang, Chris Ding, Haesun Park, Symmetric Nonnegative Matrix Factorization for Graph Clustering, The 12th SIAM International Conference on Data Mining (SDM '12), pp. 106--117, 2012.

Georgia Tech

Random Compression for SymNMF

1. $\Omega_{ij} = \mathcal{N}(0,1)$, *a random matrix*
2. $QR = A\Omega$, compute orthonormal basis
3. $\mathrm{T} = Q^T A Q \leftarrow$ small
4. $A \approx Q T Q^T$



Q is an approximate basis for A

$$\min_{\{H\}\geq 0} ||A - HH^T||_F \rightarrow \min_{\{H\}\geq 0} ||QTQ^T - HH^T||_F$$

$$QT(Q^T v) \text{ is faster than } Av$$

N. Halko, P.-G. Martinsson, and J. A. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, (2009)

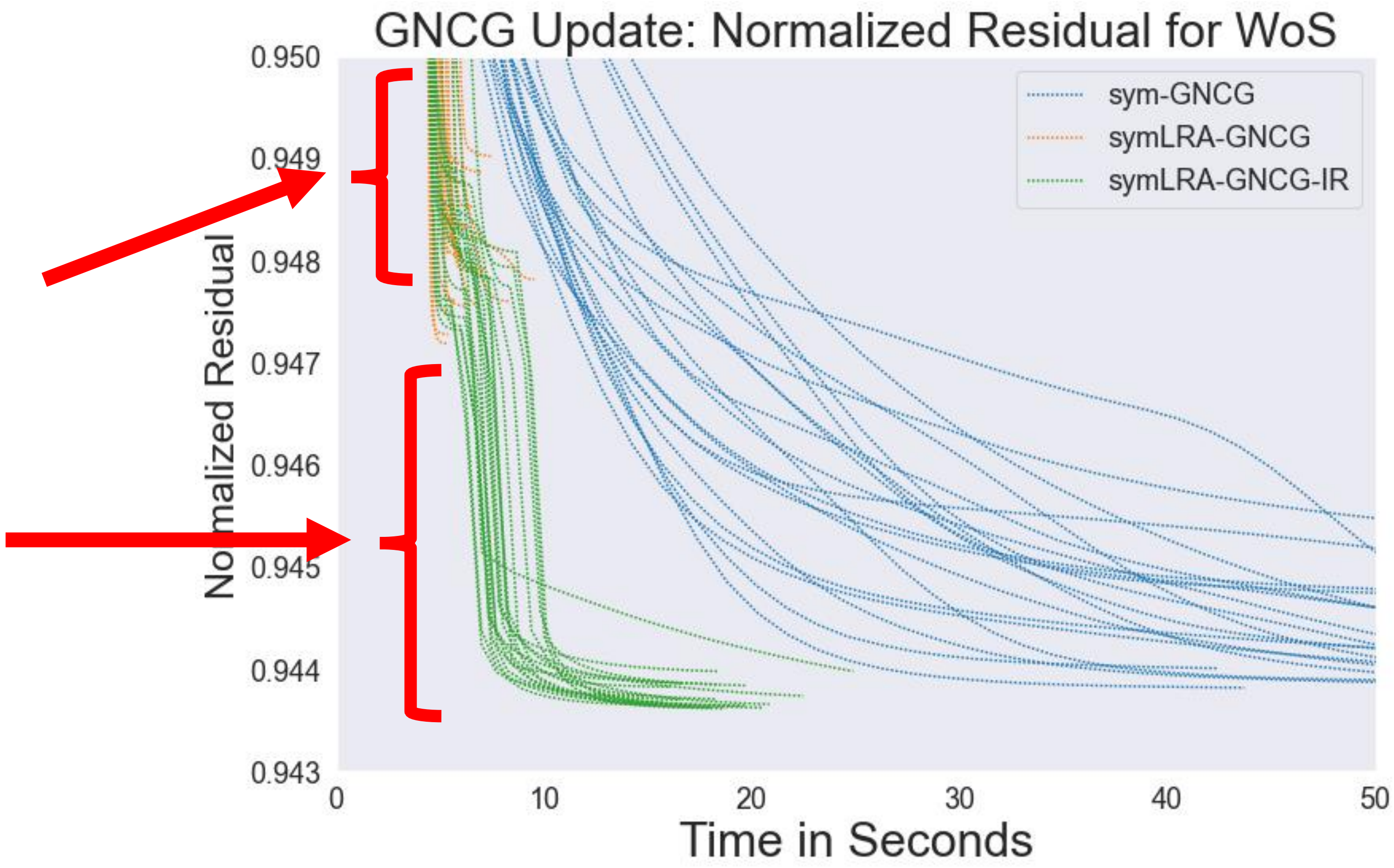
Dense Problem: Hypergraph

Start with:

$$\min_{\{H\}\geq 0} ||QTQ^T - HH^T||_F$$

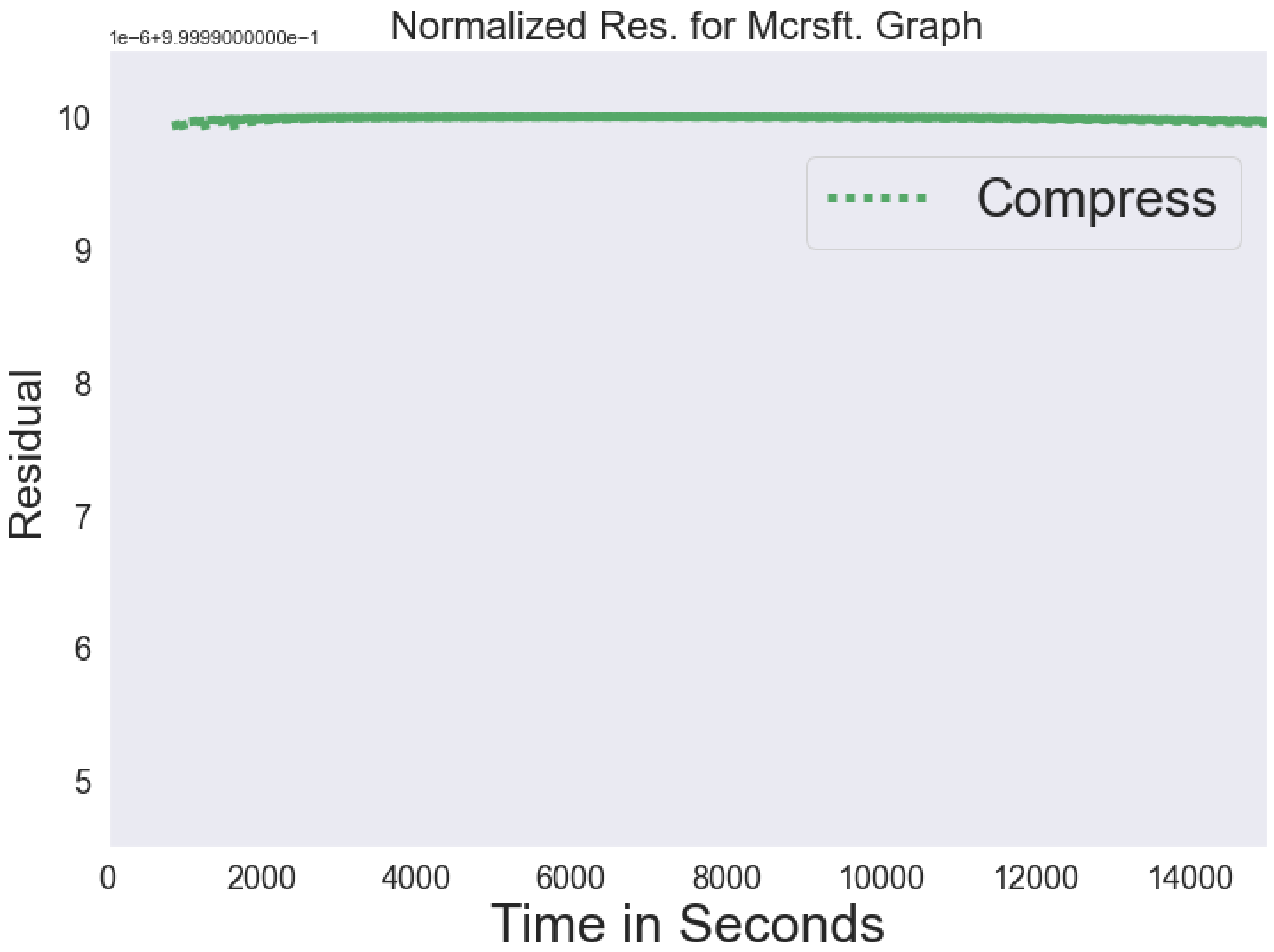
Switch to:

$$\min_{\{H\}\geq 0} ||X - HH^T||_F$$



- 4x speed up
- Preserve solution quality

Sparse Problem : Microsoft Academic Graph



Microsoft Academic Graph

1. ~37 million vertices
2. ~ 1 billion edges

Row Sampling Least Squares

$$\min_x \|Ax - b\|_2^2$$

A is $m \times k$ and full rank
x is length k
b is length m

$\rightarrow$

Obtain a distribution

$$p_1, \ldots, p_i \ldots, p_m$$

For the rows of A and b

$\rightarrow$

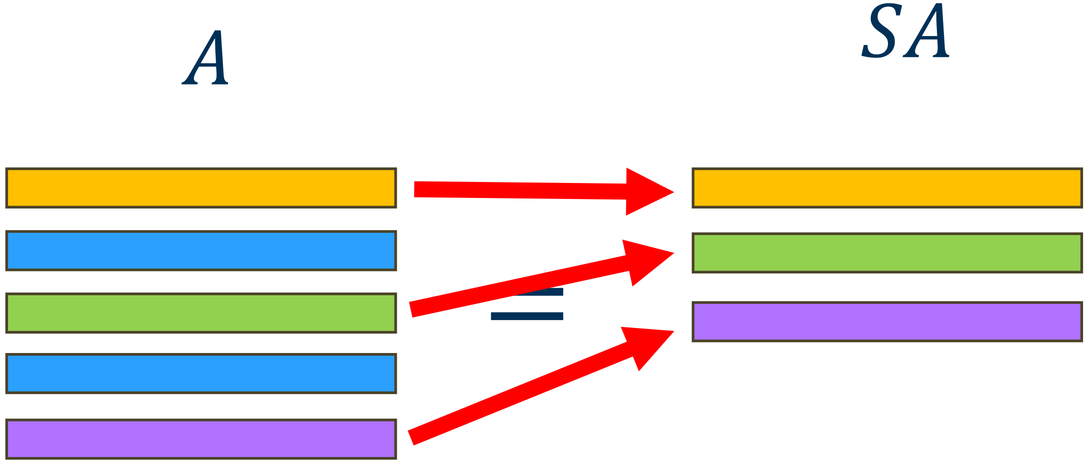
$$\min_x \|SAx - Sb\|_2^2$$

S is $s \times m, s < m$

S samples row of A and b

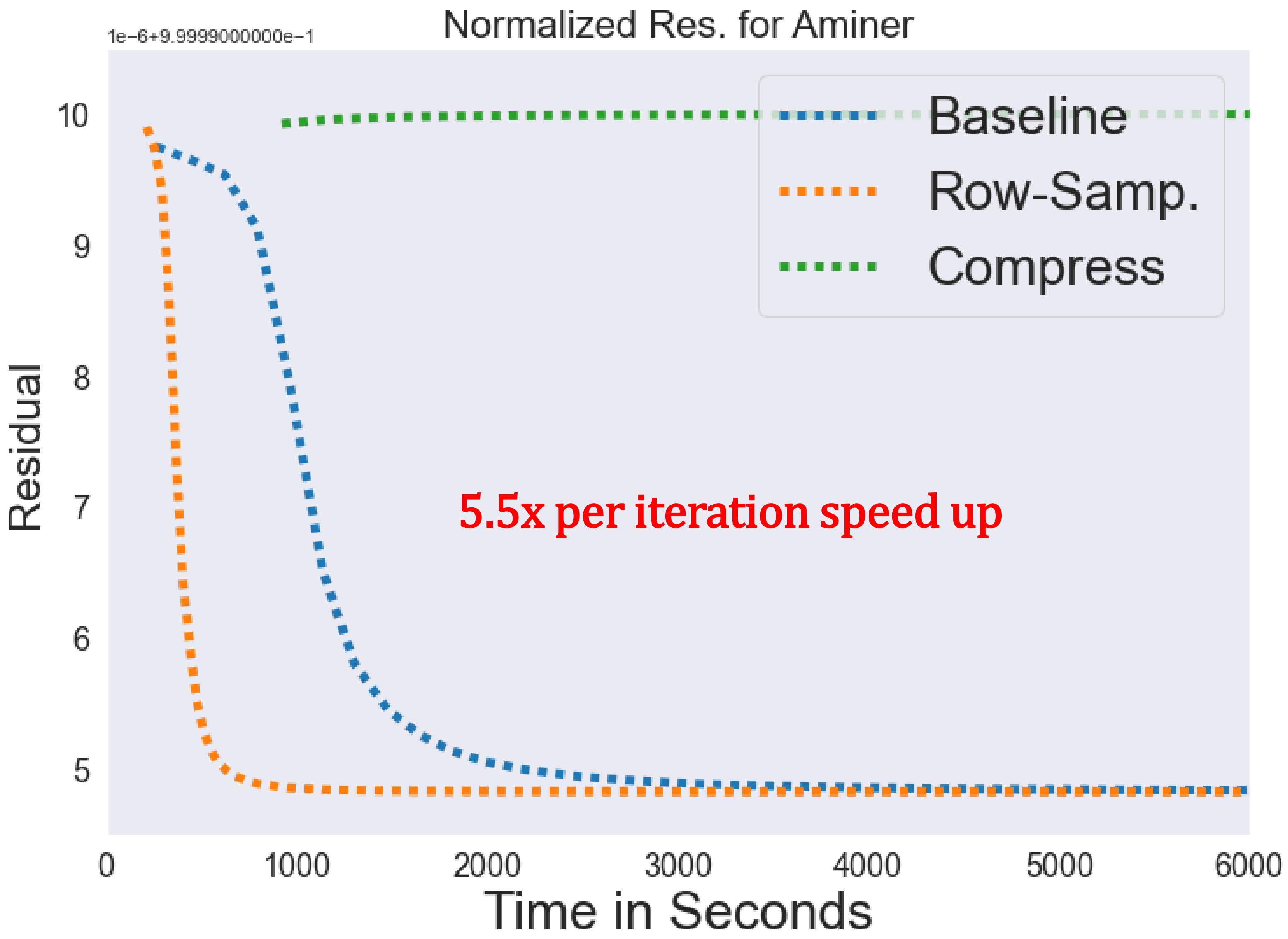
$$w_i \propto p_i^{-1}$$

$$S = \begin{pmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \times A = SA$$



M. W. Mahoney, Randomized algorithms for matrices and data, CoRR

Sparse Problem : Microsoft Academic Graph

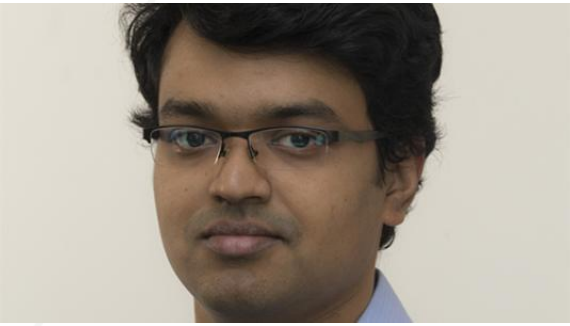


Microsoft Academic Graph

1. ~37 million vertices
2. ~ 1 billion edges

Different types of randomization work for different problem!

End



Srinivas Eswar

Ramakrishnan Kannan

Haesun Park

Richard Vuduc

Grey Ballard

Ben Cobb

DOE CSGF

Georgia Tech