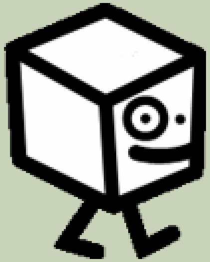


Multilinear Algebra for Analyzing Data with Multiple Linkages

Tamara G. Kolda

In collaboration with:
Brett Bader, Danny Dunlavy, Philip Kegelmeyer

Sandia National Labs

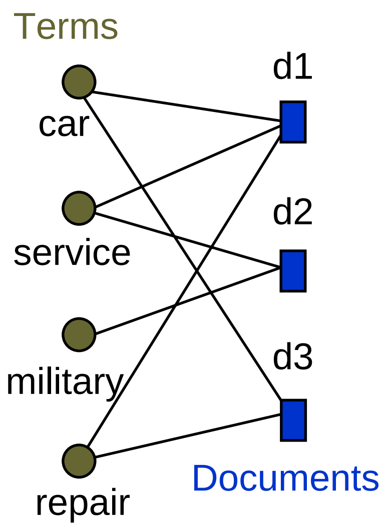
MMDS, Stanford, CA, June 21-24, 2006

Sandia National Laboratories

Linear Algebra plays an important role in Graph Analysis

$$\mathbf{A} \approx \mathbf{T}\boldsymbol{\Sigma}\mathbf{D}^{\mathsf{T}} = \sum_r \sigma_r \, \mathbf{t}_{\bullet r} \circ \mathbf{d}_{\bullet r}$$

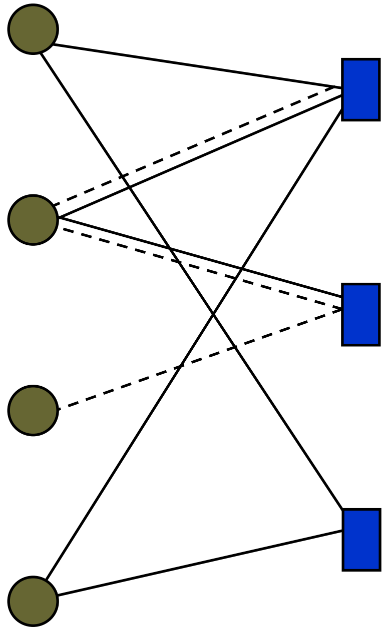
- PageRank
  - Brin & Page (1998)
  - Page, Brin, Motwani, Winograd (1999)
- HITS (hubs and authorities)
  - Kleinberg (1998/99)
- Latent Semantic Indexing (LSI)
  - Dumais, Furnas, Landauer, Deerwester, and Harshman (1988)
  - Deerwester, Dumais, Landauer, Furnas, and Harshman (1990)



One Use of LSI: Maps terms and documents to the "same" k-dimensional space.

Sandia National Laboratories

Multi-Linear Algebra can be used in more complex graph analyses

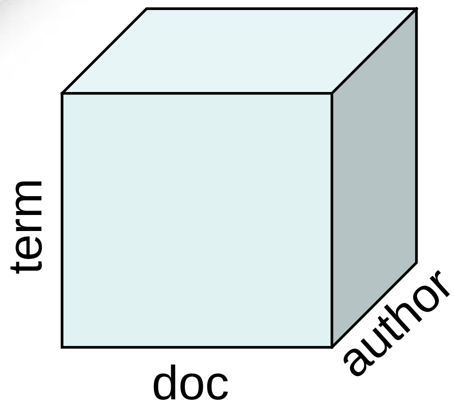


- Nodes (one type) connected by multiple types of links
  - Node x Node x Connection
- Two types of nodes connected by multiple types of links
  - Node A x Node B x Connection
- Multiple types of nodes connected by a single link
  - Node A x Node B x Node C
- Multiple types of nodes connected by multiple types of links
  - Node A x Node B x Node C x Connection
- Etc…

Sandia National Laboratories

Analyzing Publication Data: Term x Doc x Author

1999-2004
SIAM Journal Data
(except SIREV)



6928 terms
4411 documents
6099 authors
464645 nonzeros

$$\mathbf{A} = \text{term-document matrix}$$

$$a_{ij} = \frac{(1 + \log_2 f_{ij}) \log_2(N/n_i)}{d_j}$$

$$\mathbf{B} = \text{author-document matrix}$$

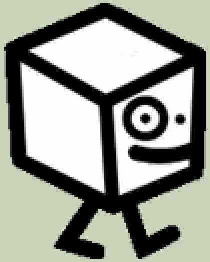
$$b_{kj} = \begin{cases} 1/\sqrt{m_j} & \text{if author k wrote document j} \\ o & \text{otherwise} \end{cases}$$

Terms must appear in at least 3 documents and no more than 10% of all documents. Moreover, it must have at least 2 characters and no more than 30.

Form tensor $\mathcal{X}$ as: $$x_{ijk} = a_{ij} b_{jk}$$

Element (i,j,k) is nonzero only if author k wrote document j using term i.

$$\mathcal{X} \approx \sum_r \lambda_r \, \mathbf{t}_{\bullet r} \circ \mathbf{d}_{\bullet r} \circ \mathbf{a}_{\bullet r}$$

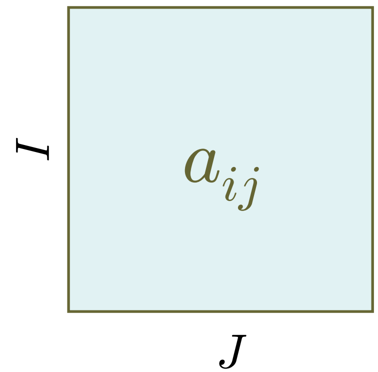


A tensor is a multidimensional array

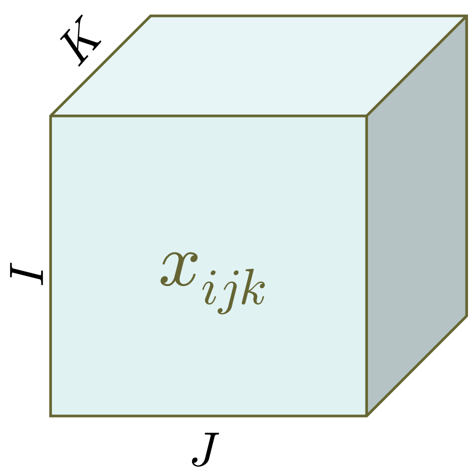
NOTATION

s	scalar
$\mathbf{a}$	vector
$\mathbf{B}$	matrix
$\mathcal{X}$	tensor

An I x J matrix



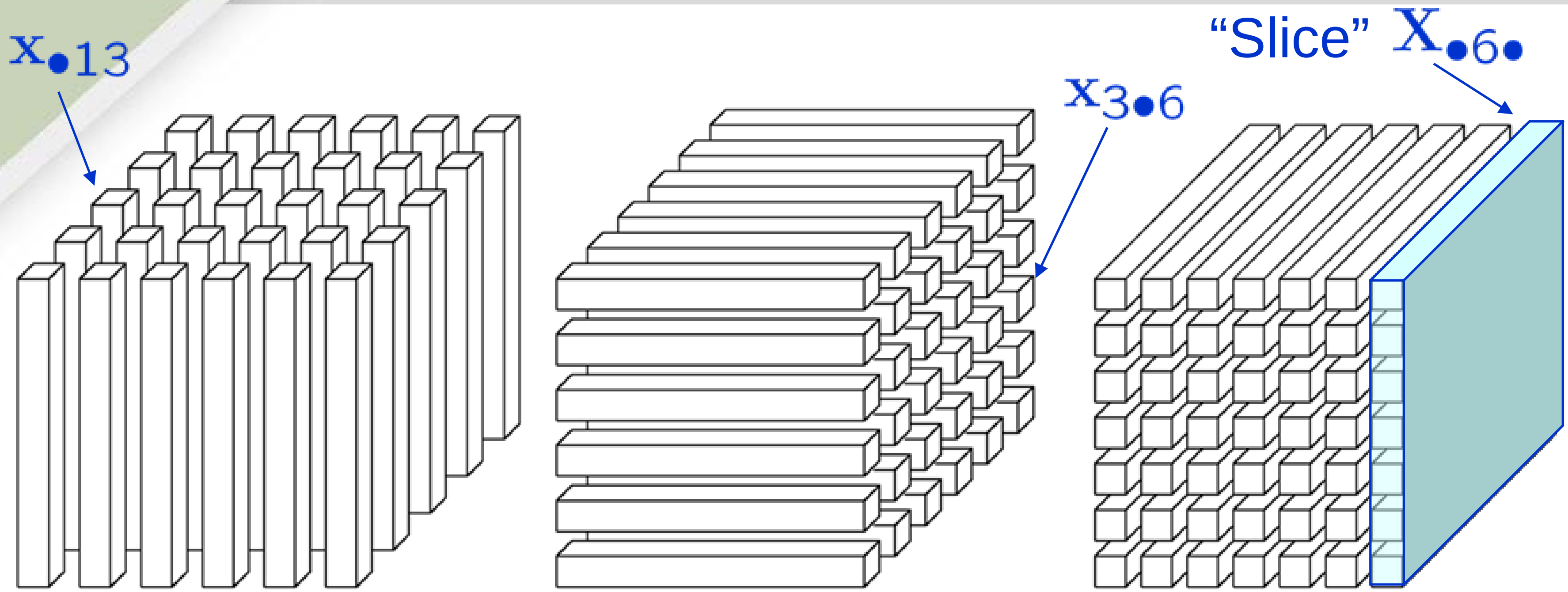
An $I \times J \times K$ tensor



- Other names for tensors…
  - Multi-way array
  - N-way array
- The "order" of a tensor is the number of dimensions
- Other names for dimension…
  - Mode
  - Way
- Example
  - The matrix **A** (at left) has order 2.
  - The tensor $\mathcal{X}$ (at left) has order 3 and its 3rd mode is of size K.

Sandia National Laboratories

Tensor “fibers” generalize the concept of rows and columns



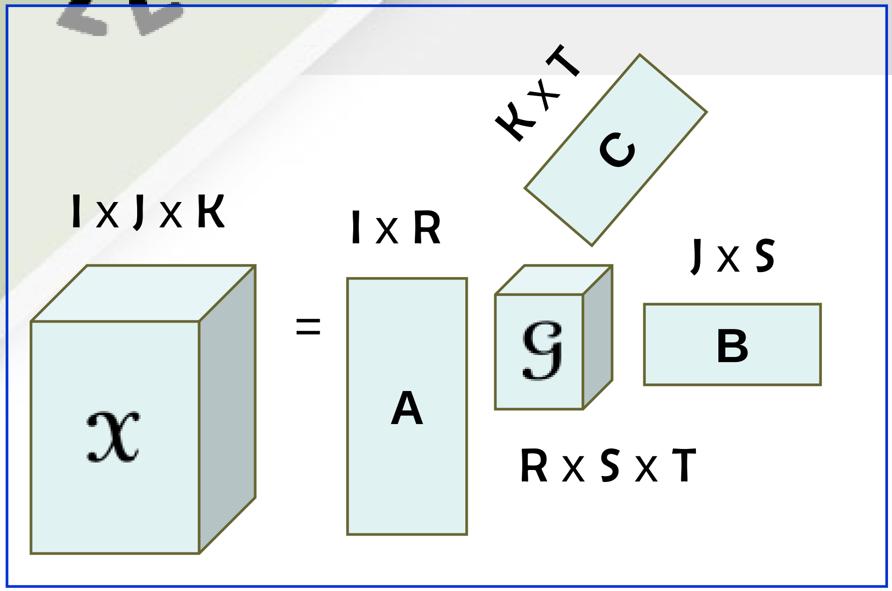
Column Fibers	Row Fibers	Tube Fibers
$\mathbf{x}_{\bullet jk}$	$\mathbf{x}_{i \bullet k}$	$\mathbf{x}_{ij \bullet}$

NOTE

There’s no naming scheme past 3 dimensions; instead, we just say, e.g., the 4th-mode fibers.

Sandia National Laboratories

Tucker Decomposition



$$\mathcal{X} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{rst} \, \mathbf{a}_{\bullet r} \circ \mathbf{b}_{\bullet s} \circ \mathbf{c}_{\bullet t}$$

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

- Proposed by Tucker (1966)
- Also known as: Three-mode factor analysis, three-mode PCA, orthogonal array decomposition
- **A**, **B**, and **C** may be orthonormal (generally assume they have full column rank)
- $\mathcal{G}$ is <u>not</u> diagonal
- Not unique

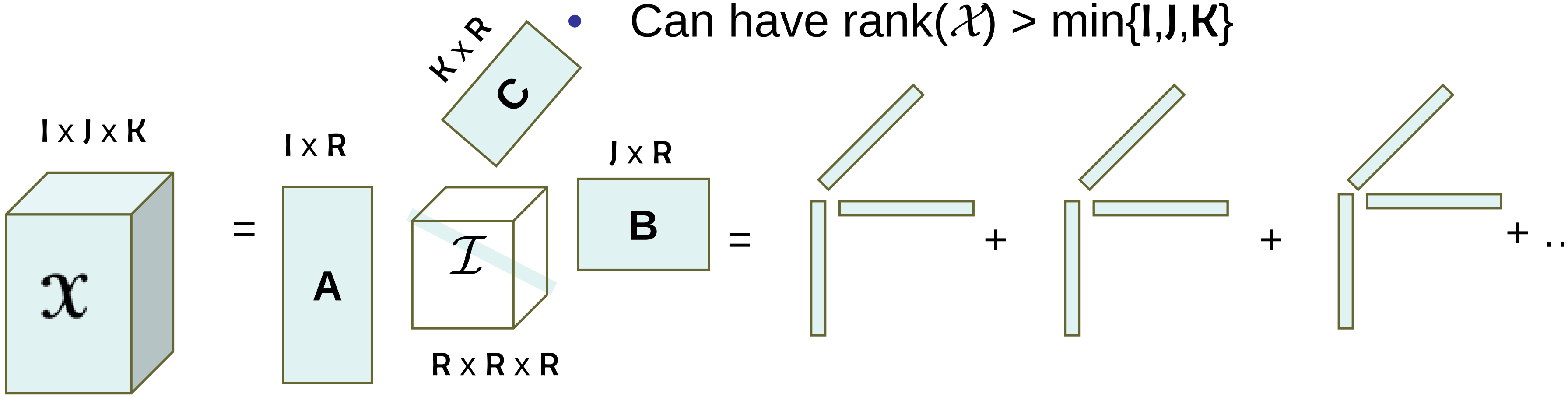
$$\mathcal{G} = [\![\mathcal{X} \, ; \mathbf{A}^{\dagger}, \mathbf{B}^{\dagger}, \mathbf{C}^{\dagger}]\!]$$

CANDECOMP/PARAFAC

- CANDECOMP = Canonical Decomposition (Carroll and Chang, 1970)
- PARAFAC = Parallel Factors (Harshman, 1970)
- Columns of **A**, **B**, and **C** are <u>not</u> orthonormal
- If **R** is *minimal*, then **R** is called the rank of the tensor (Kruskal 1977)
- Can have rank($\mathcal{X}$) > min{**I**,**J**,**K**}

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

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



Combining Tucker and PARAFAC

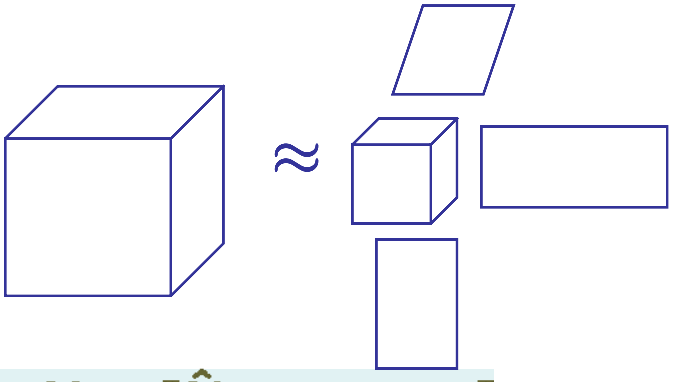
Have: Tensor $\mathfrak{X}$ of size $M \times N \times P$ Want: $\mathfrak{X} \approx \lambda[\![\mathbf{T}, \mathbf{D}, \mathbf{A}]\!]$

Step 1: Choose orthonormal compression matrices for each dimension:

$\mathbf{U}$ of size $M \times I$

$\mathbf{V}$ of size $N \times J$

$\mathbf{W}$ of size $P \times K$

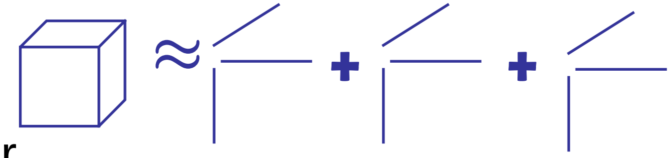


Step 2: Form reduced tensor (implicitly)

$$\hat{\mathfrak{X}} = [\![\mathfrak{X}\,; \mathbf{U}^{\mathsf{T}}, \mathbf{V}^{\mathsf{T}}, \mathbf{W}^{\mathsf{T}}]\!] \quad \Rightarrow \mathfrak{X} \approx [\![\hat{\mathfrak{X}}\,; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

Step 3: Compute PARAFAC on reduced tensor

$$\hat{\mathfrak{X}} \approx \hat{\lambda}[\![\hat{\mathbf{T}}, \hat{\mathbf{D}}, \hat{\mathbf{A}}]\!]$$



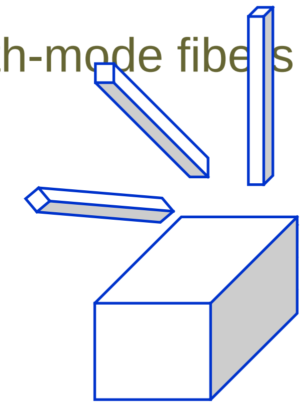
Step 4: Convert to PARAFAC of full tensor

$$\mathfrak{X} \approx \hat{\lambda}\,[\![\mathbf{U}\hat{\mathbf{T}}, \mathbf{V}\hat{\mathbf{D}}, \mathbf{W}\hat{\mathbf{A}}]\!] \equiv \lambda\,[\![\mathbf{T}, \mathbf{D}, \mathbf{A}]\!]$$

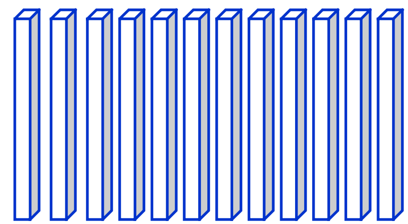
Sandia National Laboratories

Matricize: $\mathbf{X}_{(n)}$

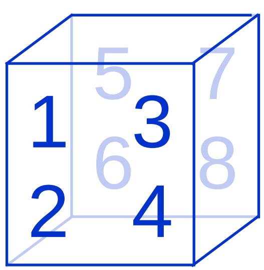
The nth-mode fibers are rearranged to be the columns of a matrix



$\mathcal{X}$



$\mathbf{X}_{(3)}$

$\mathcal{X} =$ 

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Sandia National Laboratories

Tucker and PARAFAC Matrix Representations

Fact 1:

$$(\llbracket \mathcal{G} \,; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket)_{(1)} = \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}}$$

Fact 2:

$$(\llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket)_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$

Khatri-Rao Matrix Product (Columnwise Kronecker Product):

$$\mathbf{C} \odot \mathbf{B} = \begin{bmatrix} \mathbf{c}_{\bullet 1} \otimes \mathbf{b}_{\bullet 1} & \mathbf{c}_{\bullet 2} \otimes \mathbf{b}_{\bullet 2} & \cdots & \mathbf{c}_{\bullet R} \otimes \mathbf{b}_{\bullet R} \end{bmatrix}$$

Special pseudu-inverse structure:

$$((\mathbf{C} \odot \mathbf{B})^{\mathsf{T}})^{\dagger} = (\mathbf{C} \odot \mathbf{B})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{-1}$$

Sandia National Laboratories

Implicit Compressed PARAFAC ALS

Have: $\hat{\mathcal{X}} = [\![\mathcal{X} \,; \mathbf{U}^{\mathsf{T}}, \mathbf{V}^{\mathsf{T}}, \mathbf{W}^{\mathsf{T}}]\!]$ Want: $\hat{\mathcal{X}} \approx [\![\hat{\mathbf{T}}, \hat{\mathbf{D}}, \hat{\mathbf{A}}]\!]$

Consider the problem of fixing the 2nd and 3rd factors and solving just for the 1st.

$$\min_{\hat{\mathbf{T}}} \| \hat{\mathcal{X}} - [\![\hat{\mathbf{T}}, \hat{\mathbf{D}}, \hat{\mathbf{A}}]\!] \| \qquad \min_{\hat{\mathbf{T}}} \| \hat{\mathbf{X}}_{(1)} - \hat{\mathbf{T}} (\hat{\mathbf{A}} \odot \hat{\mathbf{D}})^{\mathsf{T}} \|$$

$$\hat{\mathbf{T}} = \hat{\mathbf{X}}_{(1)} ((\hat{\mathbf{A}} \odot \hat{\mathbf{D}})^{\mathsf{T}})^{\dagger}$$

$$\hat{\mathbf{T}} = \hat{\mathbf{X}}_{(1)} (\hat{\mathbf{A}} \odot \hat{\mathbf{D}}) \mathbf{Z}^{-1}$$ with $$\mathbf{Z} = \hat{\mathbf{A}}^{\mathsf{T}} \hat{\mathbf{A}} * \hat{\mathbf{D}}^{\mathsf{T}} \hat{\mathbf{D}}$$

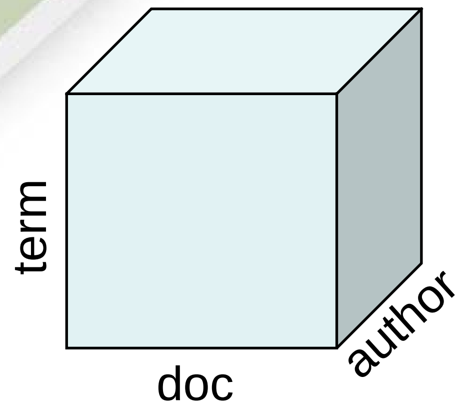
$$\hat{\mathbf{T}} = \boxed{\mathbf{U}^{\mathsf{T}} \mathbf{X}_{(1)} (\mathbf{W} \otimes \mathbf{V})} (\hat{\mathbf{A}} \odot \hat{\mathbf{D}}) \mathbf{Z}^{-1}$$

$$\hat{\mathbf{T}} = \mathbf{U}^{\mathsf{T}} \mathbf{X}_{(1)} (\mathbf{W}\hat{\mathbf{A}} \odot \mathbf{V}\hat{\mathbf{D}}) \mathbf{Z}^{-1}$$

$$\left(\hat{\mathbf{T}}\mathbf{Z}\right)_{\bullet r} = \mathbf{U}^{\mathsf{T}} \mathbf{X}_{(1)} \left[\left(\mathbf{W}\hat{\mathbf{A}}\right)_{\bullet r} \otimes \left(\mathbf{V}\hat{\mathbf{D}}\right)_{\bullet r} \right]$$

Update columnwise

Back to the Problem: Term x Doc x Author



6928 documents
4411 terms
6099 authors
464645 nonzeros

$$\mathbf{A} = \text{term-document matrix}$$

$$a_{ij} = \frac{(1 + \log_2 f_{ij}) \log_2(N/n_i)}{d_j}$$

$$\mathbf{B} = \text{author-document matrix}$$

$$b_{kj} = \begin{cases} 1/\sqrt{m_j} & \text{if author k wrote document j} \\ o & \text{otherwise} \end{cases}$$

Terms must appear in at least 3 documents and no more than 10% of all documents. Moreover, it must have at least 2 characters and no more than 30.

Form tensor $\mathcal{X}$ as: $x_{ijk} = a_{ij} b_{jk}$

Element (i,j,k) is nonzero only if author k wrote document j using term i.

$$\mathcal{X} \approx \sum_r \lambda_r \, \mathbf{t}_{\bullet r} \circ \mathbf{d}_{\bullet r} \circ \mathbf{a}_{\bullet r}$$

Sandia National Laboratories

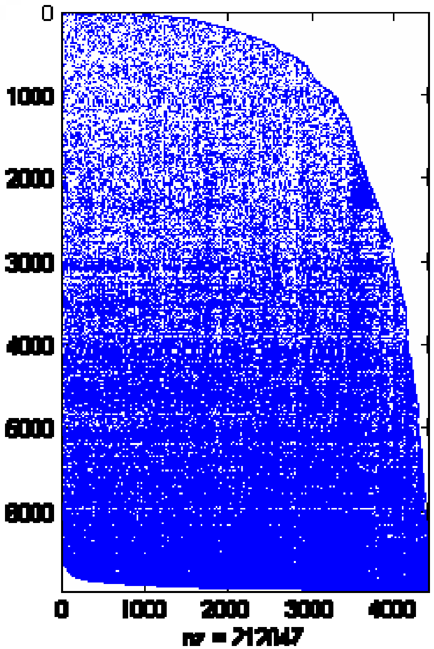
Original problem is "overly" sparse

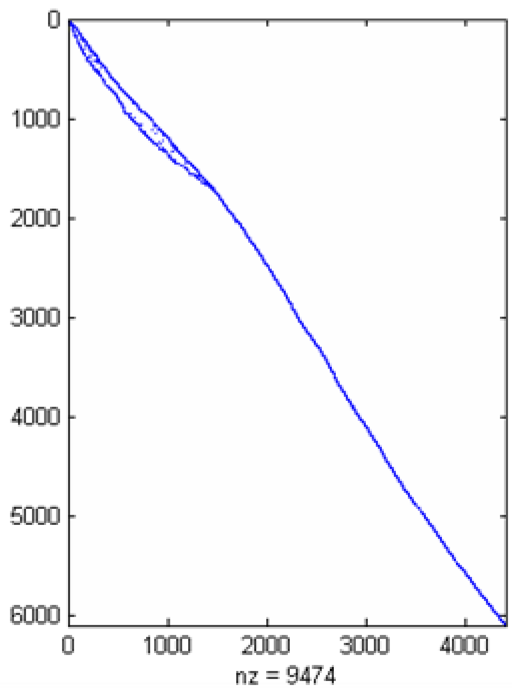
$\mathbf{A} = \text{term-document matrix}$

$$a_{ij} = \frac{(1 + \log_2 f_{ij}) \log_2(N/n_i)}{d_j}$$

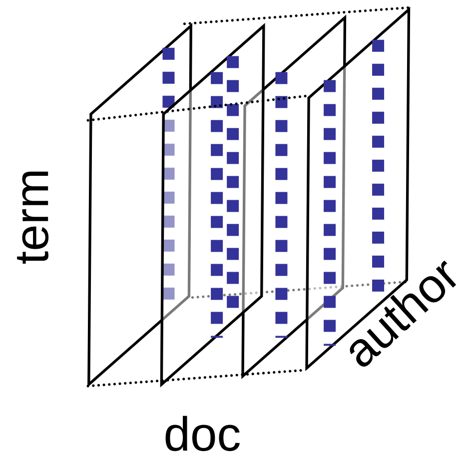
$\mathbf{B} = \text{author-document matrix}$

$$b_{ij} = \begin{cases} 1/\sqrt{m_j} & \text{if author i wrote document j} \\ o & \text{otherwise} \end{cases}$$





Result: Resulting tensor has just a few nonzero columns in each lateral slice.



Experimentally, PARAFAC seems to overfit such data and not do a good job of "mixing" different authors.

Sandia National Laboratories

Compression Matrices & PARAFAC

$$\mathcal{X} \approx [\![\hat{\mathcal{X}} \,; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

$$\mathbf{A} = \text{term-document matrix}$$

$$\mathbf{A} \approx \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{V}_A^T$$ (rank 100)

$$\mathbf{U} = \mathbf{U}_A^{\mathsf{T}}, \mathbf{V} = \mathbf{V}_A^{\mathsf{T}},$$

$$\mathbf{C} = \text{term-author matrix}$$

$$c_{ik} = \sum_j x_{ijk}$$

$$\mathbf{C} \approx \mathbf{U}_C \mathbf{\Sigma}_C \mathbf{V}_C^T$$ (rank 100)

$$\mathbf{W} = \mathbf{V}_C^{\mathsf{T}},$$

Run rank-100 PARAFAC on compressed tensor.
Reassemble results.

Three-Way Fingerprints

- Each of the Terms, Docs, and Authors has a rank-k (k=100) fingerprint from the PARAFAC approximation
- All items can be directly compared in "concept space"
- Thus, we can compare any of the following
  - Term-Term
  - Doc-Doc
  - Term-Doc
  - Author-Author
  - Author-Term
  - Author-Doc

$$\mathcal{X} \approx \lambda [\![\mathbf{T}, \mathbf{D}, \mathbf{A}]\!]$$

$$\text{score} = \mathbf{u}^{\mathsf{T}} \Lambda \mathbf{v}$$

- The fingerprints can be used as inputs for clustering, classification, etc.

MATLAB Results

- Go to MATLAB

Command Window

File Edit Debug Desktop Window Help

```
~~~~~~~~~~~~ Group 1 ~~~~~~~~~~

Weight = 0.649794
 0.2291772  3474 Vortex motion law for the Schrodinger-Ginzburg-Landua equations
 0.2280338  1633 Vortex state of d-wave superconductors in the Ginzburg-Landau energy
 0.2233726   320 Studies of a Ginzburg-Landau model for d-wave superconductors
 0.2183914  3340 Vortices in p-wave superconductivity
 0.2056138   485 Numerical solution of the three-dimensional Ginzburg-Landau models using artificial boundary
-0.0130460   463 Layer stripping for a transversely isotropic elastic medium
-0.0132632  1151 Scattering of time-harmonic electromagnetic waves by anisotropic inhomogeneous scatterers or impenetra
-0.0133375  1206 Phase equations for relaxation oscillators
-0.0135059  2592 On the two-dimensional gas expansion for compressible Euler equations
-0.0141843  3091 A thermomechanical model for energetic materials with phase transformations
 0.4828654  3387 landau
 0.4489465  2614 ginzburg
 0.2688777  6130 superconductivity
 0.2611251  6771 vortex
 0.2227376  6772 vortices
-0.0120339  1964 elastic
-0.0120368  1620 design
-0.0120543  3767 mesh
-0.0144529  2554 gas
-0.0153897  5462 scattering
 0.7300468  1322 du q
 0.3112497  3142 lin tc
 0.2275581   814 chapman sj
 0.1382164  4991 spirn d
 0.1048653  3133 lin fg
-0.0182970  5898 yao pf
-0.0188236  2045 han wm
-0.0244190  2947 laurencot p
-0.0281511  2393 izhikevich em
-0.0318239  3369 manservisi s

Return to continue, jump to rank, or '0' (zero) to quit:
```

Command Window

File Edit Debug Desktop Window Help

```
Find terms similar to 'tensor'

Match 1: tensor (6261)
  No. docs in which the term appears: 61
  No. authors that use the term: 118
  Norm of matching item: 1.934519e-001
  -- Top 10 matches for PARAFAC --
  Score 2.73e-001: tensor (6261)
  Score 2.35e-001: multilinear (3955)
  Score 2.15e-001: tensors (6262)
  Score 2.06e-001: svds (6182)
  Score 2.04e-001: deficient (1520)
  Score 2.00e-001: valuable (6660)
  Score 1.97e-001: confirms (1160)
  Score 1.94e-001: hyper (2860)
  Score 1.93e-001: displacement (1787)
  Score 1.92e-001: div (1814)
  -- Top 10 matches for SVD --
  Score 1.17e-001: decomposition (1498)
  Score 1.13e-001: squares (5891)
  Score 1.07e-001: rank (4980)
  Score 9.75e-002: least (3437)
  Score 9.20e-002: singular (5724)
  Score 7.89e-002: tensor (6261)
  Score 7.21e-002: elasticity (1965)
  Score 6.22e-002: orthogonal (4327)
  Score 6.19e-002: mixed (3837)
  Score 5.71e-002: elastic (1964)
```

Command Window

File Edit Debug Desktop Window Help

```
Find documents similar to 'tensor'

Match 1: tensor (6261)
  No. docs in which the term appears: 61
  No. authors that use the term: 118
  Norm of matching item: 1.934519e-001
  -- Top 10 matches for PARAFAC --
  Score 2.21e-001: On the best rank-1 and rank-(R1R2...R-N) approximation of higher-order tensors (1224)
  Score 2.01e-001: Efficient solution of the rank-deficient linear least squares problem (148)
  Score 1.87e-001: On the best rank-1 approximation of higher-order supersymmetric tensors (2570)
  Score 1.86e-001: Orthogonal tensor decompositions (2180)
  Score 1.82e-001: A counterexample to the possibility of an extension of the Eckart-Young low-rank approximation theor
  Score 1.78e-001: Least squares solution of matrix equation AXB(*)+CYD*=E-* (3192)
  Score 1.74e-001: Least-squares methods for incompressible Newtonian fluid flow Linear stationary problems (4244)
  Score 1.74e-001: Least-squares methods for linear elasticity (4243)
  Score 1.73e-001: Tensor methods for large sparse nonlinear least squares problems (1119)
  Score 1.69e-001: Multilevel boundary functionals for least-squares mixed finite element methods (396)
  -- Top 10 matches for SVD --
  Score 5.78e-002: A counterexample to the possibility of an extension of the Eckart-Young low-rank approximation theor
  Score 5.77e-002: On the best rank-1 and rank-(R1R2...R-N) approximation of higher-order tensors (1224)
  Score 5.59e-002: Least-squares methods for linear elasticity (4243)
  Score 5.35e-002: First-order system least squares for the stress-displacement formulation Linear elasticity (3431)
  Score 4.98e-002: Rank-one approximation to high order tensors (2369)
  Score 4.72e-002: Least-squares methods for incompressible Newtonian fluid flow Linear stationary problems (4244)
  Score 4.51e-002: First-order system least squares for linear elasticity Numerical results (1178)
  Score 4.43e-002: Orthogonal tensor decompositions (2180)
  Score 4.39e-002: Layer stripping for a transversely isotropic elastic medium (463)
  Score 3.91e-002: First-order system least squares for the Stokes and linear elasticity equations Further results (117
```

Command Window

File Edit Debug Desktop Window Help

```
Find authors similar to 'tensor'

Match 1: tensor (6261)
  No. docs in which the term appears: 61
  No. authors that use the term: 118
  Norm of matching item: 1.934519e-001
  -- Top 10 matches for PARAFAC --
  Score 1.91e-001: vandewalle j (5451)
  Score 1.84e-001: delathauwer l (1181)
  Score 1.83e-001: quintanaorti g (4293)
  Score 1.83e-001: quintanaorti es (4292)
  Score 1.83e-001: petitet a (4109)
  Score 1.76e-001: chen y (873)
  Score 1.76e-001: shim sy (4846)
  Score 1.73e-001: demoor b (1199)
  Score 1.68e-001: barlow jl (288)
  Score 1.66e-001: cai zq (693)
```

Command Window

File Edit Debug Desktop Window Help

```
Find terms similar to Dhillon

Match 1: dhillon is (1239)
  No. terms used by author: 68
  No. documents written by author: 1
  Norm of matching item: 5.289941e-002
  -- Top 10 matches for PARAFAC --
  Score 2.27e-001: bidiagonal (575)
  Score 2.26e-001: qr (4907)
  Score 2.11e-001: ldlt (3424)
  Score 2.08e-001: lapack (3391)
  Score 2.07e-001: columns (1000)
  Score 2.04e-001: column (999)
  Score 2.03e-001: revealing (5308)
  Score 2.03e-001: pivoting (4579)
  Score 2.02e-001: rank (4980)
  Score 1.98e-001: bjorck (610)
Find authors similar to Dhillon

Match 1: dhillon is (1239)
  No. terms used by author: 68
  No. documents written by author: 1
  Norm of matching item: 5.289941e-002
  -- Top 10 matches for PARAFAC --
  Score 3.11e-001: dhillon is (1239)
  Score 3.11e-001: parlett bn (4024)
  Score 2.28e-001: drmac z (1315)
  Score 2.19e-001: molera jm (3625)
  Score 2.16e-001: jessup er (2437)
  Score 2.04e-001: dopico fm (1292)
  Score 2.04e-001: moro j (3661)
  Score 2.02e-001: jubete f (2495)
  Score 2.02e-001: pruneda re (4253)
  Score 2.02e-001: castillo e (761)
```

Command Window

File Edit Debug Desktop Window Help

```
Find terms similar to OLeary DP

Match 1: oleary dp (3913)
  No. terms used by author: 114
  No. documents written by author: 2
  Norm of matching item: 2.567276e-001
  -- Top 10 matches for PARAFAC --
  Score 2.35e-001: ill (2906)
  Score 2.15e-001: tikhonov (6334)
  Score 2.12e-001: posed (4667)
  Score 2.07e-001: regularization (5142)
  Score 2.05e-001: conditioned (1138)
  Score 2.02e-001: clustered (940)
  Score 2.01e-001: unmixed (6601)
  Score 2.01e-001: regularizing (5145)
  Score 1.95e-001: regularisation (5140)
  Score 1.95e-001: regularized (5144)
Find authors similar to OLeary DP

Match 1: oleary dp (3913)
  No. terms used by author: 114
  No. documents written by author: 2
  Norm of matching item: 2.567276e-001
  -- Top 10 matches for PARAFAC --
  Score 2.55e-001: oleary dp (3913)
  Score 2.37e-001: kilmer me (2645)
  Score 2.30e-001: hansen pc (2056)
  Score 2.18e-001: o'leary dp (3889)
  Score 2.10e-001: gulliksson m (1956)
  Score 2.10e-001: wedin pa (5695)
  Score 2.09e-001: maass p (3306)
  Score 2.08e-001: mante c (3372)
  Score 2.07e-001: jin qn (2458)
  Score 2.05e-001: johnston pr (2470)
```

Command Window

File Edit Debug Desktop Window Help

```
Find authors like H.Y. Zha

Match 1: zha hy (5990)
  No. terms used by author: 164
  No. documents written by author: 5
  Norm of matching item: 3.795614e-001
  -- Top 10 matches for PARAFAC --
  Score 3.55e-001: zha hy (5990)
  Score 3.46e-001: simon hd (4890)
  Score 3.36e-001: zhang zy (6025)
  Score 3.28e-001: simon h (4889)
  Score 3.19e-001: fundelic re (1645)
  Score 3.09e-001: zha h (5989)
  Score 2.94e-001: zhang t (6013)
  Score 2.81e-001: vandooren p (5453)
  Score 2.77e-001: golub g (1820)
  Score 2.75e-001: dopico fm (1292)
```

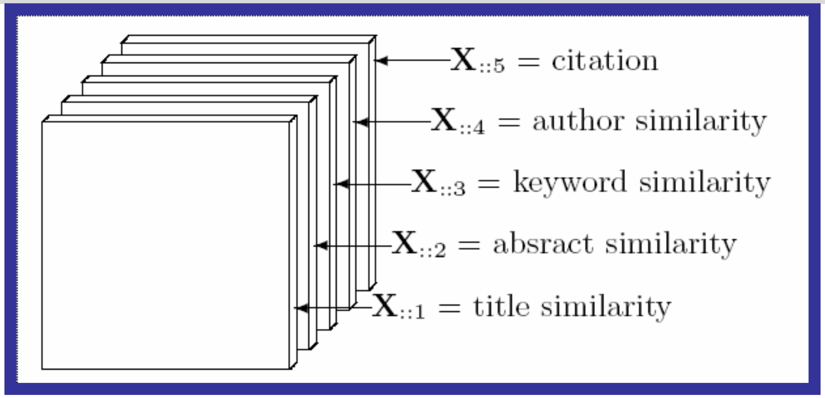
Command Window

File Edit Debug Desktop Window Help

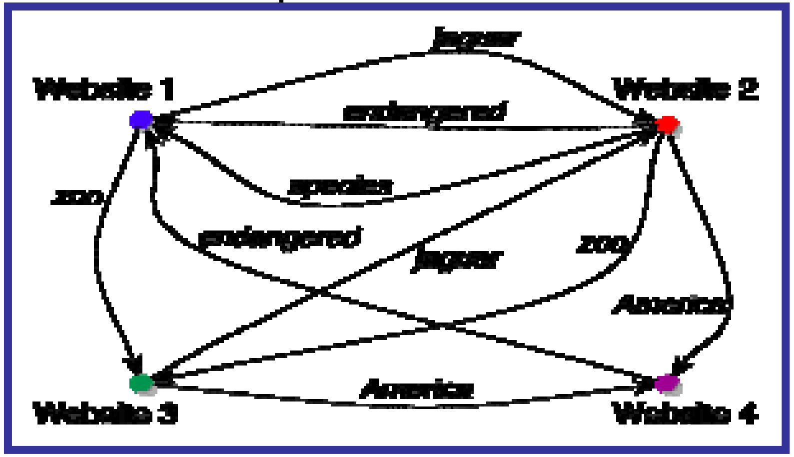
```
Find authors similar to 'svd'

Match 1: svd (6181)
  No. docs in which the term appears: 24
  No. authors that use the term: 36
  Norm of matching item: 1.789480e-001
  -- Top 10 matches for PARAFAC --
  Score 3.28e-001: delathauwer l (1181)
  Score 3.23e-001: golub g (1820)
  Score 3.23e-001: vandooren p (5453)
  Score 3.21e-001: dopico fm (1292)
  Score 3.21e-001: moro j (3661)
  Score 3.20e-001: fundelic re (1645)
  Score 3.13e-001: jessup er (2437)
  Score 3.12e-001: zha h (5989)
  Score 3.12e-001: demmel j (1197)
  Score 3.12e-001: vandewalle j (5451)
>>
```

Wrap-Up

- Higher-order LSI for term-doc-author tensor
- Tucker-PARAFAC combination for sparse tensors
  - Spasre Tensor Toolbox (release summer 2006)
- Mathematical manipulations
  - Kolda, Tech. Rep. SAND2006-2081
- Thanks to Kevin Boyack for journal data
- For more info: Tammy Kolda, tgkolda@sandia.gov



Dunlavy, Kolda, Kegelmeyer, Tech. Rep. SAND2006-2079



Kolda, Bader, Kenny, ICDM05