Sandia National Laboratories

Tensor Analysis for Evolving Networks

Tamara G. Kolda

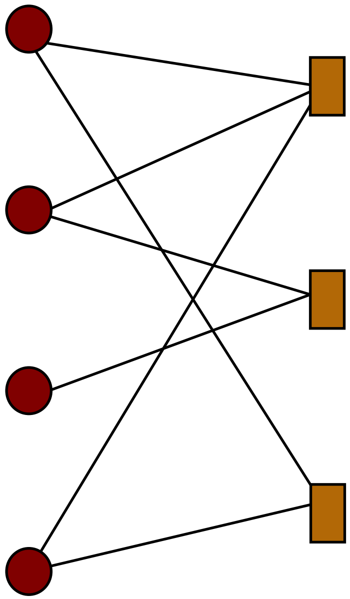
Workshop on Time-varying Complex Network Analysis
Cambridge, UK, September 19, 2012

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

U.S. Department of Energy
Office of Advanced Scientific Computing Research

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Networks, Matrices, Factor Analysis

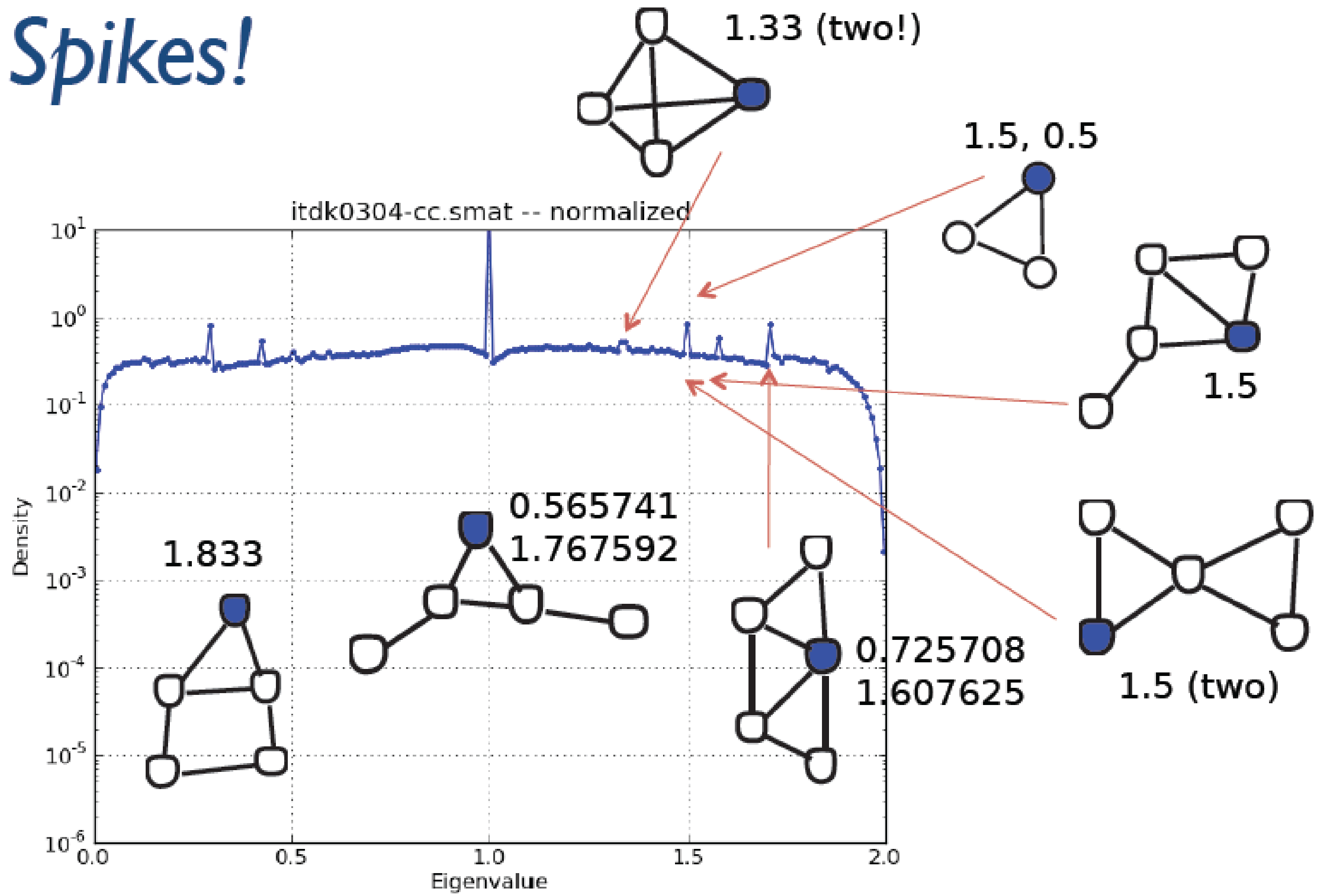


$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- Networks correspond to sparse matrices
  - Symmetric $\Rightarrow$ Undirected
  - Asymmetric $\Rightarrow$ Directed
  - Rectangular $\Rightarrow$ Bipartite
  - Binary $\Rightarrow$ Unweighted
- Matrix analysis yields insight
  - Ranking methods
    - PageRank (Page et al., 1999)
    - Hubs & Authorities (Kleinberg, 1999)
  - Eigenvalues
    - Pattern indications (Gleich, SIAM CSE 2011)
  - Eigenvectors of Laplacian
    - Partitioning (Pothen, Simon, Liou, 1990)
    - Estimating commute time (Fouss et al., 2007)
  - Matrix factorization
    - Dimension reduction
    - Unsupervised learning
    - Nonnegative, sparse, etc.

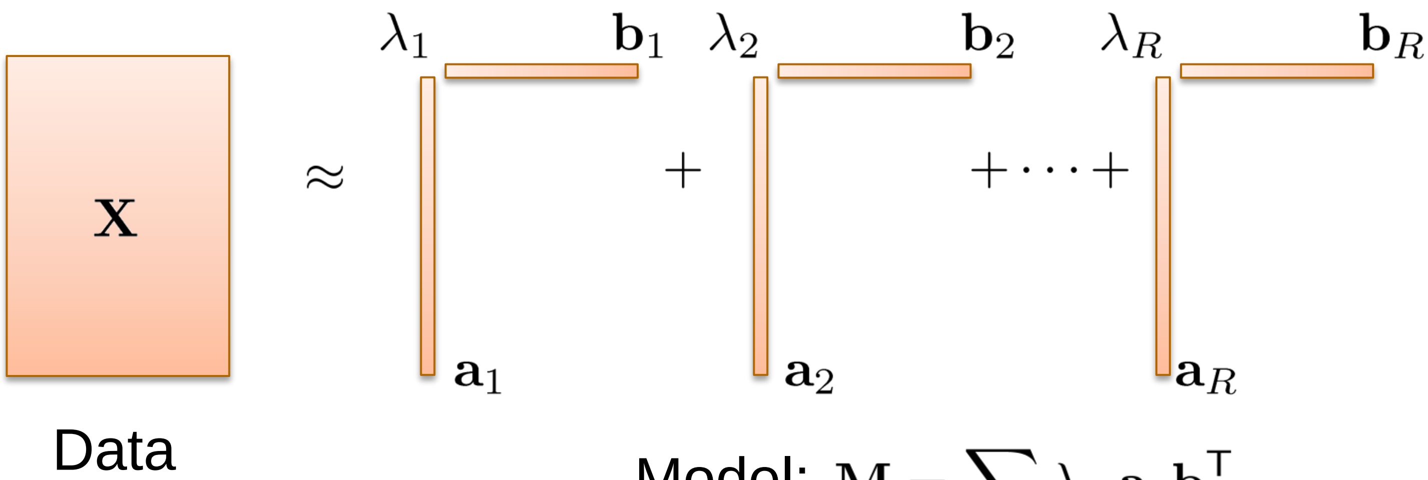
Aside: Gleich's work on Eigevalues as Sandia's Von Neumann Fellow

http://www.slideshare.net/dgleich/the-spectre-of-the-spectrum



Matrix Factorizations for Analysis

Singular Value Decomposition (SVD)

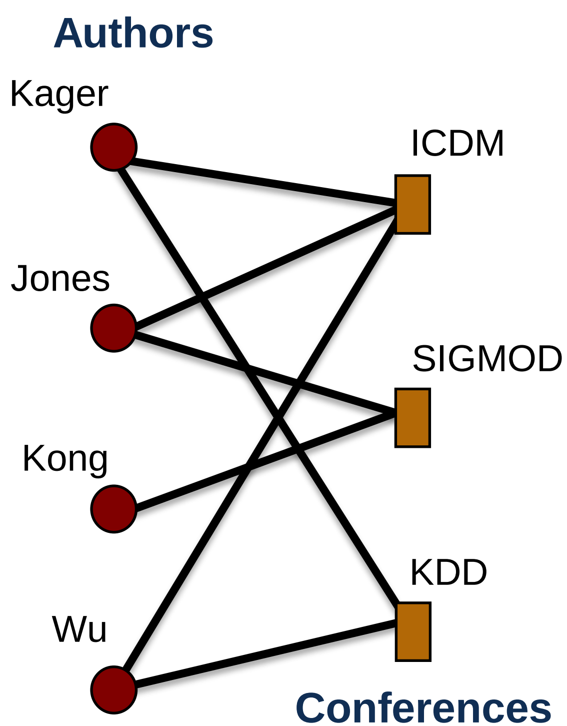


Model: $\mathbf{M} = \sum_r \lambda_r \, \mathbf{a}_r \mathbf{b}_r^{\mathsf{T}}$

$$\min \sum_{ij} (x_{ij} - m_{ij})^2 \quad \text{subject to} \quad m_{ij} = \sum_r \lambda_r \, a_{ir} \, b_{jr}$$

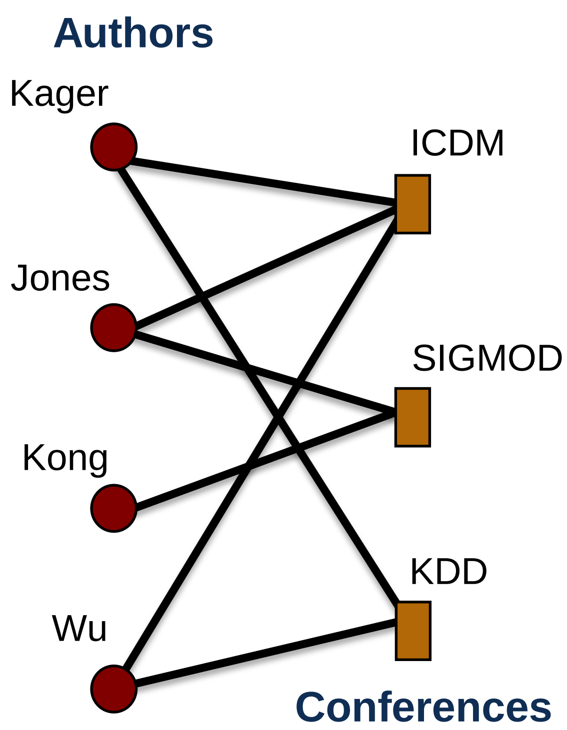
Key references: Beltrami (1873), Pearson (1901), Eckart & Young (1936)

Interpretation of 2-Way Factor Model



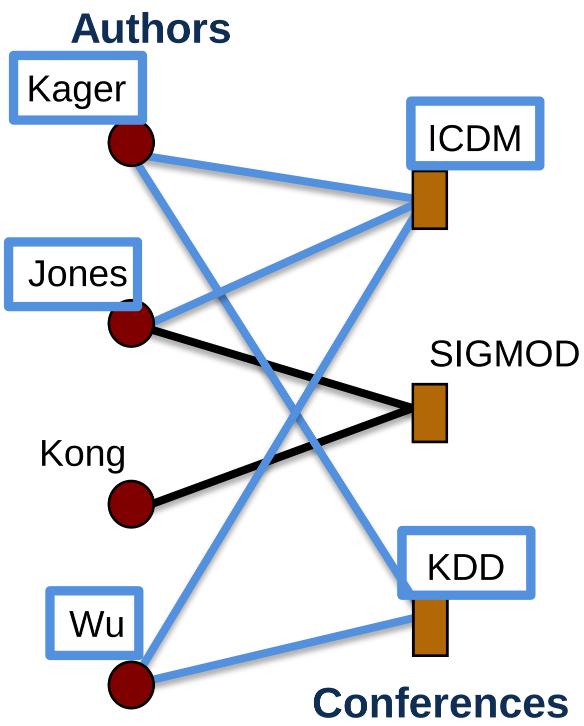
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Interpretation of 2-Way Factor Model



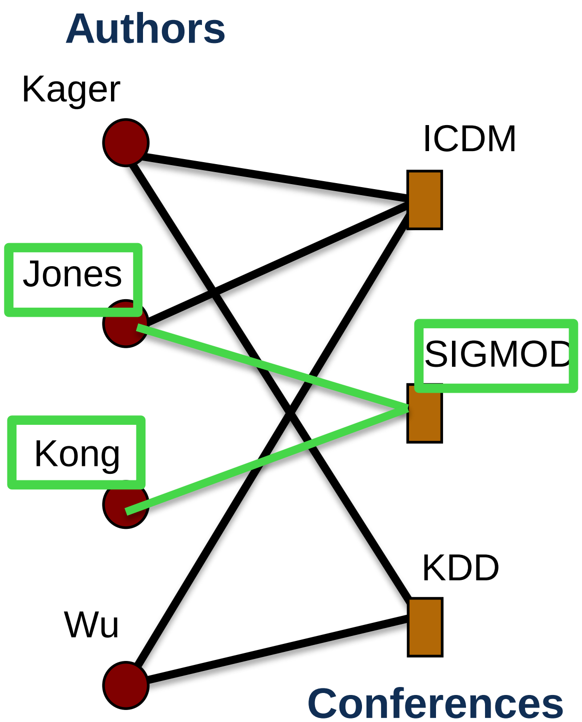
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model



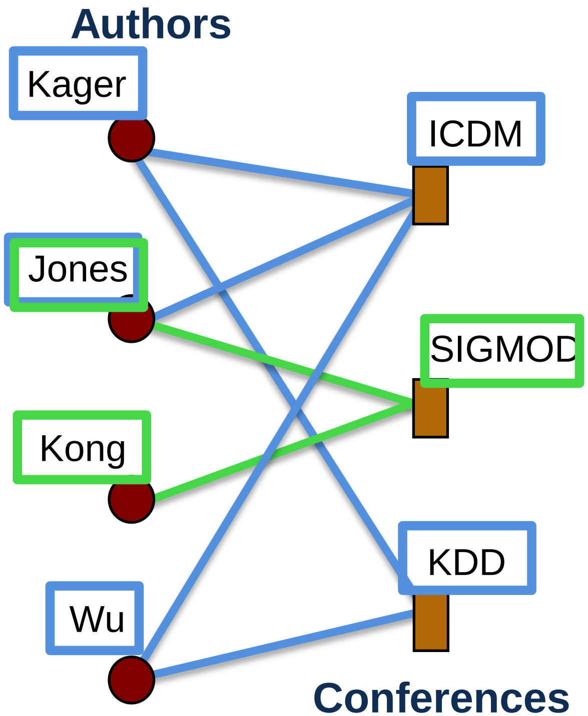
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model



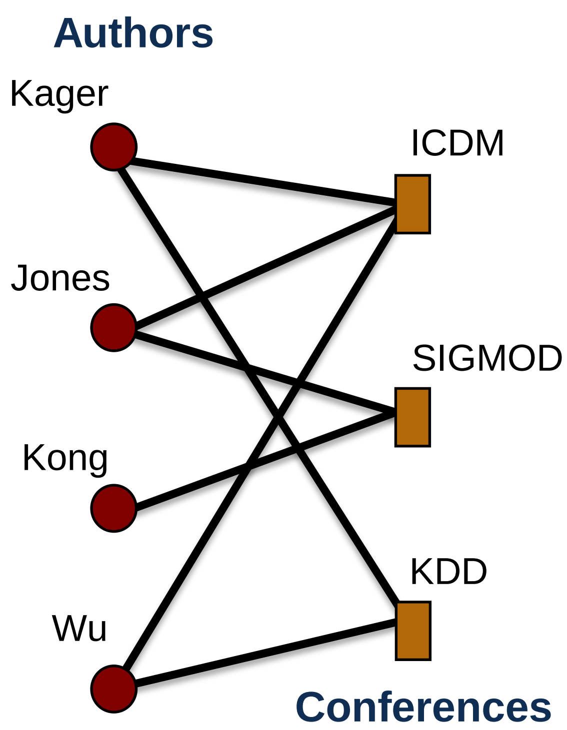
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model



$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model

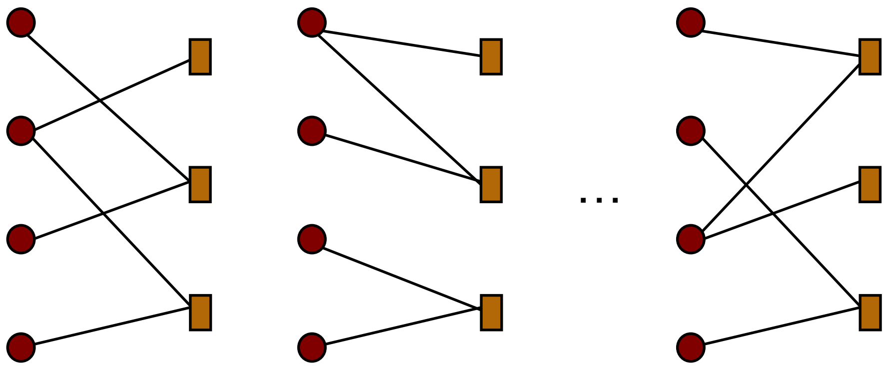


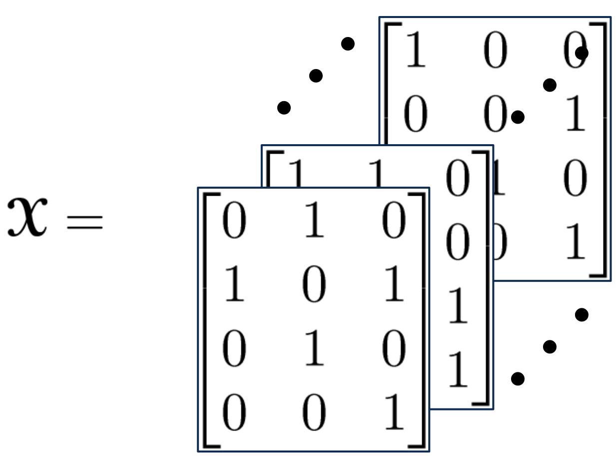
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

2-Way Models Suffer from "Gauge Freedom"

$$\mathbf{X} \approx \mathbf{A}\mathbf{B}^{\mathsf{T}} = \underbrace{\begin{bmatrix} .39 & .90 \\ 1.04 & -.04 \\ .66 & -.41 \\ .39 & .90 \end{bmatrix}}_{\hat{\mathbf{A}} = \mathbf{A}\mathbf{S}} \underbrace{\begin{bmatrix} .83 & 0.80 \\ 1.04 & -.48 \\ .23 & .96 \end{bmatrix}^{\mathsf{T}}}_{\hat{\mathbf{B}}^{\mathsf{T}} = (\mathbf{B}\mathbf{S}^{-1})^{\mathsf{T}}}$$

Time-Varying Networks & Tensors

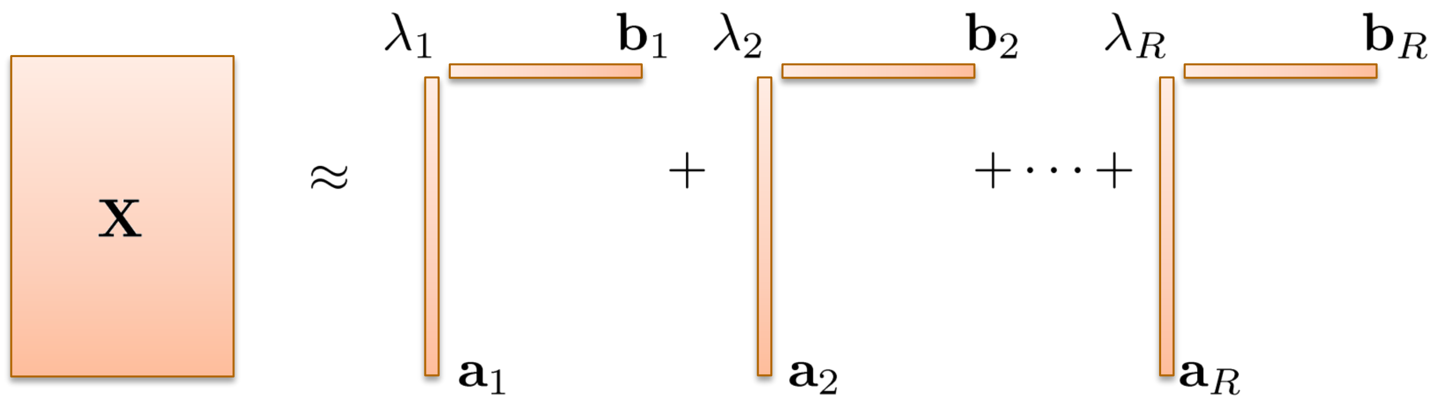




- Time-varying networks correspond naturally to 3-way tensors
  - Time must be “binned”
- Additional modes correspond to higher-order tensors
  - Link type (like, post, IM, msg)
- Tensor factorizations yield insights similar to matrix case
  - Tensor factorizations
    - Canonical decomposition
    - Poisson tensor decomposition
    - Coupled matrix/tensor
  - Other factorizations
    - Tucker2 decomposition
    - DEDICOM

Matrix Factorizations for Analysis

Think: SVD or NMF



Data

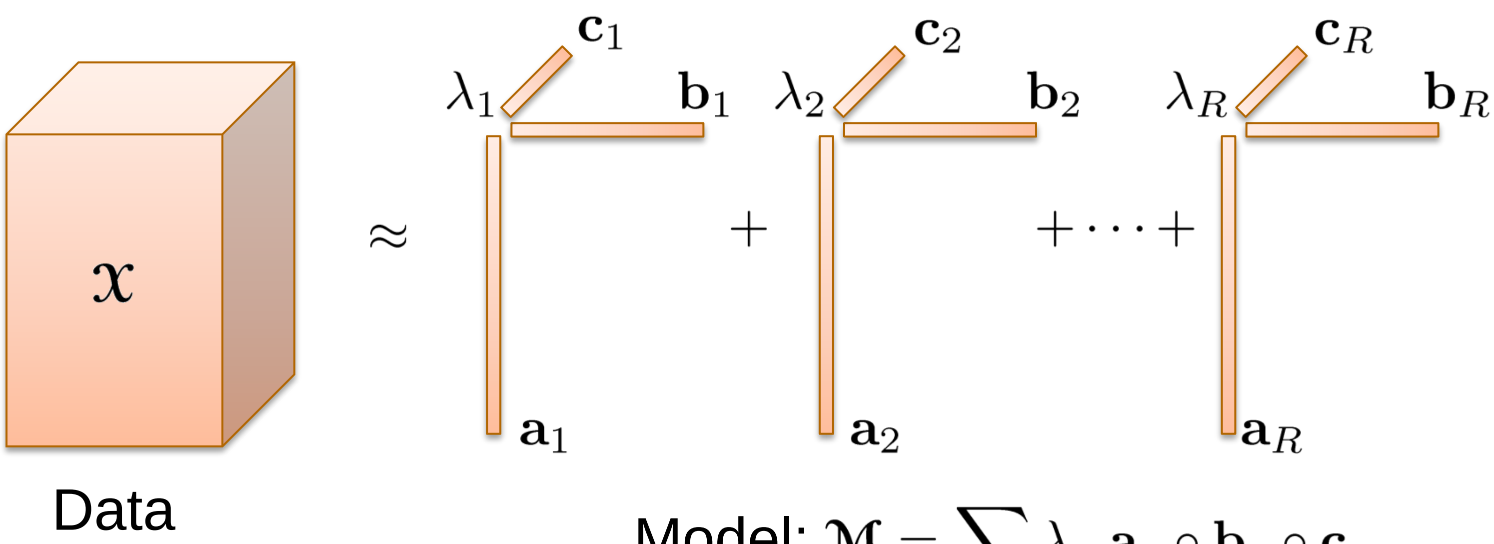
Model: $\mathbf{M} = \sum_r \lambda_r \, \mathbf{a}_r \mathbf{b}_r^{\mathsf{T}}$

$$\min \sum_{ij} (x_{ij} - m_{ij})^2 \quad \text{subject to} \quad m_{ij} = \sum_r \lambda_r \, a_{ir} \, b_{jr}$$

Key references: Beltrami (1873), Pearson (1901), Eckart & Young (1936)

Multi-way Factorizations for Analysis

CANDECOMP/PARAFAC (CP) Model

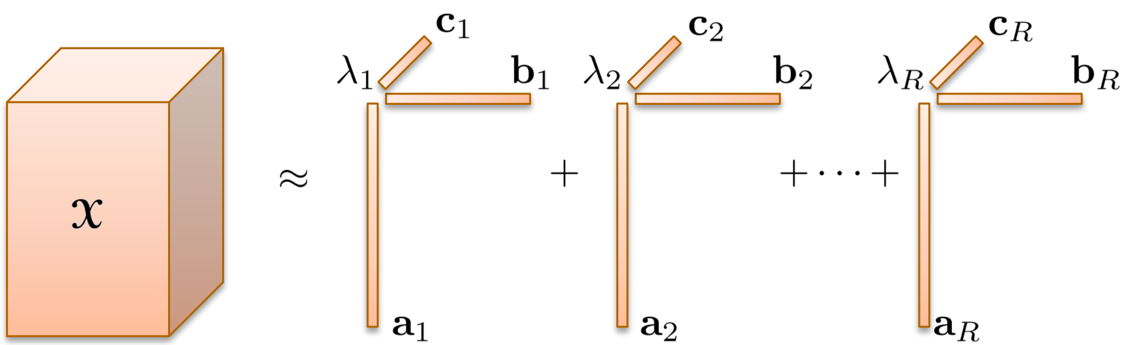


Model: $\mathcal{M} = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$

$$\min \sum_{ijk} (x_{ijk} - m_{ijk})^2 \quad \text{subject to} \quad m_{ijk} = \sum_r \lambda_r \, a_{ir} \, b_{jr} \, c_{kr}$$

Key references: Hitchcock (1927), Harshman (1970), Carroll and Chang (1970)

Uniqueness of Tensor Factorization

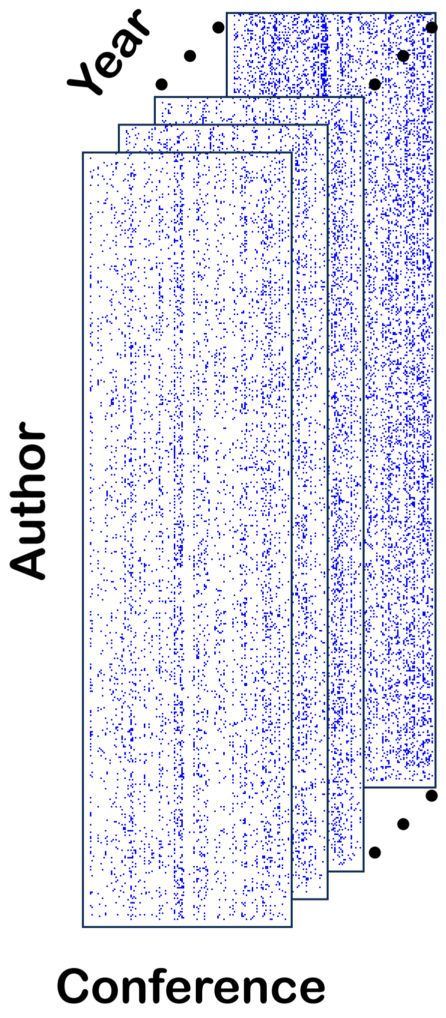


- $k_{\mathbf{A}}$ = k-rank of a matrix **A** = maximum value of k such that any k columns are linearly independent
- Factorization essentially unique if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2R + 2$$

- Essentially unique = unique up to permutation and scaling ambiguities = no gauge freedom (unlike matrix case)

Example: DBLP Data



DBLP has data from 1936-2007
(used only "inproceedings" from 1991-2000)

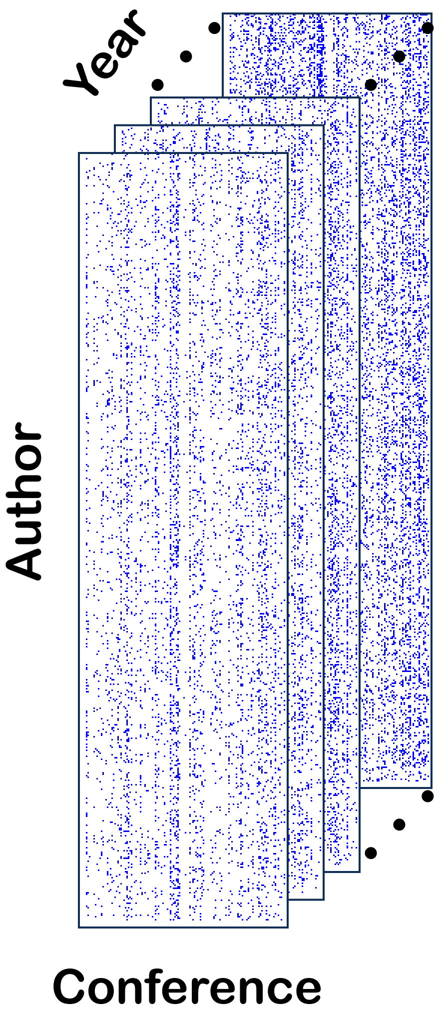
Training Data	10 Years: 1991-2000
# Authors (min 10 papers)	7108
# Conferences	1103
Links	113k (0.14% dense)

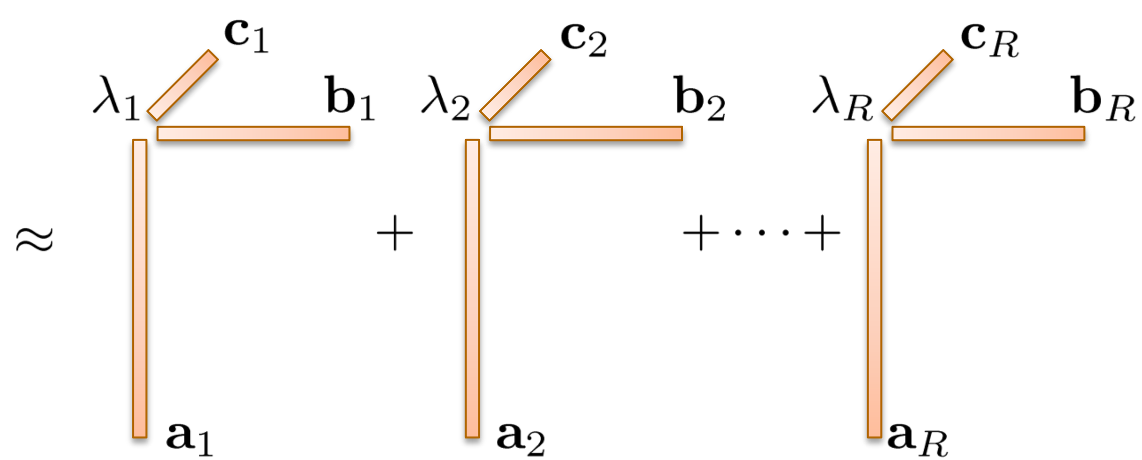
Nonzeros defined by:

$$x_{ijk} = \log(c_{ijk}) + 1 \text{ if } c_{ijk} > 0$$

Acar, Dunlavy, & Kolda, Temporal Link Prediction using Matrix and Tensor Factorizations, ACM TKDD, 2010

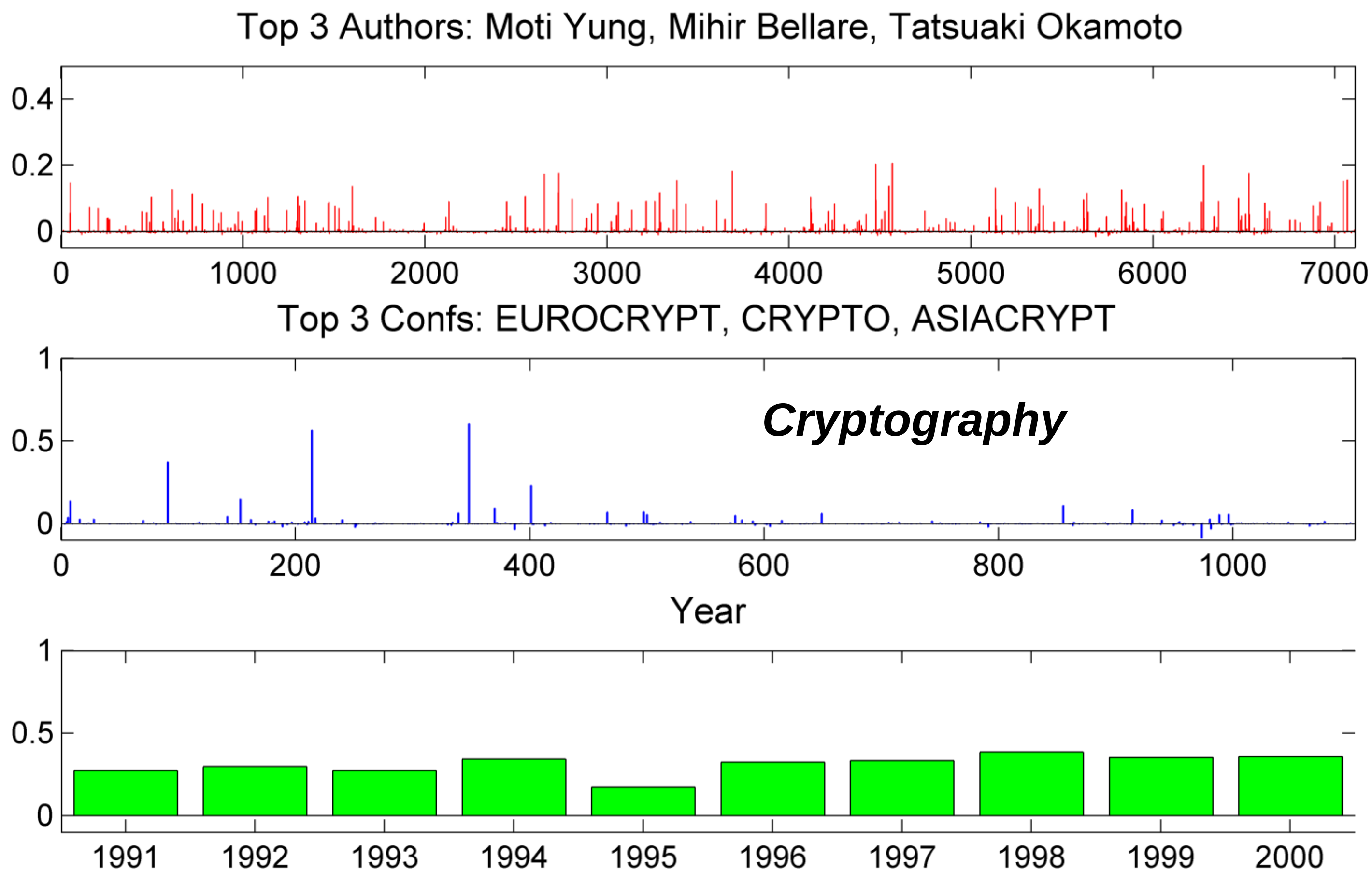
Example: DBLP Data



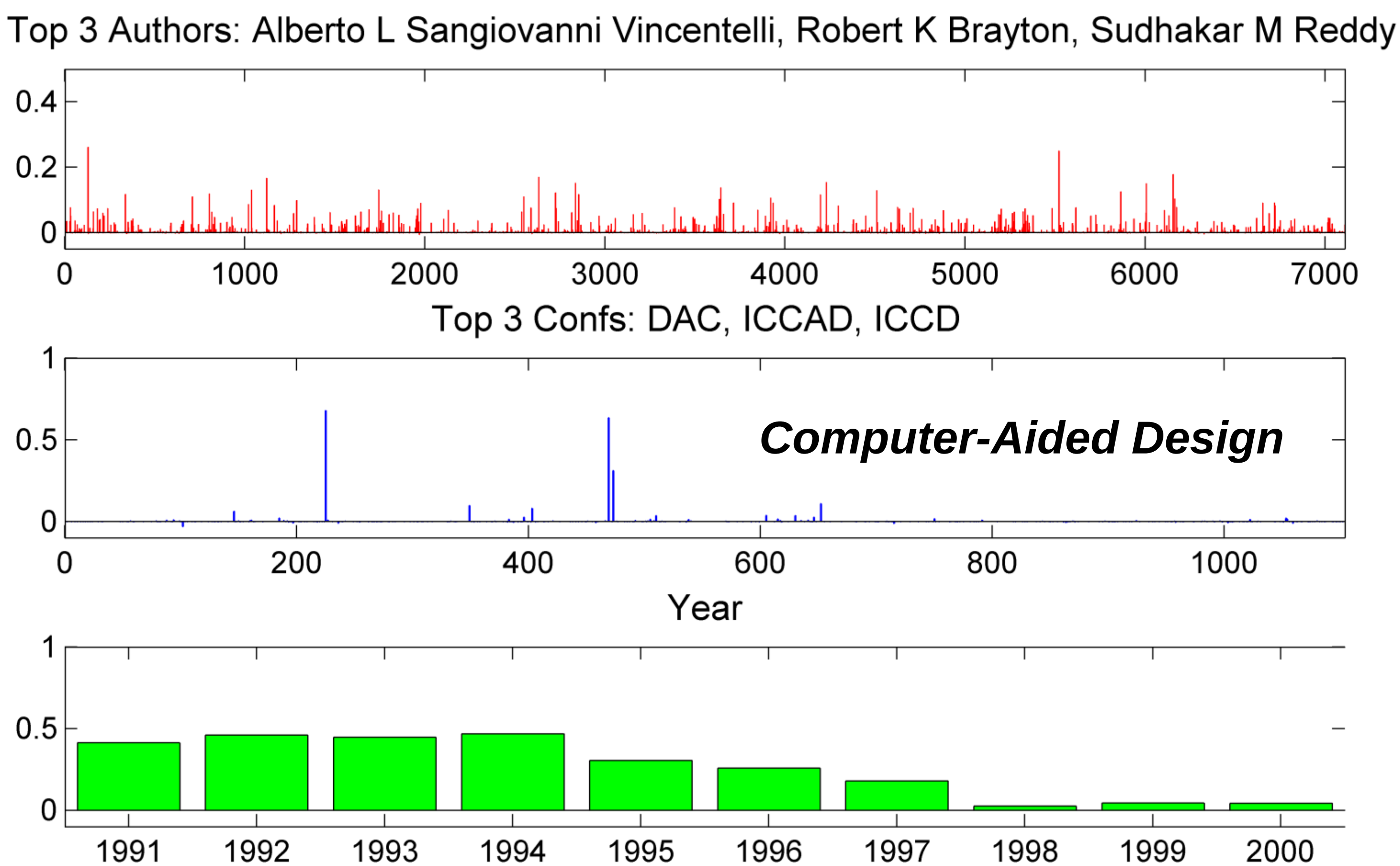


Let's look at some components from a 50-component (R=50) factorization.

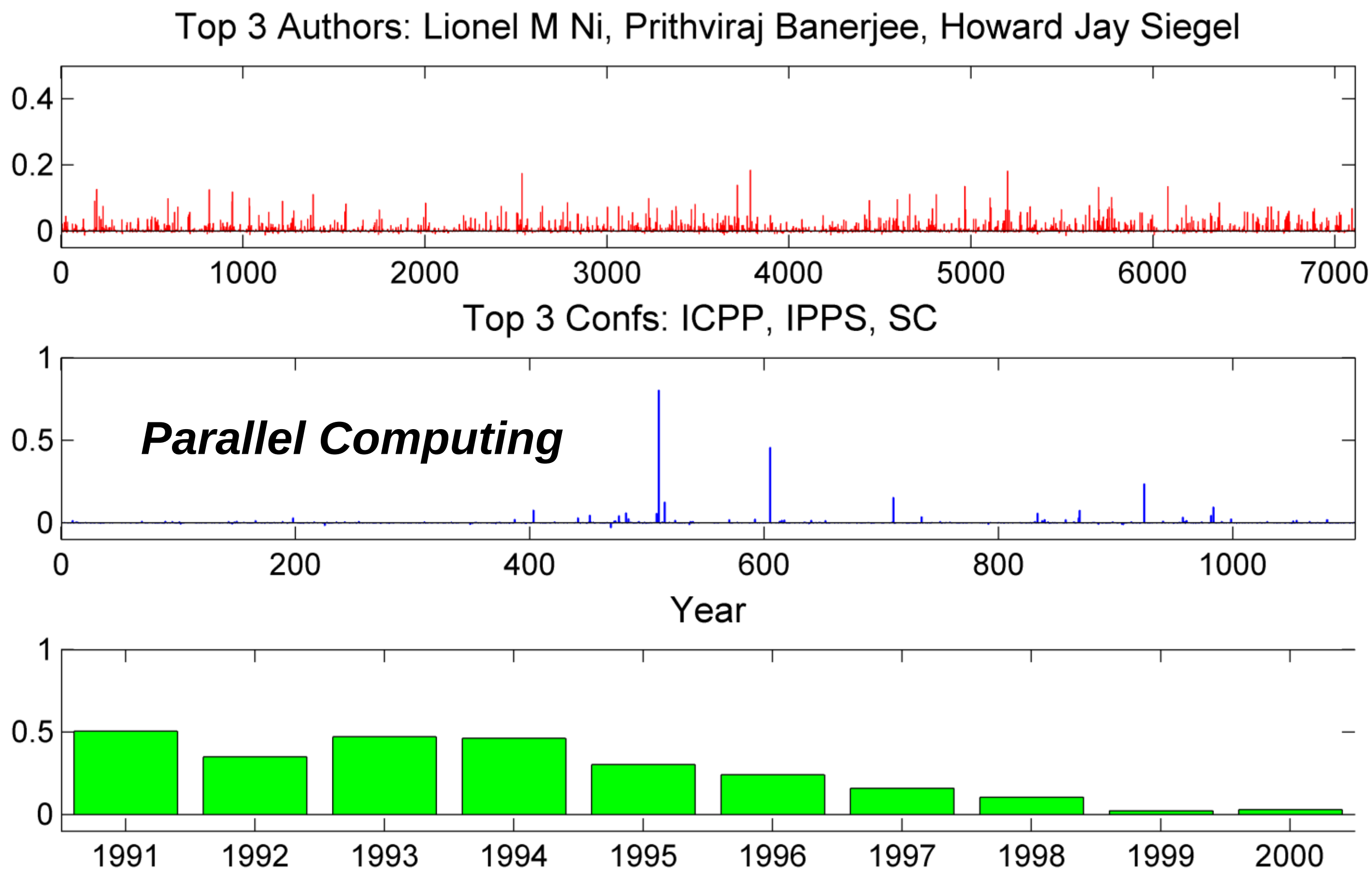
DBLP Component #30 (of 50)



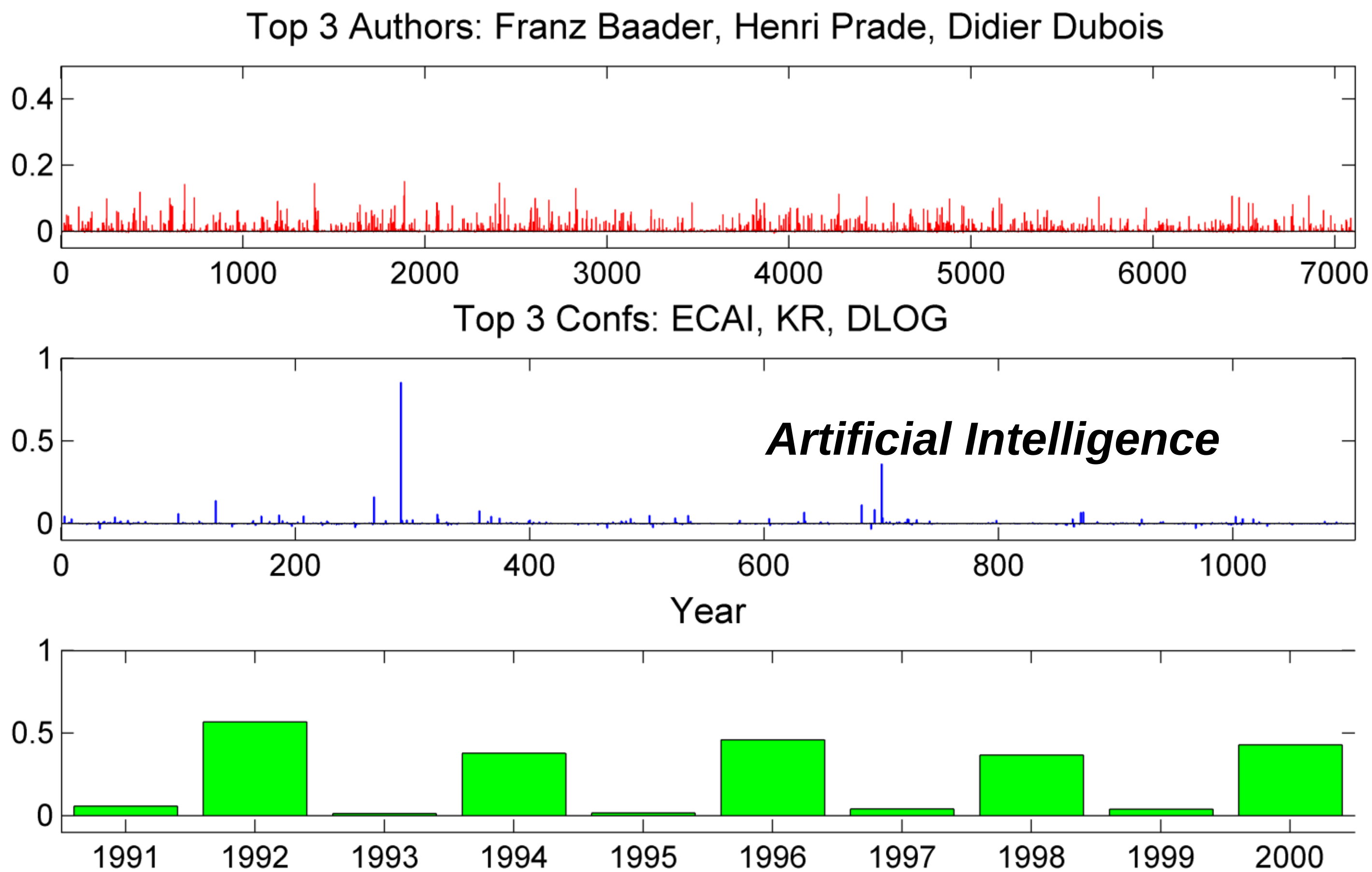
DBLP Component #5 (of 50)



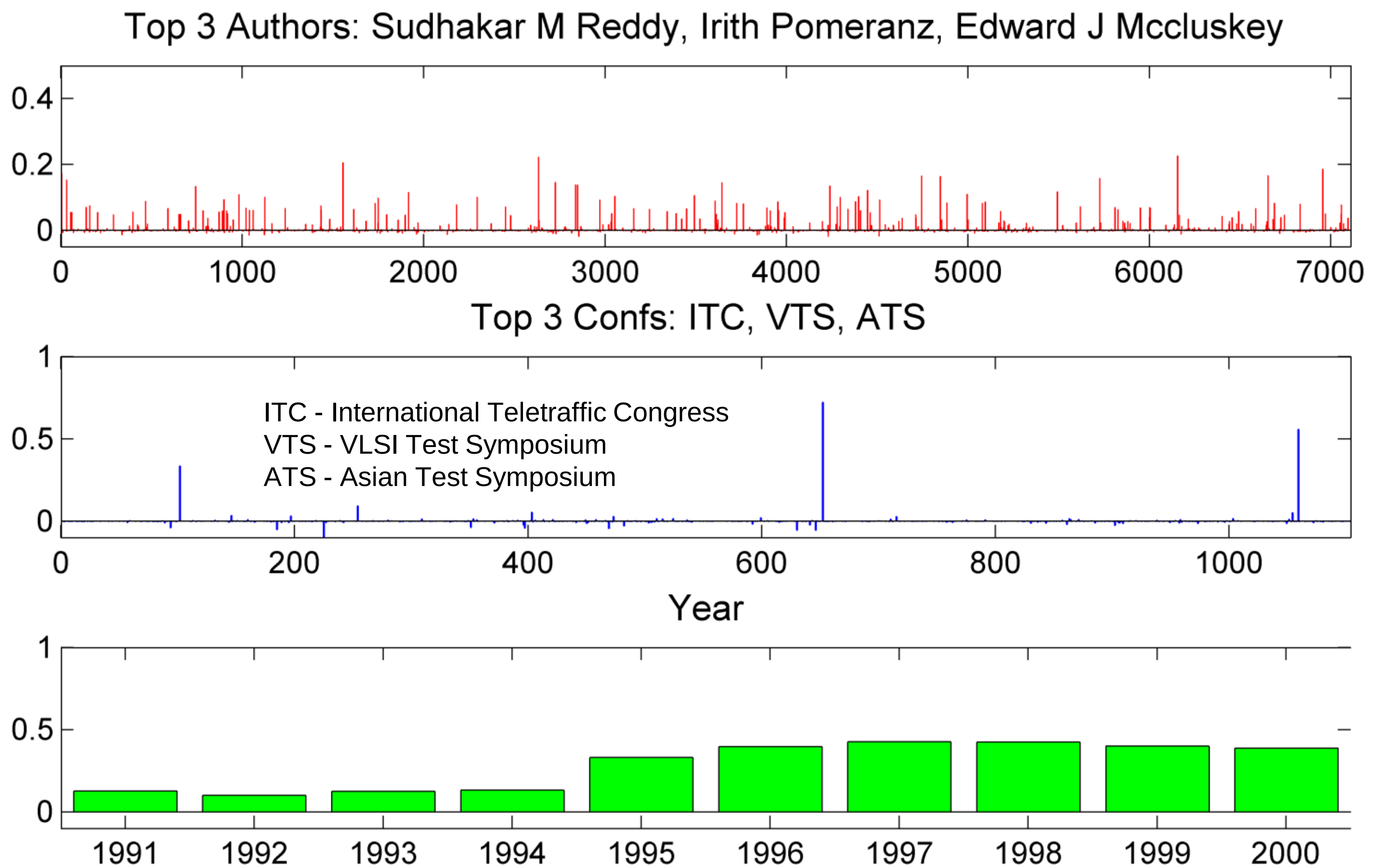
DBLP Component #19 (of 50)



DBLP Component #43 (of 50)

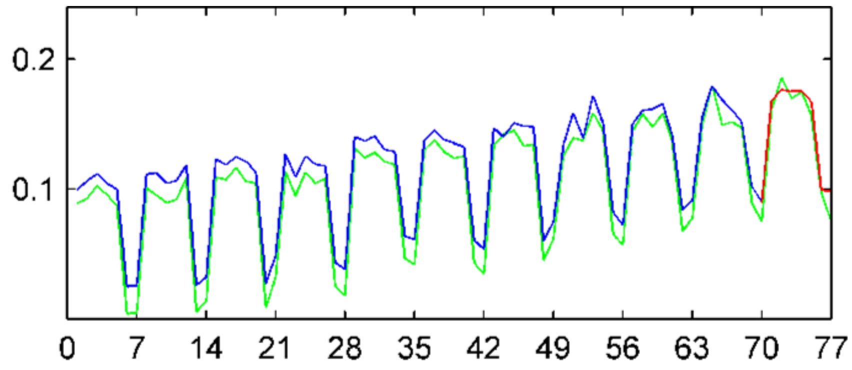


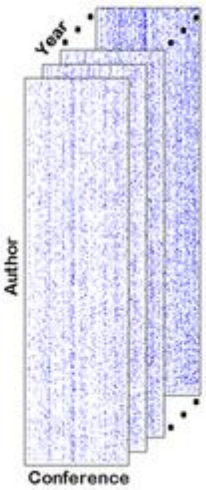
DBLP Component #10 (of 50)



Extension: Temporal Link Prediction

- Problem: Predicting future connections
  - Between computers on a network
  - Between "persons of interest" and places
  - Between buyers and products
- "Needle in the Haystack" Problem
  - # possible connections is huge!
  - # actual connections is small!
- Solution: Represent past connections as tensor
  - Example: Buyer x Object x Date
  - Factorize to look for temporal patterns
  - Use regression to predict future behavior



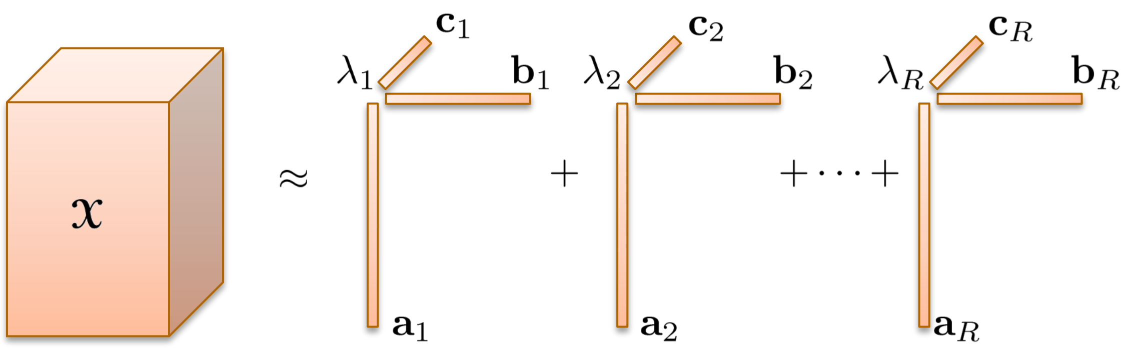


Example Prediction Results

- Predict **who** will publish at **which conference** based on 10 years past data
  - Data: DBLP 1997-2006 / 2007
    - 21K Authors x 2K Conferences
    - 1997-2006: 377K Links
    - 2007: 41K (20k New)
  - Top-1000 Predicted Links
    - Random: 1
    - Our Method: 733
  - Top-1000 New Only [Hard]
    - Random: ½
    - Our Method: 83

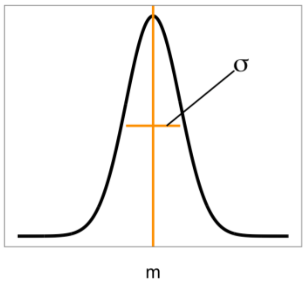
Acar, Dunlavy, & Kolda, Temporal Link Prediction using Matrix and Tensor Factorizations, ACM TKDD, 2010

What does "≈" mean?



- Typically, we minimize the least-squares error
- This corresponds to maximizing the likelihood, assuming a **Gaussian distribution**

$$x_{\mathbf{i}} \sim N(m_{\mathbf{i}}, \sigma^2)$$



Maximize this:

$$\text{likelihood}(\mathcal{M}) = \prod_{\mathbf{i}} \frac{\exp(-(x_{\mathbf{i}} - m_{\mathbf{i}})^2/2\sigma^2)}{2\pi\sigma^2}$$

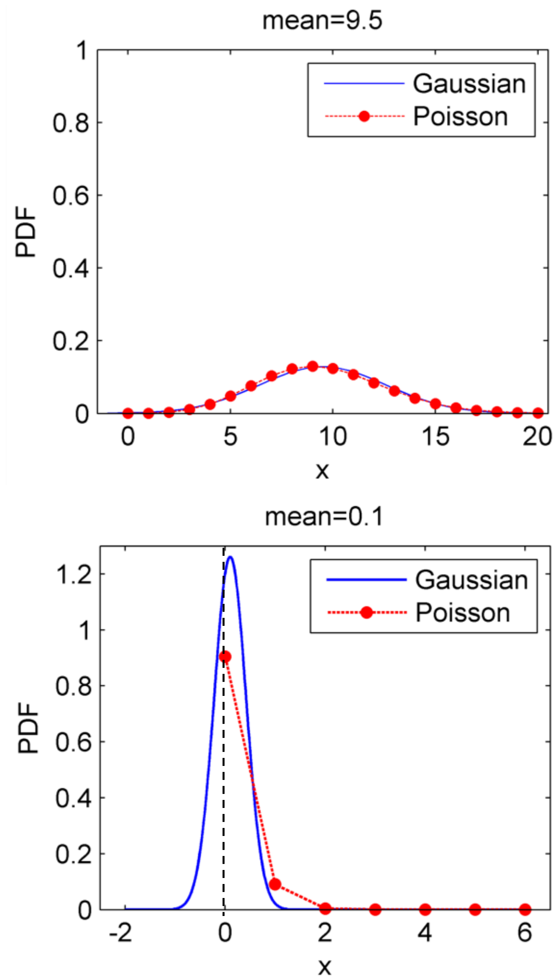
By monotonicity of log, same as maximizing this:

$$\text{log-likelihood}(\mathcal{M}) = c_1 - c_2 \sum_{\mathbf{i}} (x_{\mathbf{i}} - m_{\mathbf{i}})^2$$

Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

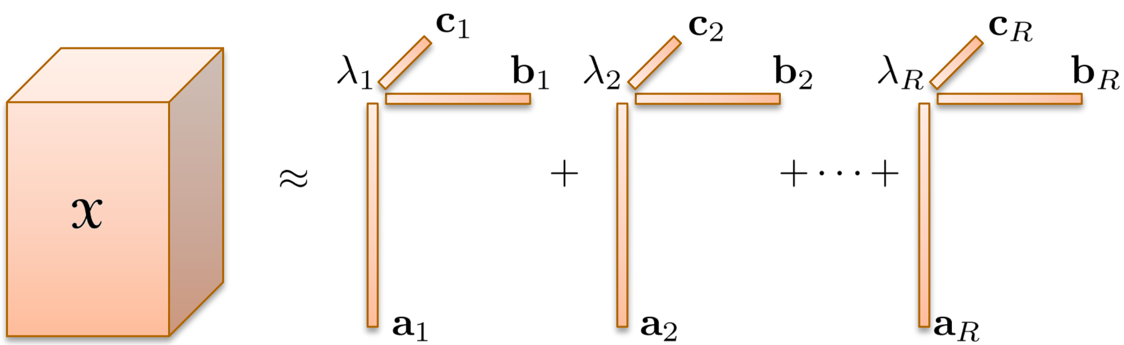
Gaussian often Works Well, But…

- Gaussian (normal) distribution
  - Default model, and for good reason
  - Limiting distribution of the sum of random variables
- Some data are better explained elsewise
  - Non-symmetric errors (e.g., data that grows exponentially)
  - Data with outliers or multiple modes
  - Etc.
- Poisson distribution
  - Associated with count data
  - Discrete, nonnegative
  - High counts can be reasonably approximated by a Gaussian



Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

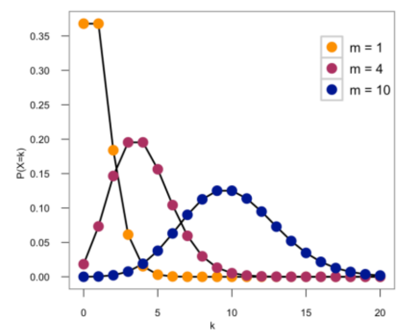
Poisson Tensor Factorization (PTF)



- Poisson preferred for sparse count data
- Automatically nonnegative
- More difficult objective function than least squares
- Note that this objective is also called Kullback-Liebler (KL) divergence

$$x_{\mathbf{i}} \sim \text{Poisson}(m_{\mathbf{i}})$$

$$P(X=x) = \frac{\exp(-m)m^x}{x!}$$



Maximize this:

$$\text{likelihood}(\mathcal{M}) = \prod_{\mathbf{i}} \frac{\exp(-m_{\mathbf{i}})\, m_{\mathbf{i}}^{x_{\mathbf{i}}}}{x_{\mathbf{i}}!}$$

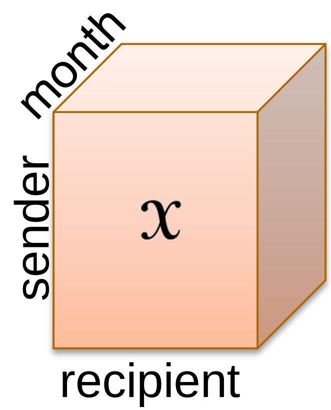
By monotonicity of log, same as maximizing this:

$$\text{log-likelihood}(\mathcal{M}) = c - \sum_{\mathbf{i}} m_{\mathbf{i}} - x_{\mathbf{i}} \log(m_{\mathbf{i}})$$

Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Motivating Example: Enron Email

- Emails from Enron FERC investigation
  - 8540 Messages
  - 28 Months (from Dec 1999 to Mar 2002)
  - 105 People (sent and received at least one email every month)
  - x_{ijk} = # emails from sender i to recipient j in month k
  - 105 x 105 x 28 = 308,700 possible entries
  - 8,500 nonzero counts
  - **3% dense**



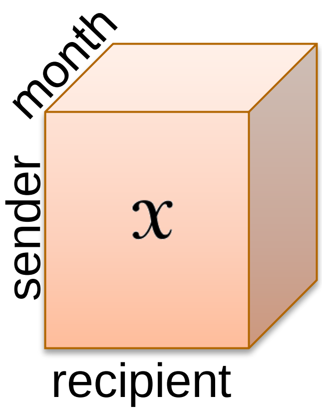
- Questions: What can we learn about this data?
  - Each person labeled by Zhou et al. (2007); see also Owen and Perry (2010)
    - Seniority: 57% senior, 43% junior
    - Gender: 67% male, 33% female
    - Department: 24% legal, 31% trading, 45% other

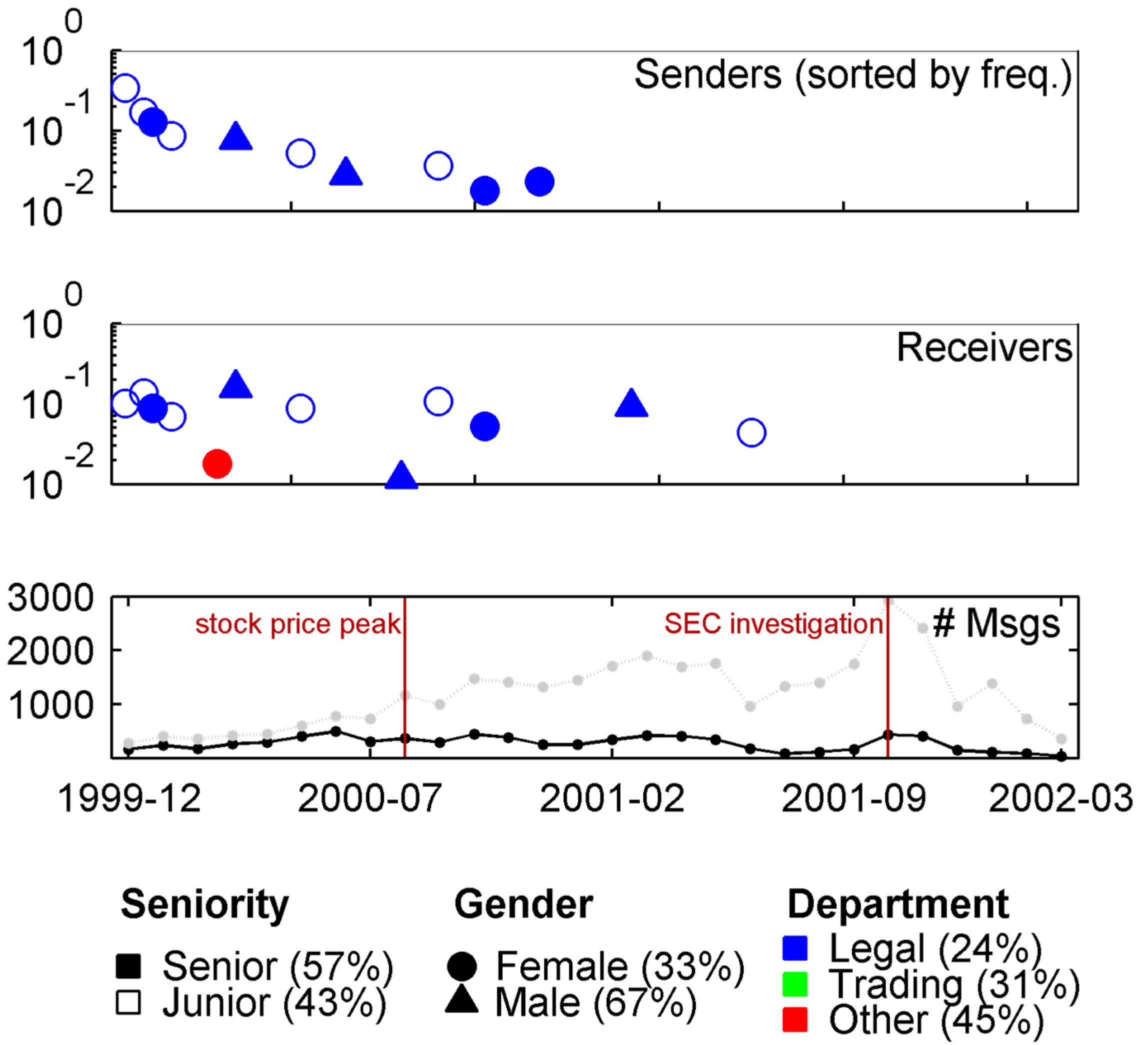
This information is not part of the tensor factorization

Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 1)

Legal Dept;
Mostly Female

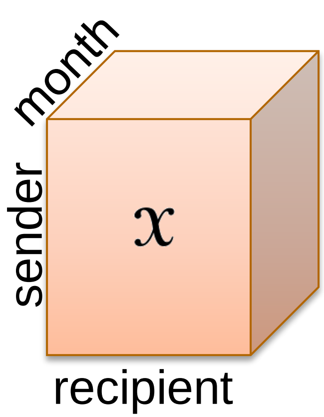


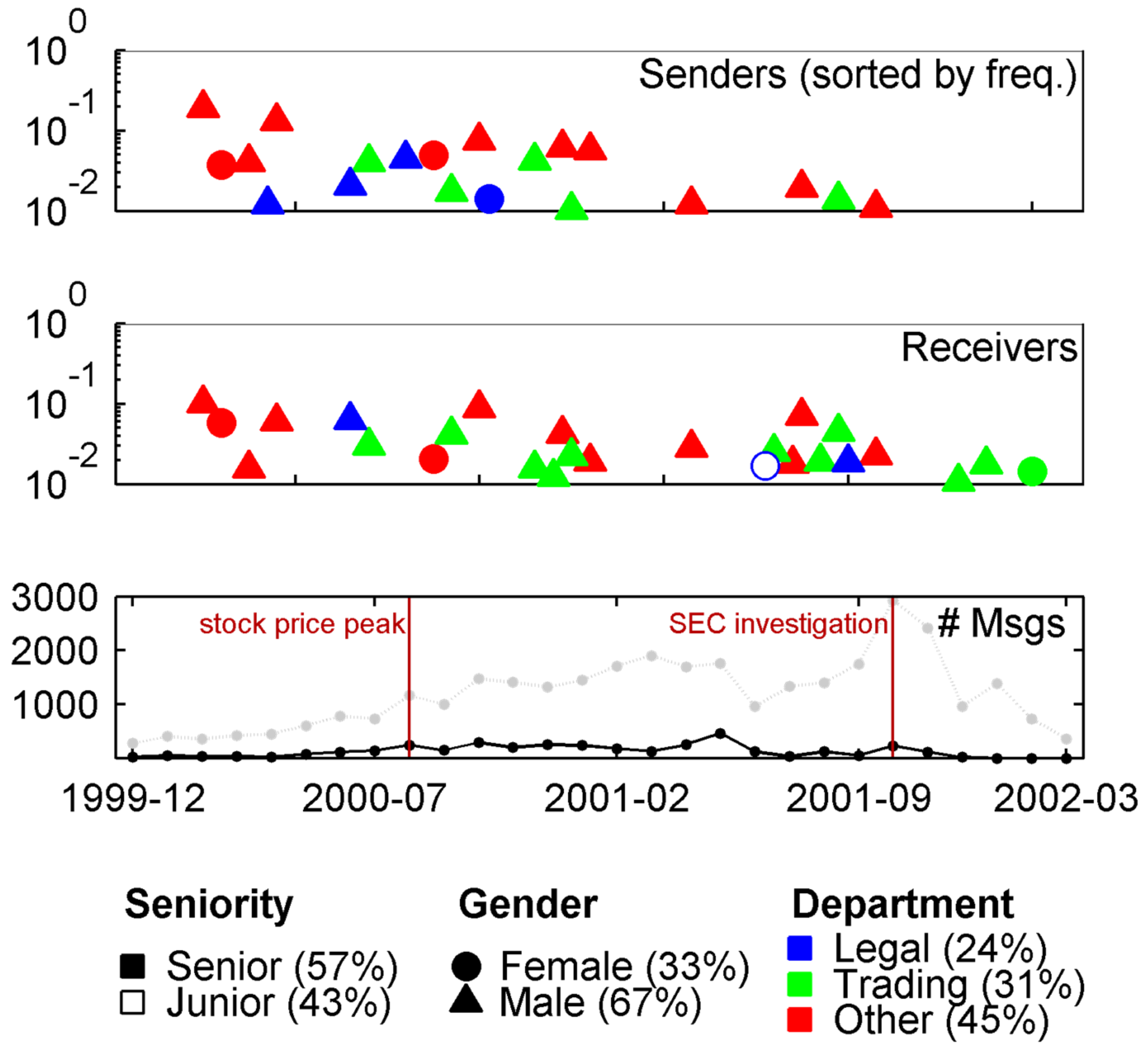


Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 3)

Senior;
Mostly Male

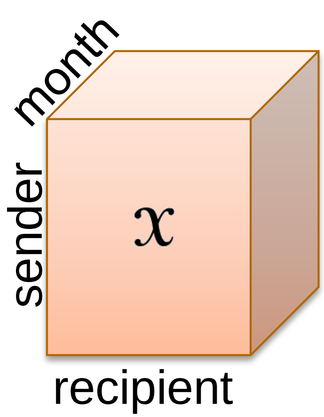


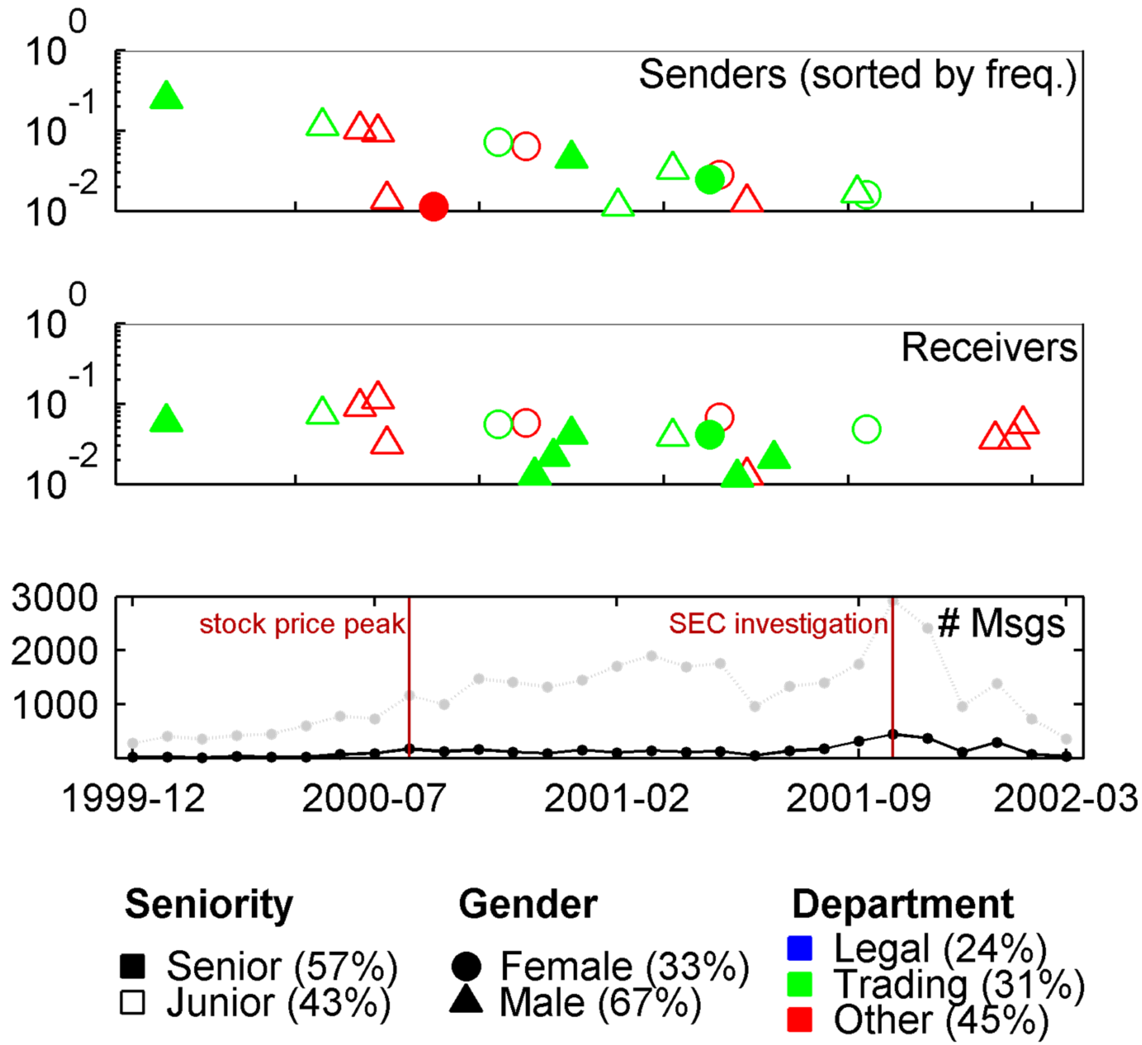


Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 4)

Not Legal

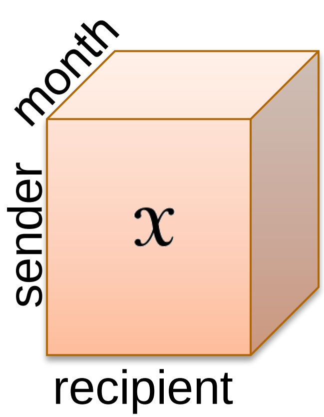


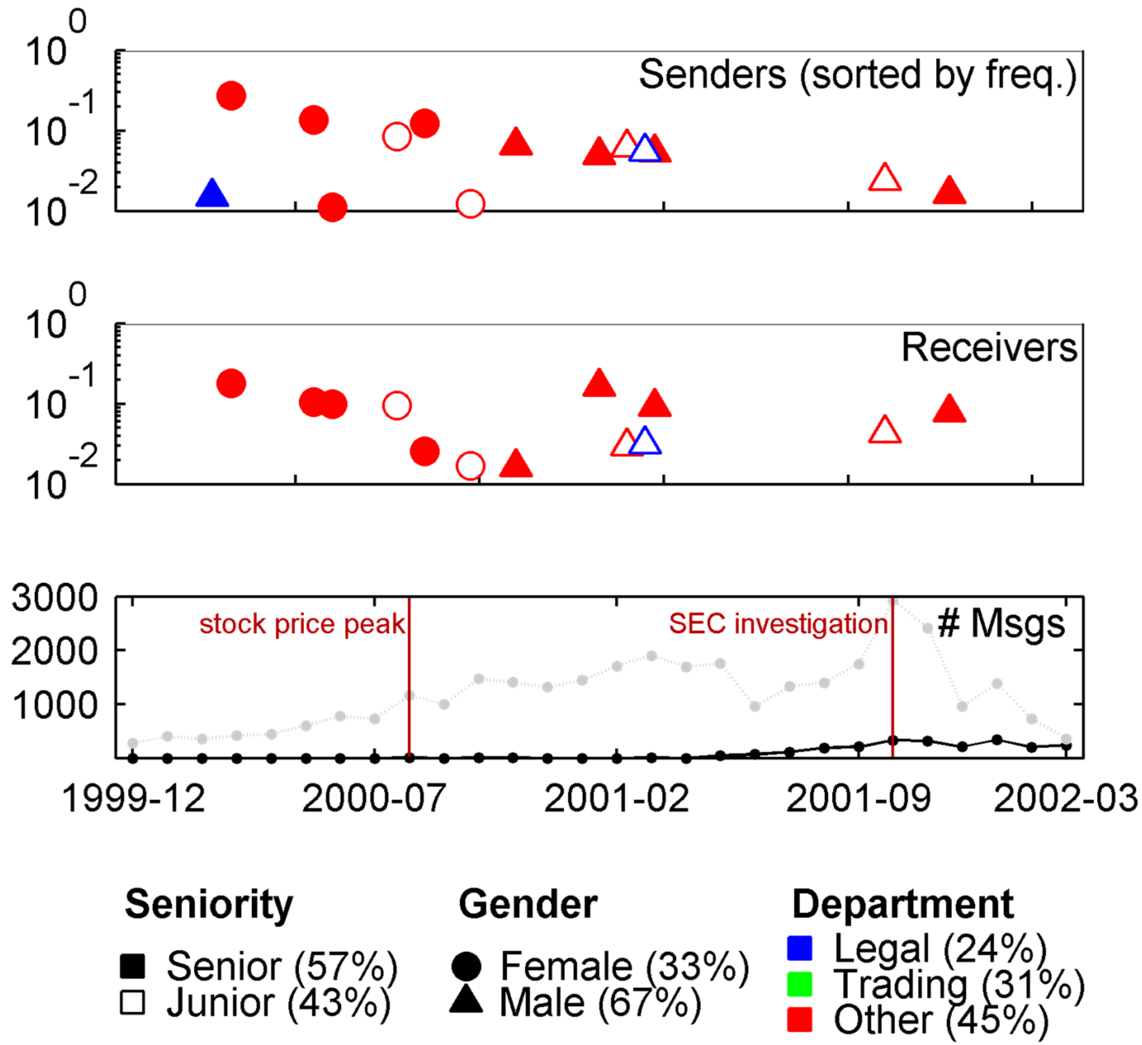


Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 5)

Other;
Mostly Female

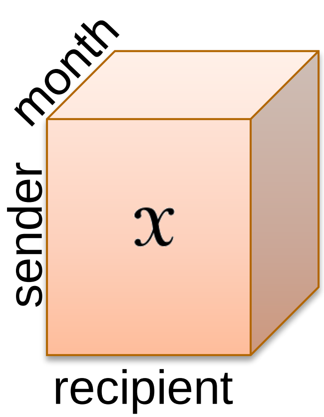


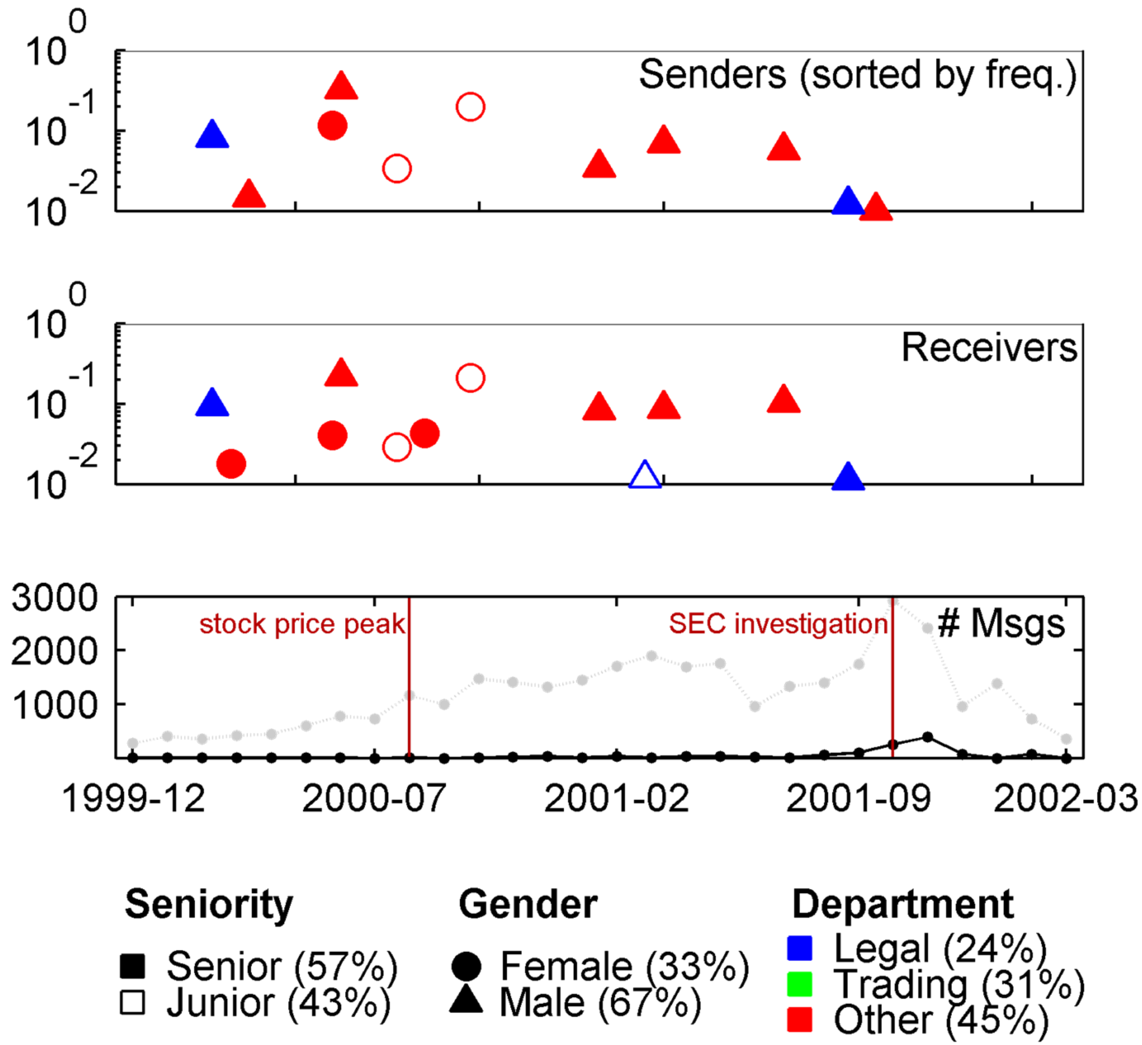


Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 10)

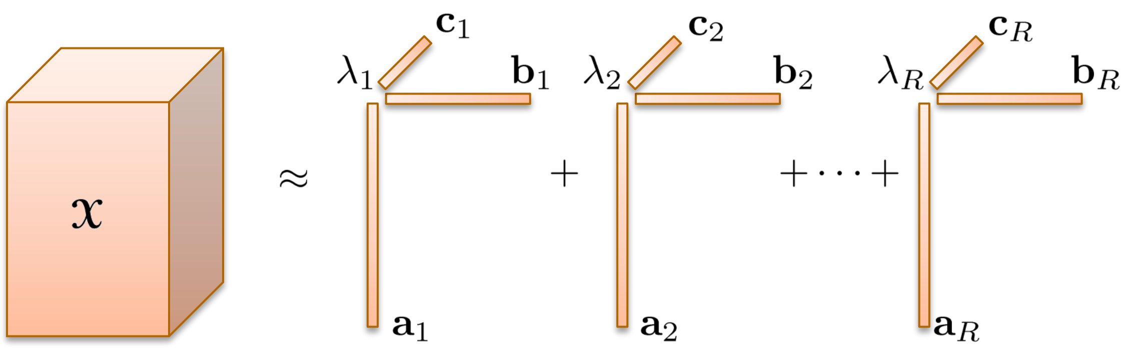
Mostly Other





Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

We define what "≈" means



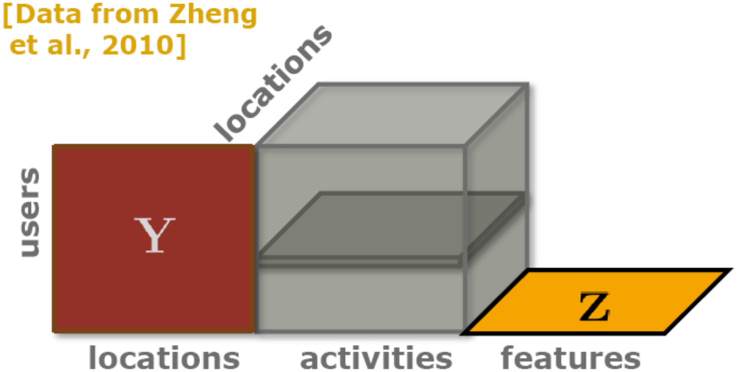
- Least squares
- Nonnegative least squares
- KL divergence
- Sparsity
- Etc.

Coupled Factorizations (Slide from Acar)

Cold-start problem in Link Prediction

Slide from Evrim Acar, TRICAP 2012, Belgium

[Data from Zheng et al., 2010]



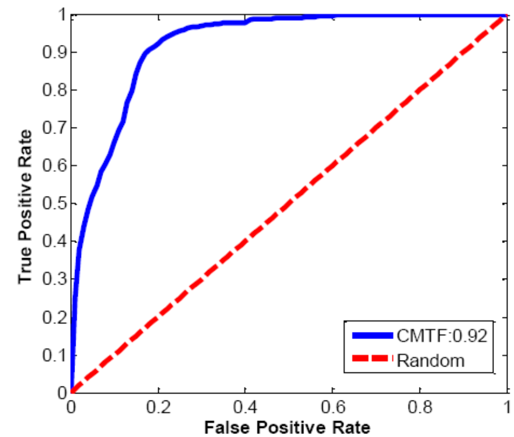
We face with the cold-start problem when a new user starts using an application, e.g., location-activity recommender system. This will correspond to a completely missing slice for the new user.

For the missing slice i (for i=1,2,..I):

Original values	Estimated values using CMTF
$\text{vec}(\mathbf{X}_i)$	$\text{vec}(\hat{\mathbf{X}}_i)$

$$x_{ijk} = \begin{cases} 1 & \text{if user i performs activity j at location k,} \\ 0 & \text{otherwise.} \end{cases}$$

Average ROC curve for I=146 users



We cannot use low-rank approximation of a tensor to fill in the missing slice. However, we can use additional sources of information through the coupled model:

$$\mathrm{Y} \approx \mathrm{A}\mathrm{D}^{\mathsf{T}}$$

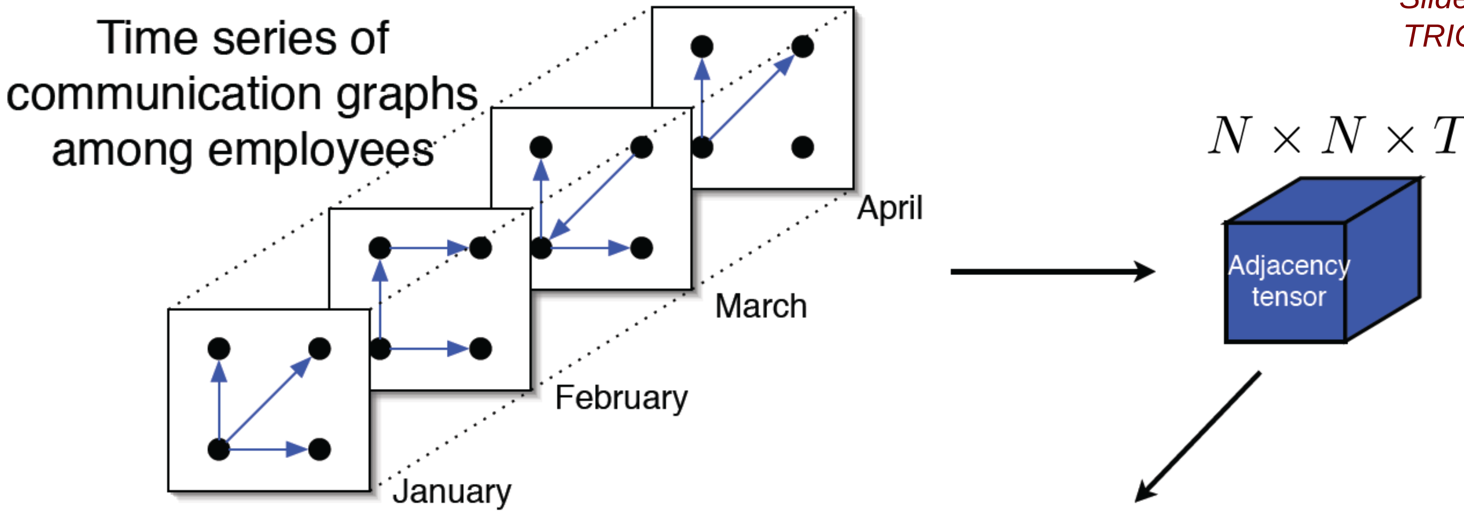
$$\mathcal{X} \approx [\![\mathrm{A}, \mathrm{B}, \mathrm{C}]\!]$$

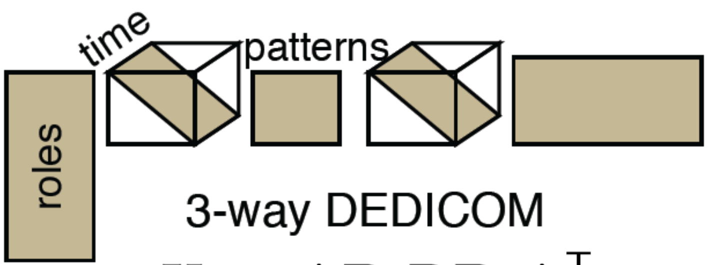
$$\mathrm{Z} \approx \mathrm{C}\mathrm{E}^{\mathsf{T}}$$

Ermis, Acar and Cemgil, ***Link Prediction via Generalized Coupled Tensor Factorisation***, ECML/PKDD 2012

Another model: DEDICOM

Slide from Brett Bader, TRICAP 2006, Greece





3-way DEDICOM

$$\mathbf{X}_t = \mathbf{A}\mathbf{D}_t\mathbf{R}\mathbf{D}_t\mathbf{A}^{\mathsf{T}}$$

$$x_{ijt} = \sum_{k\ell} a_{ik}\, a_{j\ell}\, r_{k\ell}\, d_{kkt}\, d_{\ell\ell t}$$

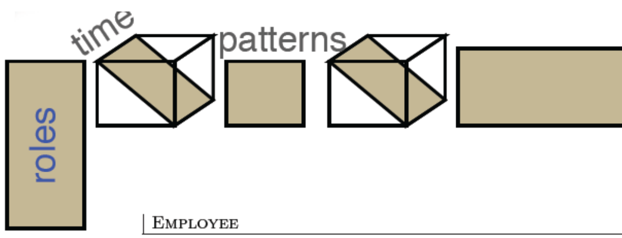
- DEDICOM = DEcomposition into DIrectional COMponents, Harshman (1978)
  - Family of models called PARATUCK2

- a_{ik} = strength of person i in group k
- r_{kl} = interaction of groups k & l
- d_{kkt} = stretch of group k at time t

Bader, Harshman and Kolda. ***Temporal Analysis of Semantic Graphs using ASALSAN***, ICDM 2007, pp. 33-42, 2007

DEDICOM Roles

Slide from Brett Bader, TRICAP 2006, Greece



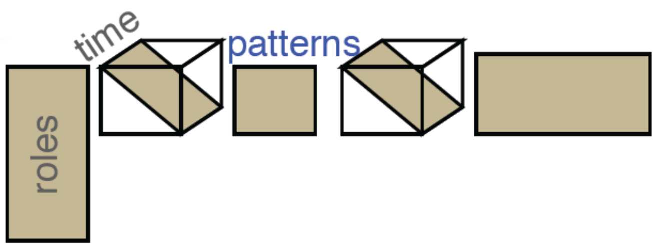
Role	Employee	1 (Legal)	2 (Gov't affairs)	3 (Trade execs)	4 (Pipeline)
Legal	T. Jones - Employee, Financial Trading Group (ENA Legal)	**0.64**	-0.01	0.02	-0.00
	S. Shackleton - Employee, ENA Legal	**0.45**	-0.00	-0.01	-0.00
	M. Taylor - Manager, Financial Trading Group ENA Legal	**0.37**	0.01	0.02	-0.00
	S. Bailey - Legal Assistant, ENA Legal	**0.26**	-0.00	-0.01	-0.00
	S. Panus - Senior Legal Specialist, ENA Legal	**0.26**	-0.00	-0.00	-0.00
	M. Heard - Senior Legal Specialist, ENA Legal	**0.23**	-0.00	0.00	-0.00
	J. Hodge - Asst General Counsel, ENA Legal	**0.13**	0.03	0.01	-0.00
	L. Kitchen - President, Enron Online	**0.11**	**-0.09**	**0.53**	0.00
	S. Dickson - Employee, ENA Legal	**0.09**	-0.00	0.00	-0.00
	E. Sager - VP and Asst Legal Counsel, ENA Legal	**0.08**	0.02	**0.07**	-0.00
Gov't affairs	J. Dasovich - Employee, Government Relationship Executive	-0.01	**0.58**	0.06	0.01
	J. Steffes - VP, Government Affairs	0.00	**0.53**	**-0.06**	-0.01
	R. Shapiro - VP, Regulatory Affairs	-0.00	**0.40**	**0.10**	-0.00
	S. Kean - VP, Chief of Staff	-0.00	**0.37**	-0.04	-0.00
	R. Sanders - VP, Enron Wholesale Services	0.03	**0.16**	-0.01	-0.00
	D. Delainey - CEO, ENA and Enron Energy Services	0.01	**0.09**	**0.09**	-0.00
	S. Corman - VP, Regulatory Affairs	-0.00	**0.08**	-0.00	**0.20**
	M. Carson - Employee, Corporate and Environmental Policy	-0.00	**0.08**	-0.02	-0.00
	S. Scott - Employee, Transwestern Pipeline Company (ETS)	-0.00	**0.08**	-0.00	0.04
Execs - trading	J. Lavorato - CEO, Enron America	0.02	-0.04	**0.49**	0.00
	M. Grigsby - Director, West Desk Gas Trading	0.00	-0.03	**0.20**	-0.00
	G. Whalley - President,	0.01	-0.01	**0.19**	0.00
	J. Steffes - VP, Government Affairs	0.00	-0.02	**0.18**	0.00
	K. Presto - VP, East Power Trading	0.01	-0.05	**0.18**	0.00
	S. Beck - COO,	0.01	-0.03	**0.17**	0.00
	B. Tycholiz - VP, Marketing	0.01	-0.02	**0.16**	0.00
	J. Arnold - VP, Financial Enron Online	0.03	-0.04	**0.16**	-0.00
	J. Williamson - Executive Assistant,	0.00	-0.02	**0.14**	0.01
Pipeline employees	K. Watson - Employee, Transwestern Pipeline Company (ETS)	-0.00	-0.00	0.01	**0.59**
	M. Lokay - Admin. Asst., Transwestern Pipeline Company (ETS)	-0.00	0.01	0.01	**0.42**
	L. Donoho - Employee, Transwestern Pipeline Company (ETS)	-0.00	0.01	0.01	**0.35**
	M. McConnell - Employee, Transwestern Pipeline Company (ETS)	0.00	-0.00	0.01	**0.26**
	L. Blair - Employee, Northern Natural Gas Pipeline (ETS)	-0.00	0.00	0.00	**0.22**
	K. Hyatt - Director, Asset Development TW Pipeline Business (ETS)	-0.00	0.01	0.00	**0.20**
	D. Schoolcraft - Employee, Gas Control (ETS)	-0.00	0.00	0.00	**0.18**
	T. Geaccone - Manager, (ETS)	0.00	-0.00	0.01	**0.17**
	R. Hayslett - VP, Also CFO and Treasurer	0.00	-0.00	0.02	**0.16**

c

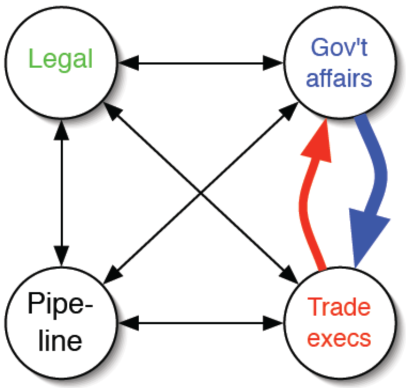
Bader, Harshman and Kolda. ***Temporal Analysis of Semantic Graphs using ASALSAN***, ICDM 2007, pp. 33-42, 2007

DEDICOM Patterns

Slide from Brett Bader, TRICAP 2006, Greece



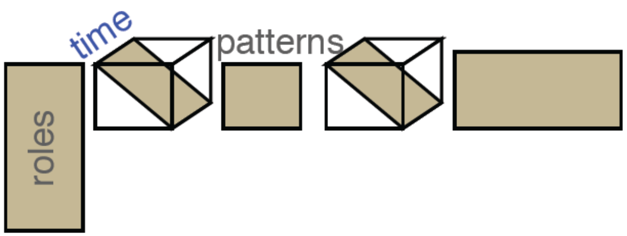
	Legal	Gov't affairs	Trade execs	Pipeline
Legal	440.2	1.6	-15.0	0.4
Government & regulatory affairs	1.6	278.3	135.4	1.6
Trade executives	-29.3	70.7	201.6	-6.2
Pipeline employees	1.4	-4.6	-7.5	172.3

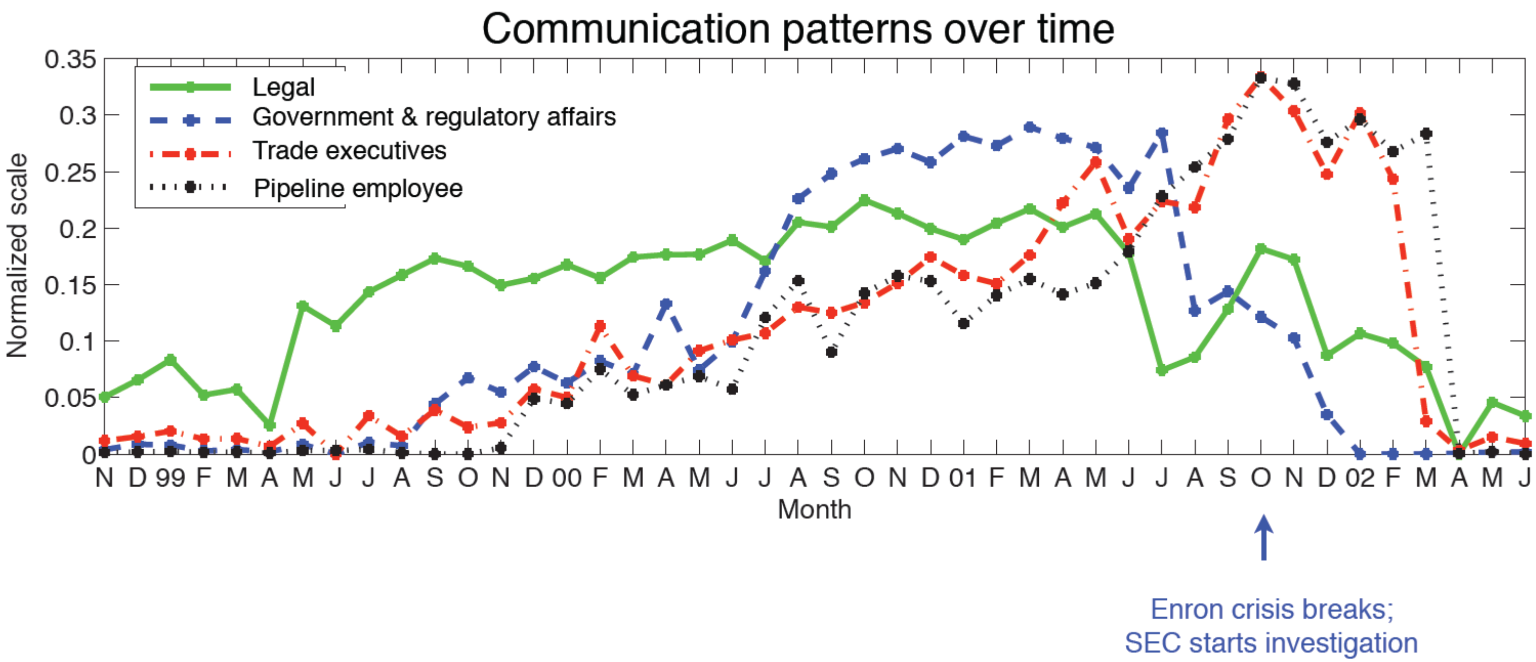


Bader, Harshman and Kolda. ***Temporal Analysis of Semantic Graphs using ASALSAN***, ICDM 2007, pp. 33-42, 2007

DEDICOM Time Profiles

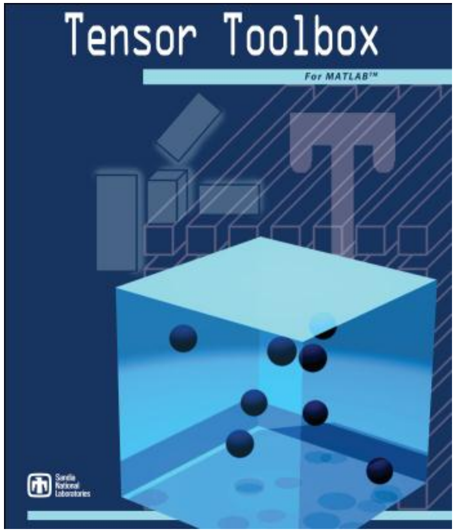
Slide from Brett Bader, TRICAP 2006, Greece





Bader, Harshman and Kolda. ***Temporal Analysis of Semantic Graphs using ASALSAN***, ICDM 2007, pp. 33-42, 2007

Sparse Tensor Computations

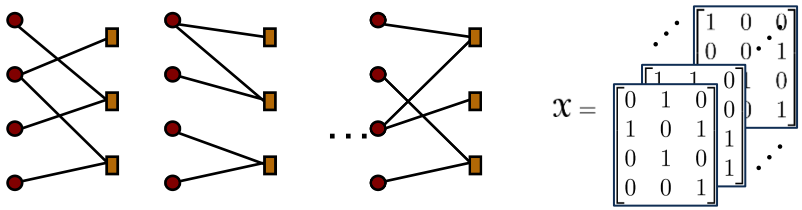


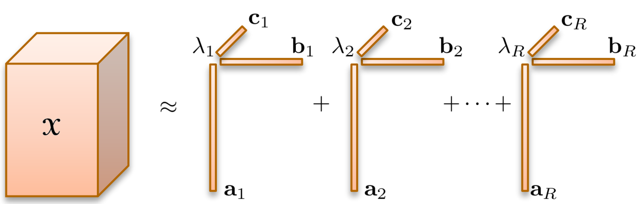
Tensor Toolbox for MATLAB
Bader & Kolda
plus
Acar, Dunlavy, Sun, et al.

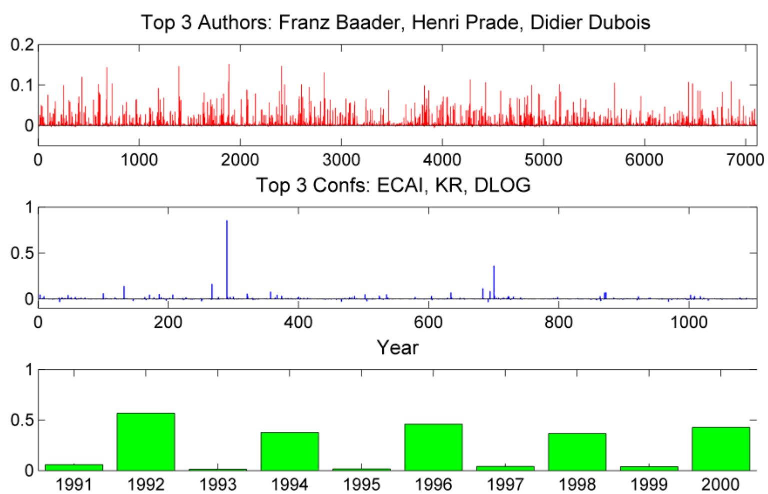
- Many real-world data analysis problems are naturally expressed as in terms of a *sparse* tensor
  - Computer traffic analysis
  - Author-keyword analysis
  - Email analysis
  - Link prediction
  - Web page analysis

- Tensor Toolbox has 5000+ users
  - Main feature is support for sparse tensors

Benefits & Shortcomings of Tensor Analysis for Complex Networks







- What Tensors Do
  - Find clique-like structure in data (similar to matrix factorization)
  - Capture temporal differences, since data is not merged
- Shortcomings
  - Picking the rank is more art than science
  - Time is just another dimension – no special treatment
- Benefits
  - Uniqueness of factorizations under mild conditions $\Rightarrow$ Interpretable results
  - “Natural” nonnegativity
  - Constraints on the factors can impose sparsity, smoothness, etc.
- Other issues
  - Partial symmetries
  - PageRank for tensors is not yet defined
  - Nothing like the Gleich eigenvalue work

For more info: Tammy Kolda
tgkolda@sandia.gov

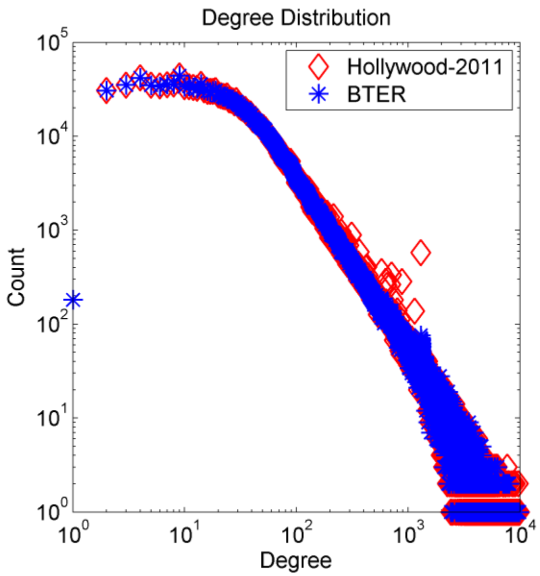
Other Work in Network Analysis

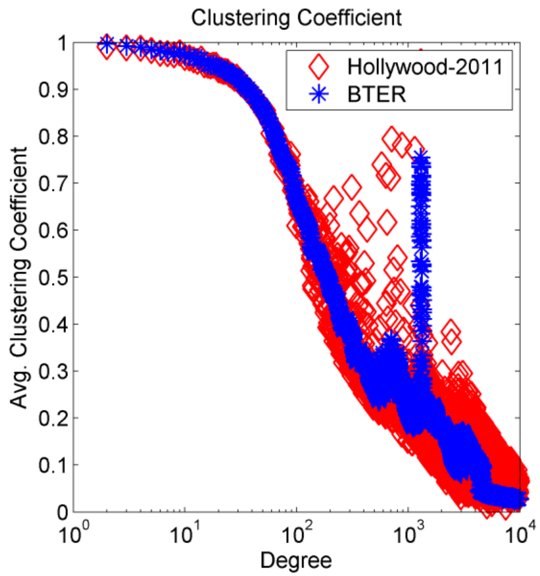
- Realistic models of large-scale networks
  - Match degree distribution
  - Match clustering coefficient (CC)
- Our model = Block Two-level Erdos-Renyi (BTER)

U.S. Department of Defense
Defense Advanced Research Projects Agency

Hollywood 2011 (sym):
2M nodes, 114M edges
(downloaded from LAW)

Total Run Time
BTER = 55 sec
CC via Sampling = 8 min (x2)
32 node MapReduce cluster





References

- Tensors & Networks
  - Acar, Kolda and Dunlavy. ***All-at-once Optimization for Coupled Matrix and Tensor Factorizations***, MLG'11: Proc. Mining and Learning with Graphs, 2011
  - Kolda, Bader and Kenny. ***Higher-Order Web Link Analysis Using Multilinear Algebra***, ICDM 2005, pp. 242-249, 2005 (doi:10.1109/ICDM.2005.77)
  - Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414
  - Dunlavy, Kolda and Acar. ***Temporal Link Prediction using Matrix and Tensor Factorizations***, ACM Trans. KDD 5(2), 2011 (doi:10.1145/1921632.1921636)
  - (*) Coupled Factorizations: Ermis, Acar and Cemgil, ***Link Prediction via Generalized Coupled Tensor Factorisation***, ECML/PKDD 2012
- General
  - Kolda and Bader. ***Tensor Decompositions and Applications***, SIAM Review 51(3):455-500, Sep 2009. (doi:10.1137/07070111X)
  - Bader and Kolda. ***Efficient MATLAB computations with sparse and factored tensors***, SIAM J. Scientific Computing 30(1), 2007 (doi:10.1137/060676489)
- Other Work
  - DEDICOM: Bader, Harshman and Kolda. ***Temporal Analysis of Semantic Graphs using ASALSAN***, ICDM 2007, pp. 33-42, 2007 (doi:10.1109/ICDM.2007.54)
  - (*) Tucker: Sun, Tao, Faloutsos, ***Beyond Streams and Graphs: Dynamic Tensor Analysis***, KDD’06, pp. 374-383, 2006 (doi:10.1145/1150402.1150445)

All available on my web page except those marked with asterisks.