Improvements to 3-way DEDICOM for Applications in Social Network Analysis
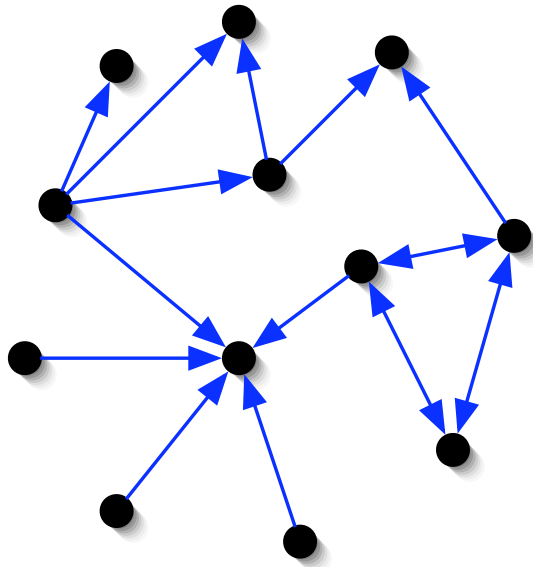
Brett Bader*, Richard Harshman** & Tamara Kolda*
*Sandia National Laboratories
**University of Western Ontario
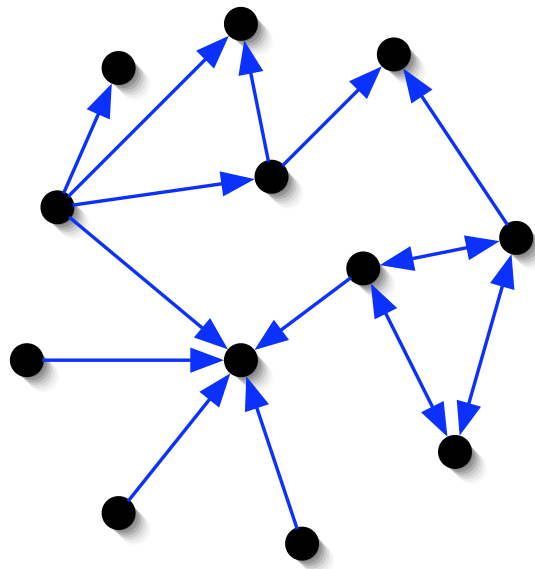
TRICAP
June 5, 2006

Graph Analysis

- Graphs or networks
- Nodes / vertices = objects
  - people, places, things
- Edges / links = relationships

- Examples
  - Citations in bibliometrics
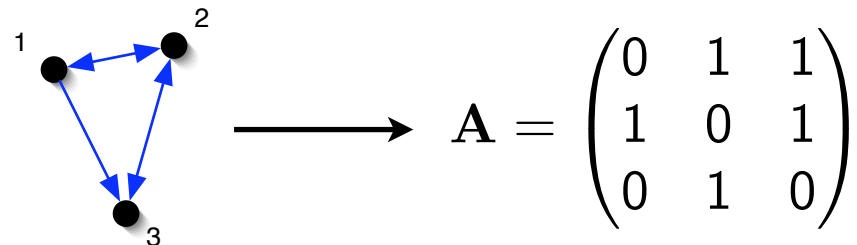 - WWW
  - Communications

Adjacency Matrix

Adjacency matrix

$$a_{ij} = \begin{cases} 1 & \text{if edge } (i,j) \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$
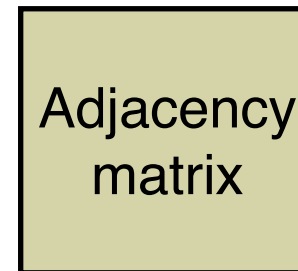
For example:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$\mathbf{A}$ is asymmetric for a directed graph
Could also have weighted adjacency matrix

Common Graph Analysis Technique

For example:

Search on the WWW: HITS (Kleinberg, 1998)

Adjacency matrix
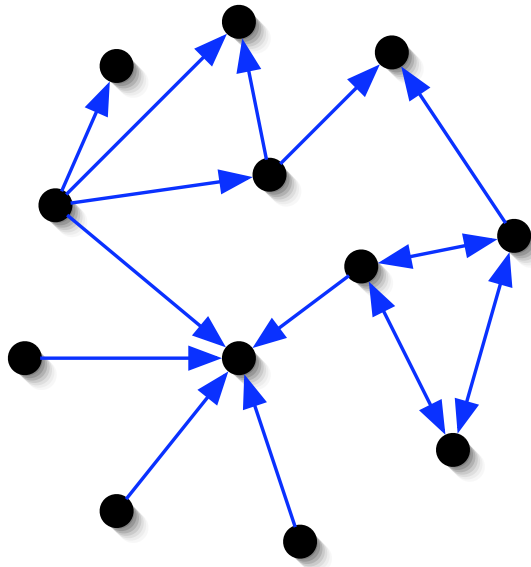
Truncated SVD
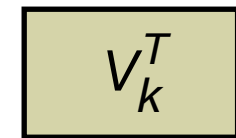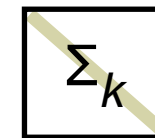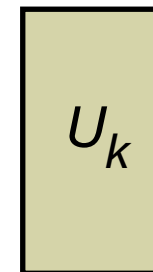
Best rank-*k* matrix filters out noise and captures "latent" information, which improves certain data mining tasks

$$A_k = U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

U_k Σ_k V_k^T

But we may have lost critical information by ignoring edge metadata!

Sandia National Laboratories

Semantic Graphs
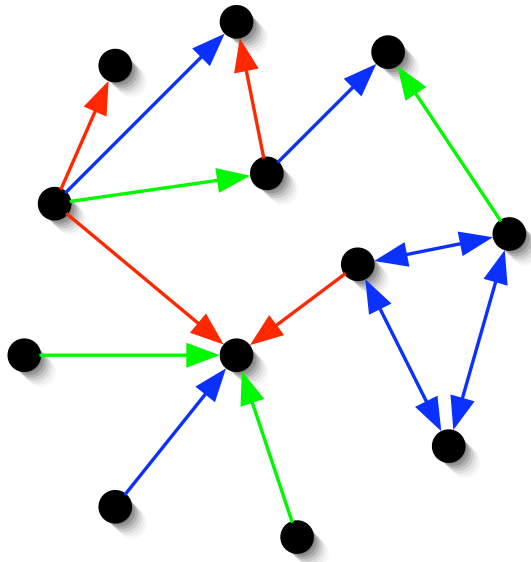
- Different types of edges

- Attribute or relationship labels on edges

- Examples
  - WWW (anchor text)
  - Email communications (time stamp, to/cc)

New Paradigm: "Multidimensional Data Mining"

Build an "adjacency tensor" such that there is an adjacency matrix for each type of edge.

Adjacency matrix

Adjacency tensor

Multilinear algebra

DEDICOM

Tucker

PARAFAC

$+$ $+ ...$
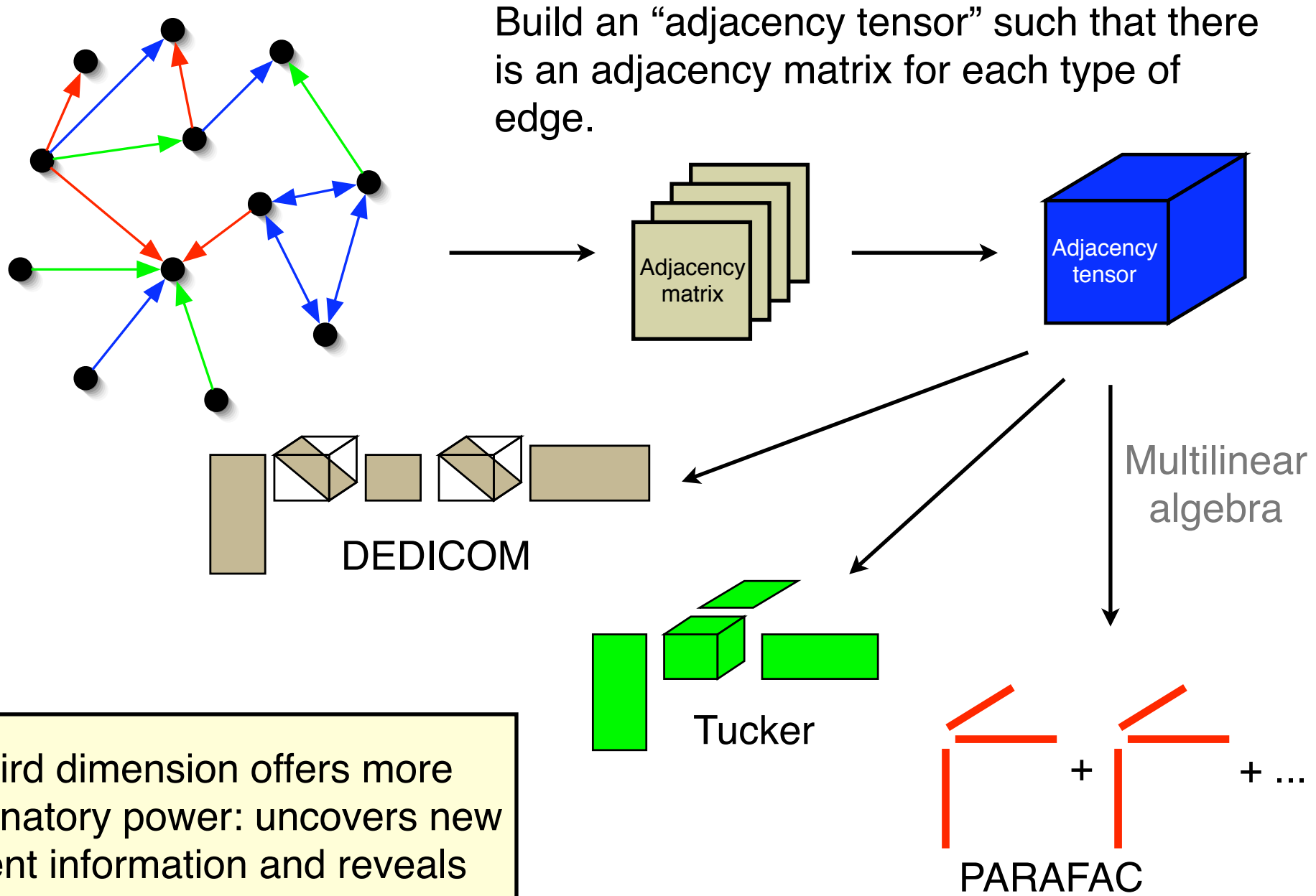
Third dimension offers more explanatory power: uncovers new latent information and reveals subtle relationships

Sandia National Laboratories

Using Tensors for Graph Analysis

Use 3-way DEDICOM to analyze complex social networks that change over time

3-way DEDICOM
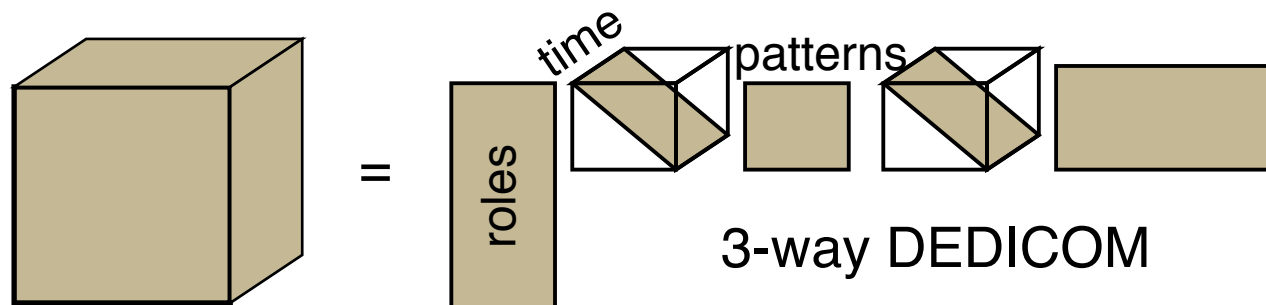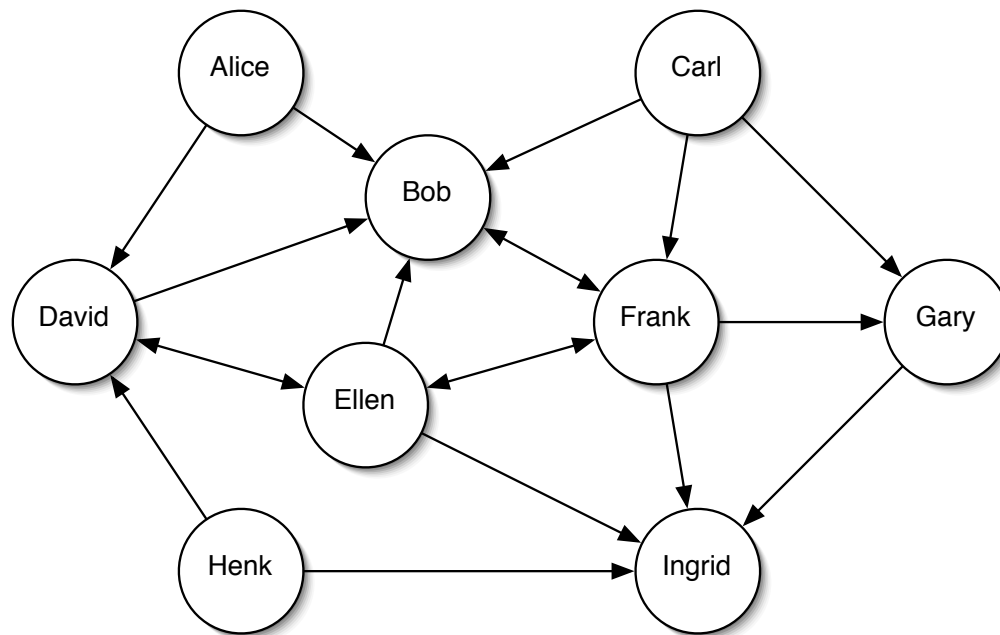
DEDICOM

- DEcomposition into DIrectional COMponents
- Introduced in 1978 by Harshman
- Part of the family of models called PARATUCK2
- Past applications
  - Study asymmetries in telephone calls among cities
  - Marketing research
    - car "switching" - car owners and what they buy next
    - free associations of words (e.g., shampoo: "body" evokes "fullness")
  - Asymmetric measures of world trade (import/export)
- Variations
  - Three-way DEDICOM
  - Constrained DEDICOM
  - Skew-symmetric data

DEDICOM Algorithms

$$X = A \quad R \quad A^T$$

$$X = A \quad R \quad A^T$$

Solve by "alternating" algorithms
- Generalized Takane method
- Kiers' method
- New algorithm

Mathematical Notation

- Scalars a
- Vectors $\mathbf{a}$
- Matrices $\mathbf{A}$
- Tensors (3-way array) $\mathcal{D}$ $\mathcal{X}$
  - frontal slices of $\mathcal{X}$: $\mathbf{X}_i$

- Special symbols
  - Kronecker product

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \ldots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \ldots & a_{mn}\mathbf{B} \end{bmatrix}$$

  - Hadamard product (elementwise)

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & \ldots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & \ldots & a_{mn}b_{mn} \end{bmatrix}$$

Sandia National Laboratories

DEDICOM
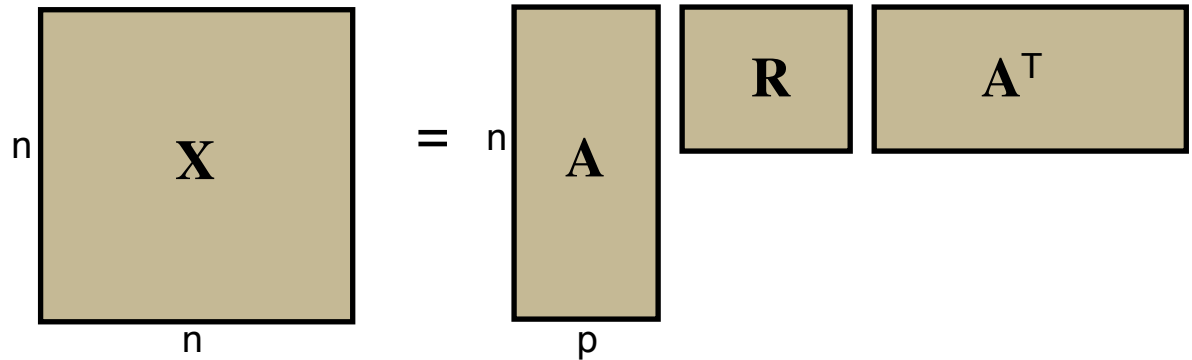
$$X = ARA^T + E$$

$$X \approx ARA^T$$

$$\min_{A,R} \left\| X - ARA^T \right\|_F^2$$

s.t. A orthogonal

- A (n x p) is a matrix of loadings or weights
- R (p x p) is a matrix that captures asymmetric relationships
- A is not unique
  - A can be transformed with no loss of fit to the data
  - Nonsingular transformation Q:

$$ARA^T = (AQ)(Q^{-1}RQ^{-T})(AQ)^T$$

  - Usually "fix" A with some standard rotation (e.g., VARIMAX)

Sandia National Laboratories

Generalized Takane Method

$$\mathbf{X} \approx \mathbf{ARA}^T$$
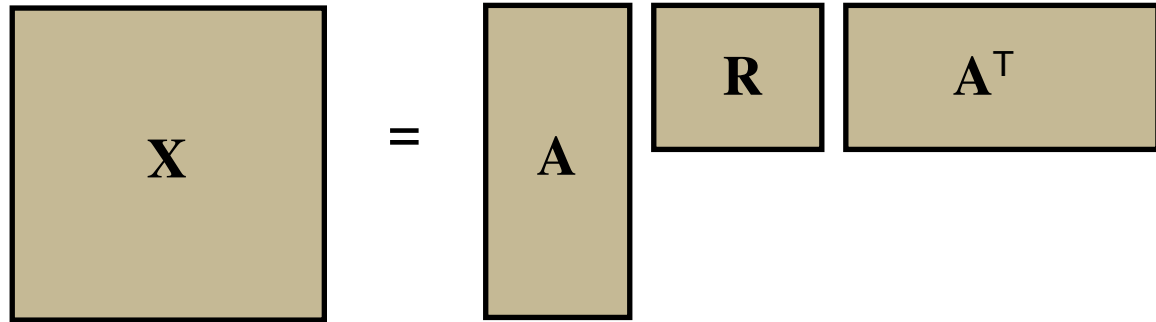
$$\min_{\mathbf{A},\mathbf{R}} \left\| \mathbf{X} - \mathbf{ARA}^T \right\|_F^2$$

$$\mathbf{X} = \mathbf{A} \ \mathbf{R} \ \mathbf{A}^T$$

Loss function: $\sigma(\mathbf{A}, \mathbf{R}) = \left\| \mathbf{X} - \mathbf{ARA}^T \right\|_F^2$

Solving for $\mathbf{A}$:

Minimizing σ wrt $\mathbf{A}$, for fixed $\mathbf{R}$,
is equivalent to maximizing: $\quad f(\mathbf{A}) = tr(\mathbf{A}^T \mathbf{XAA}^T \mathbf{X}^T \mathbf{A})$

Compute $\mathbf{A}$ via Gram-Schmidt
orthonormalization: $\quad (\mathbf{XAA}^T \mathbf{X}^T \mathbf{A} + \mathbf{X}^T \mathbf{AA}^T \mathbf{XA} + 2\alpha \mathbf{A})$
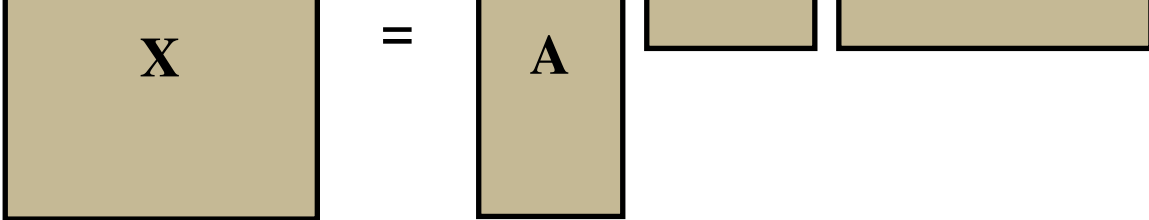
where α is greater than largest eigenvalue of
symmetric part of $(-\mathbf{X} \otimes \mathbf{A}^T \mathbf{XA})$

Practical method:
- Compute an update for $\mathbf{A}$ using $\alpha = 0$
- check if $f(\mathbf{Anew}) > f(\mathbf{A})$
- If not, compute $\mathbf{A}$ using nonzero α

Sandia
National
Laboratories

Generalized Takane Method

$$\mathbf{X} \approx \mathbf{A}\mathbf{R}\mathbf{A}^T$$

$$\min_{\mathbf{A},\mathbf{R}} \left\| \mathbf{X} - \mathbf{A}\mathbf{R}\mathbf{A}^T \right\|_F^2$$

Loss function: $\sigma(\mathbf{A},\mathbf{R}) = \left\| \mathbf{X} - \mathbf{A}\mathbf{R}\mathbf{A}^T \right\|_F^2$

Solving for $\mathbf{R}$:

Least squares update:

$$\mathbf{R}_{new} = \mathbf{A}^\dagger \mathbf{X}(\mathbf{A}^T)^\dagger$$

But because $\mathbf{A}$ is orthonormal...

$$\mathbf{R}_{new} = \mathbf{A}^T \mathbf{X} \mathbf{A}$$

New Algorithm

Solving for $\mathbf{A}$:

Stack data and model "side by side" in a single equation

$$\begin{pmatrix} \mathbf{X} & \mathbf{X}^T \end{pmatrix} = \begin{pmatrix} \mathbf{ARA}^T & \mathbf{AR}^T\mathbf{A}^T \end{pmatrix}$$

$$= \mathbf{A}\left(\begin{pmatrix} \mathbf{R} & \mathbf{R}^T \end{pmatrix} \begin{pmatrix} \mathbf{A}^T & 0 \\ 0 & \mathbf{A}^T \end{pmatrix} \right)$$

...and solve extended LS problem, i.e., $\min_{\mathbf{A}} \| \mathbf{Y} - \mathbf{AB} \|_F^2$

$$\mathbf{A}_{new} \leftarrow \begin{pmatrix} \mathbf{X} & \mathbf{X}^T \end{pmatrix} \left(\begin{pmatrix} \mathbf{R} & \mathbf{R}^T \end{pmatrix} \begin{pmatrix} \mathbf{A}^T & 0 \\ 0 & \mathbf{A}^T \end{pmatrix} \right)^{\dagger}$$

or

$$\mathbf{A}_{new} = \begin{pmatrix} \mathbf{XAR}^T + \mathbf{X}^T\mathbf{AR} \end{pmatrix} \begin{pmatrix} \mathbf{R}(\mathbf{A}^T\mathbf{A})\mathbf{R}^T + \mathbf{R}^T(\mathbf{A}^T\mathbf{A})\mathbf{R} \end{pmatrix}^{-1}.$$

Solving for $\mathbf{R}$:

$$\mathbf{R}_{new} = \mathbf{A}^{\dagger}\mathbf{X}(\mathbf{A}^T)^{\dagger}$$

Sandia National Laboratories

Three-way DEDICOM
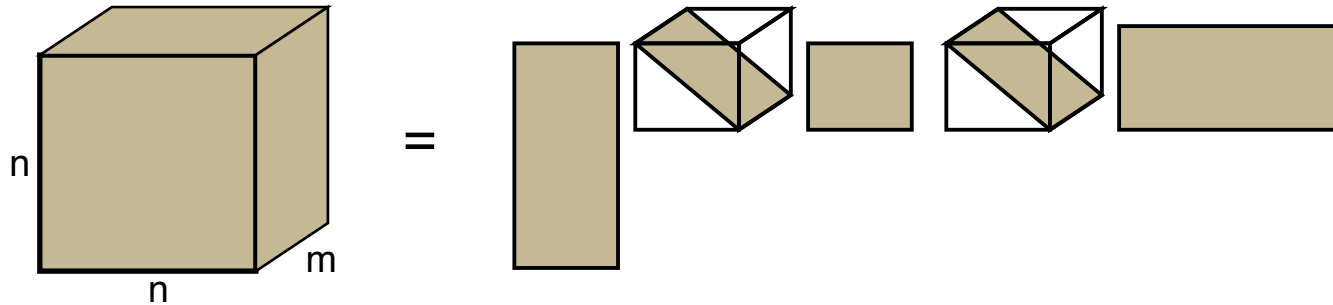
$$\mathbf{X}_i = \mathbf{A}\mathbf{D}_i\mathbf{R}\mathbf{D}_i\mathbf{A}^T + \mathbf{E}_i \quad \text{for } i = 1, \ldots, m,$$

$$\min_{\mathbf{A},\mathbf{R},\mathcal{D}} \sum_{i=1}^{m} \left\| \mathbf{X}_i - \mathbf{A}\mathbf{D}_i\mathbf{R}\mathbf{D}_i\mathbf{A}^T \right\|_F^2$$

- $\mathbf{A}$ (n x p) is a matrix of loadings or weights (not necessarily orthogonal)
- $\mathbf{R}$ (p x p) is a matrix that captures asymmetric relationships
- D (p x p x m) gives the weights of the columns of A for each level in third mode
- Unique solution with enough levels of $\mathbf{X}$

Kiers' Algorithm

$$\min_{\mathbf{A,R,\mathcal{D}}} \sum_{i=1}^{m} \left\| \mathbf{X}_i - \mathbf{A}\mathbf{D}_i\mathbf{R}\mathbf{D}_i\mathbf{A}^T \right\|_F^2$$

Alternating Least Squares

1) Column-wise minimization to find $\mathbf{A}$
   (potentially slow convergence)

2) Least-squares problem for $\mathbf{R}$

$$\text{minimize: } f(\mathbf{R}) = \left\| \begin{pmatrix} \mathsf{Vec}(\mathbf{X}_1) \\ \vdots \\ \mathsf{Vec}(\mathbf{X}_m) \end{pmatrix} - \begin{pmatrix} \mathbf{A}\mathbf{D}_1 \otimes \mathbf{A}\mathbf{D}_1 \\ \vdots \\ \mathbf{A}\mathbf{D}_m \otimes \mathbf{A}\,\mathbf{D}_m \end{pmatrix} \mathsf{Vec}(\mathbf{R}) \right\|$$

$$\mathsf{Vec}(\mathbf{R}) = \left(\sum_{i=1}^{m} (\mathbf{D}_i\mathbf{A}^T\mathbf{A}\mathbf{D}_i) \otimes (\mathbf{D}_i\mathbf{A}^T\mathbf{A}\mathbf{D}_i) \right)^{-1} \sum_{i=1}^{m} \mathsf{Vec}(\mathbf{D}_i\mathbf{A}^T\mathbf{X}_i\mathbf{A}\mathbf{D}_i)$$

3) Element-wise minimization to find $\mathbf{D}$
   (potentially slow convergence)

Sandia
National
Laboratories

Solving for $\mathbf{A}$:

$$\begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_1^T & \cdots & \mathbf{X}_m & \mathbf{X}_m^T \end{pmatrix} = \mathbf{A} \begin{pmatrix} \mathbf{D}_1\mathbf{R}\mathbf{D}_1 & \mathbf{D}_1\mathbf{R}^T\mathbf{D}_1 & \cdots & \mathbf{D}_m\mathbf{R}\mathbf{D}_m & \mathbf{D}_m\mathbf{R}^T\mathbf{D}_m \end{pmatrix} \begin{pmatrix} \mathbf{I}_{2m} \otimes \mathbf{A}^T \end{pmatrix}$$

$$\mathbf{A} = \left[\sum_{i=1}^{m} \left(\mathbf{X}_i\mathbf{A}\mathbf{D}_i\mathbf{R}^T\mathbf{D}_i + \mathbf{X}_i^T\mathbf{A}\mathbf{D}_i\mathbf{R}\mathbf{D}_i \right) \right] \left[\sum_{i=1}^{m} (\mathbf{B}_i + \mathbf{C}_i) \right]^{-1}$$

$$\text{where} \quad \mathbf{B}_i \quad \equiv \quad \mathbf{D}_i\mathbf{R}\mathbf{D}_i(\mathbf{A}^T\mathbf{A})\mathbf{D}_i\mathbf{R}^T\mathbf{D}_i,$$
$$\mathbf{C}_i \quad \equiv \quad \mathbf{D}_i\mathbf{R}^T\mathbf{D}_i(\mathbf{A}^T\mathbf{A})\mathbf{D}_i\mathbf{R}\mathbf{D}_i.$$

Sandia National Laboratories

New Algorithm - updating D

$$\min_{\mathbf{D}_i} \left\| \mathbf{X}_i - \mathbf{A}\mathbf{D}_i\mathbf{R}\mathbf{D}_i\mathbf{A}^T \right\|_F^2$$

Solving for **D**:

Use Newton's method to solve the optimization problem for $d = \text{diag}(\mathbf{D}_i)$

$$d_{new} = d - H^{-1}g$$

$$g_k = -\sum_{i,j} \left[2(\mathbf{X} - \mathbf{A}\mathbf{D}\mathbf{R}\mathbf{D}\mathbf{A}^T) * (\mathbf{A}\mathbf{D}r_k\mathbf{a}_k^T + \mathbf{a}_k r_{k,:}\mathbf{D}\mathbf{A}^T) \right]_{i,j}$$

$$h_{st} = -2\sum_{i,j} \left[(\mathbf{X} - \mathbf{A}\mathbf{D}\mathbf{R}\mathbf{D}\mathbf{A}^T) * (\mathbf{a}_s r_{st}\mathbf{a}_t^T + \mathbf{a}_t r_{ts}\mathbf{a}_s^T) \right.$$

$$\left. - (\mathbf{A}\mathbf{D}r_s\mathbf{a}_s^T + \mathbf{a}_s r_{s:}\mathbf{D}\mathbf{A}^T) * (\mathbf{A}\mathbf{D}r_t\mathbf{a}_t^T + \mathbf{a}_t r_{t:}\mathbf{D}\mathbf{A}^T) \right]_{i,j}$$

Use compression

QR factorization: $\mathbf{A} = \mathbf{Q}\tilde{\mathbf{A}},$

$$\min_{\mathbf{D}_i} \left\| \mathbf{Q}^T\mathbf{X}_i\mathbf{Q} - \tilde{\mathbf{A}}\mathbf{D}_i\mathbf{R}\mathbf{D}_i\tilde{\mathbf{A}}^T \right\|_F^2$$ Smaller problem (p x p)

MATLAB Tensor Toolbox

(Bader and Kolda)

- Toolbox extends functionality of Matlab's MDA datatype:
  - Basic operations
  - Convert to/from a matrix
  - Multiplication
    - Tensor
    - Matrix
    - Vector
- Facilitates rapid prototyping of algorithms
  - PARAFAC/CANDECOMP
  - Tucker
  - DEDICOM
- Extensions for a sparse tensor format (in development)
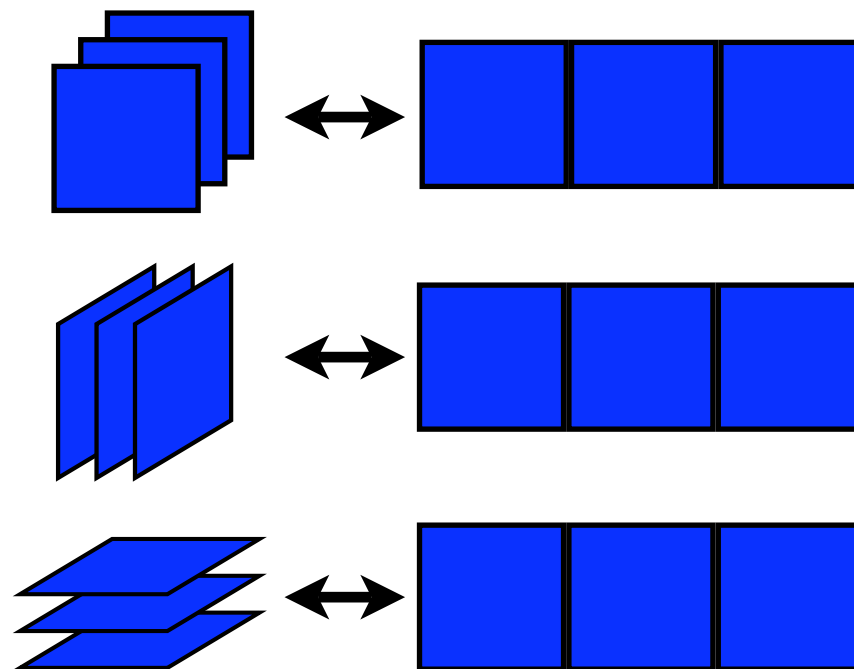
$$\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}$$
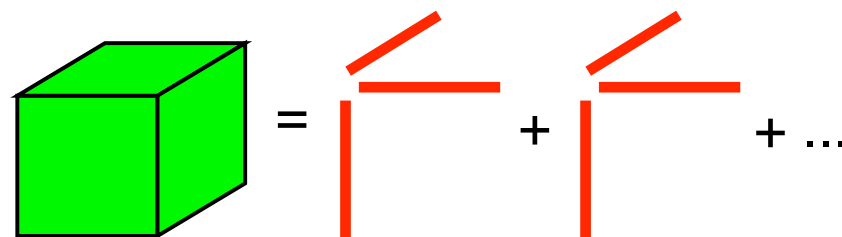
Note: not intended to replace Andersson & Bro's N-way Toolbox

http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox
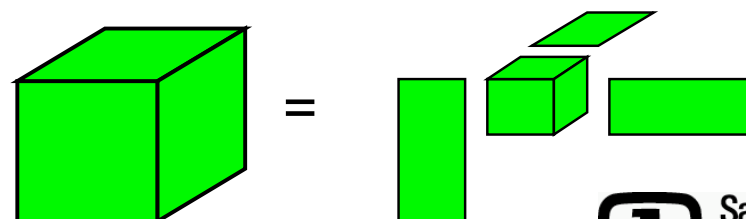
New `sparse_tensor` Class

sparse_tensor class

- Coordinate based storage: (i,j,k) indices & values

- Implements many of the same functions as `tensor` class

- Reshape and permute operations handled implicitly with index

- Row / column / slice operations are easy

- Large problem to date
  - 73,000 x 73,000 x 40,000 sparse tensor
  - 469,000 nonzeros (*very* sparse)
  - Approximate rank-10 PARAFAC model computed in 12 minutes

Sandia National Laboratories

Application: Social Network Analysis
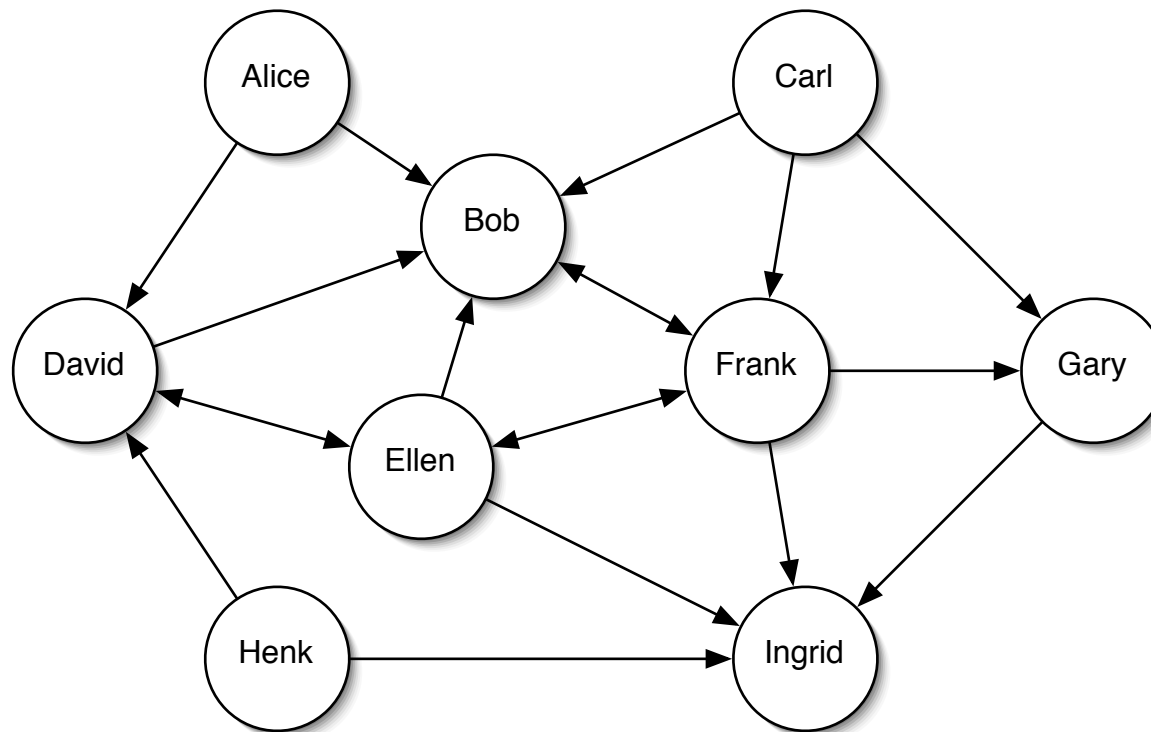
- Links may consist of
  - Relationships (e.g., friends, family, co-workers, co-authors)
  - Communications
- What can we learn about this network strictly from these connections?

Enron Corp.

- U.S. corporation involved with creating energy markets
  - 7th largest by revenue
- EnronOnline: e-trading business
  - natural gas
  - electric power

```
                              ┌──────────────┐
                              │  Enron Corp  │
                              └──────┬───────┘
        ┌───────────┬───────────┬────┴──────┬───────────────┐
   ┌────┴────┐ ┌────┴────┐ ┌────┴────┐ ┌────┴────┐ ┌─────────┴──────┐
   │ Enron   │ │ Enron   │ │ Enron   │ │ Enron   │ │ Enron          │
   │ Networks│ │ North   │ │ Energy  │ │Broadband│ │ Transportation │
   │         │ │ America │ │ Services│ │         │ │ Services       │
   └────┬────┘ └────┬────┘ └─────────┘ └─────────┘ └────────┬───────┘
   ┌────┴────┐ ┌────┴────────┐                       ┌───────┴──────┐
   │EnronOnLine│ │ Enron       │                      │ Enron        │
   └─────────┘ │ Power       │                       │ Pipelines    │
               │ Marketing   │                       └──────────────┘
               └─────────────┘
               ┌─────────────┐
               │ Enron       │
               │ Gas Marketing│
               └─────────────┘
               ┌─────────────┐
               │ Enron       │
               │ Generation  │
               └─────────────┘
```
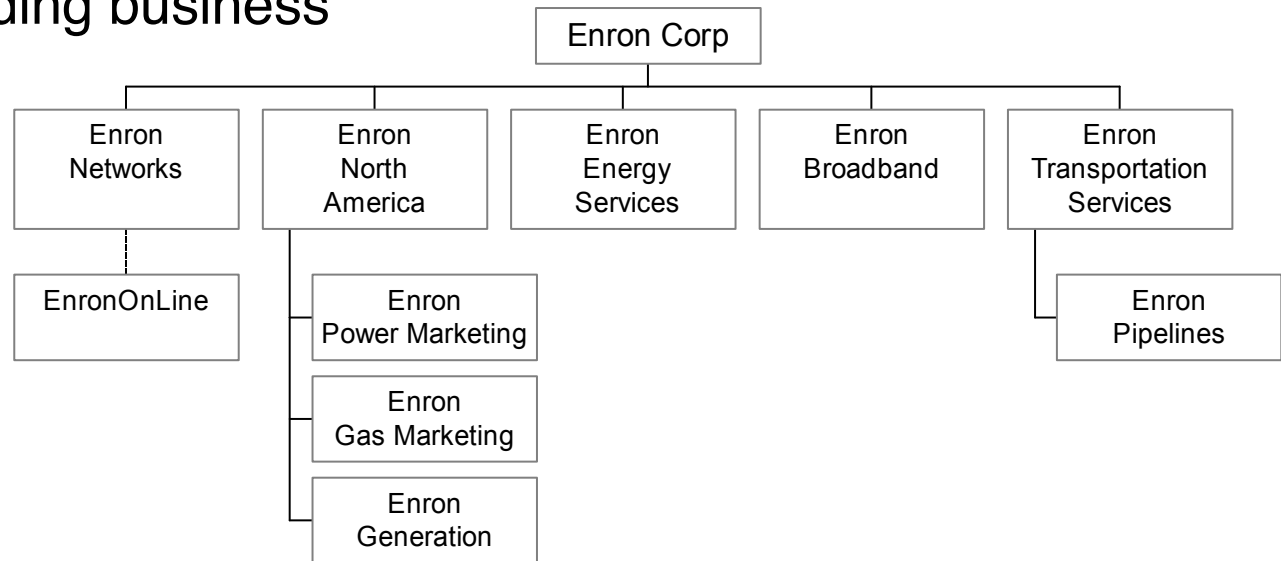
- Investigations
  - U.S. Federal Energy Regulatory Commission (FERC)
    - energy market manipulation
    - involved energy traders
  - U.S. Securities and Exhchange Commission (SEC)
    - accounting fraud
    - insider trading

Enron Email Data

- FERC collected email of ~150 employees as evidence
  - Included emails saved in inbox, sent items, deleted items, and all other folders

- Released to the public in 2002 by FERC as part of their investigation
  - To/from, date, subject, body
  - Attachments and some names/emails removed
  - Approx. 500,000 email messages

- Research uses:
  - Email classification
  - Natural language processing
  - Organizational theory/behavior
  - Social network analysis

Sandia National Laboratories

Smaller Data Set

We used a smaller dataset prepared by Priebe et al.
34,427 emails among 184 employees over 44 months

Email communications at Enron (1998-2002)

Messages / Month
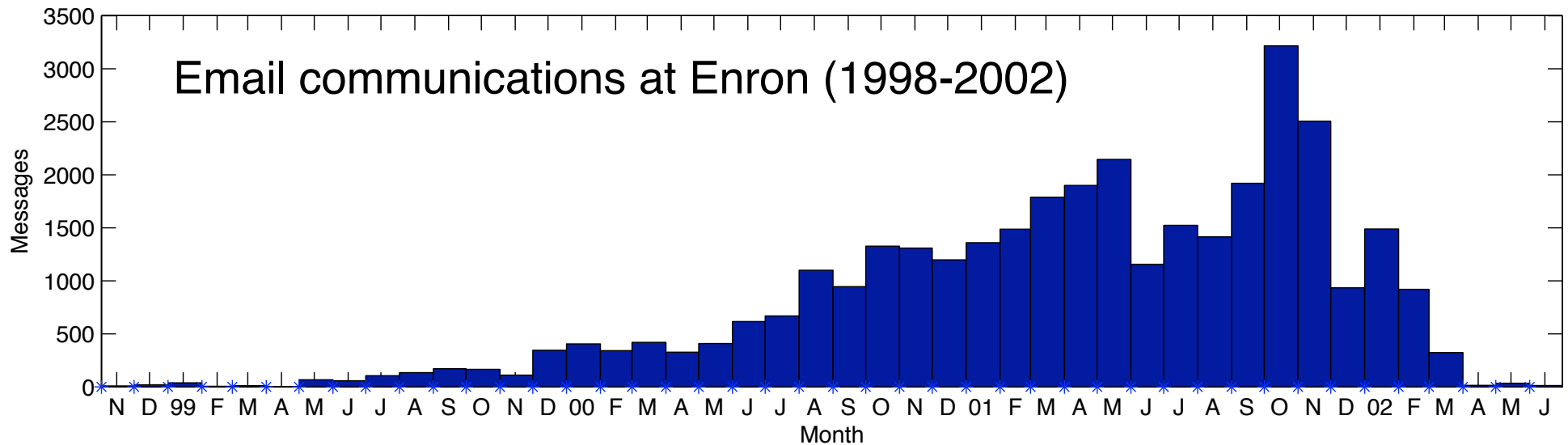
- Email folders collected at one point in time.
- Shape of histogram depends on:
  - How far back employees kept emails
  - Employment history of individual
- Limited information on former employees
  - No org chart

Sandia
National
Laboratories

DEDICOM Experiment

- Time series of communication graphs

- Sparse tensor of size 184 x 184 x 44 (9838 nnz)

- Scaling: x number of messages scaled by $\log(x)+1$

- Models:
  - SVD
  - 2-way DEDICOM
  - 3-way DEDICOM
  - PARAFAC

Social Network Analysis
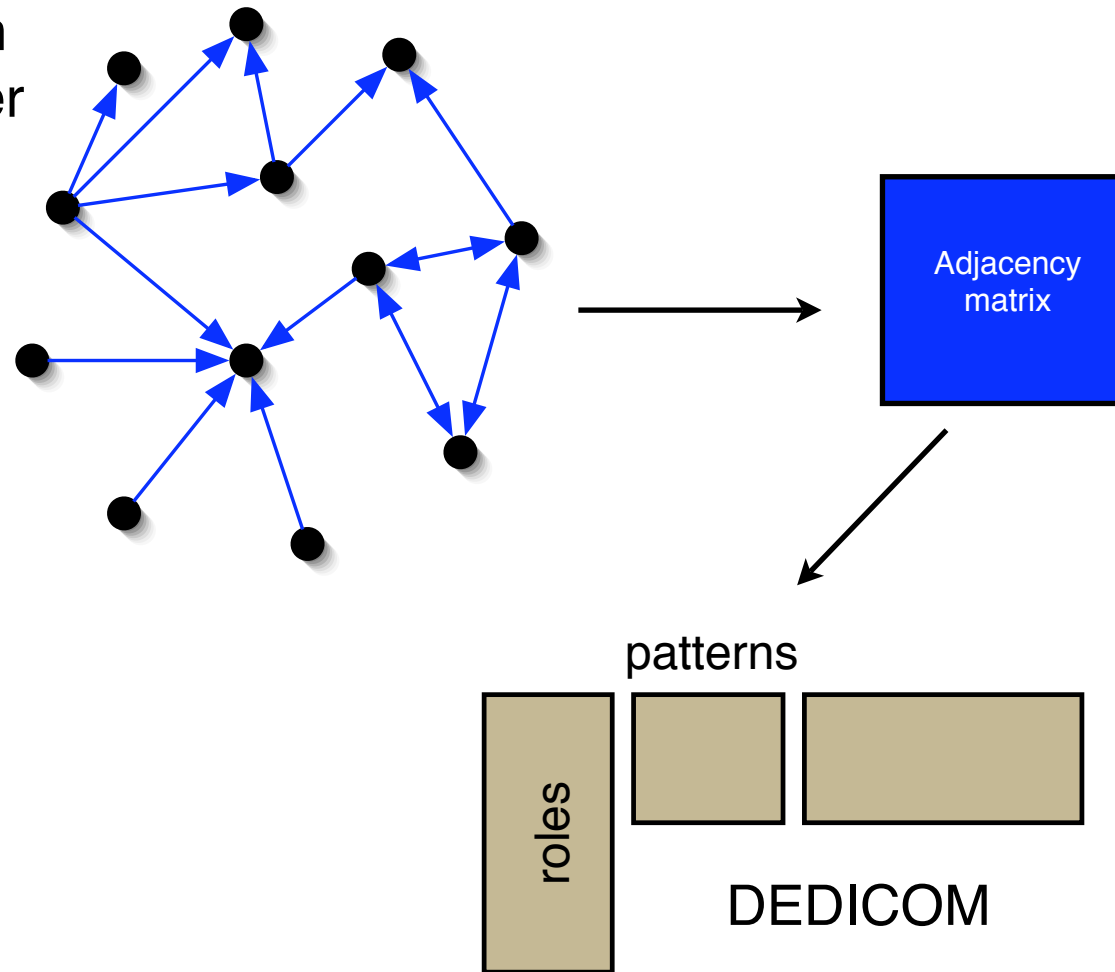
Communication graph among employees over all times

Adjacency matrix

patterns

roles

DEDICOM

- Description of employees by their roles
- Aggregate communication patterns among roles

Sandia National Laboratories

DEDICOM Results

patterns / roles

Identify shared characteristics to label group

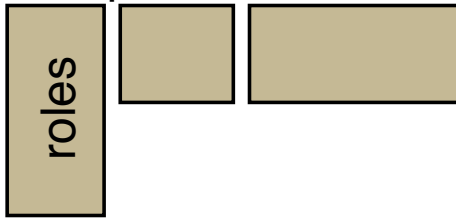| EMPLOYEE	DEDICOM Solution		
	1	2	3
J. Lavorato - CEO, Enron America	**0.41**	0.07	0.04
L. Kitchen - President, Enron Online	**0.26**	**0.21**	0.04
M. Grigsby - Director, West Desk Gas Trading	**0.22**	-0.01	-0.01
D. Delainey - CEO, ENA and Enron Energy Services	**0.20**	0.06	0.06
G. Whalley - President,	**0.17**	0.05	0.04
L. Taylor - Executive Assistant to Greg Whalley,	**0.17**	0.06	0.03
T. Jones - Employee, Financial Trading Group (ENA Legal)	**-0.12**	**0.38**	-0.02
M. Taylor - Manager, Financial Trading Group ENA Legal	-0.10	**0.35**	-0.01
S. Shackleton - Employee, ENA Legal	**-0.13**	**0.31**	-0.02
S. Panus - Senior Legal Specialist, ENA Legal	**-0.11**	**0.26**	-0.02
M. Heard - Senior Legal Specialist, ENA Legal	-0.10	**0.24**	-0.02
E. Sager - VP and Asst Legal Counsel, ENA Legal	-0.01	**0.24**	0.02
S. Corman - VP, Regulatory Affairs	-0.04	-0.01	**0.33**
K. Watson - Employee, Transwestern Pipeline Company (ETS)	-0.08	-0.03	**0.32**
L. Donoho - Employee, Transwestern Pipeline Company (ETS)	-0.08	-0.03	**0.30**
D. Fossum - VP, Transwestern Pipeline Company (ETS)?	-0.06	-0.00	**0.30**
M. Lokay - Admin. Asst., Transwestern Pipeline Company (ETS)	-0.07	-0.02	**0.28**
K. Hyatt - Director, Asset Development TW Pipeline Co. (ETS)	-0.06	-0.02	**0.25**
R. Hayslett - VP, Also CFO and Treasurer	-0.04	-0.01	**0.23**
R matrix	70.3	11.6	6.7
	15.4	68.2	5.0
	9.9	6.7	59.5

Column headers: Executives / Legal / Pipeline

Row groups: Execs / Legal / Pipeline

Executives

Legal

Pipeline employees

Some employees have dual roles

Pattern of communications in **R** matrix
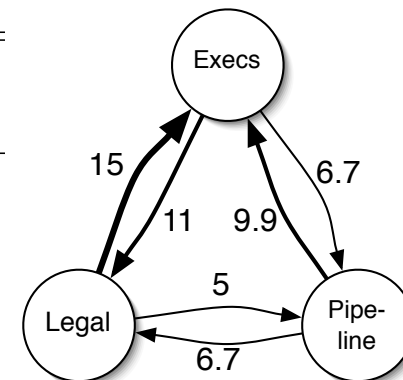
Sandia National Laboratories

Social Network Analysis

Communication graph among employees over all times

Adjacency matrix

$$U_k \quad \Sigma_k \quad V_k^T$$

SVD
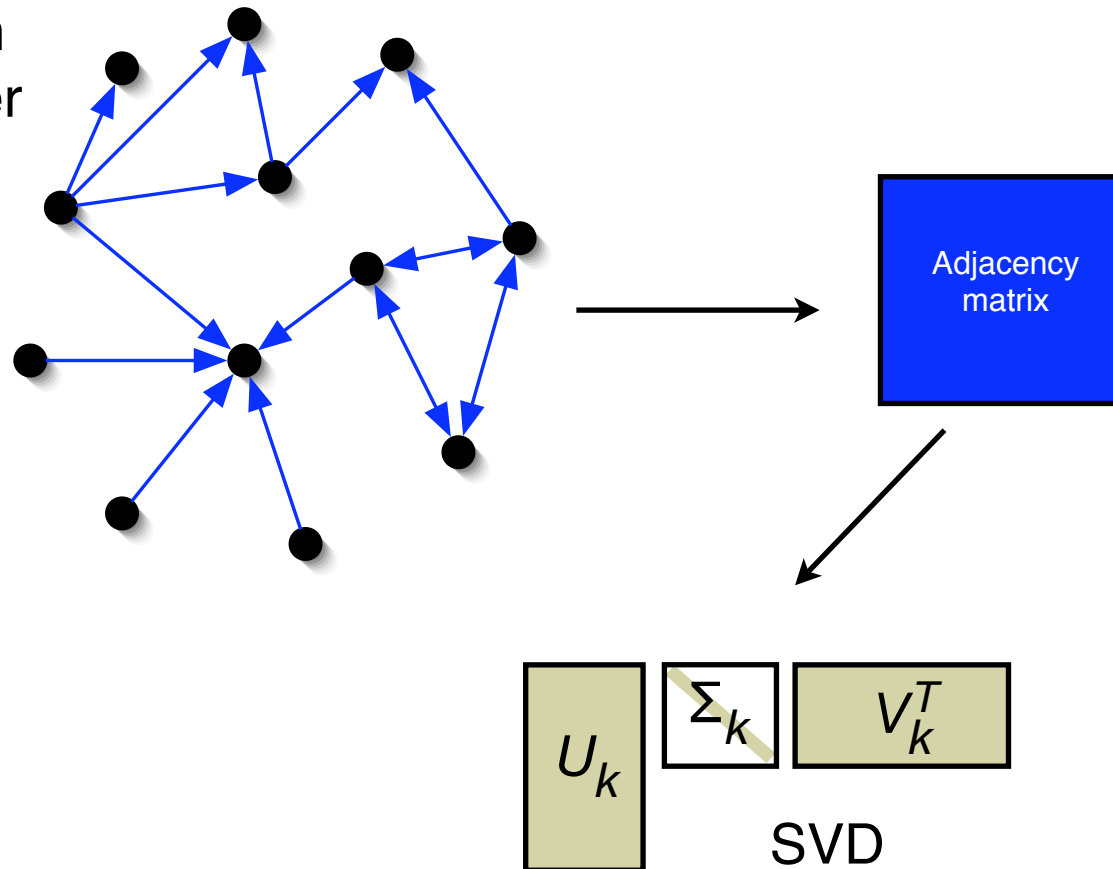
- "Hubs" and "authorities" for different roles
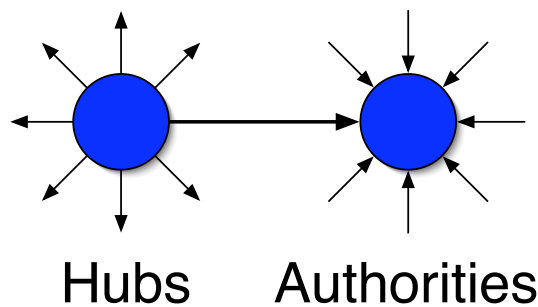
Hubs   Authorities

DEDICOM & SVD Results

patterns

roles | | | U_k | Σ_k | V_k^T
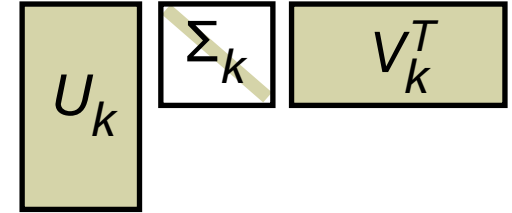
		DEDICOM Solution			U (hubs)			V (authorities)		
	Executives Legal Pipeline									
EMPLOYEE		1	2	3	1	2	3	1	2	3
Execs J. Lavorato - CEO, Enron America		**0.41**	0.07	0.04	**0.30**	-0.07	**-0.21**	**0.31**	-0.09	-0.07
L. Kitchen - President, Enron Online		**0.26**	**0.21**	0.04	**0.31**	0.07	-0.05	**0.29**	0.02	0.04
M. Grigsby - Director, West Desk Gas Trading		**0.22**	-0.01	-0.01	**0.16**	-0.09	**-0.33**	**0.14**	-0.06	**-0.20**
D. Delainey - CEO, ENA and Enron Energy Services		**0.20**	0.06	0.06	**0.20**	-0.05	-0.00	**0.20**	-0.05	0.03
G. Whalley - President,		**0.17**	0.05	0.04	0.08	-0.02	-0.02	0.24	-0.07	0.02
L. Taylor - Executive Assistant to Greg Whalley,		**0.17**	0.06	0.03	**0.24**	-0.05	-0.08	0.09	-0.01	-0.02
Legal T. Jones - Employee, Financial Trading Group (ENA Legal)		**-0.12**	**0.38**	-0.02	**0.17**	**0.36**	**0.13**	0.10	**0.24**	**0.10**
M. Taylor - Manager, Financial Trading Group ENA Legal		-0.10	**0.35**	-0.01	**0.13**	**0.27**	**0.13**	**0.13**	**0.26**	**0.12**
S. Shackleton - Employee, ENA Legal		**-0.13**	**0.31**	-0.02	0.08	**0.26**	**0.10**	0.08	**0.26**	**0.10**
S. Panus - Senior Legal Specialist, ENA Legal		**-0.11**	**0.26**	-0.02	0.09	**0.27**	**0.10**	0.05	**0.20**	0.08
M. Heard - Senior Legal Specialist, ENA Legal		-0.10	**0.24**	-0.02	0.06	**0.20**	0.09	0.08	**0.22**	0.09
E. Sager - VP and Asst Legal Counsel, ENA Legal		-0.01	**0.24**	0.02	**0.12**	**0.13**	**0.10**	**0.15**	**0.21**	**0.12**
Pipeline S. Corman - VP, Regulatory Affairs		-0.04	-0.01	**0.33**	0.08	**-0.18**	**0.22**	0.07	**-0.18**	**0.21**
K. Watson - Employee, Transwestern Pipeline Company (ETS)		-0.08	-0.03	**0.32**	0.03	**-0.16**	**0.19**	0.04	**-0.18**	**0.22**
L. Donoho - Employee, Transwestern Pipeline Company (ETS)		-0.08	-0.03	**0.30**	0.03	**-0.16**	**0.18**	0.03	**-0.17**	**0.20**
D. Fossum - VP, Transwestern Pipeline Company (ETS)?		-0.06	-0.00	**0.30**	0.07	**-0.18**	**0.23**	0.05	**-0.13**	**0.16**
M. Lokay - Admin. Asst., Transwestern Pipeline Company (ETS)		-0.07	-0.02	**0.28**	0.03	**-0.14**	**0.17**	0.04	**-0.17**	**0.20**
K. Hyatt - Director, Asset Development TW Pipeline Co. (ETS)		-0.06	-0.02	**0.25**	0.03	**-0.13**	**0.17**	0.04	**-0.14**	**0.17**
R. Hayslett - VP, Also CFO and Treasurer		-0.04	-0.01	**0.23**	0.04	**-0.13**	**0.16**	0.05	**-0.14**	**0.16**
R matrix / singular values		70.3	11.6	6.7	86.3			86.3		
		15.4	68.2	5.0		54.1			54.1	
		9.9	6.7	59.5			52.6			52.6

SVD: Hubs and Authorities in U and V

Roles more difficult to identify in singular vectors

No patterns of communication

Sandia National Laboratories

Temporal Social Network Analysis

Time series of communication graphs among employees

January
February
March
April

Adjacency tensor

roles

time

patterns

3-way DEDICOM

- Unique description of employees by their roles
- Aggregate communication patterns among roles
- Behavior over time

Sandia National Laboratories

Roles of Employees

roles · time · patterns

Legal

Gov't affairs

Execs - trading

Pipeline employees

Identify shared characteristics to label group

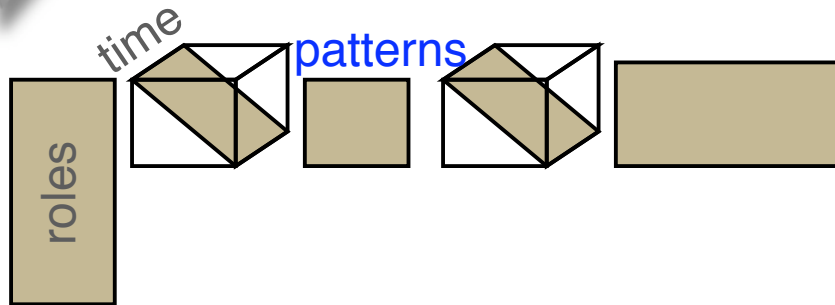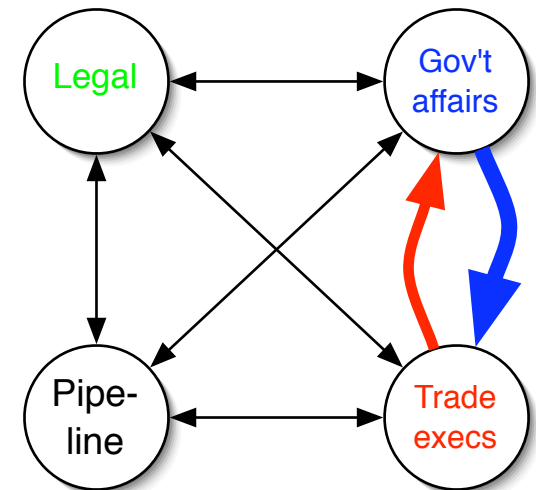| EMPLOYEE	Legal 1	Gov't affairs 2	Trade execs 3	Pipeline 4
T. Jones - Employee, Financial Trading Group (ENA Legal)	**0.64**	-0.01	0.02	-0.00
S. Shackleton - Employee, ENA Legal	**0.45**	-0.00	-0.01	-0.00
M. Taylor - Manager, Financial Trading Group ENA Legal	**0.37**	0.01	0.02	-0.00
S. Bailey - Legal 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
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
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
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**

Sandia National Laboratories

Communication Patterns

roles | time | patterns

	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

Legal — Gov't affairs — Pipe-line — Trade execs

Mostly communication within roles

Negative values complicates interpretation

Sandia National Laboratories

Temporal Patterns

time patterns

roles

Communication patterns over time

Legend:
- Legal
- Government & regulatory affairs
- Trade executives
- Pipeline employee

Month: N D 99 F M A M J J A S O N D 00 F M A M J J A S O N D 01 F M A M J J A S O N D 02 F M A M J

Normalized scale: 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35

Enron crisis breaks;
SEC starts investigation

Sandia National Laboratories

Temporal Social Network Analysis

Time series of communication graphs among employees

April

March

February

January

Adjacency tensor

time

authorities

hubs

+

+ ...
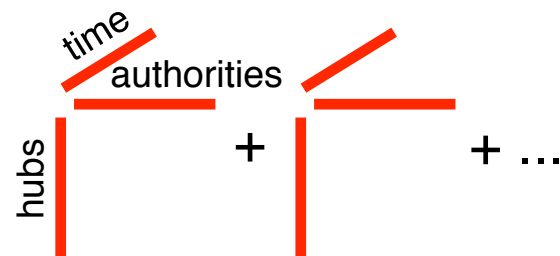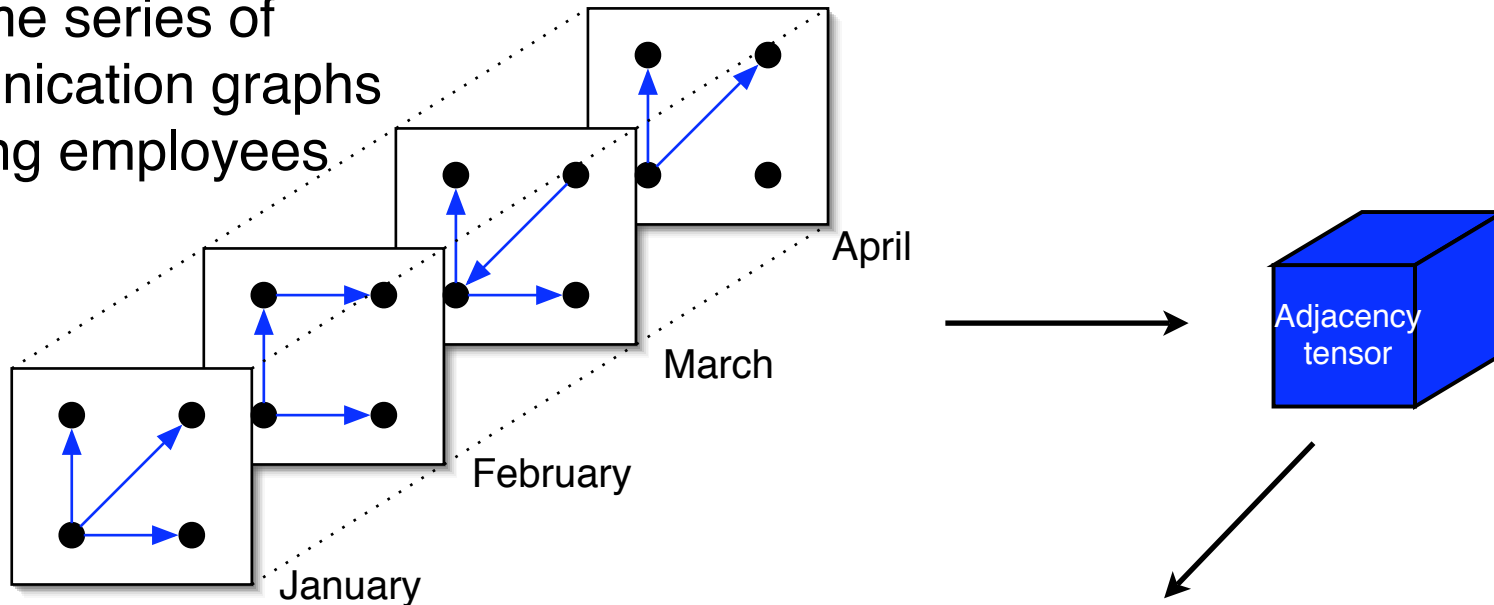
PARAFAC

- Unique description of employees by their roles
- "Hubs" and "authorities"
- Behavior over time

Sandia National Laboratories

Roles of Employees
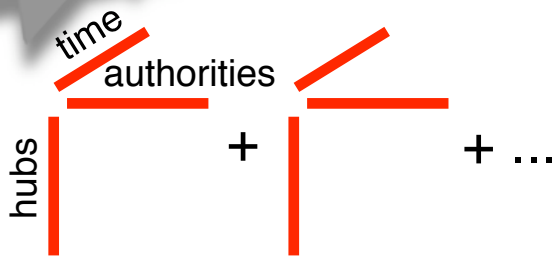
	Hubs				Authorities			
	Legal	Gov't affairs	Execs	Gas traders	Legal	Gov't affairs	Execs	Gas traders
EMPLOYEE	1	2	3	4	1	2	3	4
Legal								
T. Jones - Employee, Financial Trading Group (ENA Legal)	**0.76**	-0.01	-0.00	0.02	**0.25**	0.00	0.01	-0.00
S. Shackleton - Employee, ENA Legal	**0.46**	-0.03	-0.02	0.01	**0.36**	0.00	0.01	-0.00
M. Taylor - Manager, Financial Trading Group ENA Legal	**0.26**	0.01	-0.01	0.00	**0.48**	0.02	-0.00	-0.00
S. Bailey - Legal Assistant, ENA Legal	**0.12**	-0.01	-0.00	0.00	**0.42**	-0.00	-0.00	-0.00
S. Panus - Senior Legal Specialist, ENA Legal	**0.25**	-0.02	0.01	0.00	**0.25**	-0.00	-0.00	0.00
M. Heard - Senior Legal Specialist, ENA Legal	**0.12**	-0.01	0.02	0.01	**0.36**	0.00	0.00	-0.00
J. Hodge - Asst General Counsel, ENA Legal	0.04	0.02	-0.00	0.00	**0.31**	**0.09**	-0.01	0.00
L. Kitchen - President, Enron Online	**0.08**	0.01	<mark>**0.40**</mark>	0.03	<mark>**0.14**</mark>	<mark>**0.11**</mark>	<mark>**0.37**</mark>	-0.05
S. Dickson - Employee, ENA Legal	0.03	0.00	-0.00	0.00	**0.21**	0.01	-0.00	0.00
E. Sager - VP and Asst Legal Counsel, ENA Legal	0.05	0.04	0.03	-0.01	**0.16**	**0.07**	**0.08**	-0.01
Gov't affairs								
J. Dasovich - Employee, Government Relationship Executive	-0.04	**0.66**	**0.34**	-0.00	0.00	**0.31**	**0.13**	-0.04
J. Steffes - VP, Government Affairs	0.00	**0.51**	-0.01	-0.00	-0.00	**0.45**	0.01	-0.01
R. Shapiro - VP, Regulatory Affairs	-0.01	**0.26**	**0.23**	**-0.06**	0.00	**0.54**	**0.22**	-0.05
S. Kean - VP, Chief of Staff	-0.02	**0.37**	0.02	-0.02	0.01	**0.37**	0.00	-0.01
R. Sanders - VP, Enron Wholesale Services	0.04	**0.14**	-0.00	-0.00	0.04	**0.28**	-0.00	-0.00
D. Delainey - CEO, ENA and Enron Energy Services	0.01	**0.17**	0.04	0.00	0.01	**0.14**	**0.15**	-0.03
J. Lavorato - CEO, Enron America	0.02	**0.08**	**0.25**	0.00	0.02	**0.17**	**0.44**	**0.07**
Execs								
G. Whalley - President,	0.00	0.01	**0.07**	-0.01	0.01	**0.09**	<mark>**0.32**</mark>	-0.05
J. Steffes - VP, Government Affairs	-0.00	-0.00	**0.29**	**-0.07**	-0.00	0.02	**0.20**	-0.02
S. Kean - VP, Chief of Staff	-0.00	0.00	**0.21**	-0.06	-0.00	0.00	**0.27**	-0.06
J. Williamson - Executive Assistant,	-0.00	-0.00	<mark>**0.46**</mark>	**-0.12**	0.00	0.01	0.01	-0.00
K. Presto - VP, East Power Trading	0.00	0.00	**0.19**	-0.01	0.01	-0.00	**0.13**	-0.01
S. Beck - COO,	0.01	0.04	**0.17**	0.00	0.01	0.01	**0.10**	-0.01
Gas traders								
M. Grigsby - Director, West Desk Gas Trading	0.00	0.03	**0.13**	**0.91**	0.00	0.05	**0.09**	**0.09**
J. Tholt - VP, West Desk Gas Trading	-0.00	0.00	0.05	**0.12**	-0.00	-0.00	**0.07**	**0.25**
M. Lenhart - Analyst, West Desk Gas Trading	-0.00	0.03	0.00	**0.12**	-0.00	0.01	0.03	**0.30**
S. Scott - Employee, Transwestern Pipeline Company (ETS)	-0.00	0.01	0.01	**0.12**	-0.00	-0.00	0.02	**0.22**
M. Sanchez - Associate, West Desk Gas Trader (EWS)	-0.00	0.01	0.01	**0.08**	-0.00	0.01	0.02	**0.24**
K. Holst - Director, West Desk Gas Trading	0.00	-0.00	0.02	0.02	-0.00	0.05	0.04	**0.30**

Sandia National Laboratories

Temporal Patterns
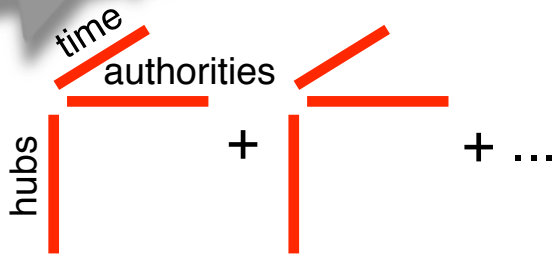
Communication patterns over time
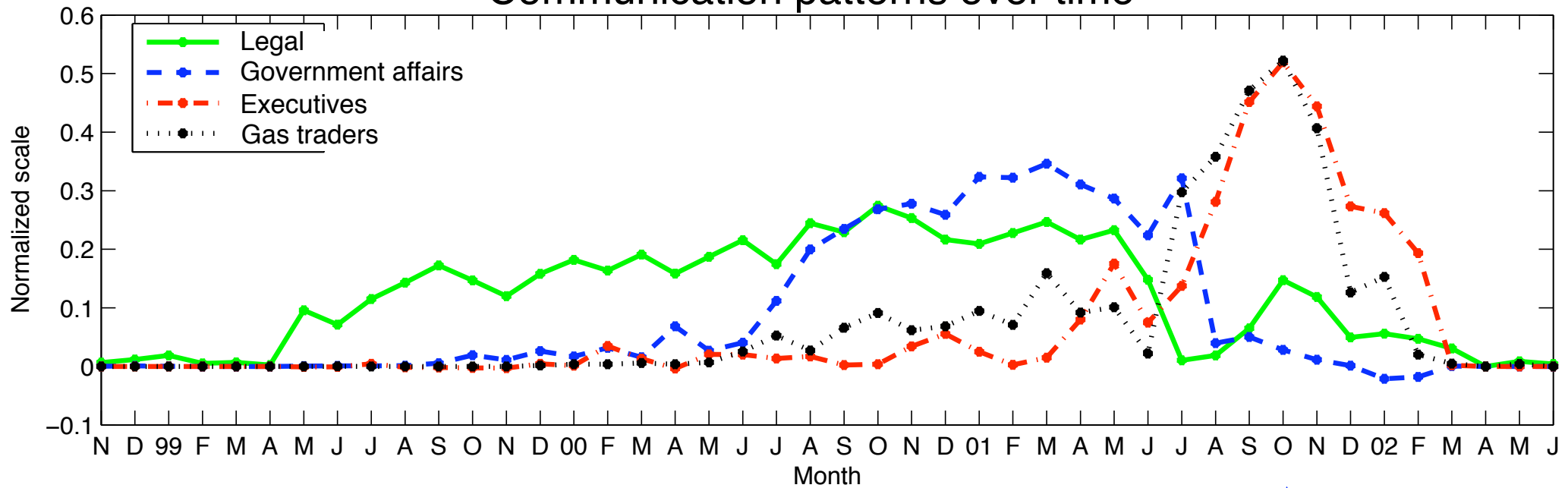
Enron crisis breaks;
SEC starts investigation

Summary

- Improvements to DEDICOM
  - New procedure for finding A
  - Newton step for finding D

- Modifications to handle sparse data arrays
  - Least squares problem
  - Compression

- Novel approach to social network analysis using DEDICOM
  - Roles of employees
  - Communication patterns among roles and over time

- Many future research directions!
  - Constrained DEDICOM
  - Nonnegative DEDICOM
  - PARAFAC

More Information

bwbader@sandia.gov
http://www.cs.sandia.gov/~bwbader/

- Tensor Classes:
  - Tech report SAND2004-5189 available on website
  - Paper to appear in ACM Trans. Math. Softw.
  - sparse_tensor to be released soon

- DEDICOM paper on Social Network Analysis:
  - Tech report SAND2006-2161 available

Sandia National Laboratories