Analysis of Latent Relationships in Semantic Graphs using DEDICOM

Brett Bader*, Richard Harshman** & Tamara Kolda*
*Sandia National Laboratories
**University of Western Ontario

Workshop for Algorithms on Modern Massive Data Sets
June 24, 2006

National Nuclear Security Administration

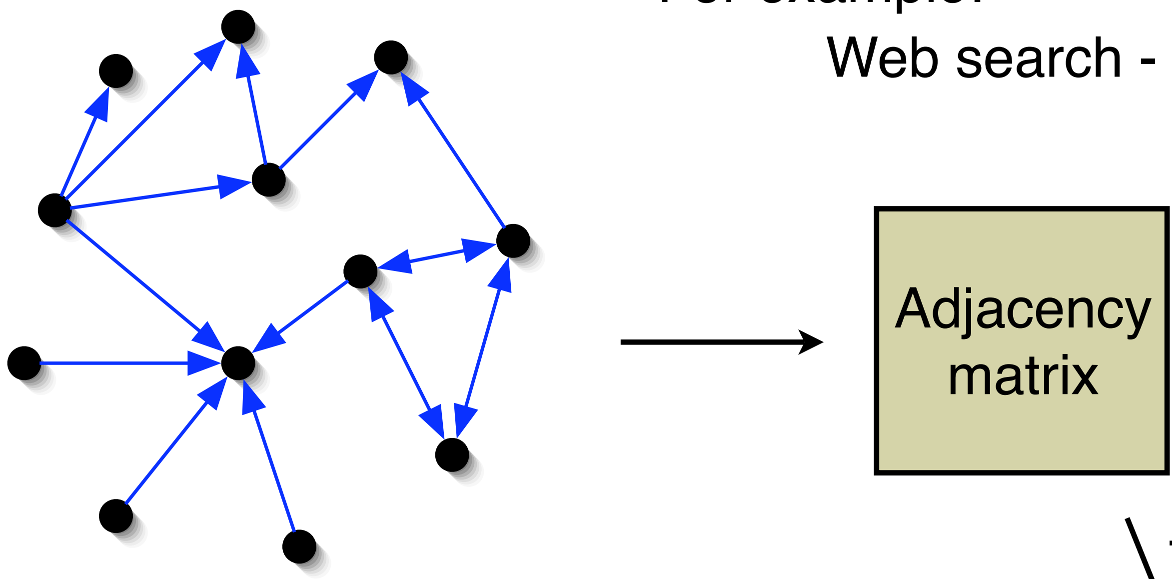
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Common Graph Analysis Technique

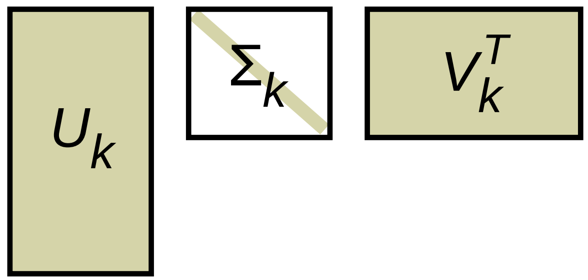
For example:

Web search - HITS (Kleinberg, 1998)



Truncated SVD

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

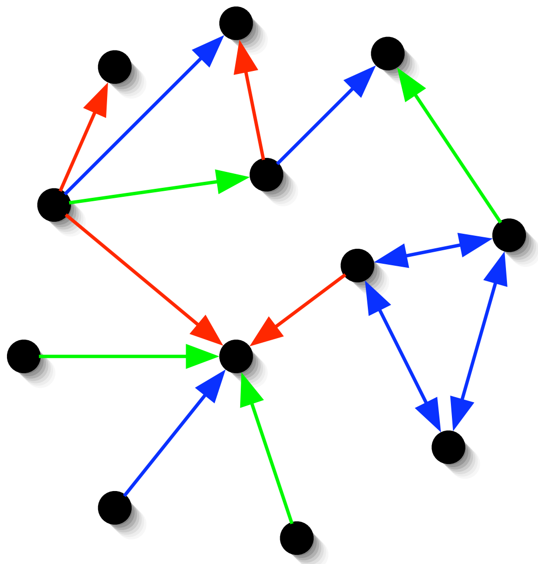


Best rank-*k* matrix filters out noise and captures “latent” information, which improves certain data mining tasks

But we may have ignored critical information by not considering edge metadata!

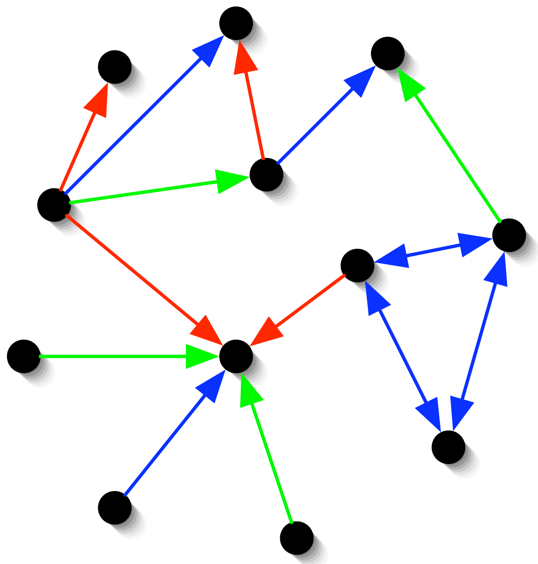
Sandia National Laboratories

Semantic Graphs

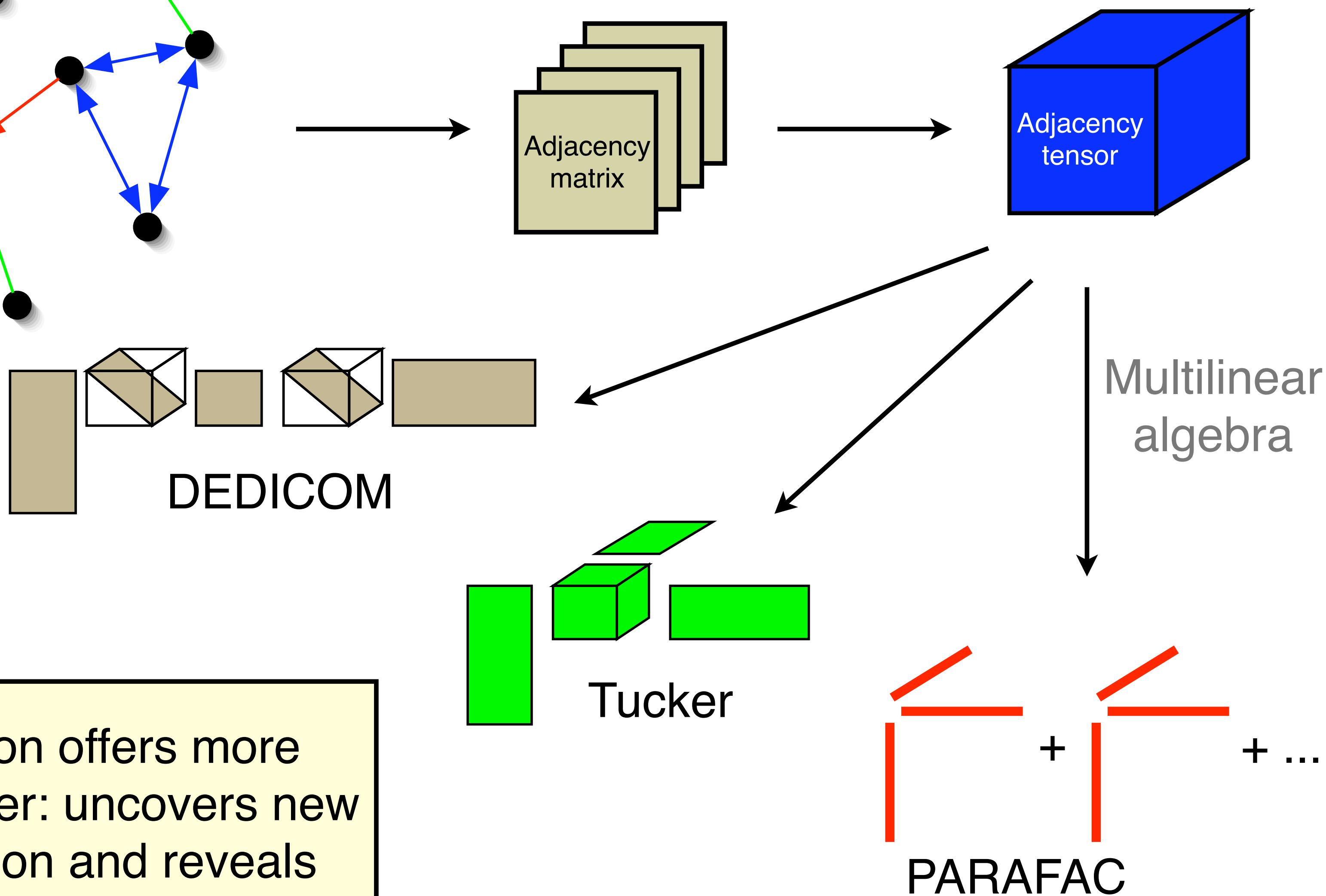


- Different types of edges
- Examples
  - WWW (anchor text)
  - Subway map [thanks Orly!]
  - Email communications (time stamp, to/cc)

New Paradigm: "Multidimensional Data Mining"



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

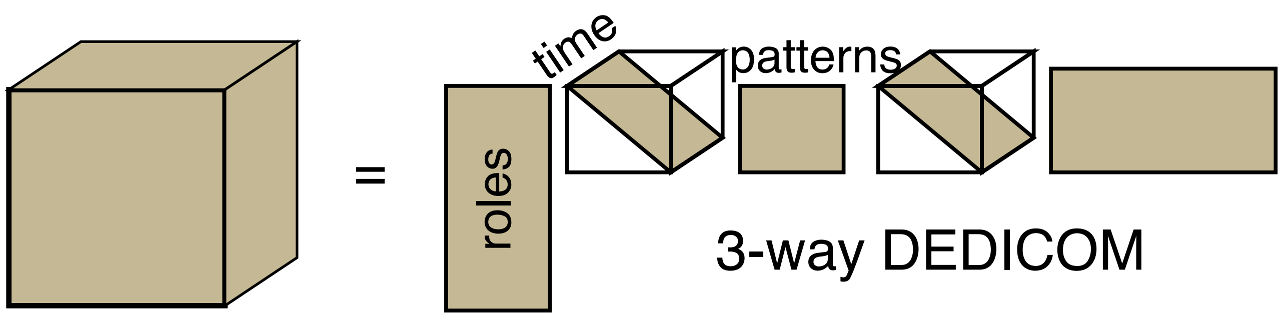


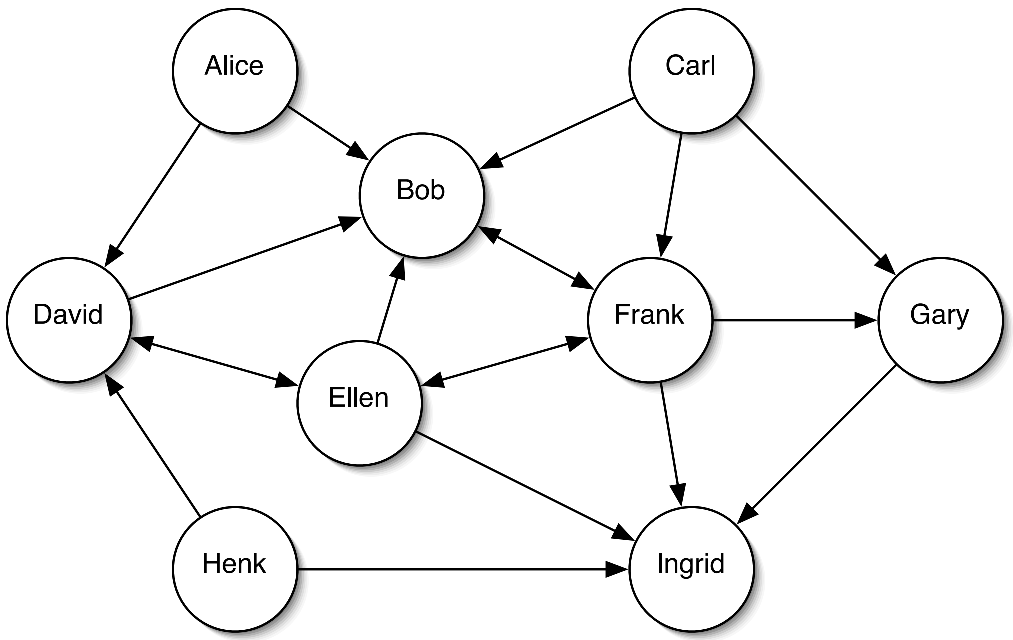
Third dimension offers more explanatory power: uncovers new latent information and reveals subtle relationships

Sandia National Laboratories

Objective

Use DEDICOM to analyze a semantic graph of email communications changing over time

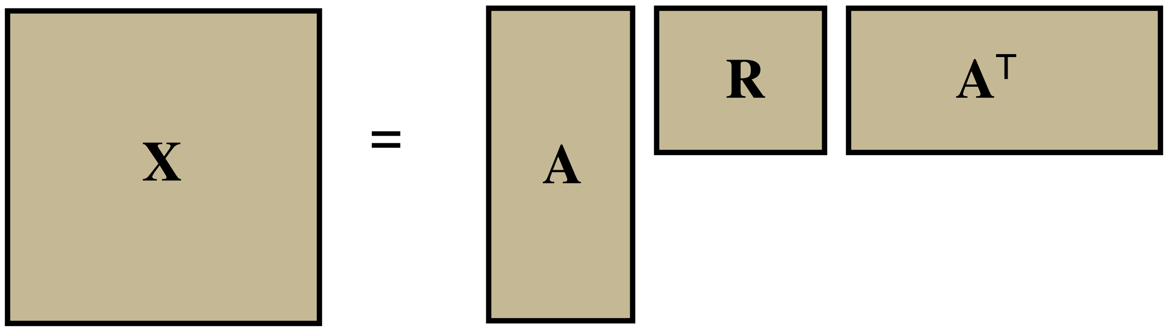




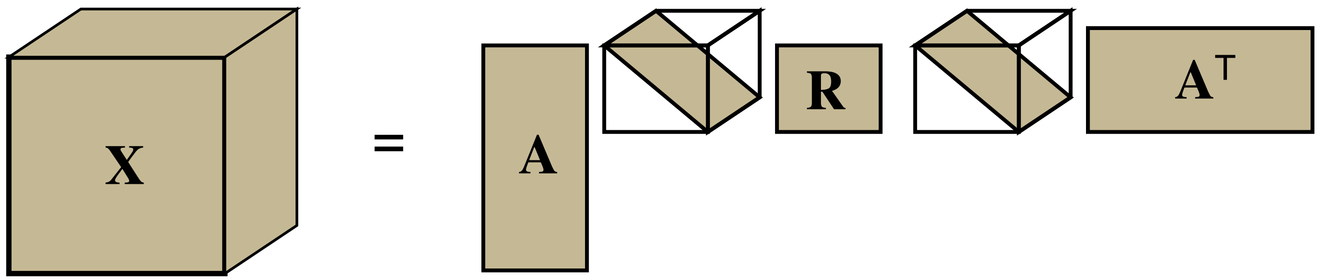
DEDICOM

- DEcomposition into DIrectional COMponents
- Introduced in 1978 by Harshman
- Past applications
  - Study asymmetries in telephone calls among cities
  - Marketing research
    - car switching: car owners and what they buy next
    - free associations of words
      - words to describe hair in advertising shampoo: “body” evokes “fullness” more often than “fullness” evokes “body”
  - Asymmetric measures of world trade (import/export)
- Variations
  - Three-way DEDICOM
  - Constrained DEDICOM

DEDICOM Models & Algorithms



- Generalized Takane method (Takane, 1985; Kiers et al., 1990)
- New algorithm

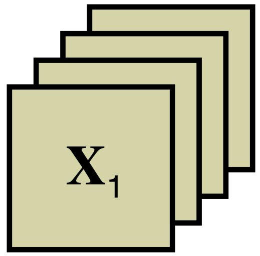


- Kiers' method (Kiers, 1993)
- New algorithm

All are "alternating" algorithms

Mathematical Notation

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



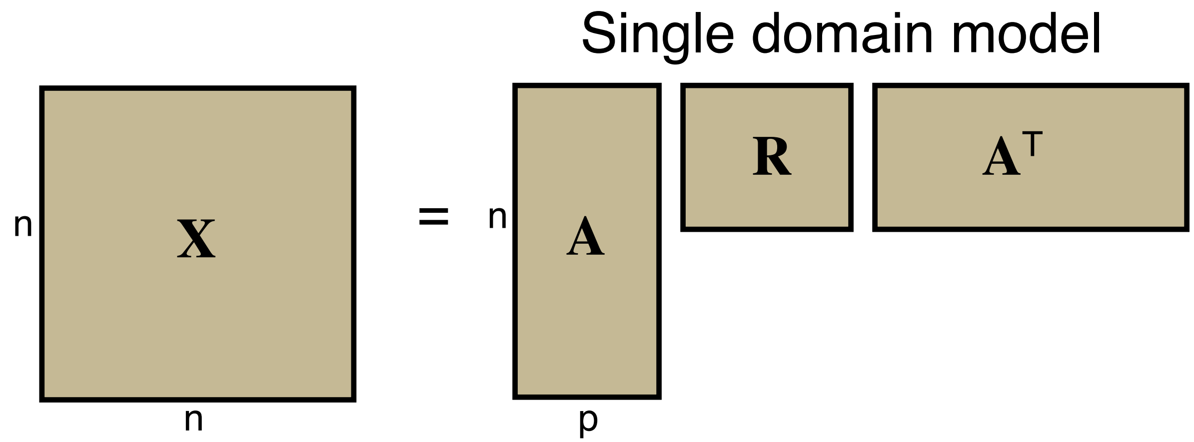
- Special symbols
  - Kronecker product

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

  - Hadamard product (elementwise)

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

Two-way DEDICOM



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

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

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

s.t. $\mathbf{A}$ orthogonal

- $\mathbf{A}$ (n x p) is an orthogonal matrix of loadings or weights
- $\mathbf{R}$ (p x p) is a dense matrix that captures asymmetric relationships

- Decomposition is not unique
  - $\mathbf{A}$ can be transformed with no loss of fit to the data
  - Nonsingular transformation $\mathbf{Q}$:

$$\mathbf{A}\mathbf{R}\mathbf{A}^T = (\mathbf{A}\mathbf{Q})(\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-T})(\mathbf{A}\mathbf{Q})^T$$

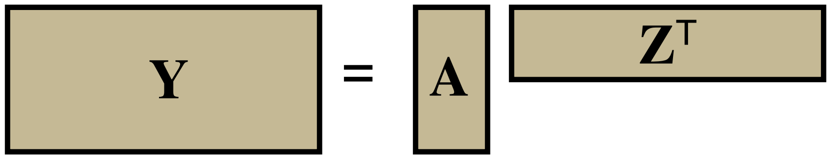
  - Usually “fix” $\mathbf{A}$ with some standard rotation (e.g., VARIMAX)

New Algorithm

Solving for $\mathbf{A}$:

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

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



...and solve least-squares problem: $\min_{\mathbf{A}} \left\| \mathbf{Y} - \mathbf{AZ}^T \right\|_F^2$

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

or

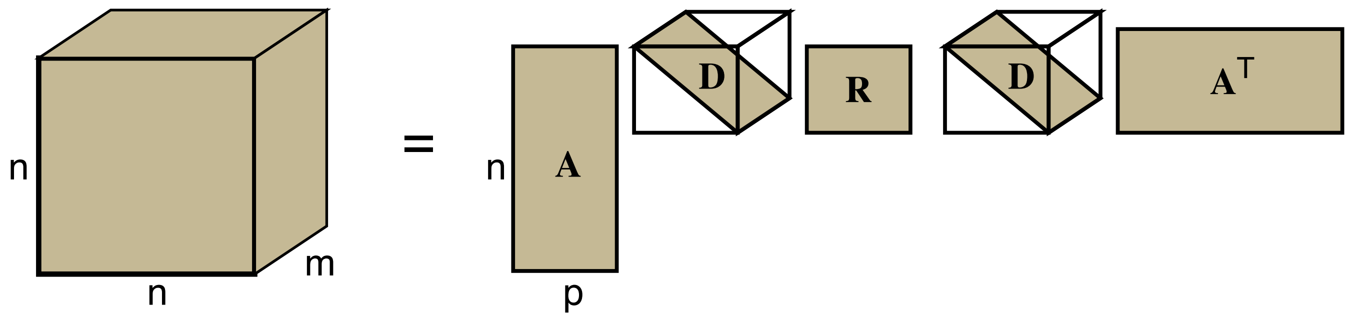
$$\mathbf{A}_{new} = \left(\mathbf{XAR}^T + \mathbf{X}^T\mathbf{AR}\right)\left(\mathbf{R}(\mathbf{A}^T\mathbf{A})\mathbf{R}^T + \mathbf{R}^T(\mathbf{A}^T\mathbf{A})\mathbf{R}\right)^{-1}.$$

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 dense matrix that captures asymmetric relationships
- $\mathbf{D}$ (p x p x m) is a tensor with diagonal frontal slices giving the weights of the columns of $\mathbf{A}$ for each slice in third mode

- *Unique* solution with enough slices of $\mathbf{X}$ with sufficient variation
  - i.e., no rotation of $\mathbf{A}$ possible
  - greater confidence in interpretation of results

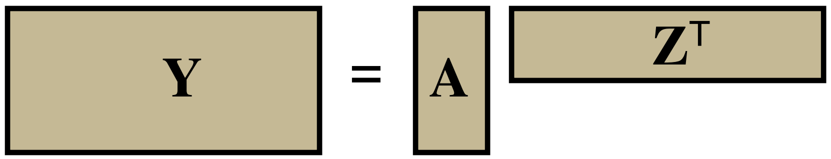
Sandia National Laboratories

New Algorithm - Updating A

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

Solving for $\mathbf{A}$:

$$\left(\mathbf{X}_1 \quad \mathbf{X}_1^T \quad \cdots \quad \mathbf{X}_m \quad \mathbf{X}_m^T\right) = \mathbf{A}\left(\mathbf{D}_1\mathbf{R}\mathbf{D}_1 \quad \mathbf{D}_1\mathbf{R}^T\mathbf{D}_1 \quad \cdots \quad \mathbf{D}_m\mathbf{R}\mathbf{D}_m \quad \mathbf{D}_m\mathbf{R}^T\mathbf{D}_m\right)\left(\mathbf{I}_{2m} \otimes \mathbf{A}^T\right)$$



$$\mathbf{A} = \mathbf{Y}\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}$$

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

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

New Algorithm - Updating D

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

Solving for $\mathbf{D}$:

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

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

Gradient: $g_k = -\sum_{i,j} \left[2(\mathbf{X} - \mathbf{ADRDA}^T) * (\mathbf{AD}\mathbf{r}_k\mathbf{a}_k^T + \mathbf{a}_k\mathbf{r}_{k,:}\mathbf{DA}^T) \right]_{i,j}$

Hessian: $h_{st} = -2\sum_{i,j} \left[(\mathbf{X} - \mathbf{ADRDA}^T) * (\mathbf{a}_s r_{st}\mathbf{a}_t^T + \mathbf{a}_t r_{ts}\mathbf{a}_s^T) - (\mathbf{AD}\mathbf{r}_s\mathbf{a}_s^T + \mathbf{a}_s\mathbf{r}_{s:}\mathbf{DA}^T) * (\mathbf{AD}\mathbf{r}_t\mathbf{a}_t^T + \mathbf{a}_t\mathbf{r}_{t:}\mathbf{DA}^T) \right]_{i,j}$

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)

Sandia National Laboratories

Our Algorithm - Updating R

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

Solving for $\mathbf{R}$:

Use the approach in (Kiers, 1993)

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

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

Algorithm Costs

Updating $\mathbf{A}$ is most expensive part

Dominant costs:

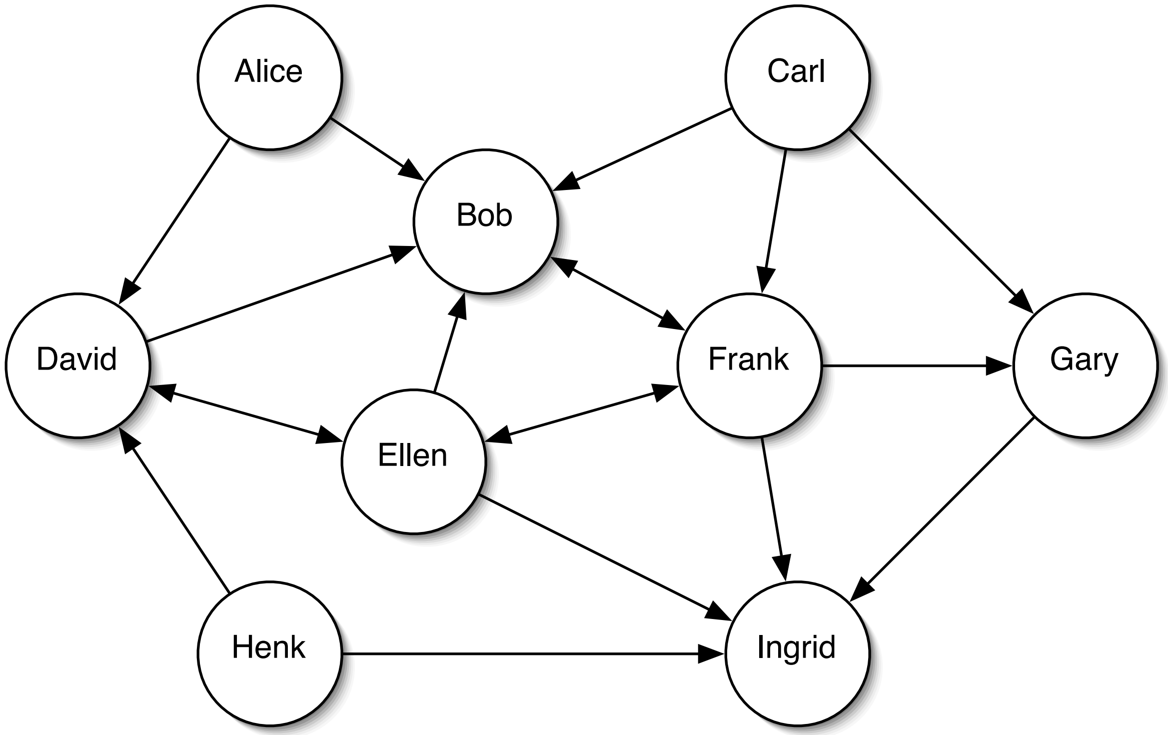
linear in nnz of $\mathbf{X}_i$:

- $\mathbf{Q}^T\mathbf{X}_i\mathbf{Q}$
- $\mathbf{X}_i\mathbf{A}\mathbf{R}^T$
- $\mathbf{X}_i^T\mathbf{A}\mathbf{R}$

$\mathcal{O}(p^2n)$:

- $\mathbf{A}^T\mathbf{A}$
- QR factorization of $\mathbf{A}$

Application: Enron Email Analysis

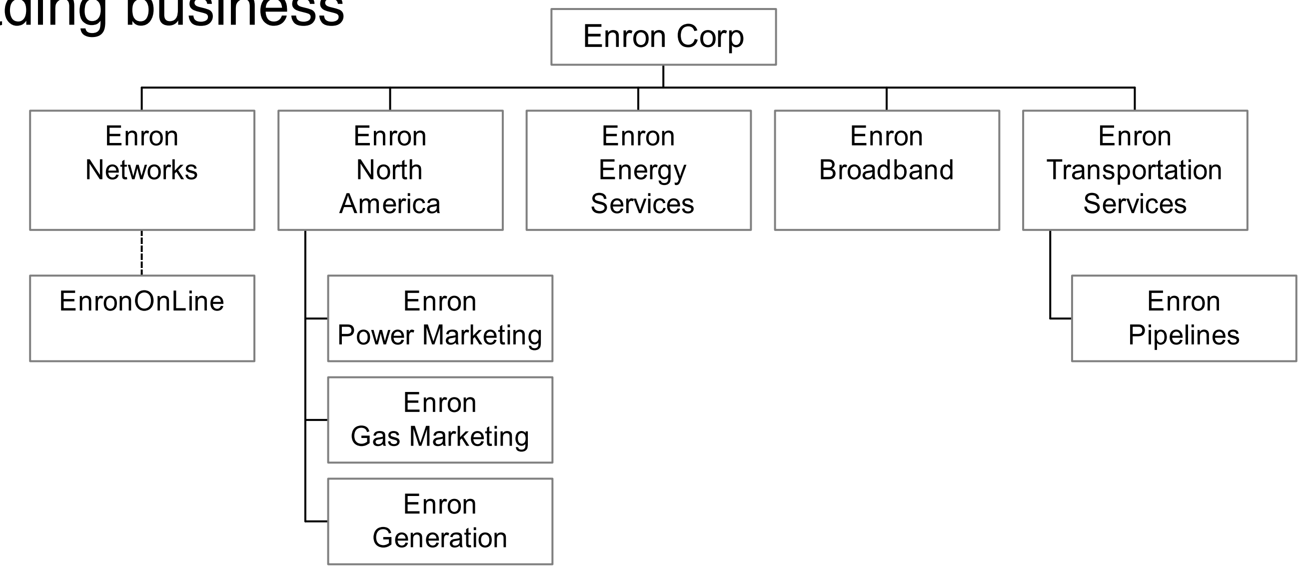


- Links consist of email communications
- What can we learn about this network strictly from their communication patterns? (Social network analysis)

Sandia National Laboratories

Enron Corp.

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



- Investigations
  - U.S. Federal Energy Regulatory Commission (FERC)
    - energy market manipulation
    - involved energy traders
  - U.S. Securities and Exchange Commission (SEC)
    - accounting fraud
    - insider trading

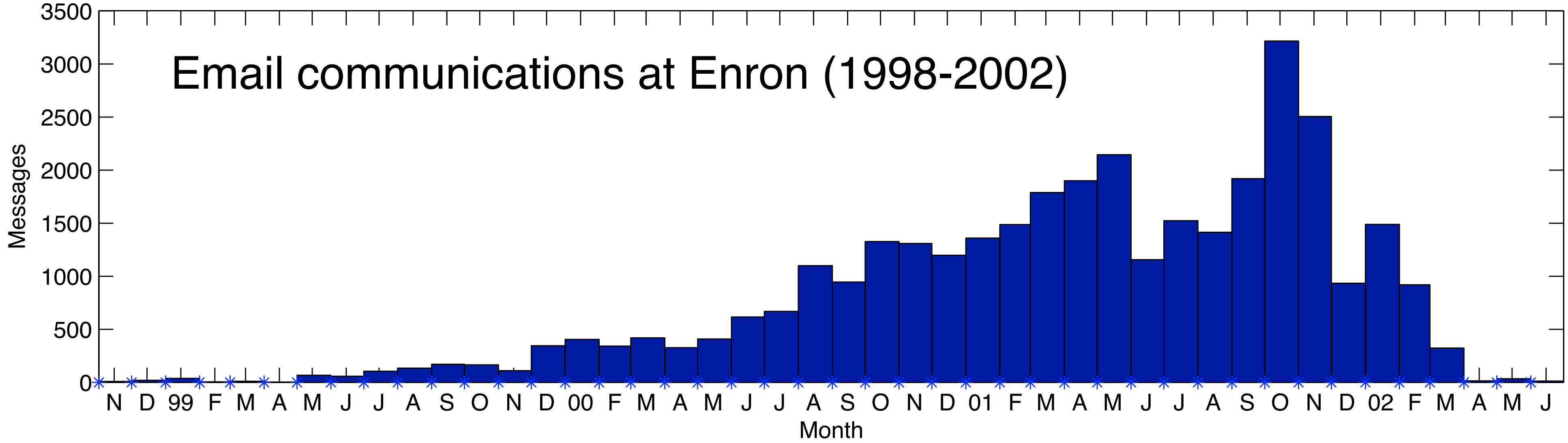
Enron Email Data

- 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

Smaller Enron Data Set

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



- Limited information on the 184 employees
- No org chart

DEDICOM Experiment

- Aggregate communications
  - Sparse matrix of size 184 x 184 (3007 nonzeros)

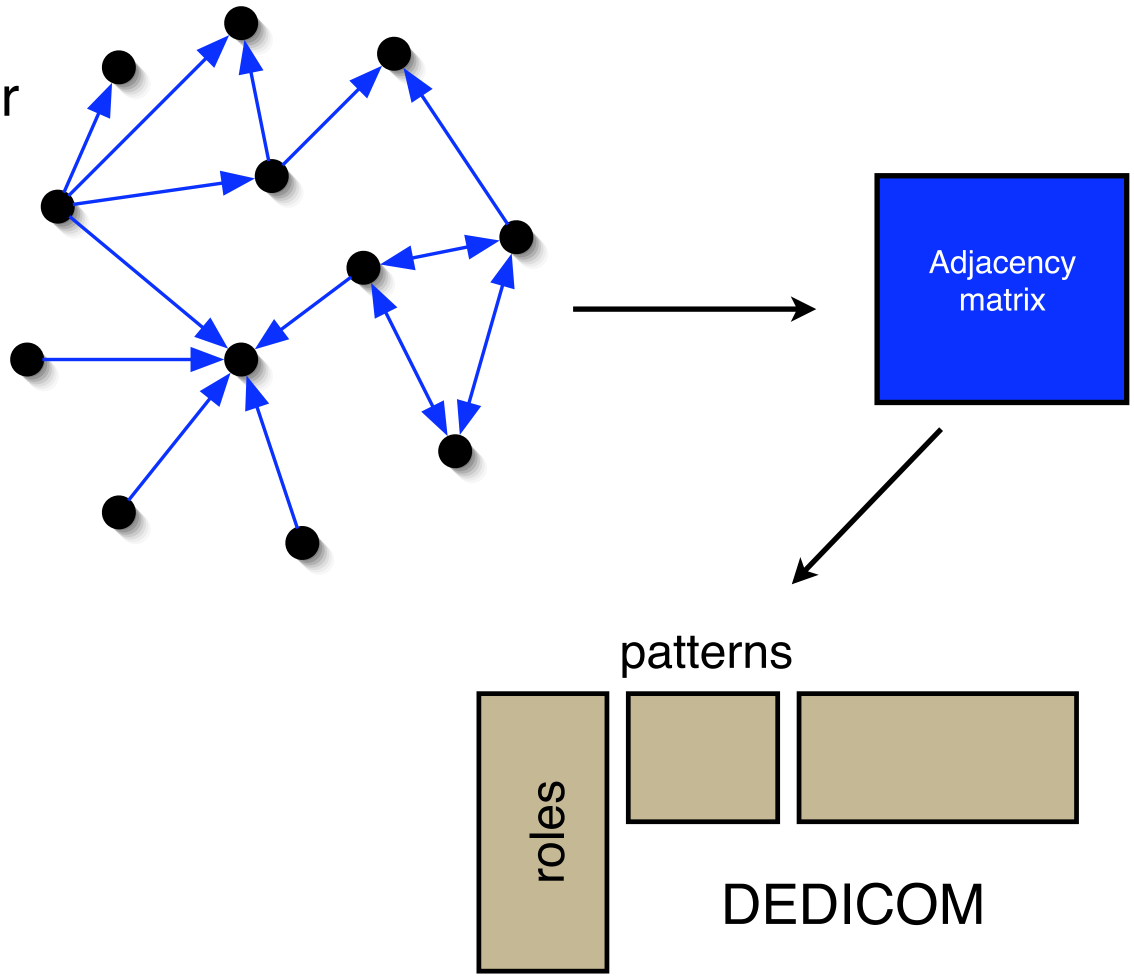
- Time series of communication graphs
  - Sparse tensor of size 184 x 184 x 44 (9838 nonzeros)

- Weighted adjacency matrix
  - scaling: x number of messages scaled by log(x)+1
  - other common choices give similar results

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

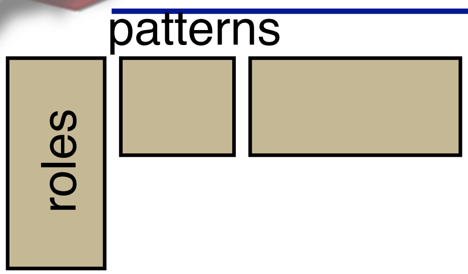
Social Network Analysis

Communication graph among employees over all times



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

DEDICOM Results



Executives Legal Pipeline

	Employee	DEDICOM Solution 1	DEDICOM Solution 2	DEDICOM Solution 3
Execs	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
Legal	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
Pipeline	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

Identify shared characteristics to label group

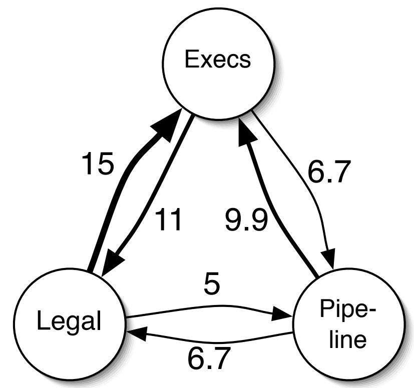
Executives

Legal

Pipeline employees

Some employees have dual roles

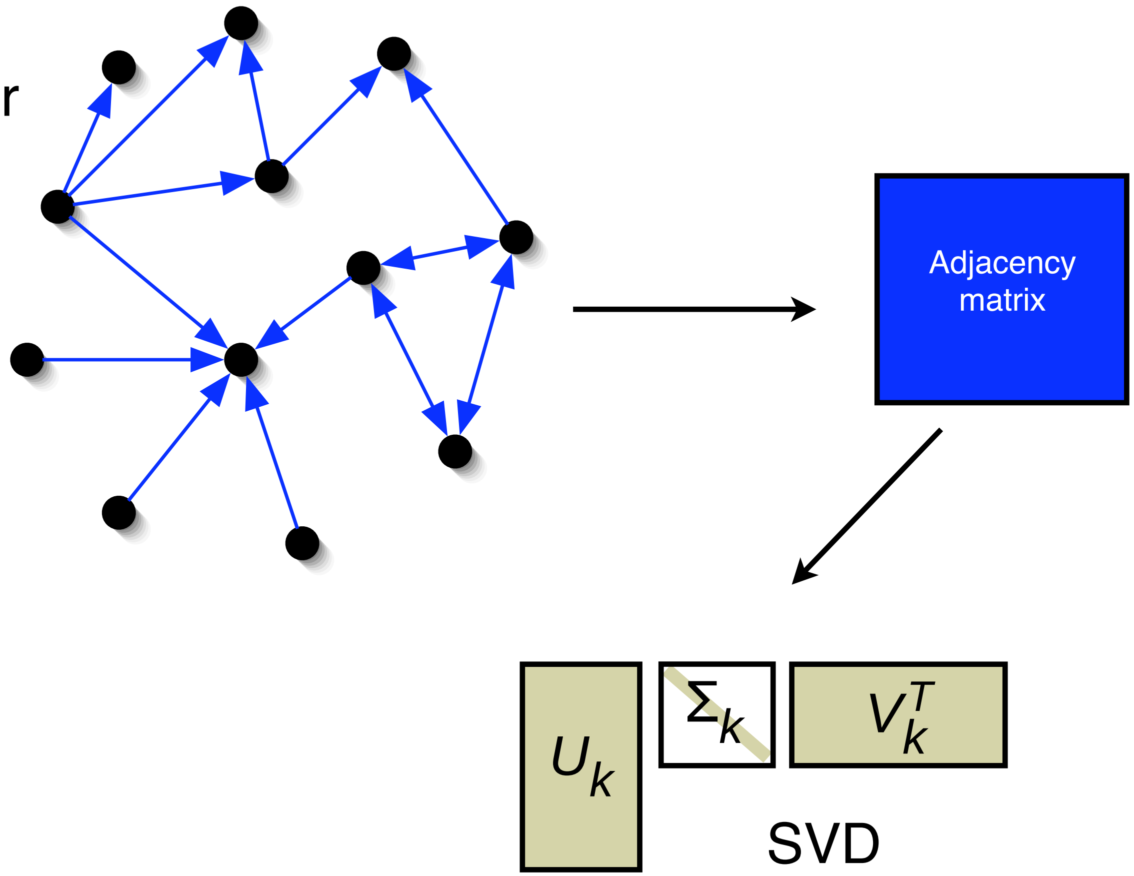
Pattern of communications in $\mathbf{R}$ matrix



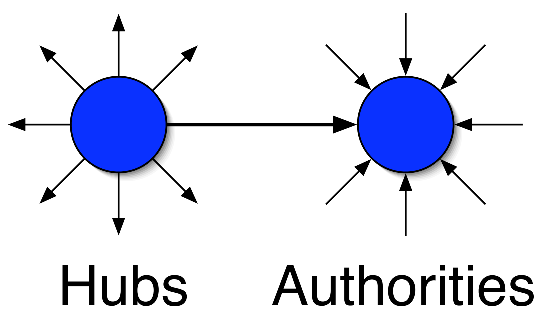
Sandia National Laboratories

Social Network Analysis

Communication graph among employees over all times

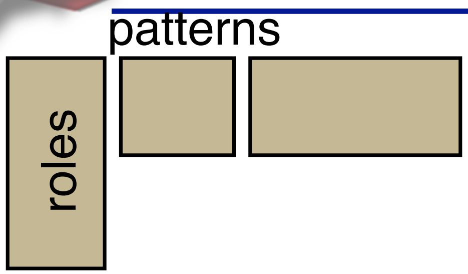


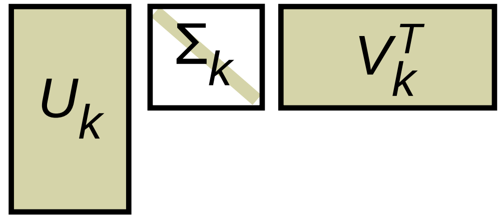
- “Hubs” and “authorities” for different roles



Sandia National Laboratories

DEDICOM & SVD Results





Executives Legal Pipeline

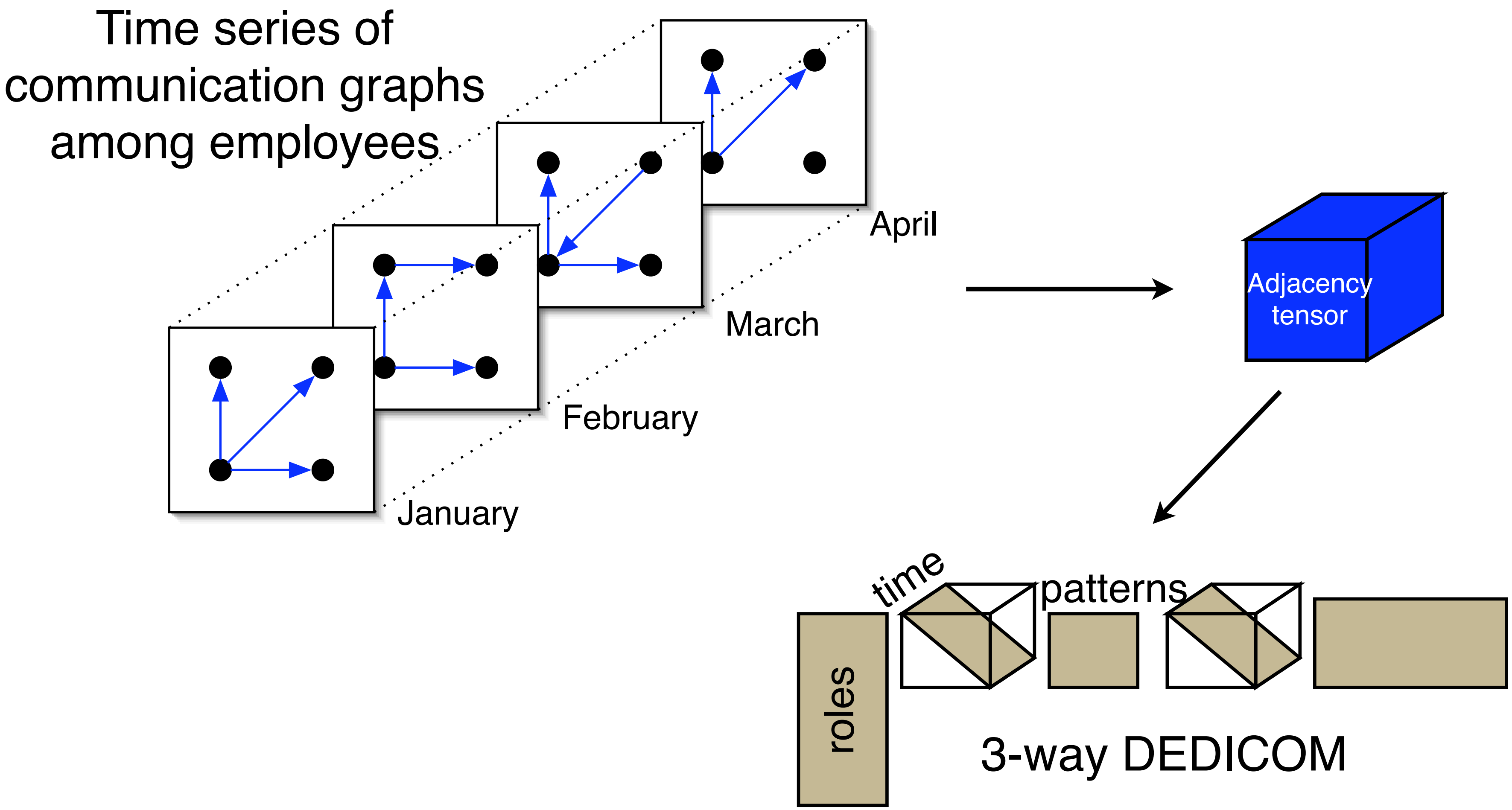
		DEDICOM Solution			U (hubs)			V (authorities)		
	Employee	1	2	3	1	2	3	1	2	3
Execs	J. Lavorato - CEO, Enron America	**0.41**	0.07	0.04	**0.30**	-0.07	**-0.21**	**0.31**	-0.09	-0.07
Execs	L. Kitchen - President, Enron Online	**0.26**	**0.21**	0.04	**0.31**	0.07	-0.05	**0.29**	0.02	0.04
Execs	M. Grigsby - Director, West Desk Gas Trading	**0.22**	-0.01	-0.01	**0.16**	-0.09	**-0.33**	**0.14**	-0.06	**-0.20**
Execs	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
Execs	G. Whalley - President,	**0.17**	0.05	0.04	0.08	-0.02	-0.02	**0.24**	-0.07	0.02
Execs	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**
Legal	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**
Legal	S. Shackleton - Employee, ENA Legal	**-0.13**	**0.31**	-0.02	0.08	**0.26**	**0.10**	0.08	**0.26**	**0.10**
Legal	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
Legal	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
Legal	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**
Pipeline	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**
Pipeline	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**
Pipeline	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**
Pipeline	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**
Pipeline	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**
Pipeline	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

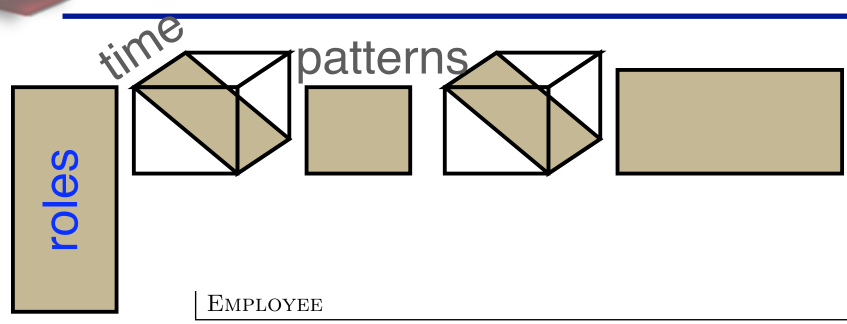
Sandia National Laboratories

Temporal Social Network Analysis



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

Roles of Employees

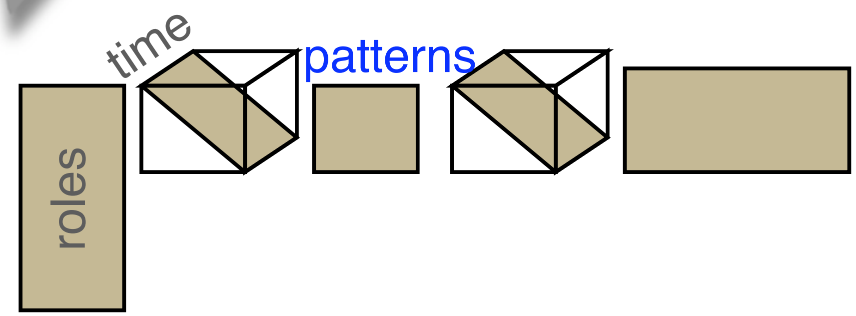


Legal | Gov't affairs | Trade execs | Pipeline

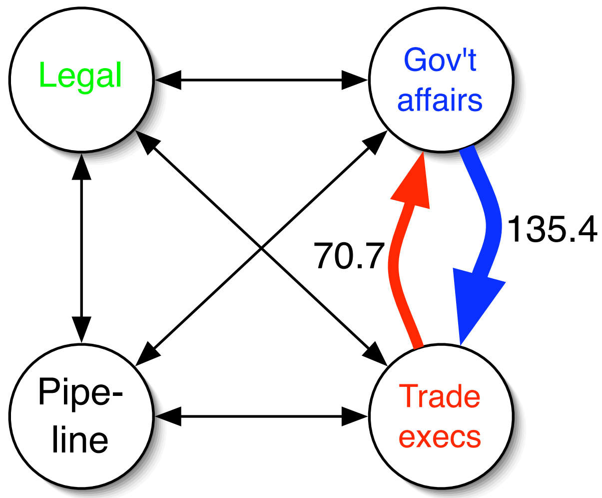
Group	Employee	1	2	3	4
Legal	T. Jones - Employee, Financial Trading Group (ENA Legal)	**0.64**	-0.01	0.02	-0.00
	S. Shackleton - Employee, ENA Legal	**0.45**	-0.00	-0.01	-0.00
	M. Taylor - Manager, Financial Trading Group ENA Legal	**0.37**	0.01	0.02	-0.00
	S. Bailey - Legal Assistant, ENA Legal	**0.26**	-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**

Identify shared characteristics to label group

Communication Patterns



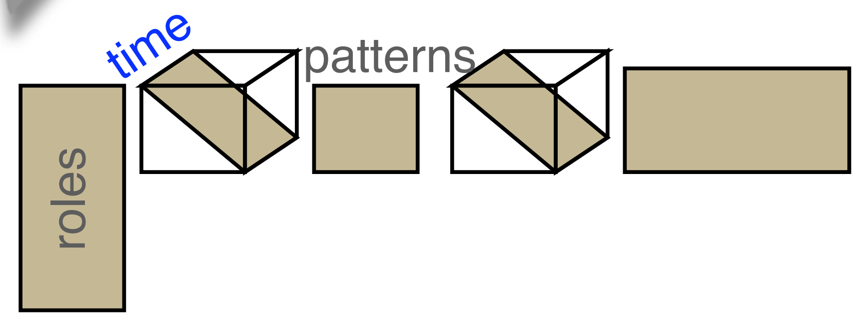
	Legal	Gov't affairs	Trade execs	Pipeline
Legal	440.2	1.6	-15.0	0.4
Government & regulatory affairs	1.6	278.3	135.4	1.6
Trade executives	-29.3	70.7	201.6	-6.2
Pipeline employees	1.4	-4.6	-7.5	172.3



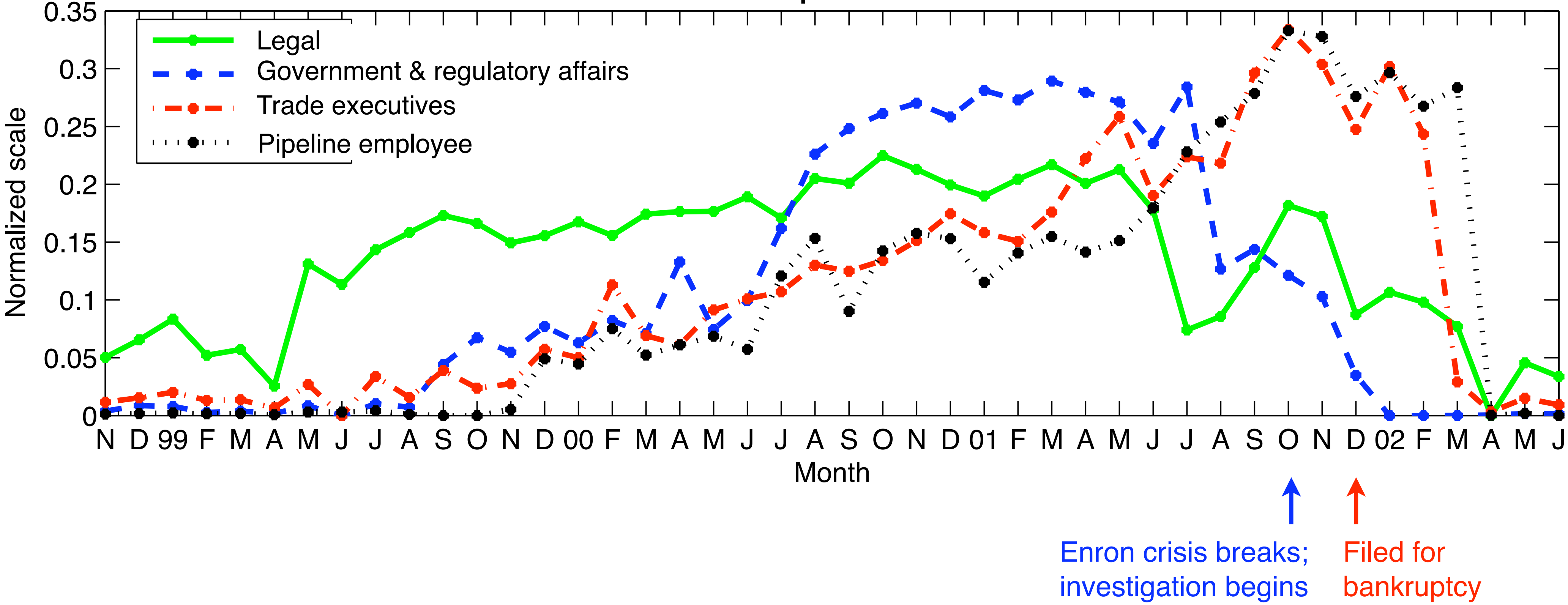
- Mostly communication within roles
- Some large exchanges
- Negative values complicates interpretation
  - Non-negative factorization being investigated

Sandia National Laboratories

Temporal Patterns



Communication patterns over time



Summary

- Improvements to DEDICOM
  - New procedure for finding $\mathbf{A}$
  - Newton step for finding $\mathbf{D}$
- Modifications to handle large data arrays
  - Compression
- Novel approach to social network analysis using DEDICOM
  - Roles of employees
  - Communication patterns among roles and over time
- Future research
  - Nonnegative DEDICOM
  - Constrained DEDICOM
  - PARAFAC

More Information

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

- DEDICOM paper on Social Network Analysis:
  - Tech report SAND2006-2161 available
- MATLAB Tensor Toolbox:
  - http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox
  - Tech report SAND2004-5189 available on website
  - Paper to appear in ACM Trans. Math. Softw.
  - sparse_tensor class to be released soon