

# 15-826: Multimedia (Databases) and Data Mining

Lecture #19: Tensor decompositions

*C. Faloutsos*



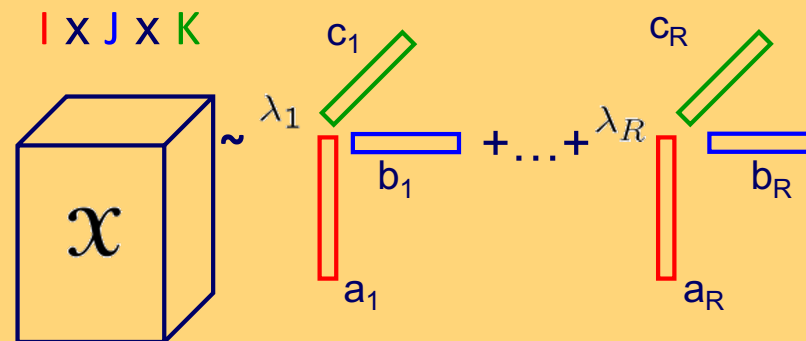
# Problem

- Q: who-calls-whom-when – patterns?
  - Triplets ( source-ip, dest-ip, port#)
  - KB (subject, verb, object)



# Conclusions

- Q: who-calls-whom-when – patterns?
  - Triplets ( source-ip, dest-ip, port#)
  - KB (subject, verb, object)
- A: Tensor analysis (PARAFAC)
  - <http://www.tensortoolbox.org/>



# Must-read Material

- [[Graph-Textbook](#)] Ch.16.
- Tensors survey: Papalexakis, Faloutsos, Sidiropoulos [Tensor for Data Mining and Data Fusion: Models, Applications, and Scalable Algorithms](#) ACM Trans. on Intelligent Systems and Technology, 8,2, Oct. 2016. ([local copy](#))



# Outline

Goal: ‘Find **similar / interesting** things’

- Intro to DB
- ➔ • Indexing - similarity search
- Data Mining



# Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
- fractals
- text
- Singular Value Decomposition (SVD)
  - ...
  - Tensors
- multimedia
- ...





# Outline - tensors

- Motivation - Definitions
- Tensor tools
- Case studies

## Most of foils by

- Dr. Tamara Kolda (Sandia N.L.)
- [csmr.ca.sandia.gov/~tgkolda](http://csmr.ca.sandia.gov/~tgkolda)
- Prof. Jimeng Sun (UIUC)
- <https://cs.illinois.edu/about/people/faculty/jimeng>



3h tutorial: [www.cs.cmu.edu/~christos/TALKS/SDM-tut-07/](http://www.cs.cmu.edu/~christos/TALKS/SDM-tut-07/)

# Motivation 1: Why “matrix”?

- Why matrices are important?

# Examples of Matrices:

## Graph - social network

|       | John | Peter | Mary | Nick | ... |
|-------|------|-------|------|------|-----|
| John  | 0    | 11    | 22   | 55   | ... |
| Peter | 5    | 0     | 6    | 7    | ... |
| Mary  | ...  | ...   | ...  | ...  | ... |
| Nick  | ...  | ...   | ...  | ...  | ... |
| ...   | ...  | ...   | ...  | ...  | ... |

# Examples of Matrices: cloud of n-d points

|       | chol# | blood# | age | ..  | ... |
|-------|-------|--------|-----|-----|-----|
| John  | 13    | 11     | 22  | 55  | ... |
| Peter | 5     | 4      | 6   | 7   | ... |
| Mary  | ...   | ...    | ... | ... | ... |
| Nick  | ...   | ...    | ... | ... | ... |
| ...   | ...   | ...    | ... | ... | ... |

# Examples of Matrices:

## Market basket

- **market basket** as in Association Rules

|       | milk | bread | choc. | wine | ... |
|-------|------|-------|-------|------|-----|
| John  | 13   | 11    | 22    | 55   | ... |
| Peter | 5    | 4     | 6     | 7    | ... |
| Mary  | ...  | ...   | ...   | ...  | ... |
| Nick  | ...  | ...   | ...   | ...  | ... |
| ...   | ...  | ...   | ...   | ...  | ... |

# Examples of Matrices:

## Documents and terms

|         | data | mining | classif. | tree | ... |
|---------|------|--------|----------|------|-----|
| Paper#1 | 13   | 11     | 22       | 55   | ... |
| Paper#2 | 5    | 4      | 6        | 7    | ... |
| Paper#3 | ...  | ...    | ...      | ...  | ... |
| Paper#4 | ...  | ...    | ...      | ...  | ... |
| ...     | ...  | ...    | ...      | ...  | ... |

# Examples of Matrices:

## Authors and terms

|       | data | mining | classif. | tree | ... |
|-------|------|--------|----------|------|-----|
| John  | 13   | 11     | 22       | 55   | ... |
| Peter | 5    | 4      | 6        | 7    | ... |
| Mary  | ...  | ...    | ...      | ...  | ... |
| Nick  | ...  | ...    | ...      | ...  | ... |
| ...   | ...  | ...    | ...      | ...  | ... |

# Examples of Matrices: sensor-ids and time-ticks

|     | temp1 | temp2 | humid. | pressure | ... |
|-----|-------|-------|--------|----------|-----|
| t1  | 13    | 11    | 22     | 55       | ... |
| t2  | 5     | 4     | 6      | 7        | ... |
| t3  | ...   | ...   | ...    | ...      | ... |
| t4  | ...   | ...   | ...    | ...      | ... |
| ... | ...   | ...   | ...    | ...      | ... |

# Motivation: Why tensors?

- Q: what is a tensor?

# Motivation 2: Why tensor?

- A: N-D generalization of matrix:

|         |      |        |          |      |     |
|---------|------|--------|----------|------|-----|
| KDD' 17 | data | mining | classif. | tree | ... |
| John    | 13   | 11     | 22       | 55   | ... |
| Peter   | 5    | 4      | 6        | 7    | ... |
| Mary    | ...  | ...    | ...      | ...  | ... |
| Nick    | ...  | ...    | ...      | ...  | ... |
| ...     | ...  | ...    | ...      | ...  | ... |

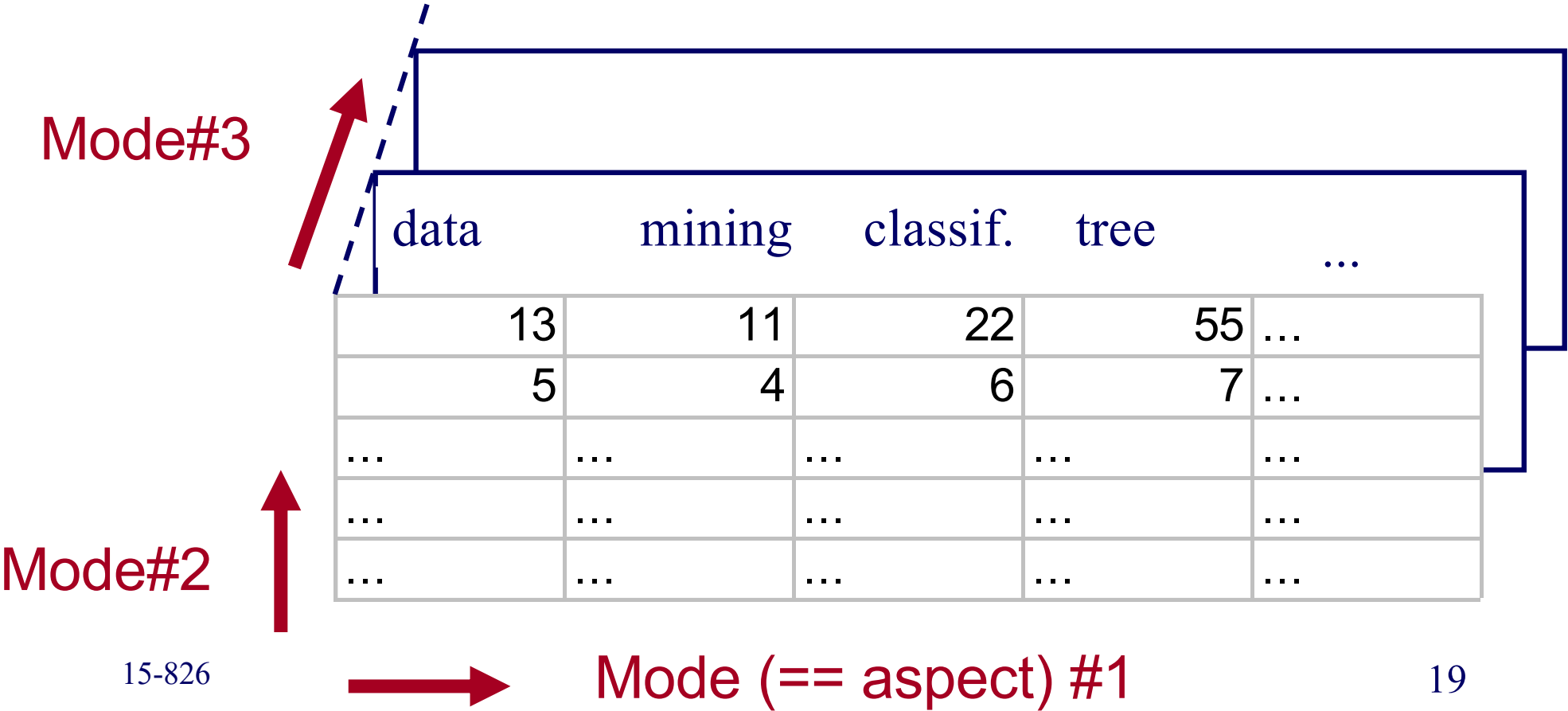
## Motivation 2: Why tensor?

- A: N-D generalization of matrix:

|         |      |        |          |      |     |
|---------|------|--------|----------|------|-----|
| KDD' 15 |      |        |          |      |     |
| KDD' 16 |      |        |          |      |     |
| KDD' 17 | data | mining | classif. | tree | ... |
| John    | 13   | 11     | 22       | 55   | ... |
| Peter   | 5    | 4      | 6        | 7    | ... |
| Mary    | ...  | ...    | ...      | ...  | ... |
| Nick    | ...  | ...    | ...      | ...  | ... |
| ...     | ...  | ...    | ...      | ...  | ... |

# Tensors are useful for 3 or more modes

Terminology: ‘mode’ (or ‘aspect’):

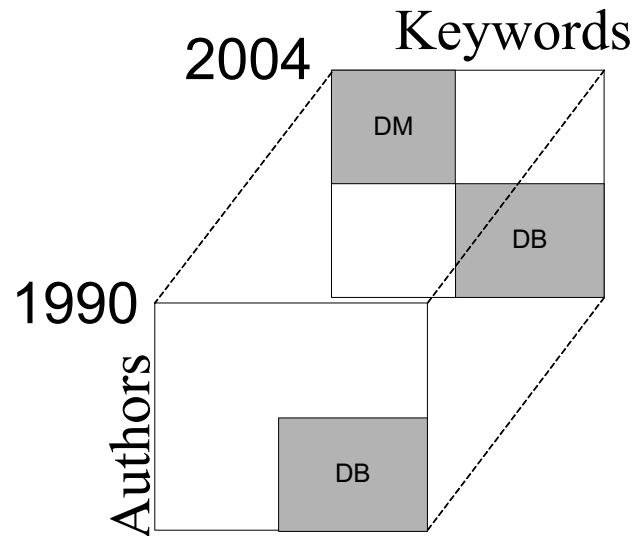


# Motivating Applications

- Why matrices are important?
- Why tensors are useful?
  - P1: social networks
  - P2: web mining

# P1: Social network analysis

- Traditionally, people focus on static networks and find community structures
- We plan to monitor the change of the community structure over time



## P2: Web graph mining

- How to order the importance of web pages?
  - Kleinberg's algorithm HITS
  - PageRank
  - Tensor extension on HITS (**TOPHITS**)
    - context-sensitive hypergraph analysis



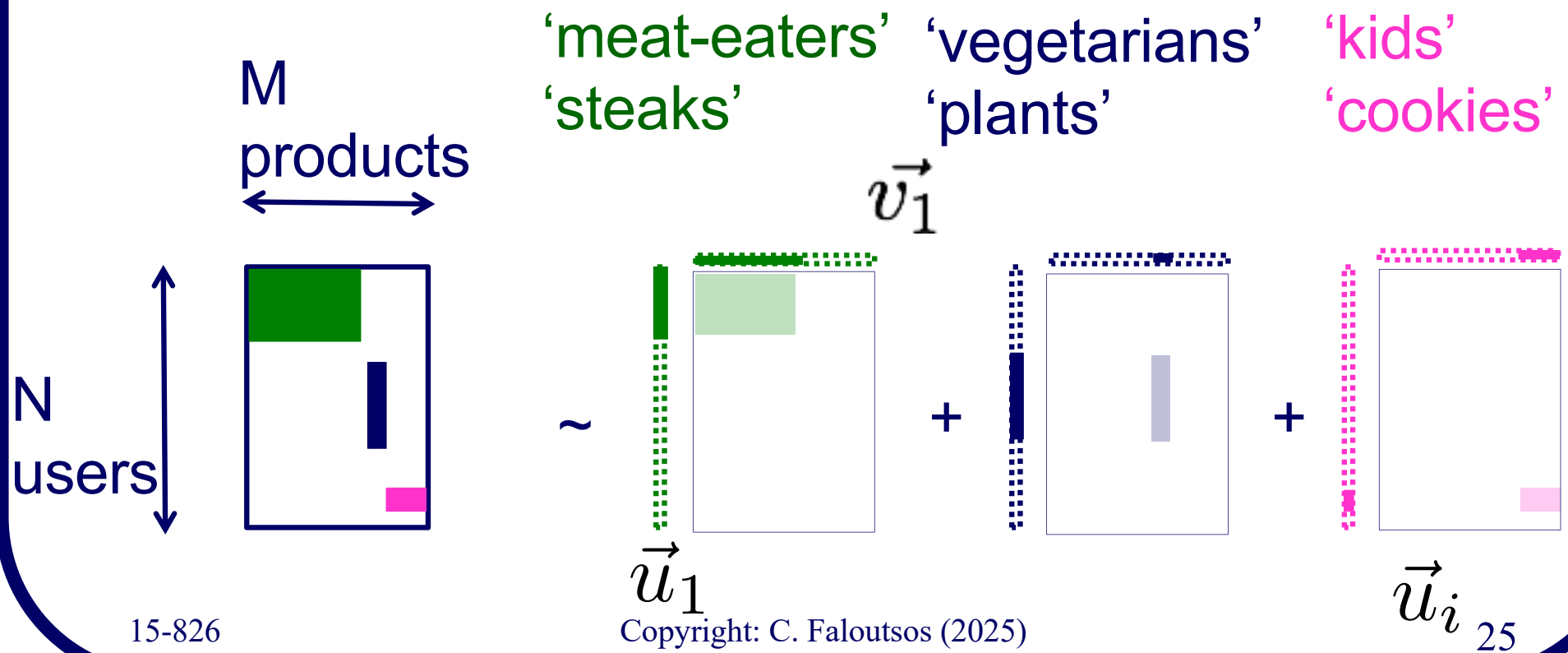
# Outline - tensors

- Motivation – Definitions
  - **Tensor tools**
  - Case studies
- Tensor Basics
  - Tucker
  - PARAFAC

# Tensor Basics

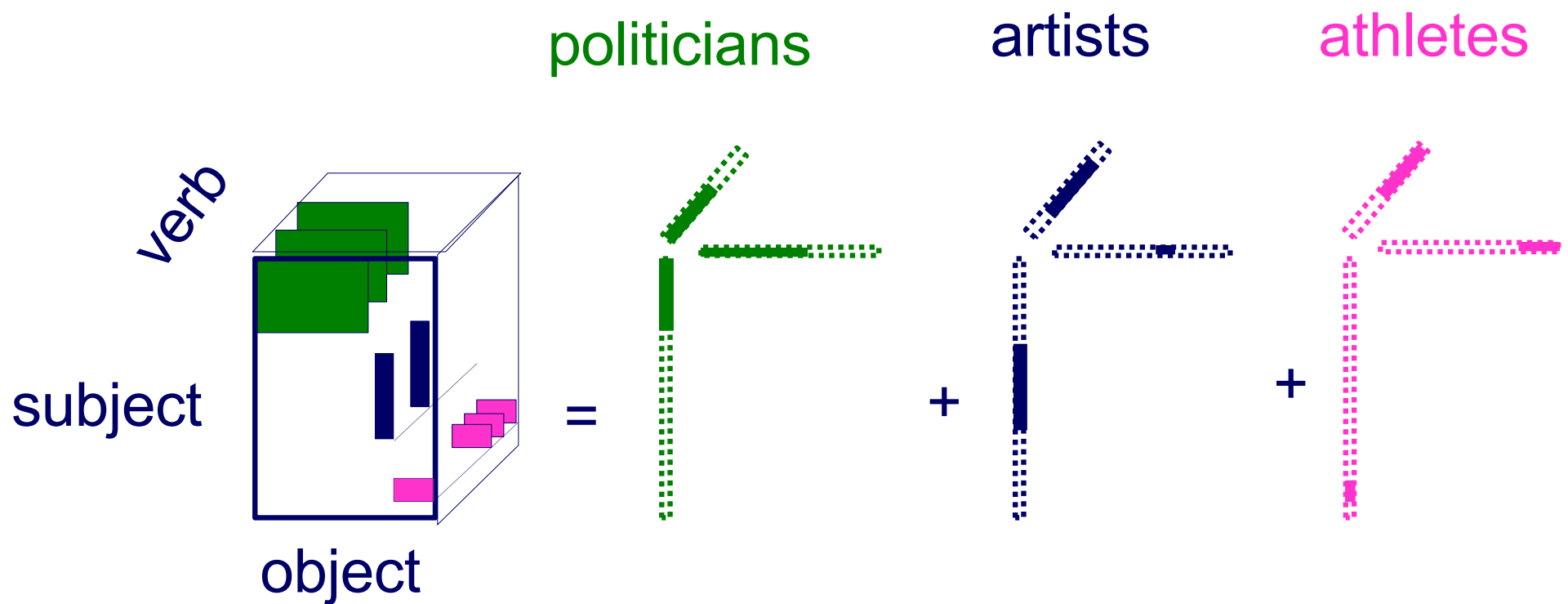
# Answer to both: tensor factorization

- Recall: (SVD) matrix factorization: finds blocks



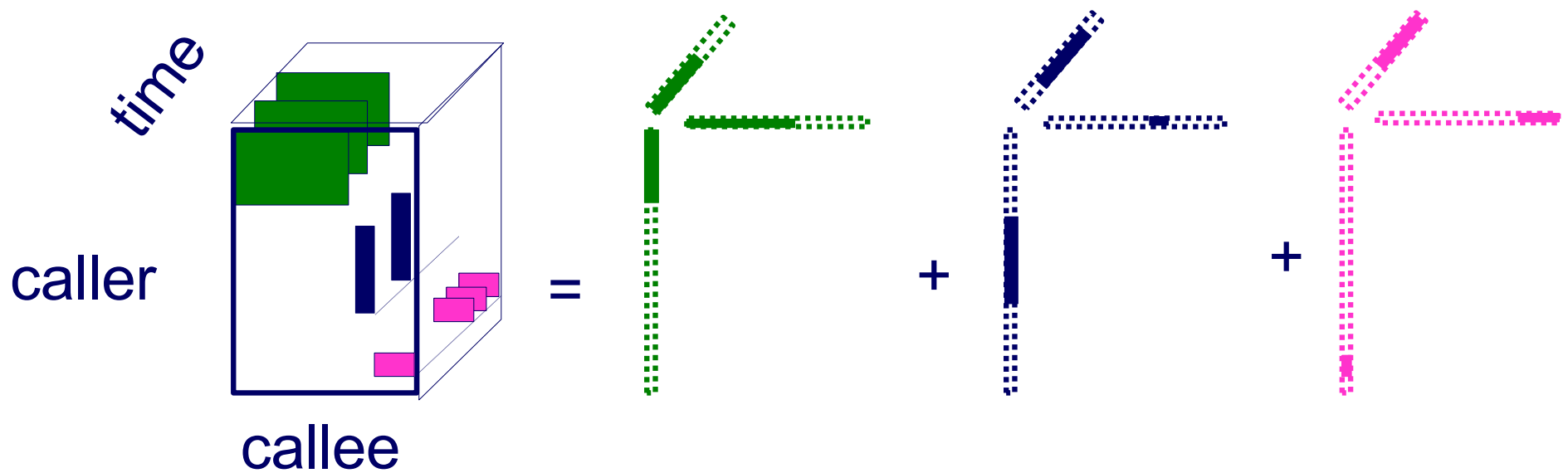
# Answer to both: tensor factorization

- PARAFAC decomposition

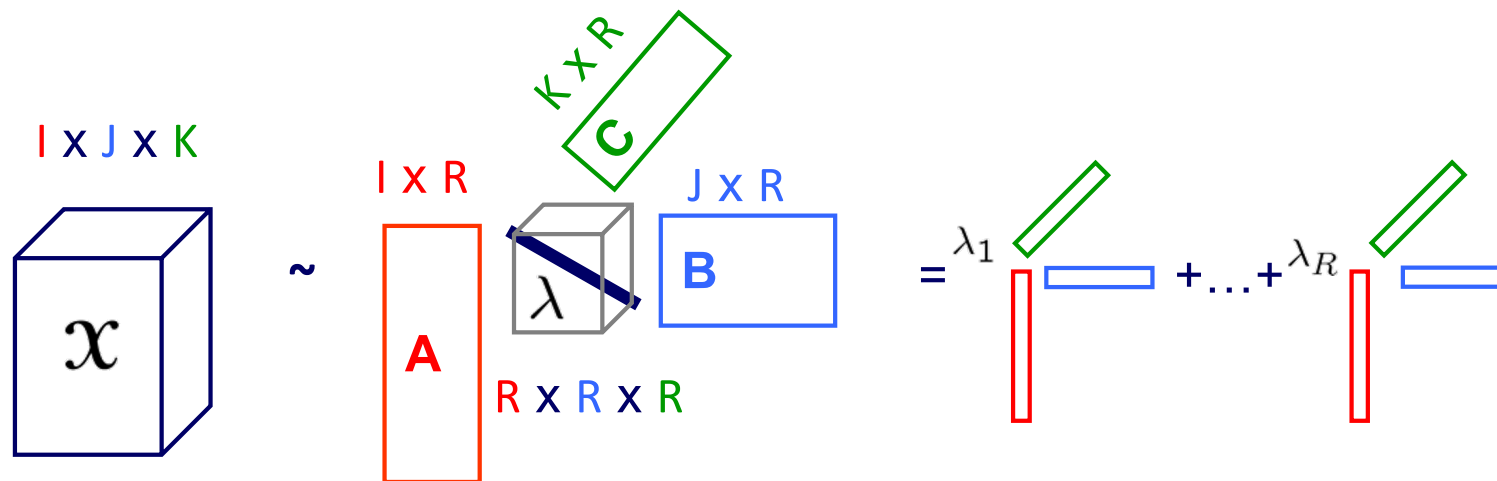


# Answer: tensor factorization

- PARAFAC decomposition
- Results for who-calls-whom-when
  - 4M x 15 days ??



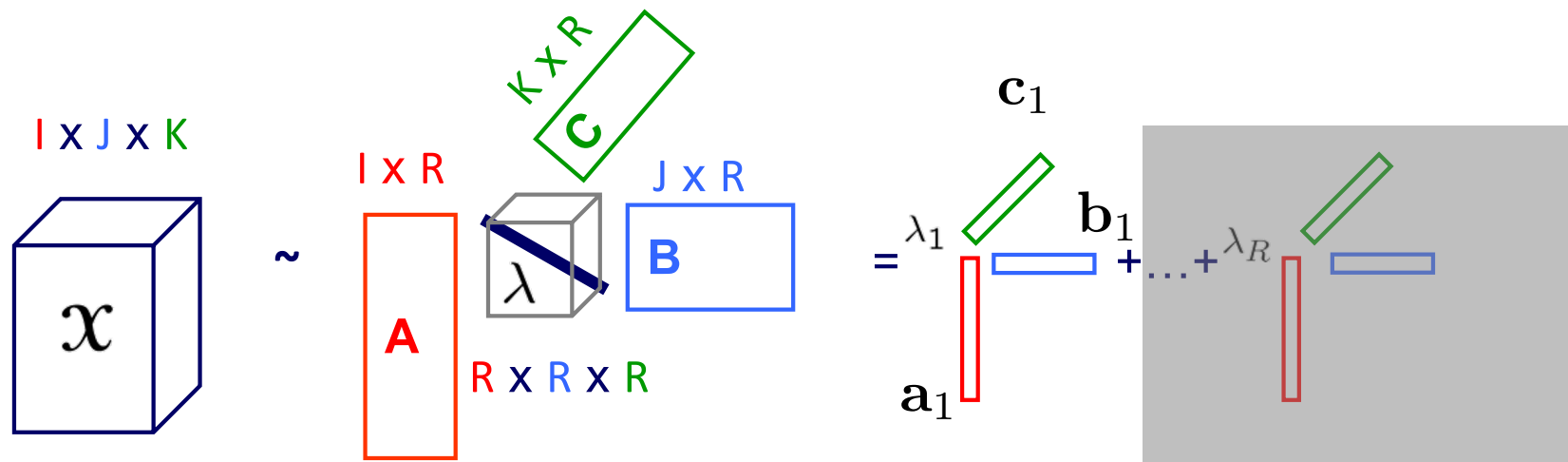
# Goal: extension to $\geq 3$ modes



$$\mathcal{X} \approx [[\lambda ; \mathbf{A}, \mathbf{B}, \mathbf{C}]] = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Example of outer product 'o':

# Goal: extension to $\geq 3$ modes



$$\mathcal{X} \approx [\lambda ; \mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Suppose

$R=1$

$\mathbf{a}_1 = (1, 2, 3, 4)$

$\mathbf{b}_1 = (2, 2, 2)$

$\mathbf{c}_1 = (10, 11)$

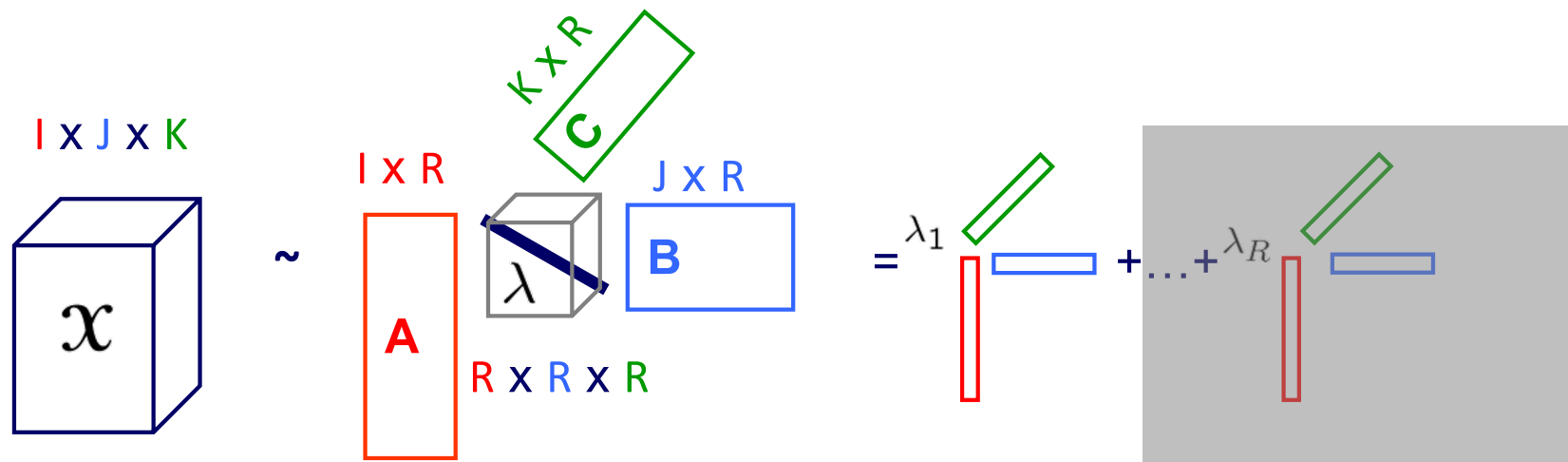
$\lambda_1 = 7$

$X(1, 1, 1) = ?$

$X(3, 1, 2) = ?$

$X(5, 1, 1) = ?$

# Goal: extension to $\geq 3$ modes



$$\mathcal{X} \approx [\lambda ; \mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Suppose

$r=1$

$\mathbf{a}_1 = (1, 2, 3, 4)$

$\mathbf{b}_1 = (2, 2, 2)$

$\mathbf{c}_1 = (10, 11)$

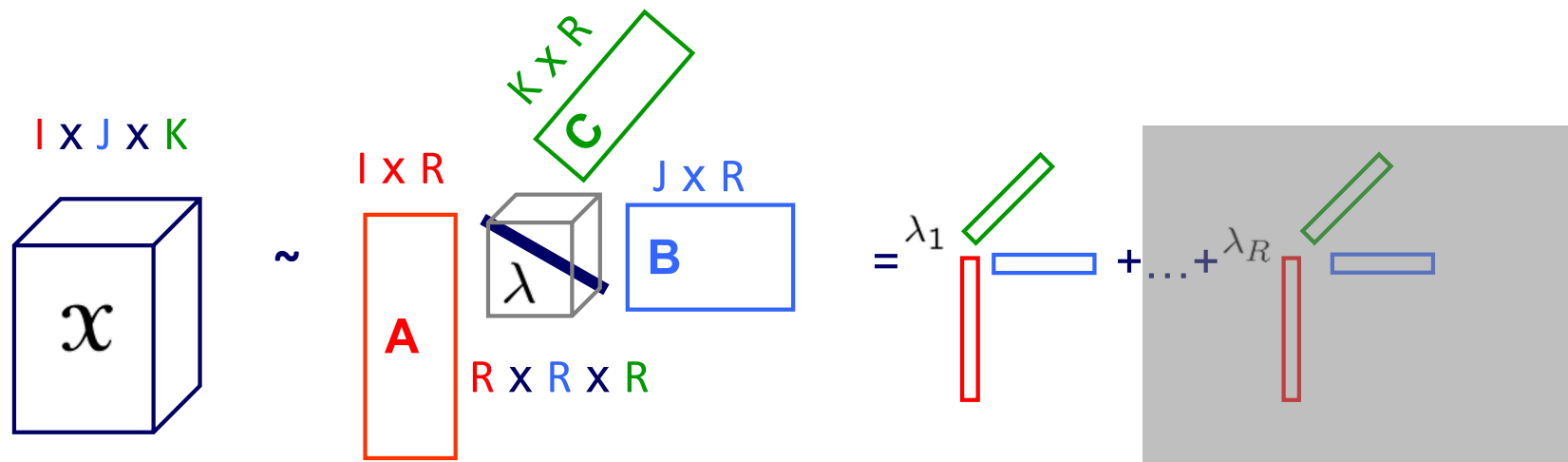
$\lambda_1 = 7$

$X(1, 1, 1) = 7 * 1 * 2 * 10$

$X(3, 1, 2) = ?$

$X(5, 1, 1) = ?$

# Goal: extension to $\geq 3$ modes



$$\mathcal{X} \approx [\lambda ; \mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Suppose

$r=1$

$\mathbf{a}_1 = (1, 2, \mathbf{3}, 4)$

$\mathbf{b}_1 = (\mathbf{2}, 2, 2)$

$\mathbf{c}_1 = (10, \mathbf{11})$

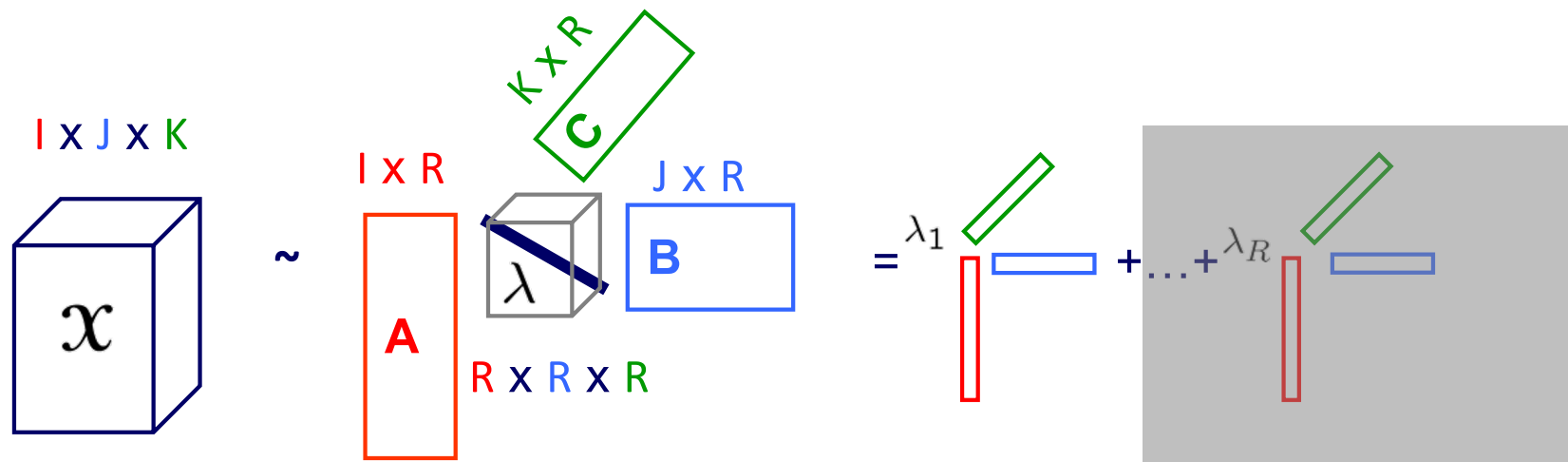
$\lambda_1 = \mathbf{7}$

$X(1, 1, 1) = 7 * 1 * 2 * 10$

$X(3, 1, 2) = 7 * 3 * 2 * 11$

$X(5, 1, 1) = ??$

# Goal: extension to $\geq 3$ modes



$$\mathcal{X} \approx [\lambda ; \mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Suppose

$r=1$

$\mathbf{a}_1 = (1, 2, 3, 4)$

$\mathbf{b}_1 = (2, 2, 2)$

$\mathbf{c}_1 = (10, 11)$

$\lambda_1 = 7$

$X(1, 1, 1) = 7 * 1 * 2 * 10$

$X(3, 1, 2) = 7 * 3 * 2 * 11$

$X(5, 1, 1) = \text{N/A} - \text{TRICK QUESTION}$

## Main points:

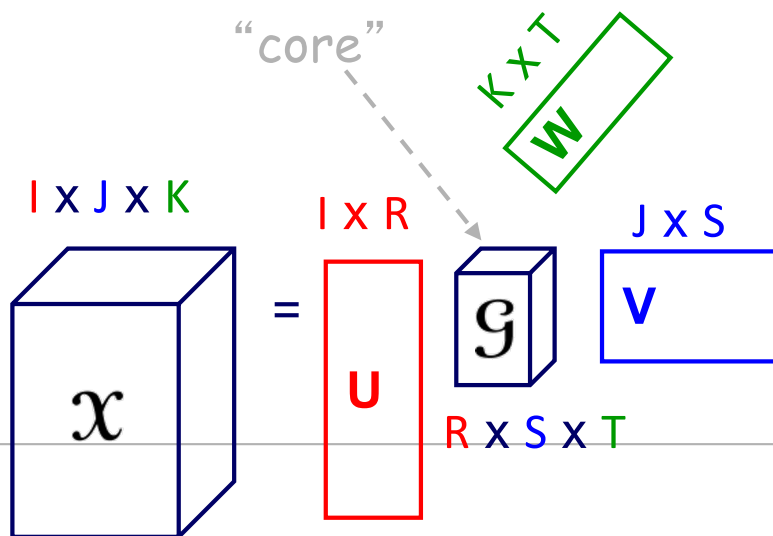
- 2 major types of tensor decompositions: PARAFAC and Tucker
- both can be solved with ``alternating least squares'' (ALS)
- Details follow

# Specially Structured Tensors

# Specially Structured Tensors

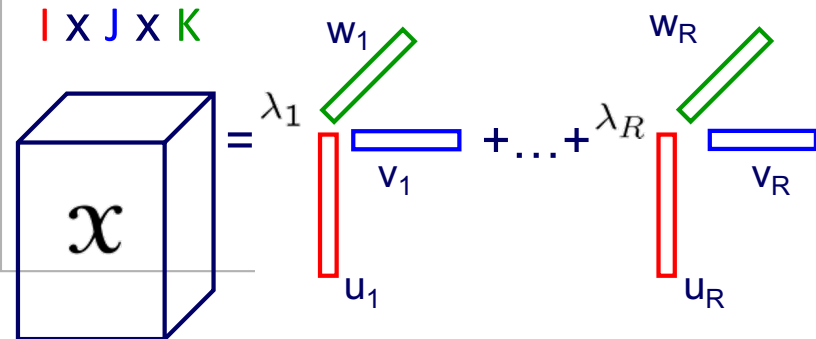
## • Tucker Tensor

$$\begin{aligned}
 \mathcal{X} &= \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} \\
 &= \sum_r \sum_s \sum_t g_{rst} \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t \\
 &\equiv [\![\mathcal{G} ; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!] \quad \left. \vphantom{\sum_r \sum_s \sum_t} \right\} \text{Our Notation}
 \end{aligned}$$



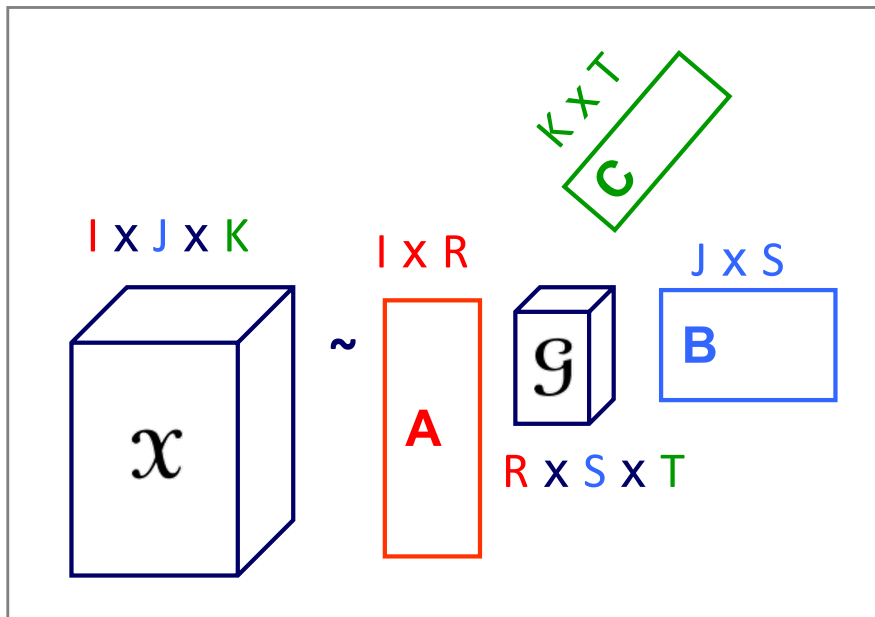
## • Kruskal Tensor

$$\begin{aligned}
 \mathcal{X} &= \sum_r \lambda_r \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r \\
 &\equiv [\![\lambda ; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!] \quad \left. \vphantom{\sum_r} \right\} \text{Our Notation}
 \end{aligned}$$



# Tensor Decompositions

# Tucker Decomposition - intuition



- author x keyword x conference
- $A$ : author x author-group
- $B$ : keyword x keyword-group
- $C$ : conf. x conf-group
- $G$ : how groups relate to each other

Needs elaboration!

## Intuition behind core tensor

- 2-d case: co-clustering
- [Dhillon et al. [Information-Theoretic Co-clustering](#), KDD' 03]

$$\overline{\hspace{1cm}}$$

$$\underline{\hspace{1cm}}$$
 $n$

$$m \begin{bmatrix} .05 & .05 & .05 & 0 & 0 & 0 \\ .05 & .05 & .05 & 0 & 0 & 0 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ .04 & .04 & 0 & .04 & .04 & .04 \\ .04 & .04 & .04 & 0 & .04 & .04 \end{bmatrix}$$

$|$   
 $|$   
 $|$

eg, terms x documents

$$\begin{matrix} & k & & & l & & n \\ m & \begin{bmatrix} .5 & 0 & 0 \\ .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & .5 \end{bmatrix} & k & \begin{bmatrix} .3 & 0 \\ 0 & .3 \\ .2 & .2 \end{bmatrix} & l & \begin{bmatrix} .36 & .36 & .28 & 0 & 0 & 0 \\ 0 & 0 & 0 & .28 & .36 & .36 \end{bmatrix} & = & \begin{bmatrix} .054 & .054 & .042 & | & 0 & 0 & 0 \\ .054 & .054 & .042 & | & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & | & .042 & .054 & .054 \\ 0 & 0 & 0 & | & .042 & .054 & .054 \\ \hline .036 & .036 & .028 & | & .028 & .036 & .036 \\ .036 & .036 & .028 & | & .028 & .036 & .036 \end{bmatrix}
 \end{matrix}$$

med. doc  
           cs doc

term group x  
 doc. group

$$\begin{bmatrix} .05 & .05 & .05 & 0 & 0 & 0 \\ .05 & .05 & .05 & 0 & 0 & 0 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ .04 & .04 & 0 & .04 & .04 & .04 \\ .04 & .04 & .04 & 0 & .04 & .04 \end{bmatrix}$$

| med. terms

| cs terms

| common terms

$$\begin{bmatrix} .5 & 0 & 0 \\ .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & .5 \end{bmatrix}$$

$$\begin{bmatrix} .3 & 0 \\ 0 & .3 \\ .2 & .2 \end{bmatrix}$$

$$\begin{bmatrix} .36 & .36 & .28 & 0 & 0 & 0 \\ 0 & 0 & 0 & .28 & .36 & .36 \end{bmatrix} =$$

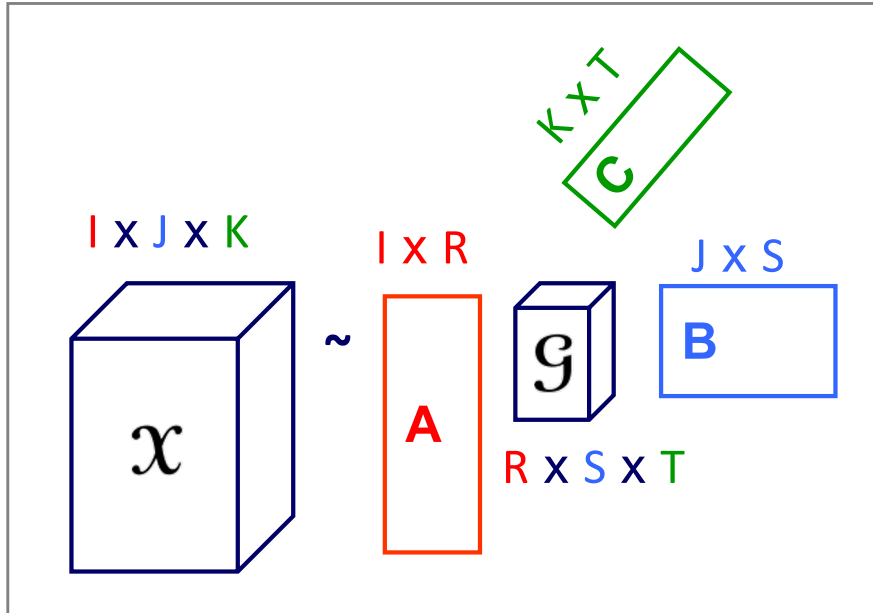
doc x  
 doc group

$$\begin{bmatrix} .054 & .054 & .042 & 0 & 0 & 0 \\ .054 & .054 & .042 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & .042 & .054 & .054 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ \hline .036 & .036 & .028 & .028 & .036 & .036 \\ .036 & .036 & .028 & .028 & .036 & .036 \end{bmatrix}$$

term x  
 term-group



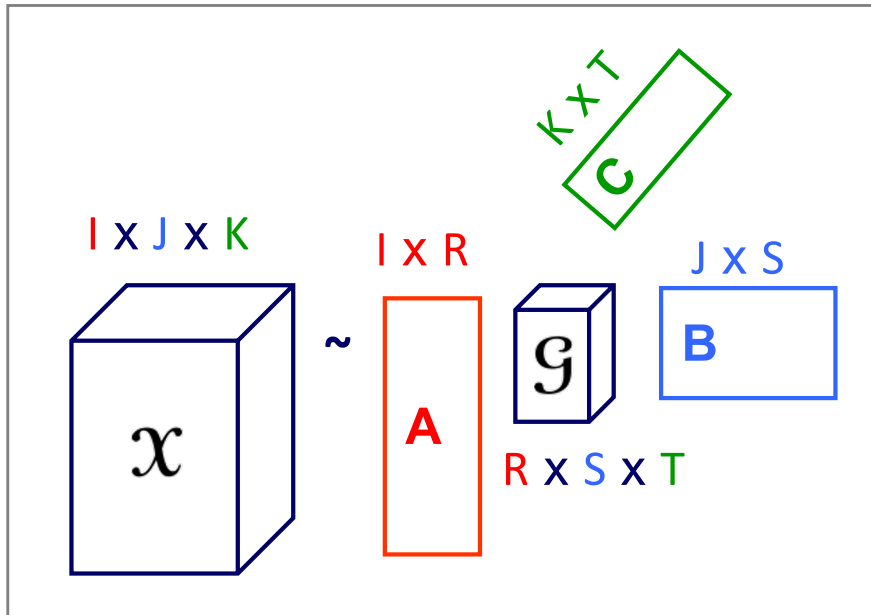
# Tucker Decomposition



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

- Proposed by Tucker (1966)
- AKA: Three-mode factor analysis, three-mode PCA, orthogonal array decomposition
- $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  generally assumed to be orthonormal (generally assume they have full column rank)
- $\mathcal{G}$  is not diagonal
- Not unique

# Tucker Decomposition



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

Given  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , the optimal core is:

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

- Proposed by Tucker (1966)
- AKA: Three-mode factor analysis, three-mode PCA, orthogonal array decomposition
- $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  generally assumed to be orthonormal (generally assume they have full column rank)
- $\mathcal{G}$  is not diagonal
- Not unique

Recall the equations for converting a tensor to a matrix

$$\mathbf{X}_{(1)} = \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^\top$$

$$\mathbf{X}_{(2)} = \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^\top$$

$$\mathbf{X}_{(3)} = \mathbf{C}\mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^\top$$

$$\text{vec}(\mathcal{X}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\text{vec}(\mathcal{G})$$

# Kronecker product

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 10 & 20 & 30 \end{bmatrix}$$

$m1 \times n1$   $m2 \times n2$

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} 1 * \mathbf{B} & 2 * \mathbf{B} \\ 3 * \mathbf{B} & 4 * \mathbf{B} \end{bmatrix}$$

$$= \begin{bmatrix} 1 * 10 & 1 * 20 & 1 * 30 & 2 * 10 & 2 * 20 & 2 * 30 \\ 3 * 10 & 3 * 20 & 3 * 30 & 4 * 10 & 4 * 20 & 4 * 30 \end{bmatrix}$$

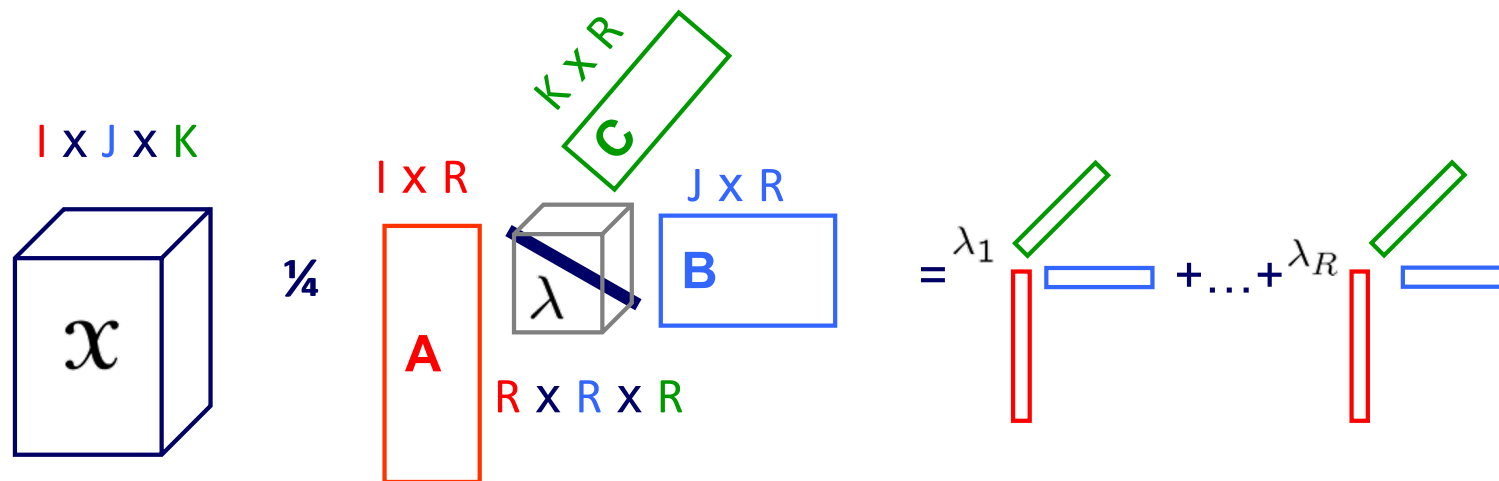
$m1*m2 \times n1*n2$



# Outline - tensors

- Motivation –  
Definitions
  - **Tensor tools**
  - Case studies
- Tensor Basics
  - Tucker
  - **PARAFAC**

# CANDECOMP/PARAFAC Decomposition



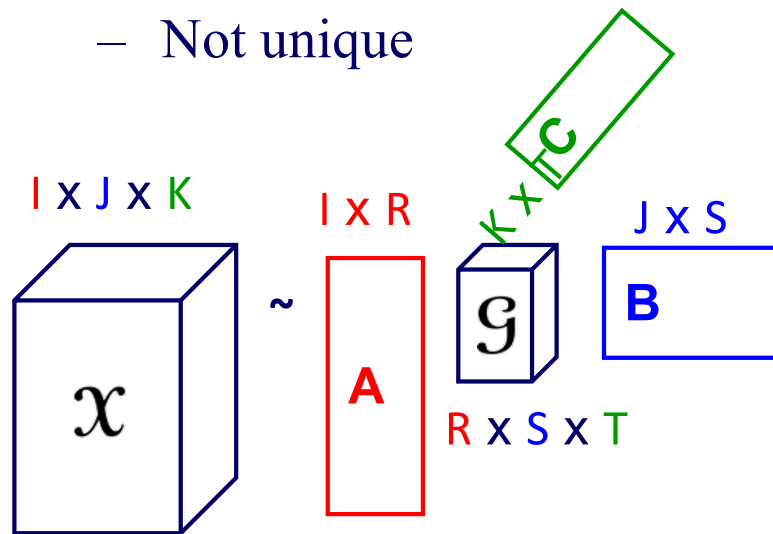
$$\mathcal{X} \approx [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]] = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

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

# Tucker vs. PARAFAC Decompositions

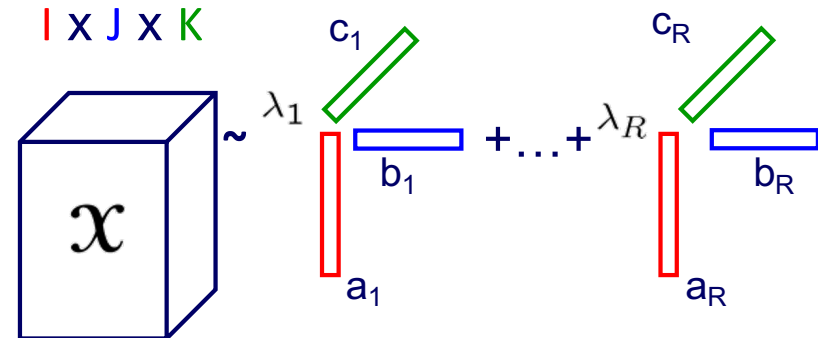
- Tucker

- Variable transformation in each mode
- Core  $G$  may be dense
- $A, B, C$  generally orthonormal
- Not unique



- PARAFAC

- Sum of rank-1 components
- No core, i.e., superdiagonal core
- $A, B, C$  may have linearly dependent columns
- Generally unique



## Tensor tools - summary

- Two main tools
  - PARAFAC
  - Tucker
- Both find row-, column-, tube-groups
  - but in PARAFAC the three groups are identical
- To solve: Alternating Least Squares
- Toolbox: from Tamara Kolda:  
<http://www.tensortoolbox.org/>

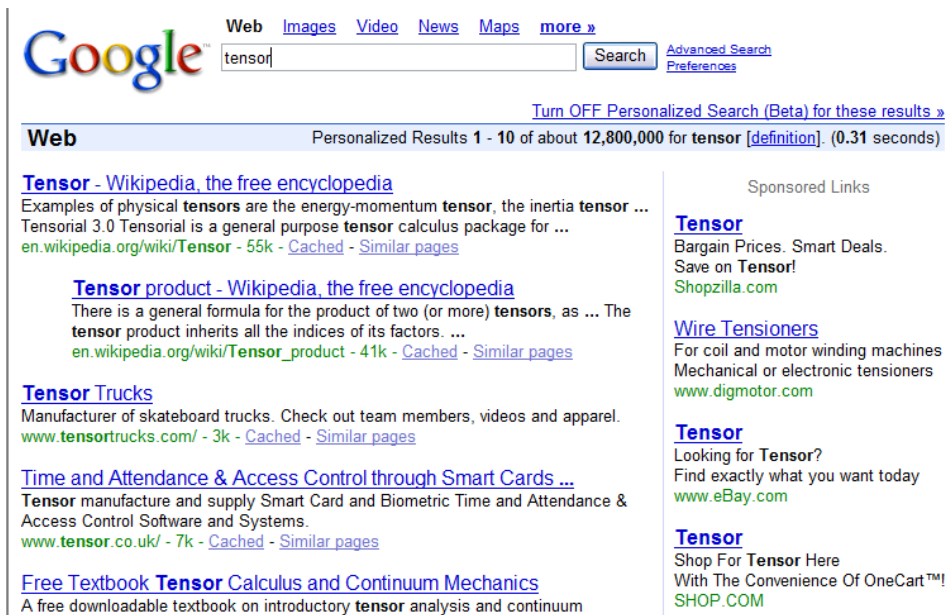


## Outline - tensors

- Motivation - Definitions
- Tensor tools
- Case studies
  - ➔ – P1: web graph mining (‘TOPHITS’)
  - P2: phone-call patterns
  - P3: network traffic
  - P4: FaceBook activity

# P1: Web graph mining

- How to order the importance of web pages?
  - Kleinberg's algorithm HITS
  - PageRank



Google Web Images Video News Maps more »

tensor Search Advanced Search Preferences

Turn OFF Personalized Search (Beta) for these results »

**Web** Personalized Results 1 - 10 of about 12,800,000 for tensor [definition]. (0.31 seconds)

**Tensor** - Wikipedia, the free encyclopedia  
Examples of physical **tensors** are the energy-momentum **tensor**, the inertia **tensor** ...  
Tensorial 3.0 Tensorial is a general purpose **tensor** calculus package for ...  
[en.wikipedia.org/wiki/Tensor](http://en.wikipedia.org/wiki/Tensor) - 55k - [Cached](#) - [Similar pages](#)

**Tensor product** - Wikipedia, the free encyclopedia  
There is a general formula for the product of two (or more) **tensors**, as ... The **tensor** product inherits all the indices of its factors. ...  
[en.wikipedia.org/wiki/Tensor\\_product](http://en.wikipedia.org/wiki/Tensor_product) - 41k - [Cached](#) - [Similar pages](#)

**Tensor Trucks**  
Manufacturer of skateboard trucks. Check out team members, videos and apparel.  
[www.tensortrucks.com/](http://www.tensortrucks.com/) - 3k - [Cached](#) - [Similar pages](#)

**Time and Attendance & Access Control through Smart Cards ...**  
Tensor manufacture and supply Smart Card and Biometric Time and Attendance & Access Control Software and Systems.  
[www.tensor.co.uk/](http://www.tensor.co.uk/) - 7k - [Cached](#) - [Similar pages](#)

**Free Textbook Tensor Calculus and Continuum Mechanics**  
A free downloadable textbook on introductory **tensor** analysis and continuum

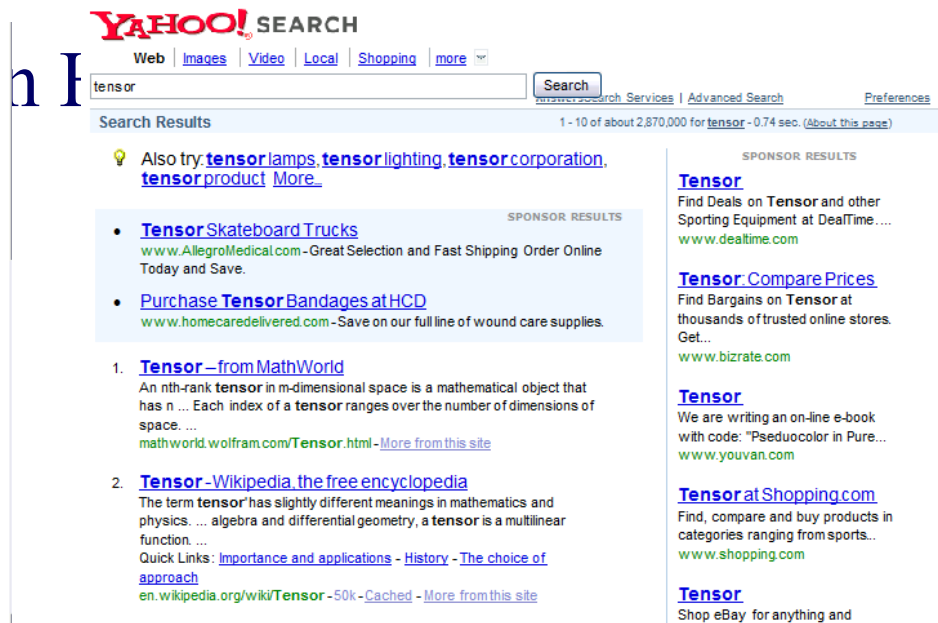
**Sponsored Links**

**Tensor**  
Bargain Prices. Smart Deals. Save on **Tensor**!  
[Shopzilla.com](http://Shopzilla.com)

**Wire Tensioners**  
For coil and motor winding machines Mechanical or electronic tensioners  
[www.digmotor.com](http://www.digmotor.com)

**Tensor**  
Looking for **Tensor**? Find exactly what you want today  
[www.eBay.com](http://www.eBay.com)

**Tensor**  
Shop For **Tensor** Here With The Convenience Of OneCart™!  
[SHOP.COM](http://SHOP.COM)



YAHOO! SEARCH

Web Images Video Local Shopping more »

tensor Search Advanced Search Preferences

Search Results 1 - 10 of about 2,870,000 for tensor - 0.74 sec. (About this page)

Also try: [tensor lamps](#), [tensor lighting](#), [tensor corporation](#), [tensor product](#) More...

**SPONSOR RESULTS**

- Tensor Skateboard Trucks**  
[www.AllegroMedical.com](http://www.AllegroMedical.com) - Great Selection and Fast Shipping Order Online Today and Save.
- Purchase Tensor Bandages at HCD**  
[www.homecaredelivered.com](http://www.homecaredelivered.com) - Save on our full line of wound care supplies.

- Tensor - from MathWorld**  
An  $n$ th-rank **tensor** in  $m$ -dimensional space is a mathematical object that has  $n$  ... Each index of a **tensor** ranges over the number of dimensions of space. ...  
[mathworld.wolfram.com/Tensor.html](http://mathworld.wolfram.com/Tensor.html) - [More from this site](#)
- Tensor - Wikipedia, the free encyclopedia**  
The term **tensor** has slightly different meanings in mathematics and physics. ... algebra and differential geometry, a **tensor** is a multilinear function. ...  
Quick Links: [Importance and applications](#) - [History](#) - [The choice of approach](#)  
[en.wikipedia.org/wiki/Tensor](http://en.wikipedia.org/wiki/Tensor) - 50k - [Cached](#) - [More from this site](#)

**SPONSOR RESULTS**

**Tensor**  
Find Deals on **Tensor** and other Sporting Equipment at DealTime....  
[www.dealtime.com](http://www.dealtime.com)

**Tensor Compare Prices**  
Find Bargains on **Tensor** at thousands of trusted online stores. Get...  
[www.bizrate.com](http://www.bizrate.com)

**Tensor**  
We are writing an on-line e-book with code: "Pseudocolor in Pure...  
[www.youvan.com](http://www.youvan.com)

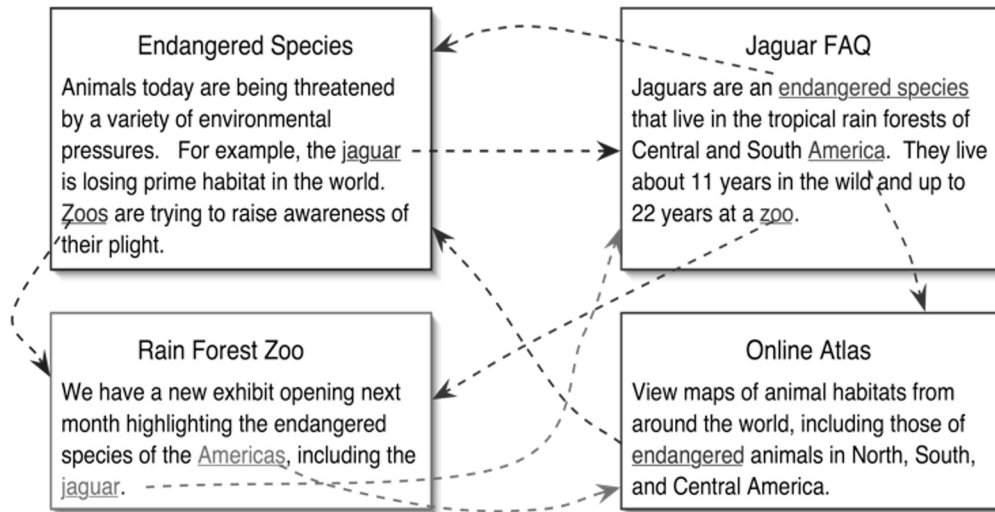
**Tensor at Shopping.com**  
Find, compare and buy products in categories ranging from sports...  
[www.shopping.com](http://www.shopping.com)

**Tensor**  
Shop eBay for anything and

## P1: Web graph mining

- T. G. Kolda, B. W. Bader and J. P. Kenny, *Higher-Order Web Link Analysis Using Multilinear Algebra*, ICDM 2005: ICDM, pp. 242-249, November 2005, [doi:10.1109/ICDM.2005.77](https://doi.org/10.1109/ICDM.2005.77). [[PDF](#)]

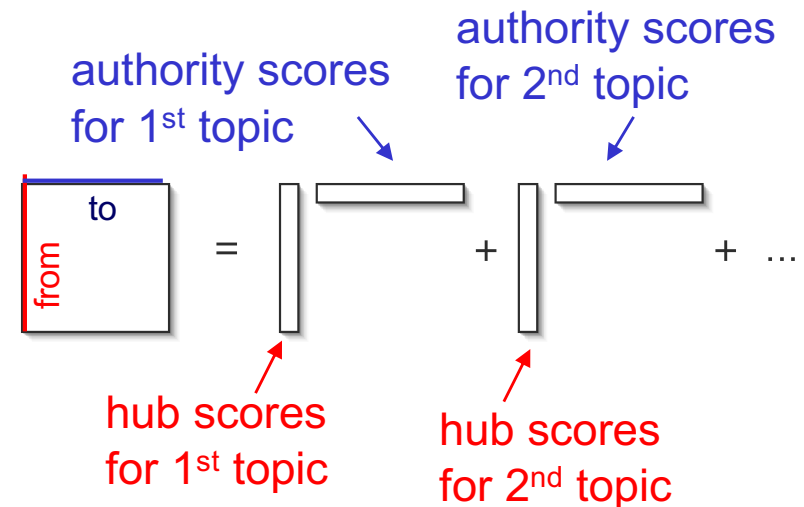
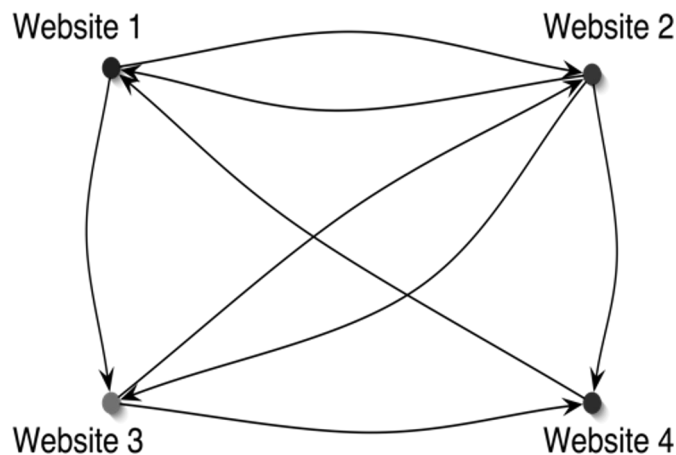
# Kleinberg's Hubs and Authorities (the HITS method)



Sparse adjacency matrix and its SVD:

$$x_{ij} = \begin{cases} 1 & \text{if page } i \text{ links to page } j \\ 0 & \text{otherwise} \end{cases}$$

$$X \approx \sum_r \sigma_r \mathbf{h}_r \circ \mathbf{a}_r$$



# HITS Authorities on Sample Data

| 1st Principal Factor |                  |
|----------------------|------------------|
| .97                  | www.ibm.com      |
| .24                  | www.alphaweb.com |
| .08                  | www-128.ibm.com  |
| .05                  | www.develop.com  |
| .02                  | www.research.com |
| .01                  | www.redbook.com  |
| .01                  | news.com.com     |

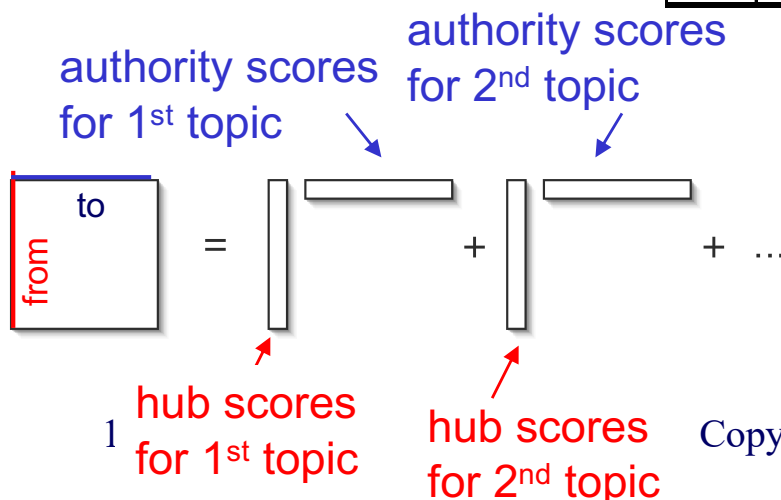
| 2nd Principal Factor |                    |
|----------------------|--------------------|
| .99                  | www.lehigh.edu     |
| .11                  | www2.lehigh.edu    |
| .06                  | www.lehigh.edu     |
| .06                  | www.lehigh.edu     |
| .02                  | www.bethlehem.edu  |
| .02                  | www.adobe.com      |
| .02                  | lewisweb.cc        |
| .02                  | www.leo.lehigh.edu |
| .02                  | www.distance.edu   |
| .02                  | fp1.cc.lehigh.edu  |

| 3rd Principal Factor |                     |
|----------------------|---------------------|
| .75                  | java.sun.com        |
| .38                  | www.sun.com         |
| .36                  | developers.sun.com  |
| .24                  | see.sun.com         |
| .16                  | www.samag.com       |
| .13                  | docs.sun.com        |
| .12                  | blogs.sun.com       |
| .08                  | sunsolve.sun.com    |
| .08                  | www.sun-catalog.com |
| .08                  | news.com.com        |

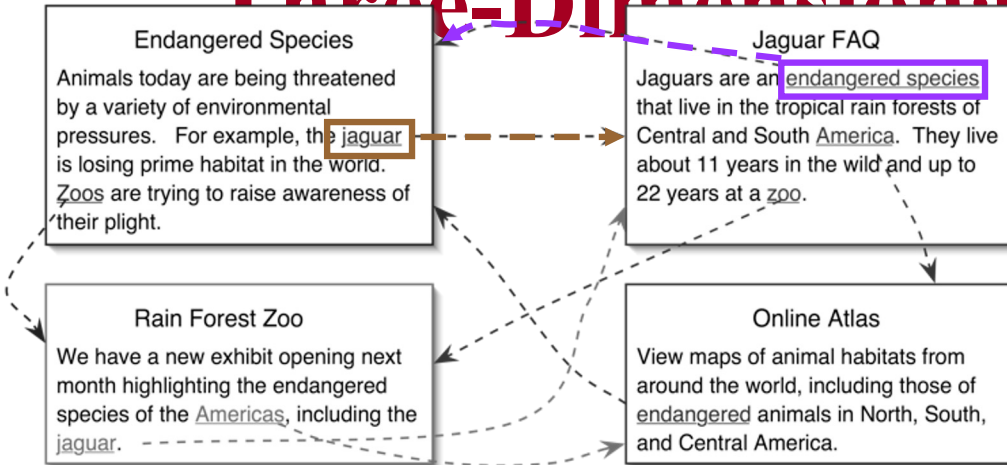
| 4th Principal Factor |                    |
|----------------------|--------------------|
| .60                  | www.pueblo.gsa.gov |
| .45                  | www.whitehouse.gov |
| .35                  | www.irs.gov        |
| .31                  | travel.state.gov   |
| .22                  | www.gsa.gov        |
| .20                  | www.ssa.gov        |
| .16                  | www.census.gov     |
| .14                  | www.govbeat.com    |
| .13                  | www.kids.gov       |
| .13                  | www.usdoj.gov      |

We started our crawl from  
<http://www-neos.mcs.anl.gov/neos>,  
 and crawled 4700 pages,  
 resulting in 560  
 cross-linked hosts.

| 6th Principal Factor |                       |
|----------------------|-----------------------|
| .97                  | mathpost.asu.edu      |
| .18                  | math.la.asu.edu       |
| .17                  | www.asu.edu           |
| .04                  | www.act.org           |
| .03                  | www.eas.asu.edu       |
| .02                  | archives.math.utk.edu |
| .02                  | www.geom.uiuc.edu     |
| .02                  | www.fulton.asu.edu    |
| .02                  | www.amstat.org        |
| .02                  | www.maa.org           |

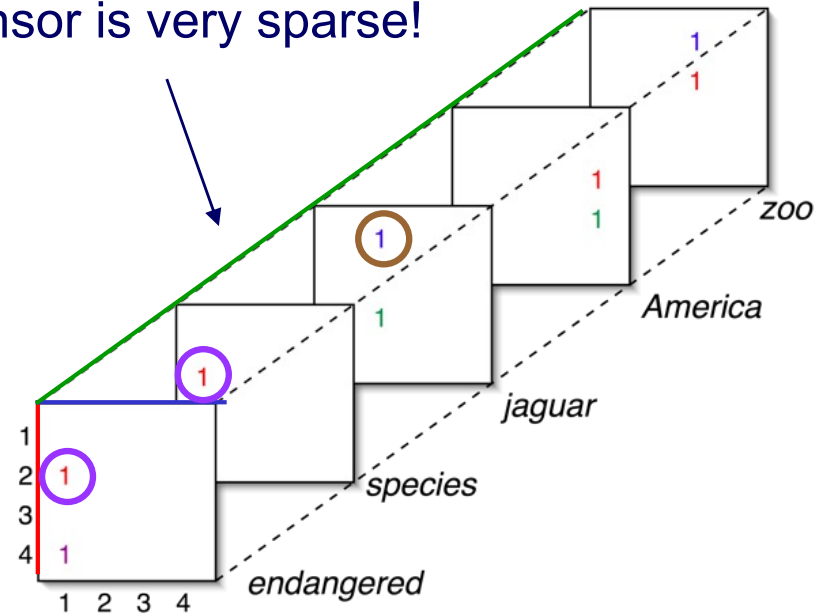
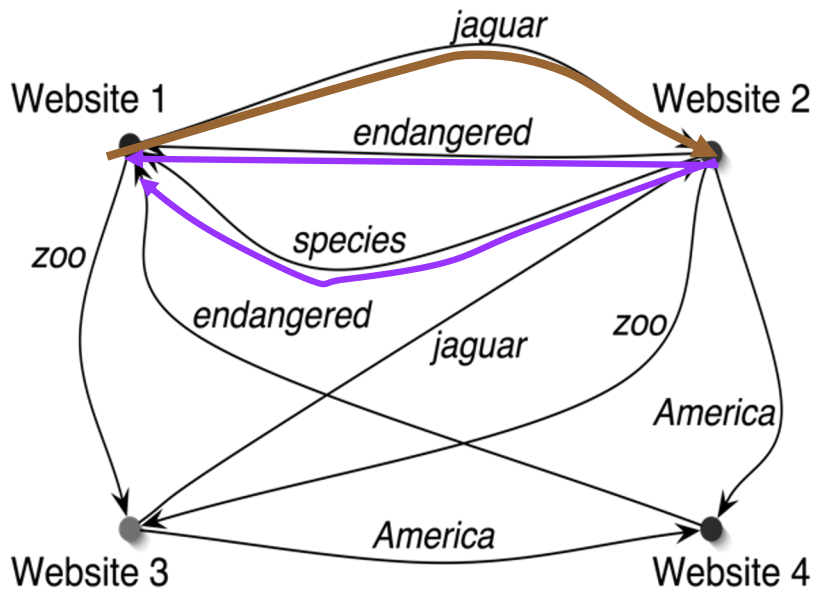


# Three-Dimensional View of the Web



$$x_{ijk} = \begin{cases} 1 & \text{if page } i \rightarrow \text{page } j \\ & \text{with term } k \\ 0 & \text{otherwise} \end{cases}$$

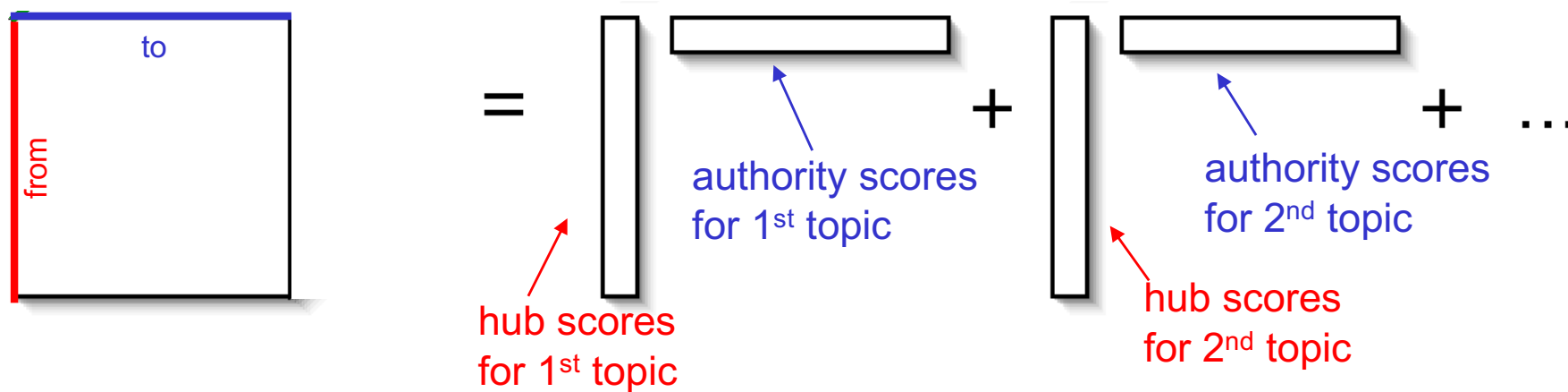
Observe that this tensor is very sparse!



# Topical HITS (TOPHITS)

**Main Idea:** Extend the idea behind the HITS model to incorporate term (i.e., topical) information.

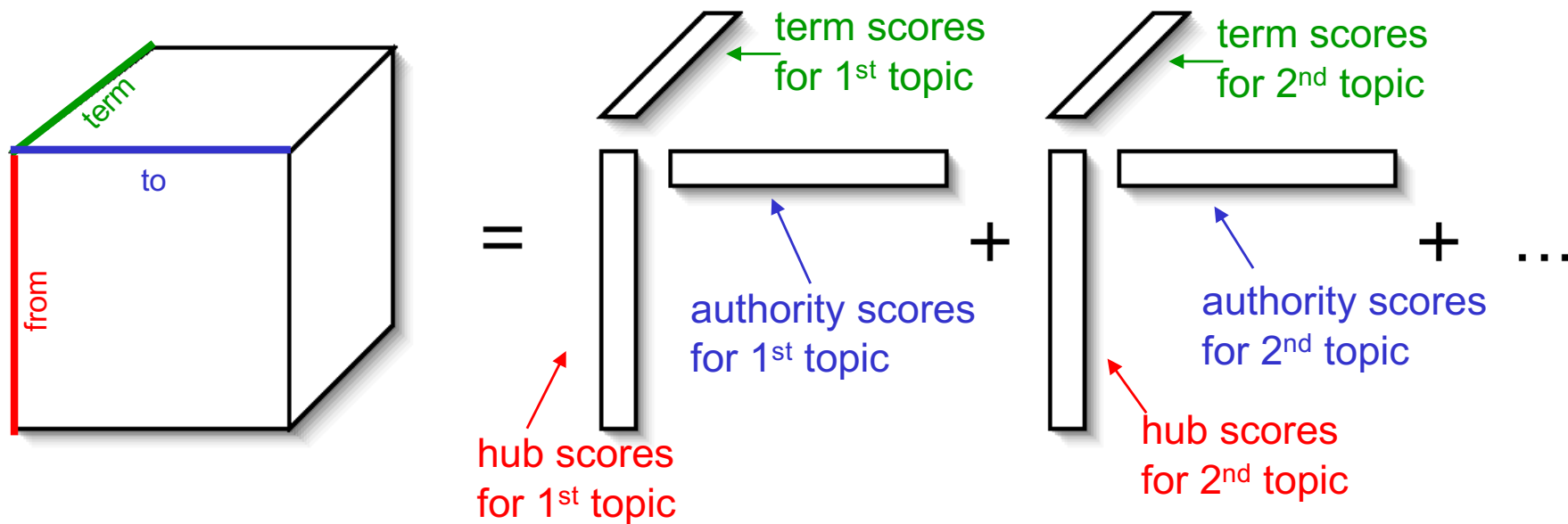
$$\mathbf{x} \approx \sum_{r=1}^R \lambda_r \mathbf{h}_r \circ \mathbf{a}_r$$



# Topical HITS (TOPHITS)

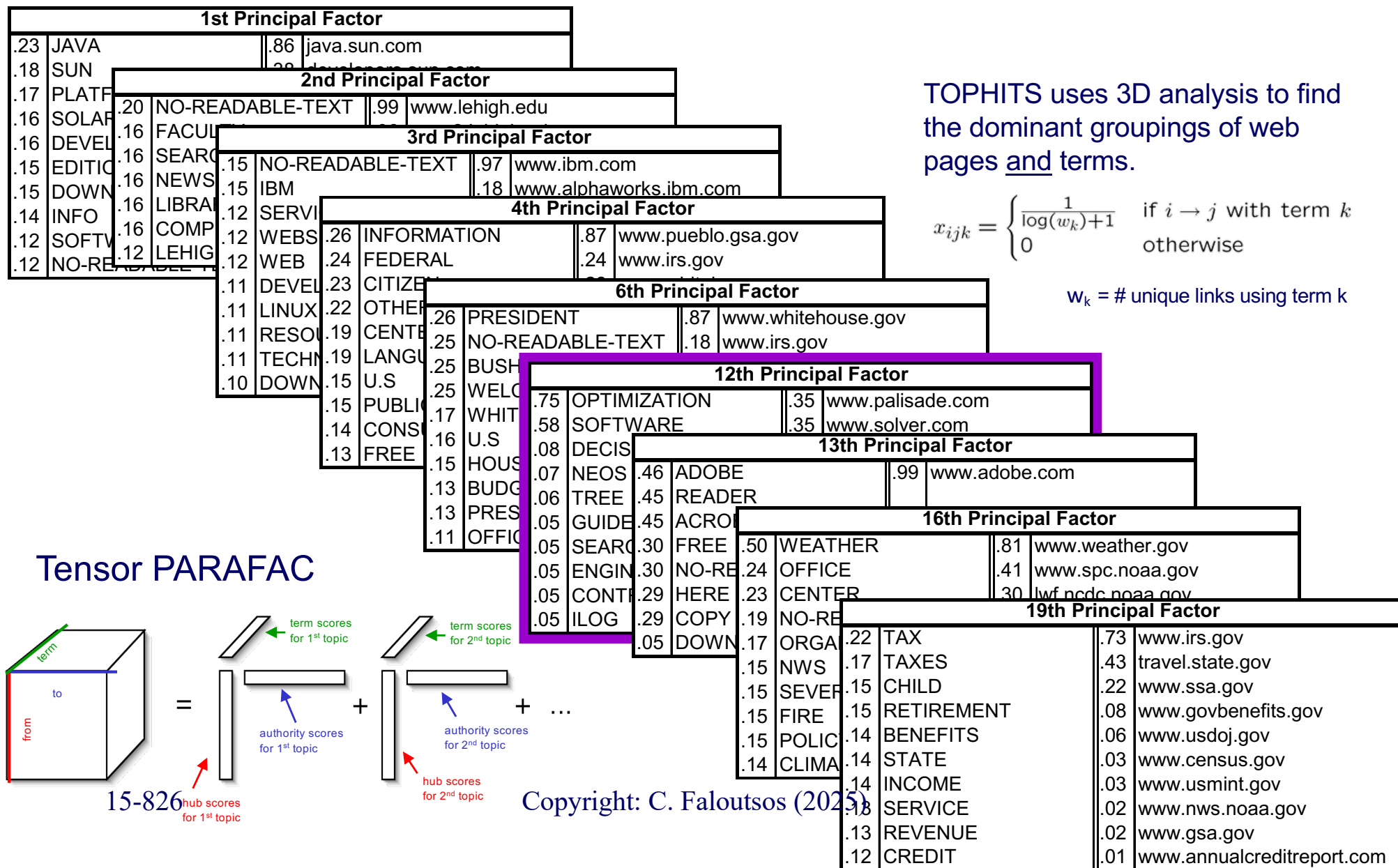
**Main Idea:** Extend the idea behind the HITS model to incorporate term (i.e., topical) information.

$$\mathbf{x} \approx \sum_{r=1}^R \lambda_r \mathbf{h}_r \circ \mathbf{a}_r \circ \mathbf{t}_r$$



# TOPHITS Terms & Authorities

## on Sample Data

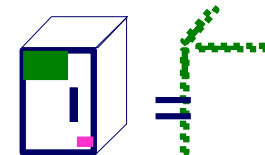




## Outline - tensors

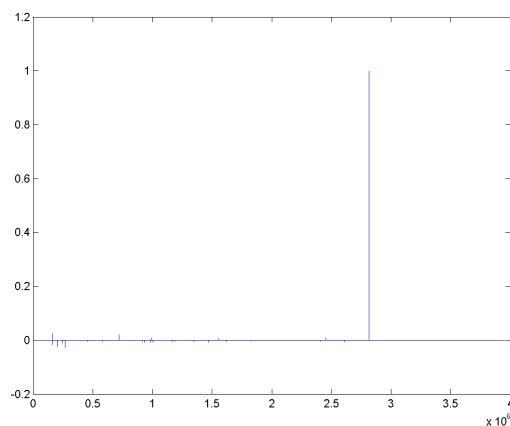
- Motivation - Definitions
- Tensor tools
- Case studies
  - P1: web graph mining ('TOPHITS')
  - ➔ – P2: phone-call patterns
  - P3: network traffic
  - P4: FaceBook activity

# P2: Anomaly detection in time-evolving graphs

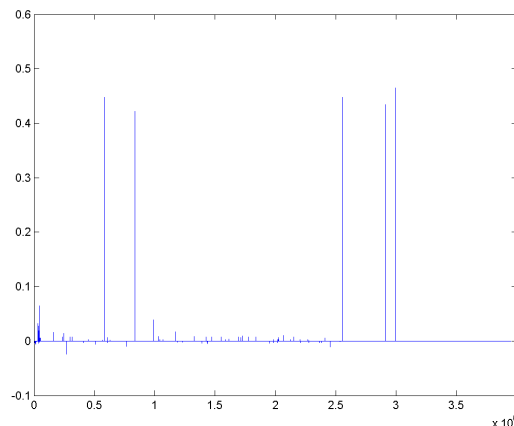


- Anomalous communities in phone call data:
  - European country, 4M clients, data over 2 weeks

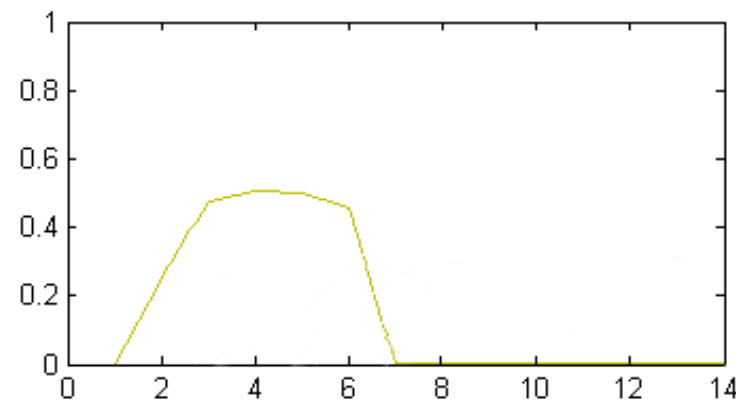
1 caller



5 receivers

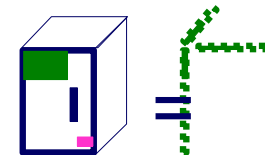


4 days of activity



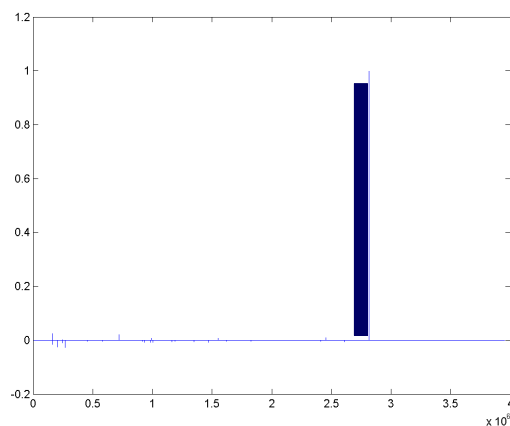
~200 calls to EACH receiver on EACH day!

# P2: Anomaly detection in time-evolving graphs

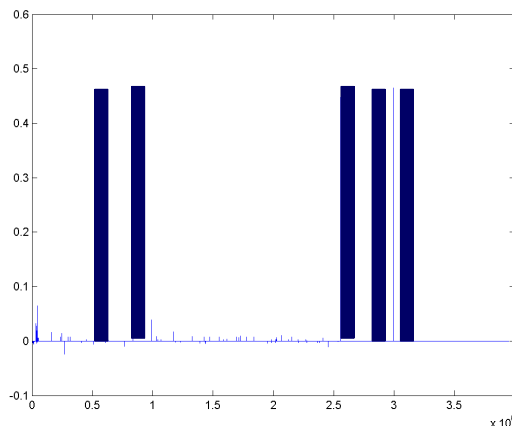


- Anomalous communities in phone call data:
  - European country, 4M clients, data over 2 weeks

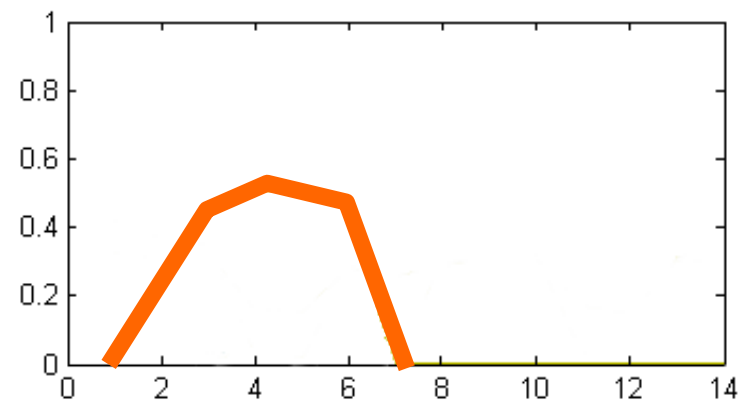
1 caller



5 receivers

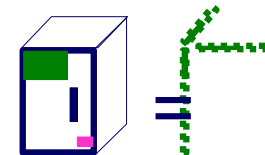


4 days of activity



~200 calls to EACH receiver on EACH day!

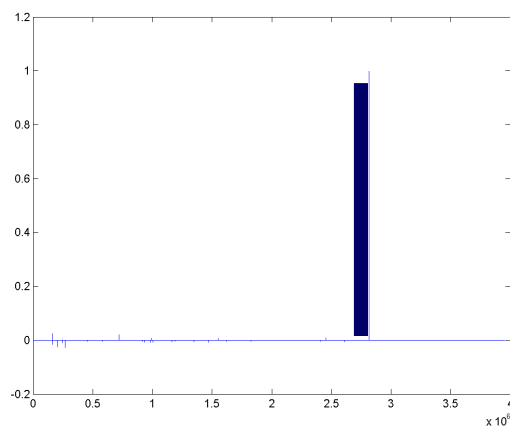
# P2: Anomaly detection in time-evolving graphs



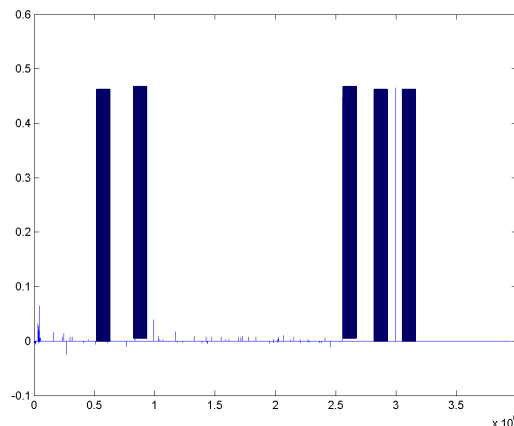
- Anomalous communities in phone call data:
  - European country, 4M clients, data over 2 weeks



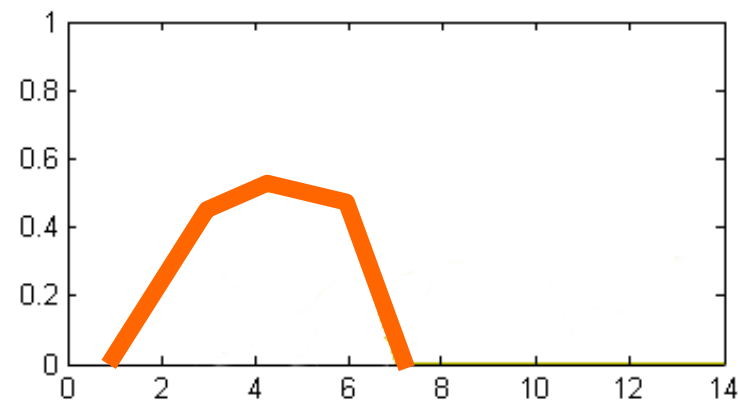
1 caller



5 receivers

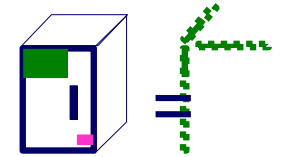


4 days of activity



~200 calls to EACH receiver on EACH day!

# P2: Anomaly detection in time-evolving graphs



- Anomalous communities in phone call data:
  - European country, 4M clients, data over 2 weeks



**Miguel Araujo, Spiros Papadimitriou, Stephan Günnemann, Christos Faloutsos, Prithwish Basu, Ananthram Swami, Evangelos Papalexakis, Danai Koutra. *Com2: Fast Automatic Discovery of Temporal (Comet) Communities.* PAKDD 2014, Tainan, Taiwan.**



## Outline - tensors

- Motivation - Definitions
- Tensor tools
- Case studies
  - P1: web graph mining ('TOPHITS')
  - P2: phone-call patterns
  - – P3: network traffic
  - P4: FaceBook activity



# ParCube: Sparse Parallelizable Tensor Decompositions

**Evangelos E. Papalexakis**, Christos Faloutsos, Nikos Sidiropoulos,  
ECML/PKDD 2012

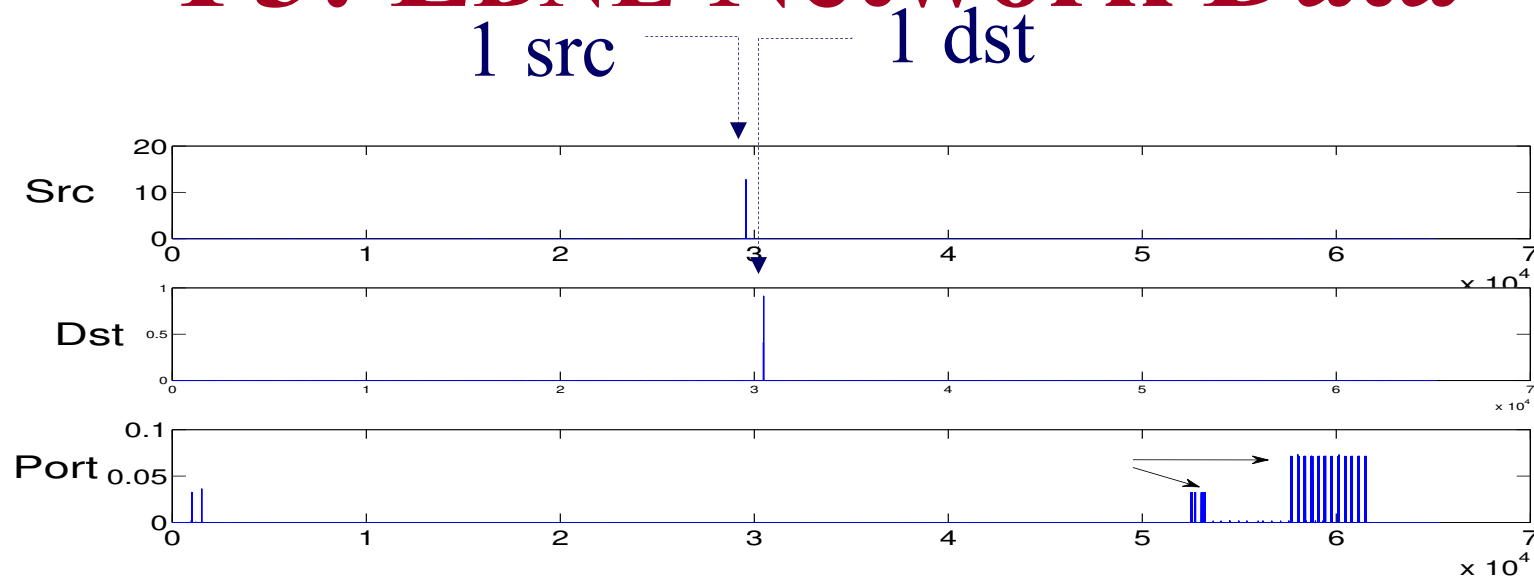
Evangelos E. Papalexakis

**Email:** [epapalex@cs.ucr.edu](mailto:epapalex@cs.ucr.edu)

**Web:** <http://www.cs.ucr.edu/~epapalex>



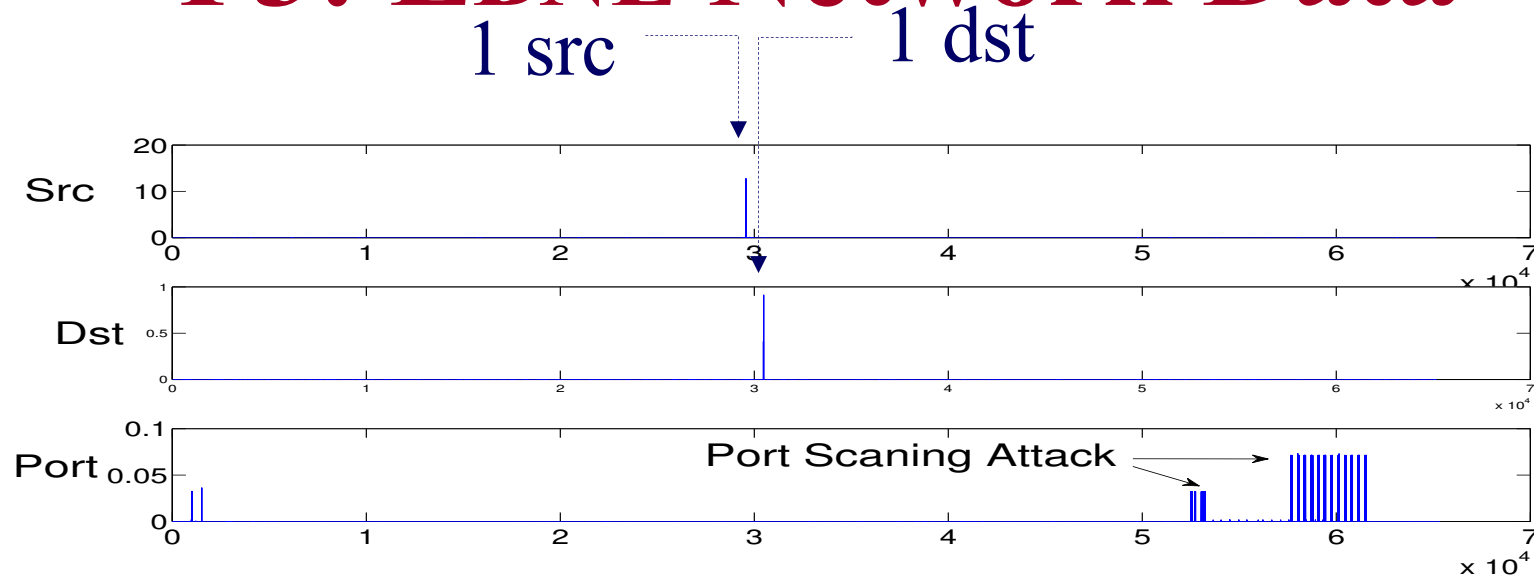
# P3: LBNL Network Data



- Modes: src IP, dst IP, port #
- Q: what could it be?
- A:



# P3: LBNL Network Data

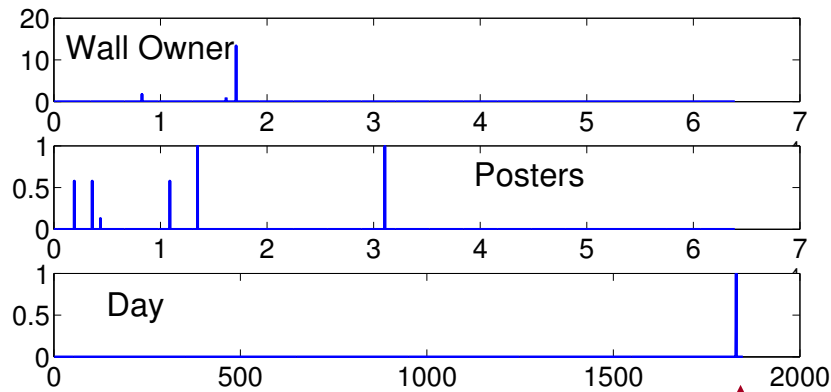


- Modes: src IP, dst IP, port #
- Q: what could it be?
- **A: ~ Port Scanning Attack**

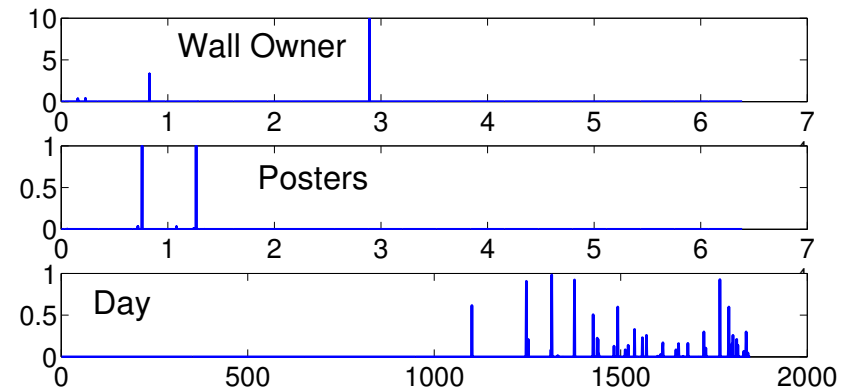


# P4: FACEBOOK Wall posts

1 Wall



(a) FACEBOOK anomaly



(b) FACEBOOK normal activity

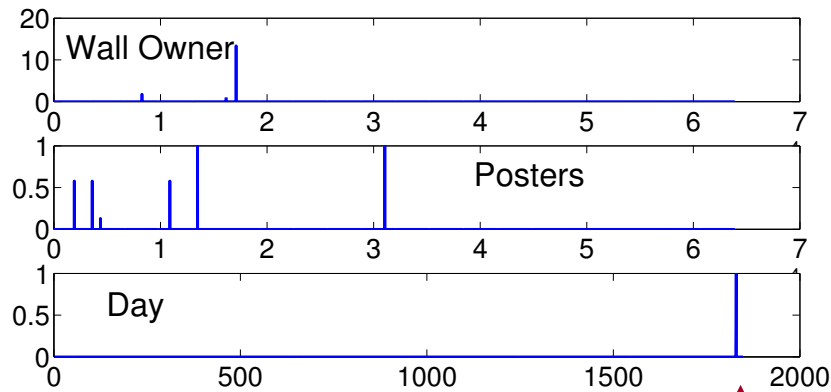
1 day

- Modes: wall-owner, poster, timestamp
- Discovery: What could it be?



# P4: FACEBOOK Wall posts

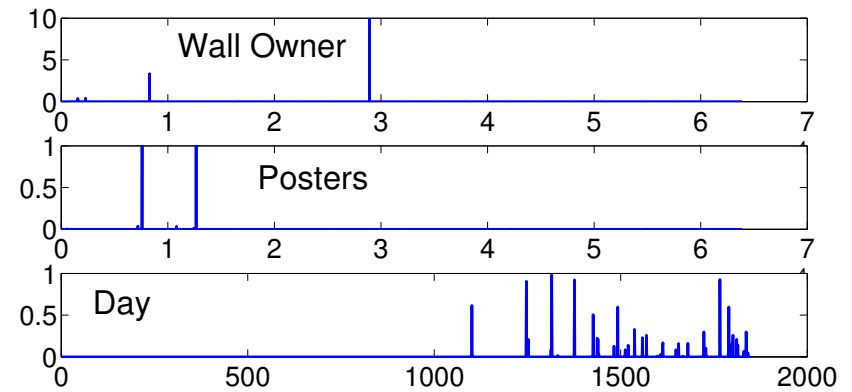
1 Wall



(a) FACEBOOK anomaly (Wall owner's birthday)



1 day



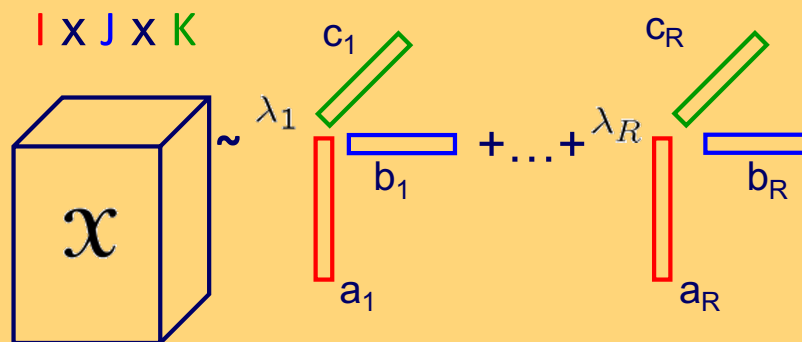
(b) FACEBOOK normal activity

- Modes: wall-owner, poster, timestamp
- Discovery: What could it be? A: Birthday



# Conclusions

- Q: who-calls-whom-when – patterns?
  - Triplets ( source-ip, dest-ip, port#)
  - KB (subject, verb, object)
- A: Tensor analysis (PARAFAC)
  - <http://www.tensortoolbox.org/>



## References

- Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha: Information-theoretic co-clustering. KDD 2003: 89-98
- T. G. Kolda, B. W. Bader and J. P. Kenny. *Higher-Order Web Link Analysis Using Multilinear Algebra*. In: ICDM 2005, Pages 242-249, November 2005.
- Jimeng Sun, Spiros Papadimitriou, Philip Yu. *Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams*, Proc. of the Int. Conf. on Data Mining (ICDM), Hong Kong, China, Dec 2006

## References

- Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, Christos Faloutsos: *Tensor Decomposition for Signal Processing and Machine Learning*. IEEE Trans. Signal Process. 65(13): 3551-3582 (2017)

# Software

- Tensorly:  
<http://tensorly.org/stable/index.html>
- Tensor toolbox (in matlab)  
<https://www.tensortoolbox.org/>
- ... and in python [pyttb](#)