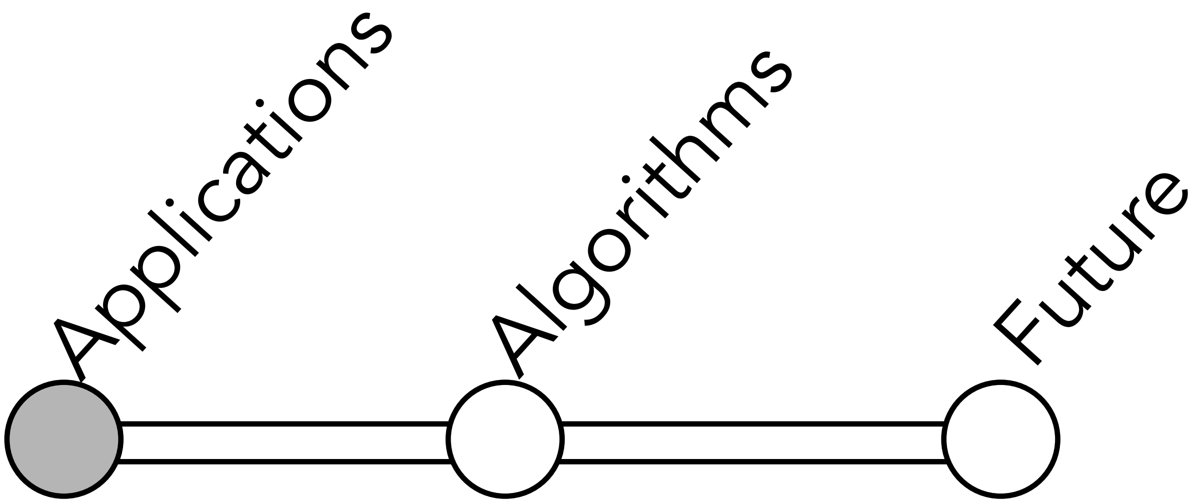
Tensor Decompositions for Big Multi-aspect Data Analytics

Evangelos (Vagelis) Papalexakis
UC Riverside

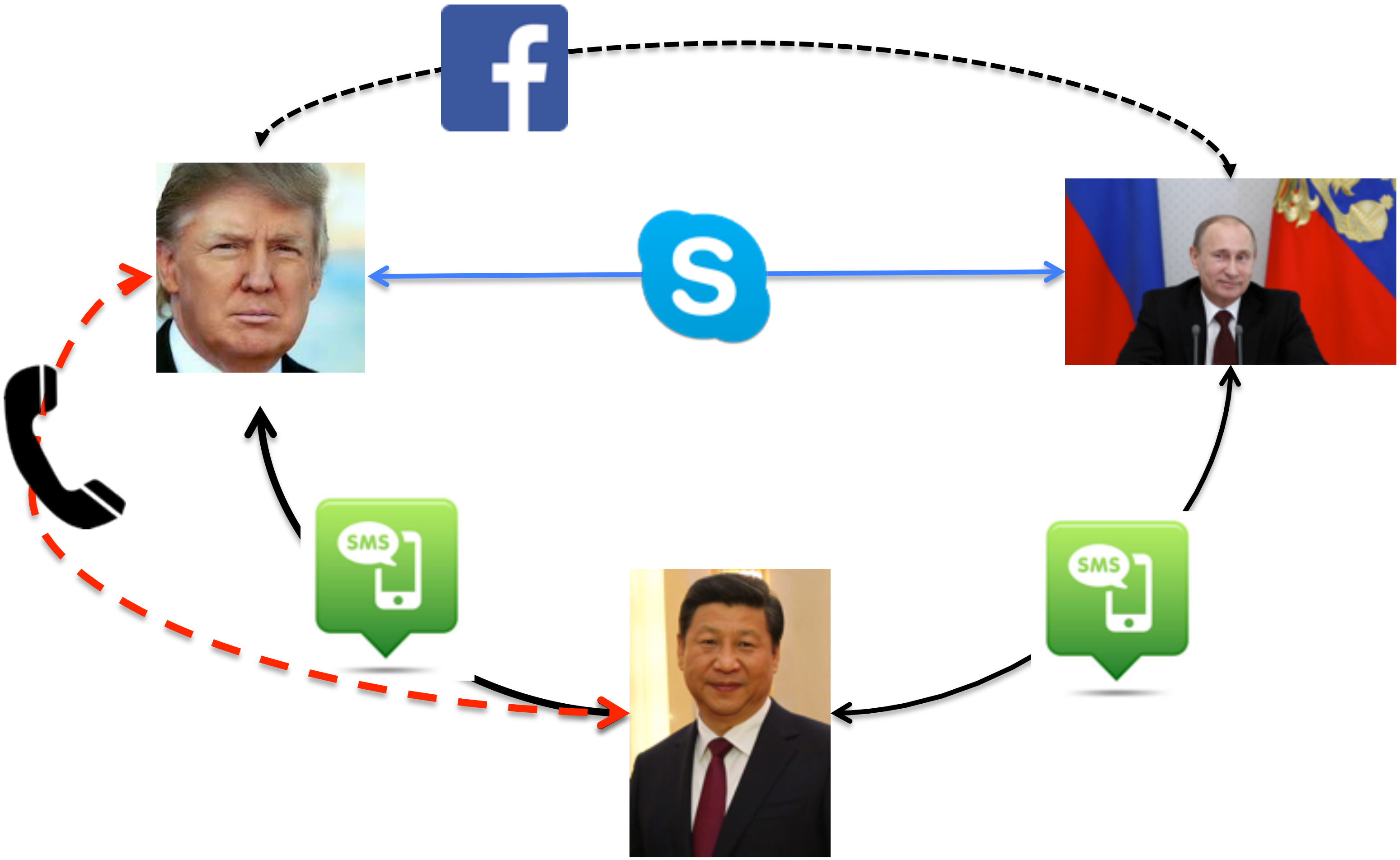
Second Workshop of Mission-Critical Big Data Analytics (MCBDA 2017)

Roadmap

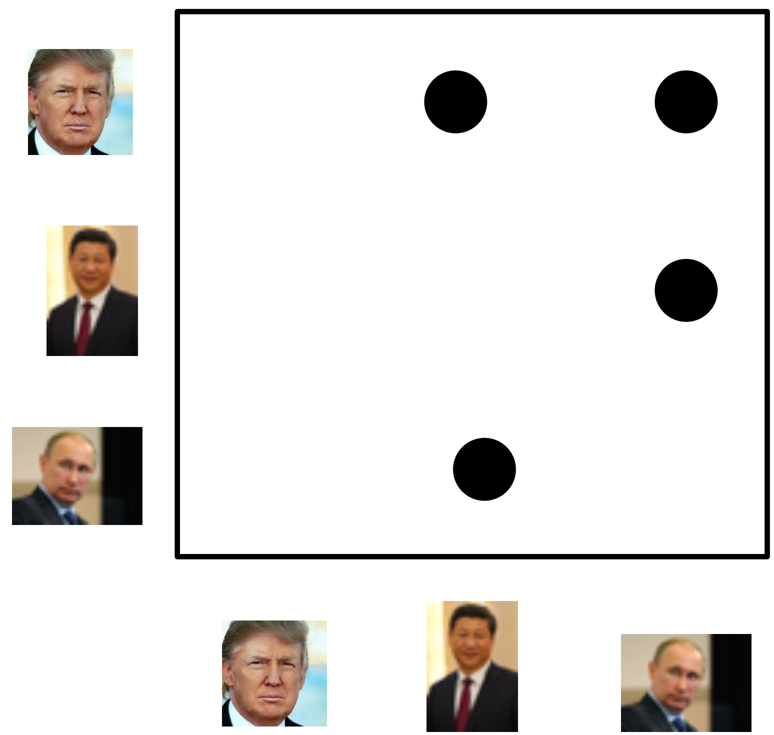


Multi-Aspect Data??

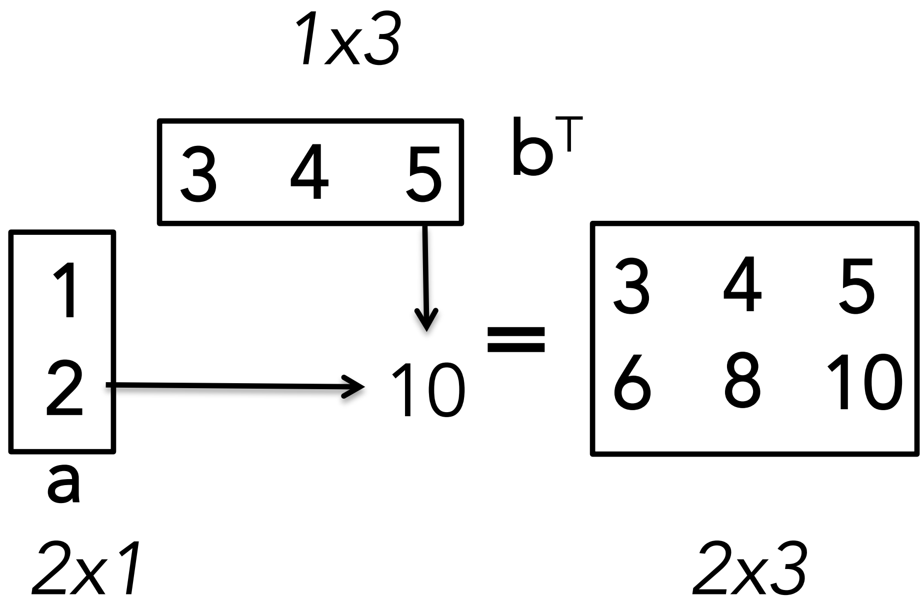
Multi-View Social Networks



Social Network Matrix



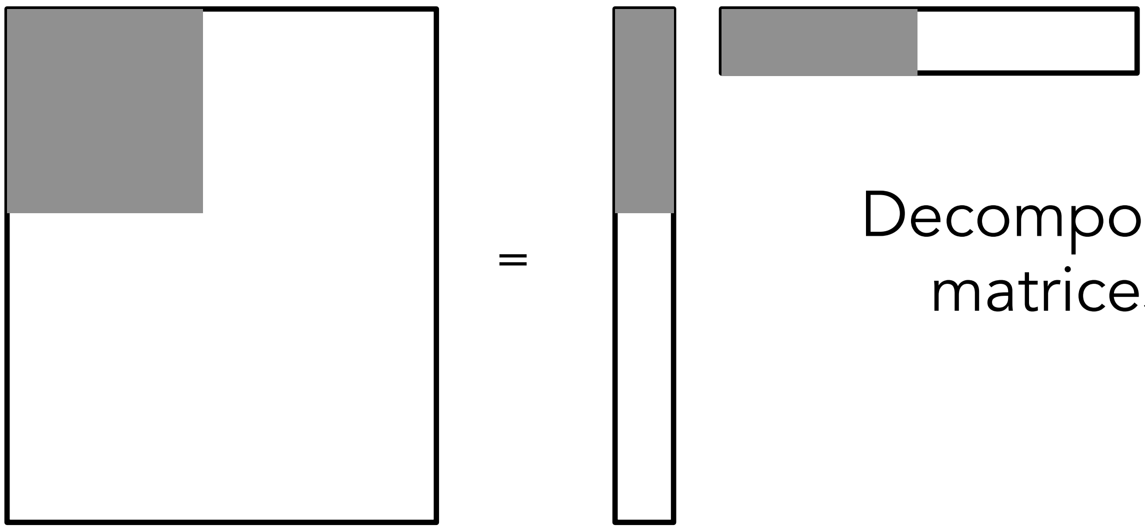
Outer Product



By definition:
Rank-one matrix

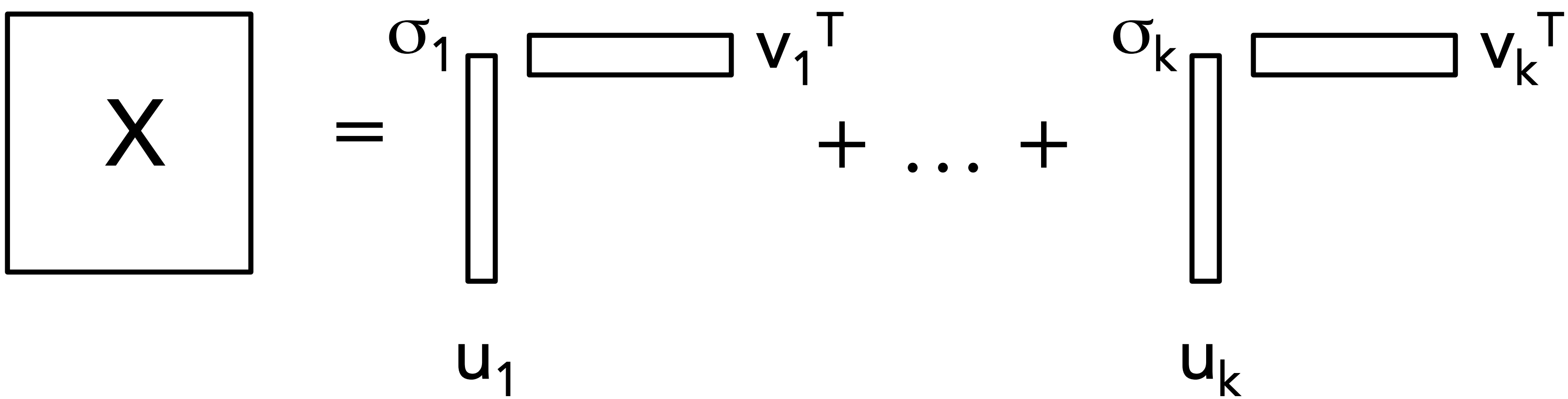
Matrix Rank = Min # of rank-one matrices that add up to that matrix

Matrix Decomposition



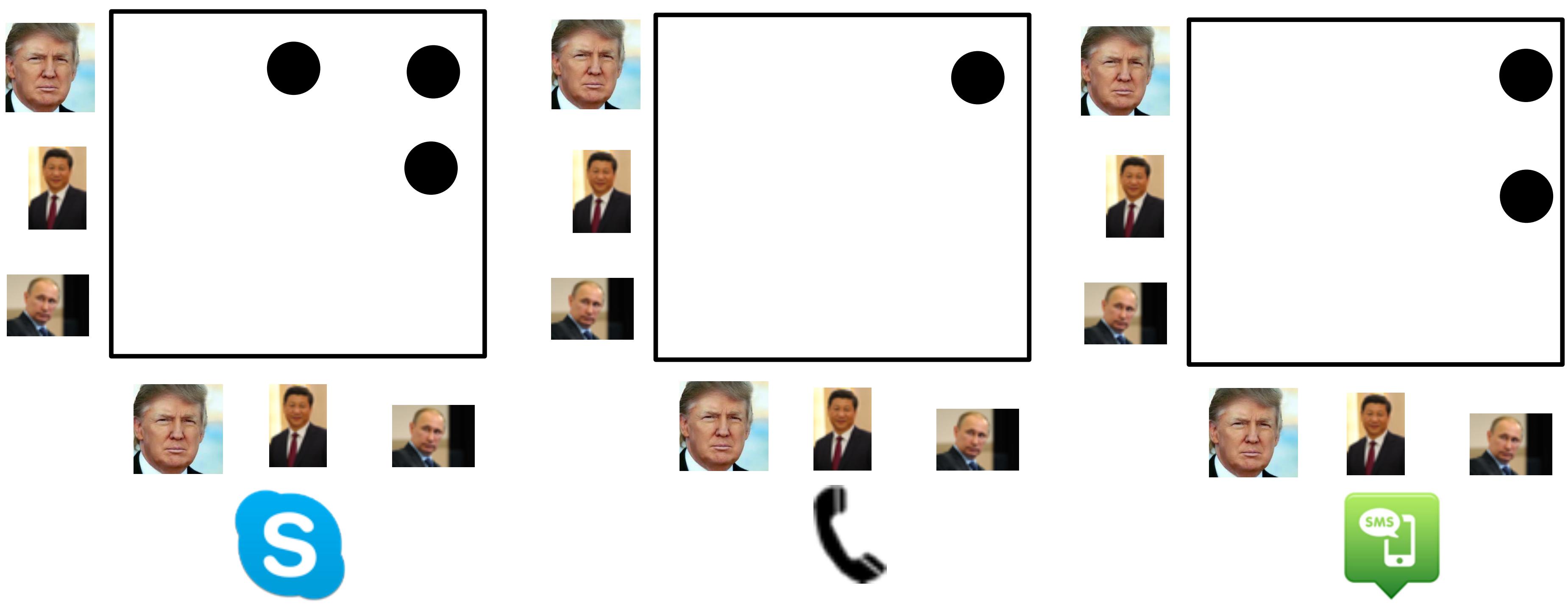
Decomposition into rank-1 matrices/components

Singular Value Decomposition



- If k = rank(X) then we have equality
- If k < rank(X) we have the best rank k approximation that minimizes the squared error (Eckart Young Theorem)

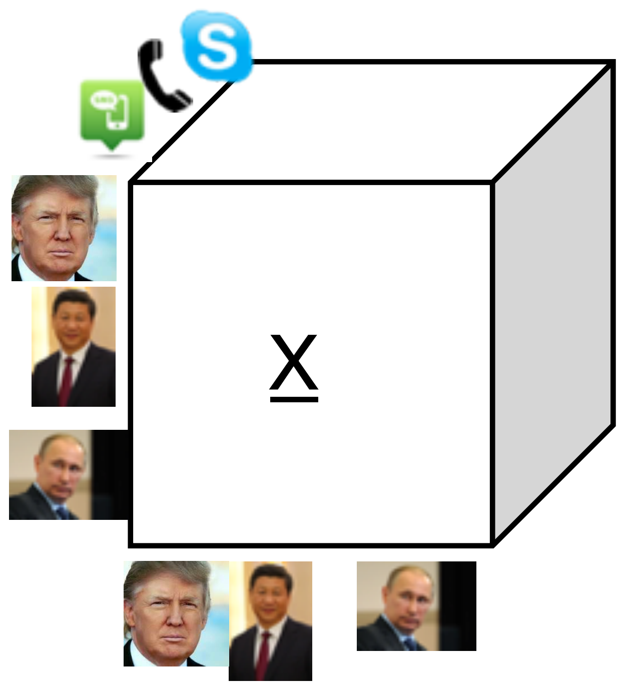
What if we have more than 1 view of the network?



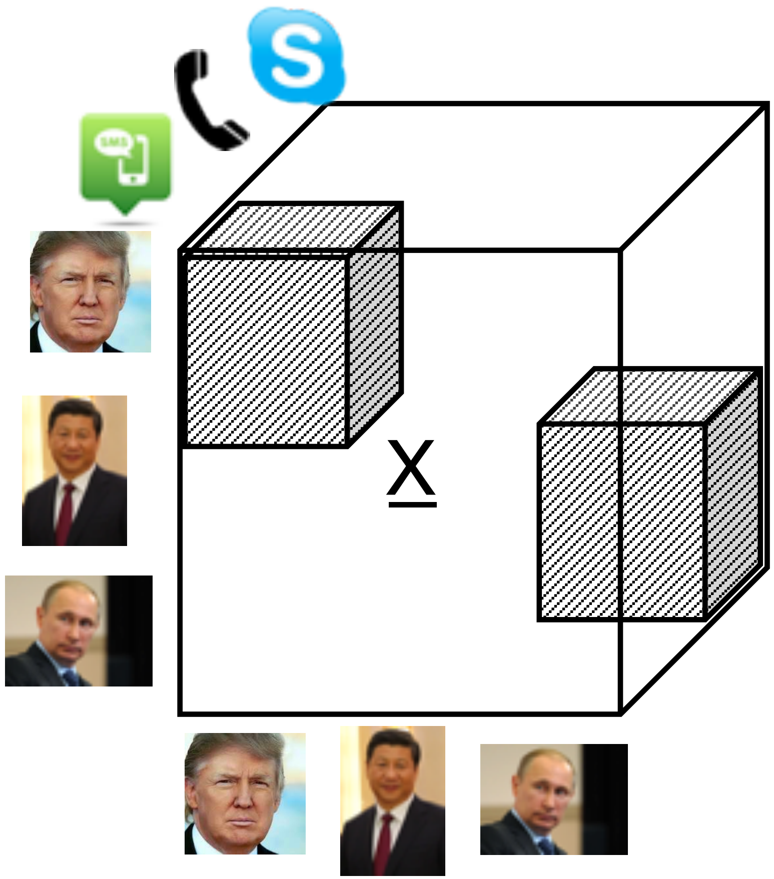
If we aggregate, we ignore important structure!!

Tensors

- Multi-dimensional matrices
- Long list of applications: Chemometrics, Psychometrics, Signal Processing, Data Mining



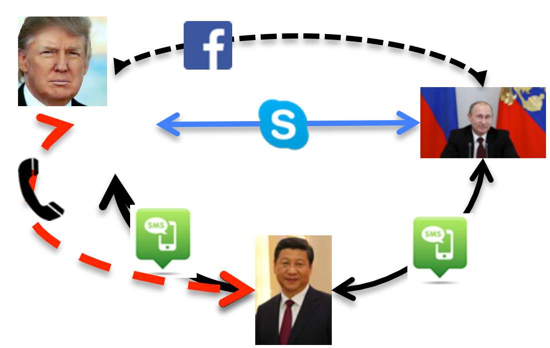
What are we looking for?



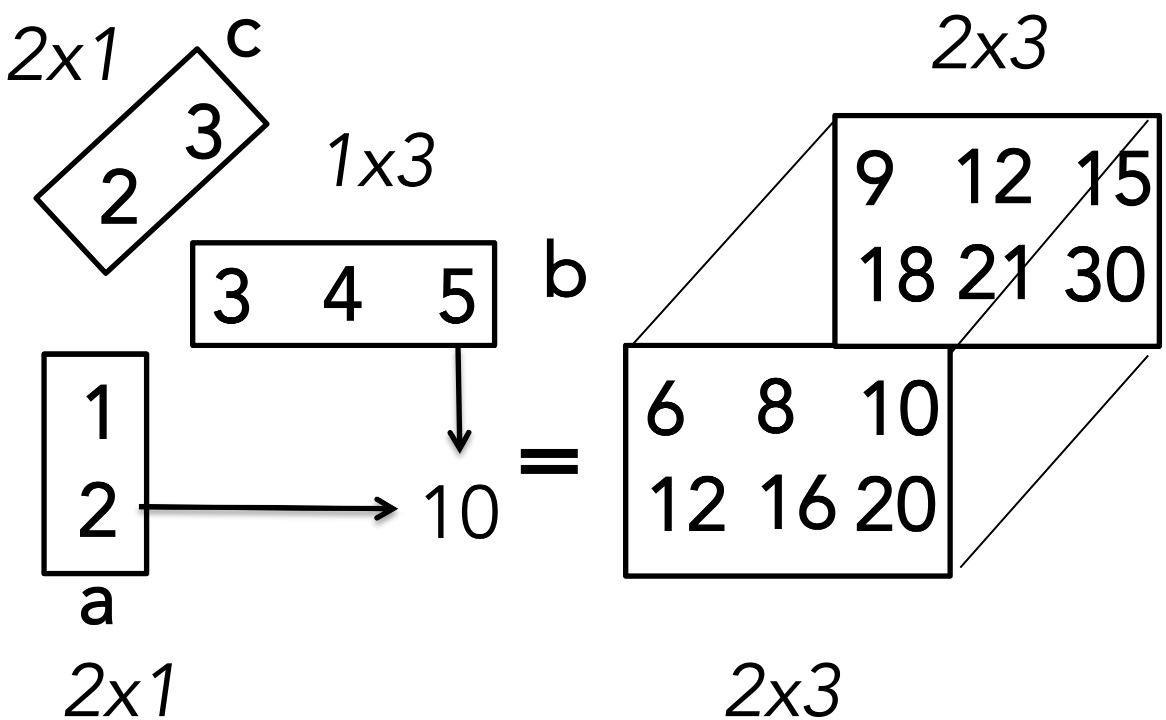
Blocks within the data

Subsets / co-clusters of:

1) Users ("senders")
2) Users ("receivers")
3) Means of communication



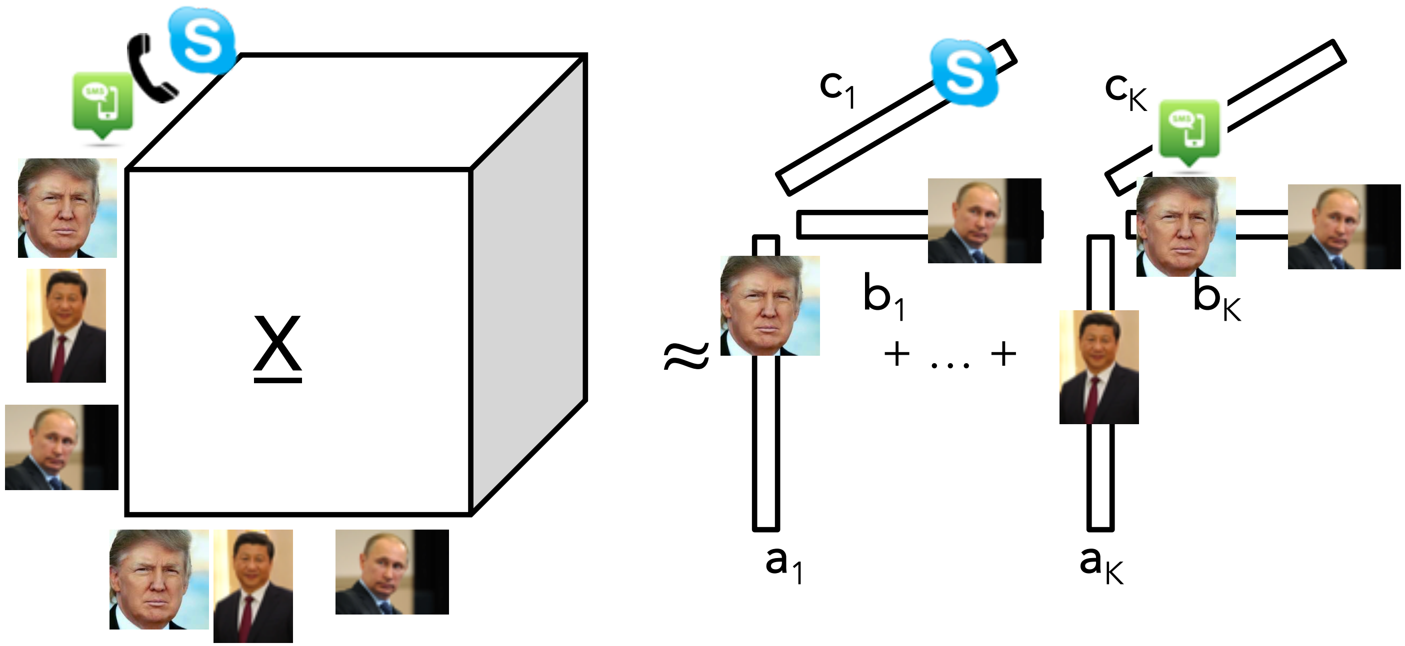
Three-way Outer Product



2x3x2 rank-1 tensor

Blocks are rank-1 tensors

CP/PARAFAC Decomposition

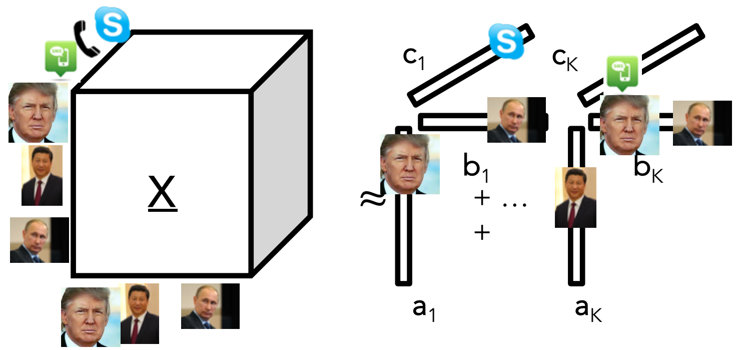


$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \|\underline{\mathbf{X}} - \sum_k \mathbf{a}_k \circ \mathbf{b}_k \circ \mathbf{c}_k\|_F^2$$

outer product

GRAPHFUSE

Step 1



PARAFAC with Sparse Latent Factors *[Papalexakis et al. IEEE ICASSP 2011, IEEE TSP 2013]*

Step 2 Assign users to communities (max component membership)

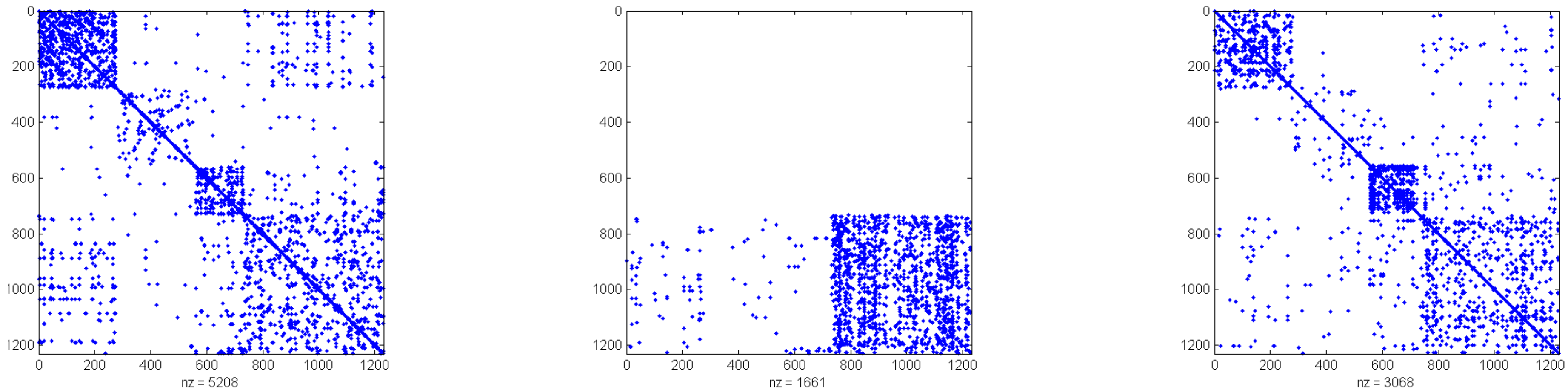
Step 3 For users with no assignment create K+1th (umbrella) community

Output:

1. Assignment of users to communities
2. Influence of a view to each community

[Papalexakis et al. **Fusion 2013**]

DBLP Multi-View Graph

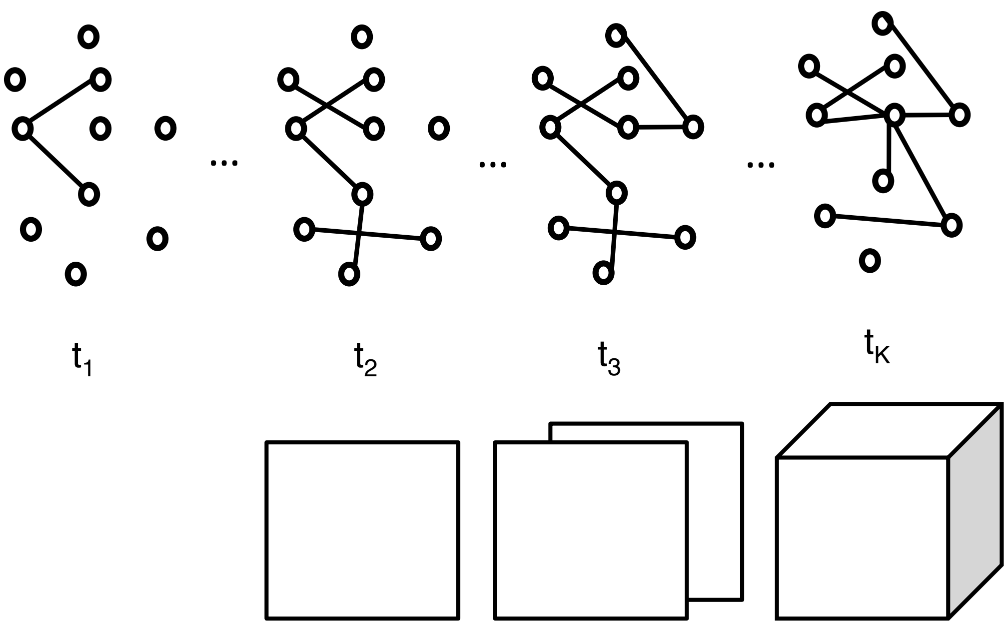


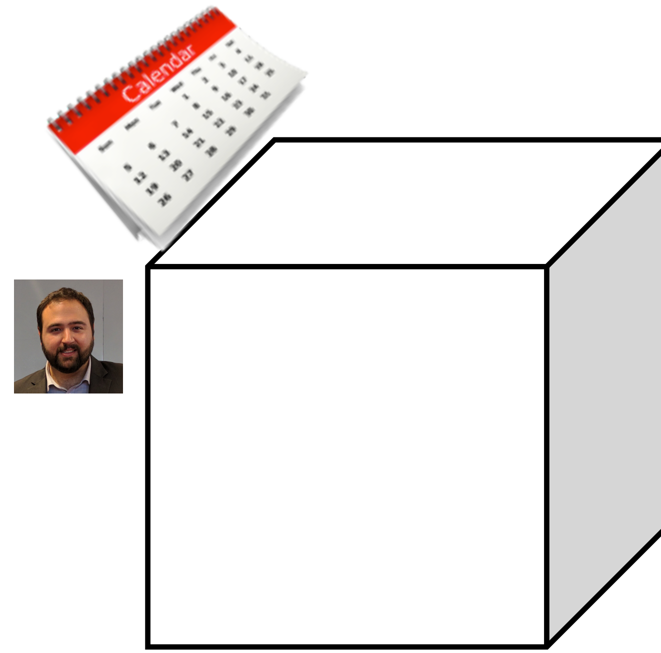
(a) citation (b) co-auth. (c) co-term

- Assignment of authors to research communities
- Measure NMI (Normalized Mutual Information)
- Baselines
  - ✧ Spectral clustering on sum of matrices / views
  - ✧ Linked Matrix Factorization [Tang et al. ICDM 2009]
- GRAPHFUSE outperforms "2D" baselines

[Papalexakis et al. **Fusion 2013**]

Time-Evolving Graphs as Tensors





Detect anomalies / real-life events

Time-Evolving Graphs as Tensors

- Tensor Decomposition will give us
  - Communities in the Graph &
  - Their evolution over time

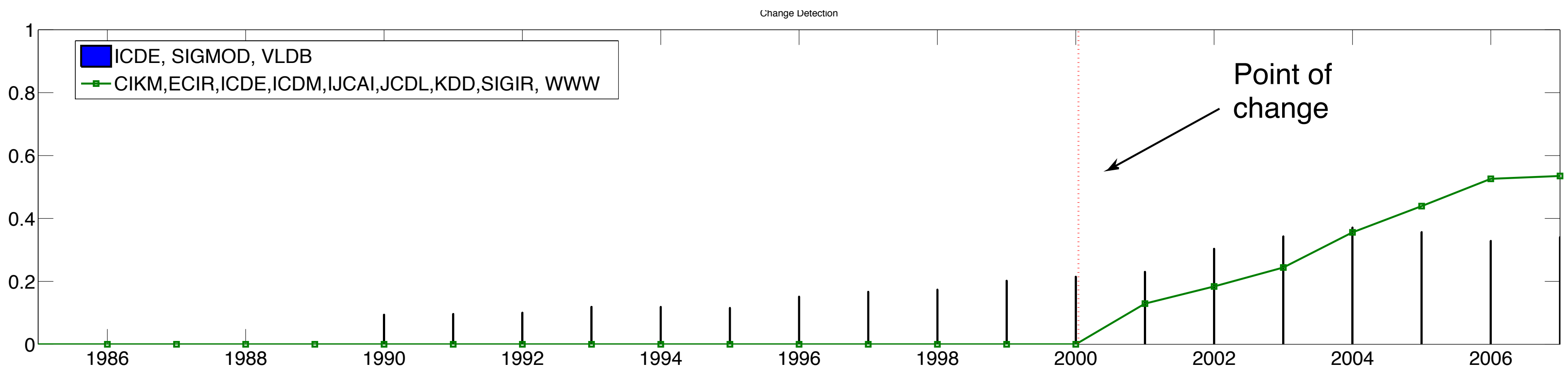
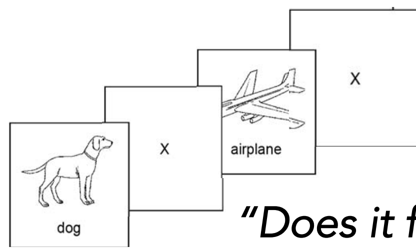


Fig. 3. In this Figure we demonstrate how TENSORSPLAT is able to perform change detection. In particular, we observe two components in which a well-known professor appears as an author; the first component mainly contains Databases conferences, whereas the second contains Data Mining conferences. The dashed red line indicates the point of change in research direction.

Koutra, Papalexakis, Faloutsos, "TENSORSPLAT: Spotting Latent Anomalies in Time"

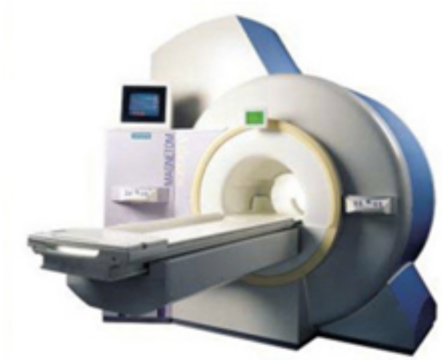
Neurosemantics

. . .

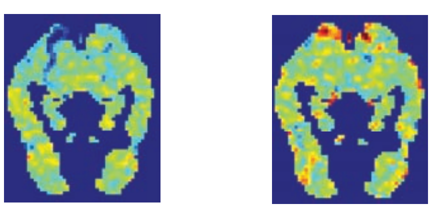


"Does it fly?" (y/n)

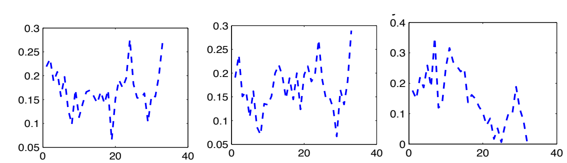
"Does it bite?" (y/n)



fMRI

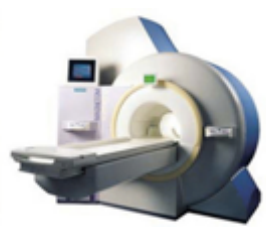


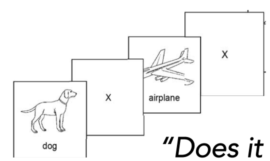
MEG



Neurosemantics

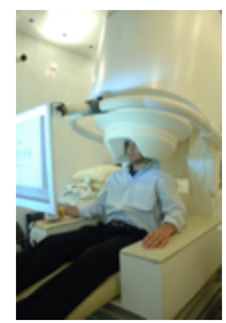
- Semantically coherent brain regions?
- Similarities/differences between subjects?
- How is language processed in the brain?

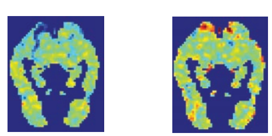


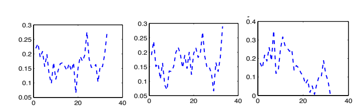


"Does it fly?" (y/n)

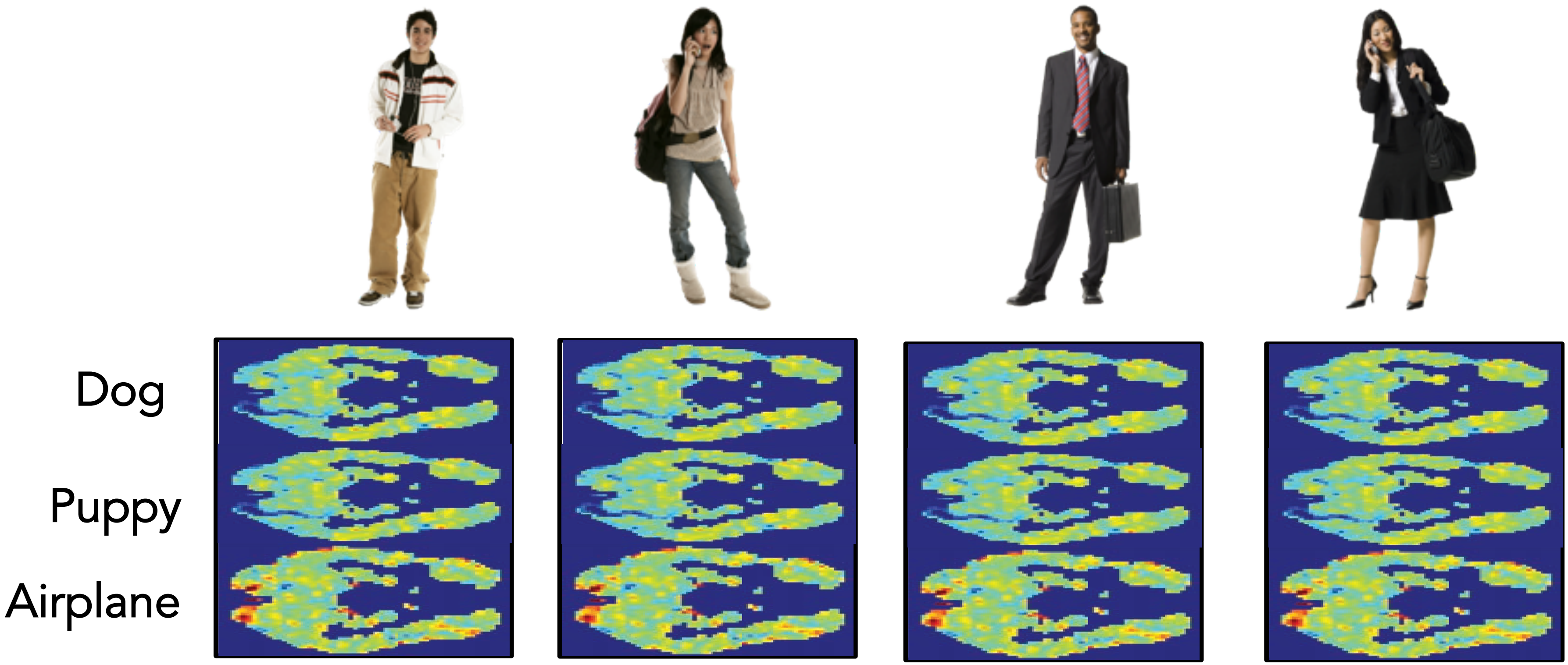
"Does it bite?" (y/n)



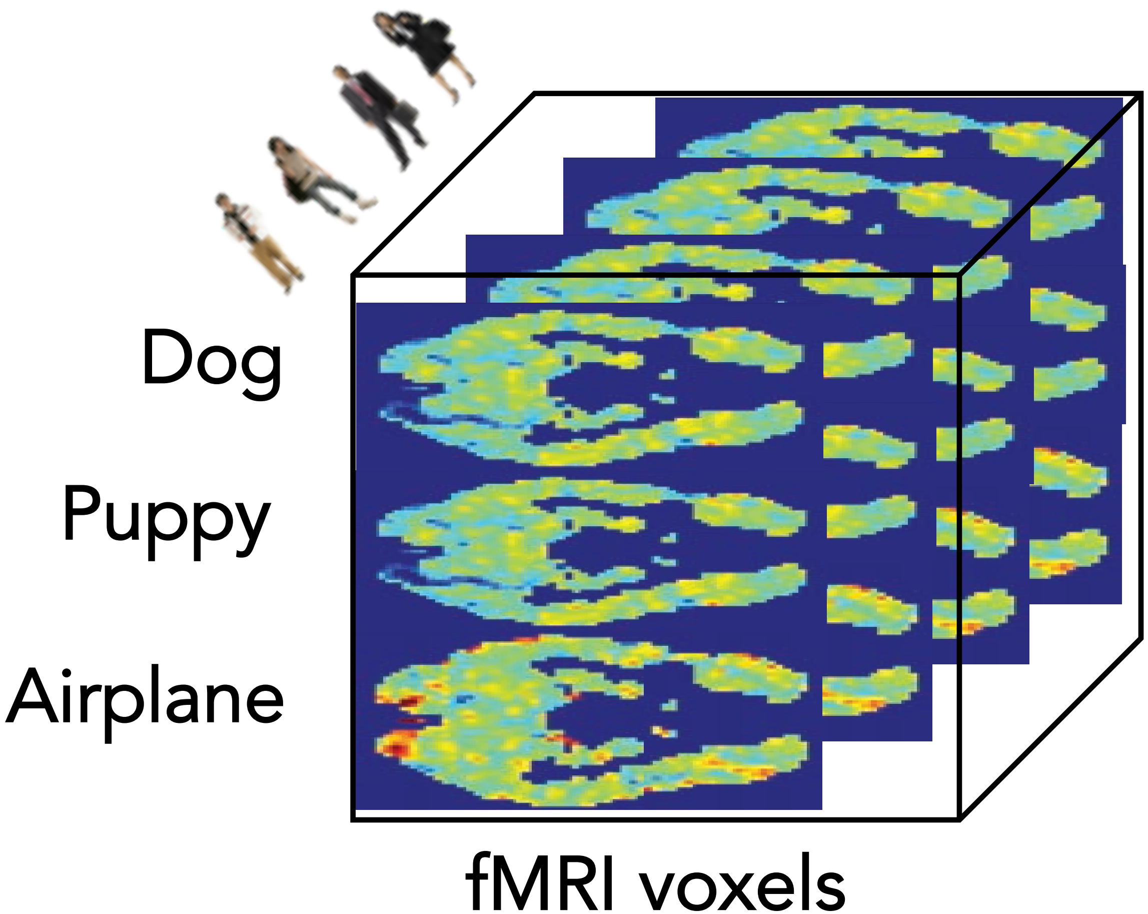




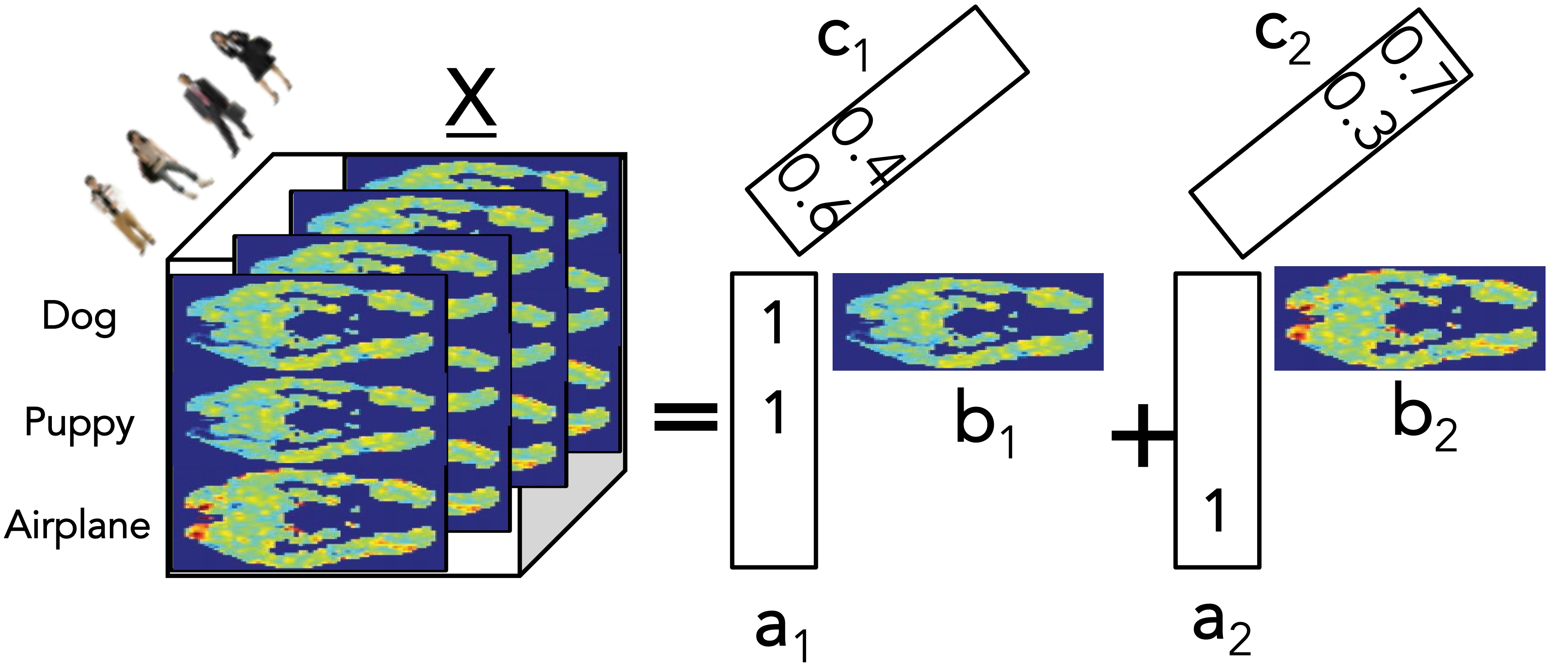
Combining measurements for multiple subjects



Modeling Brain Data as Tensor

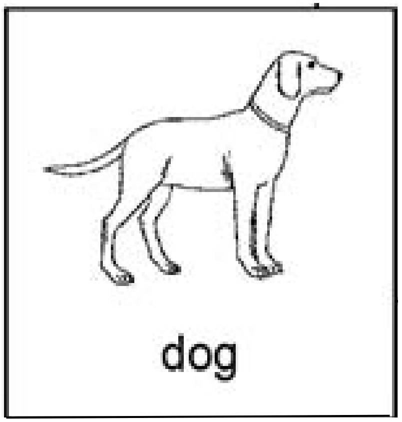


CP/PARAFAC Decomposition



$$\min_{\mathbf{a}_r,\mathbf{b}_r,\mathbf{c}_r} \|\underline{\mathbf{X}} - \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r\|_F^2$$

Semantic Information

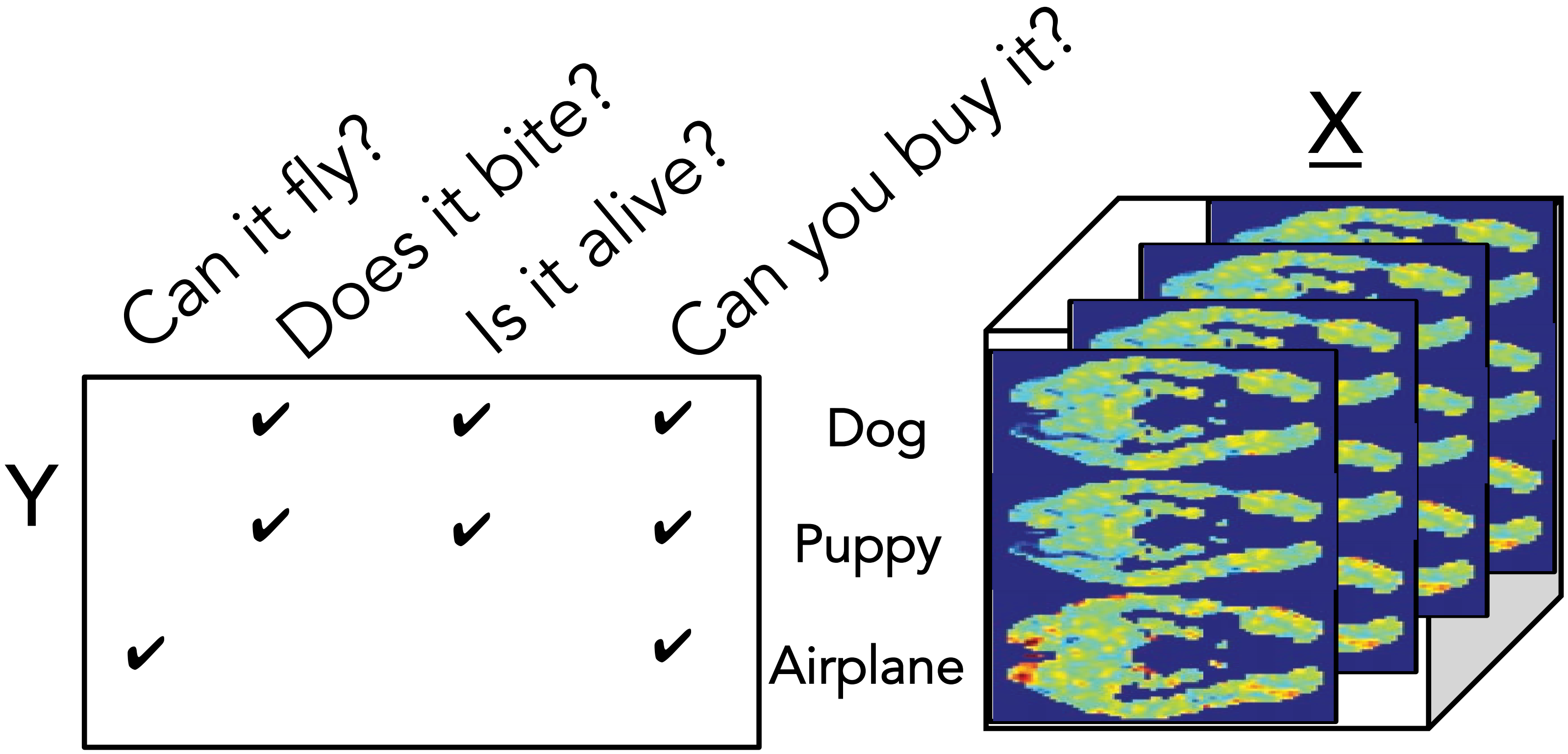


✔ Is it alive?
✔ Does it bite?
Does it fly?
✔ Can you buy it?
Is it smaller than a golf ball?

…

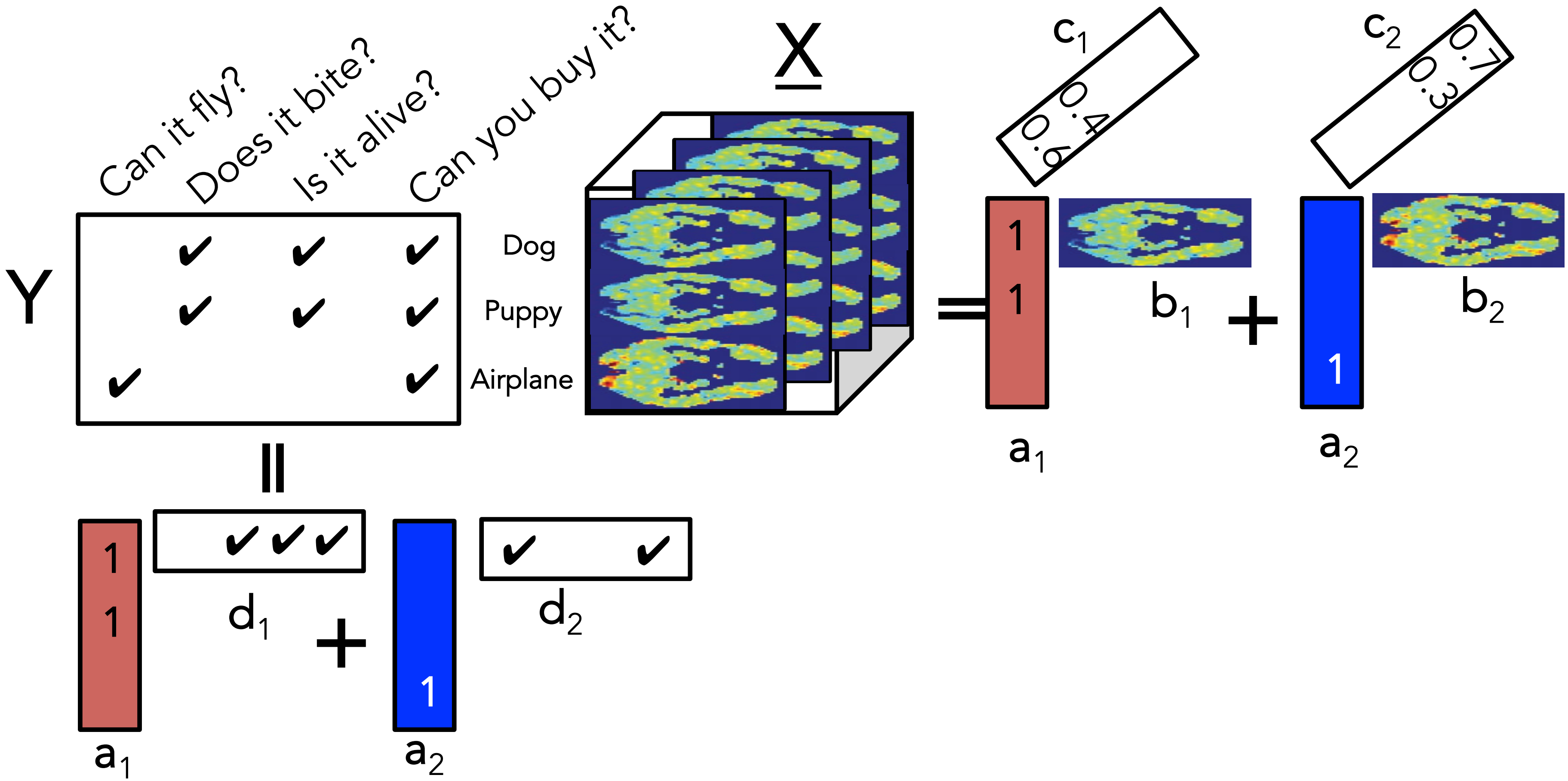
- Human readable description of the noun
- Useful information to guide the analysis
- Can have different semantic features (corpus statistics, knowledge base features)

Tensor With Side Information

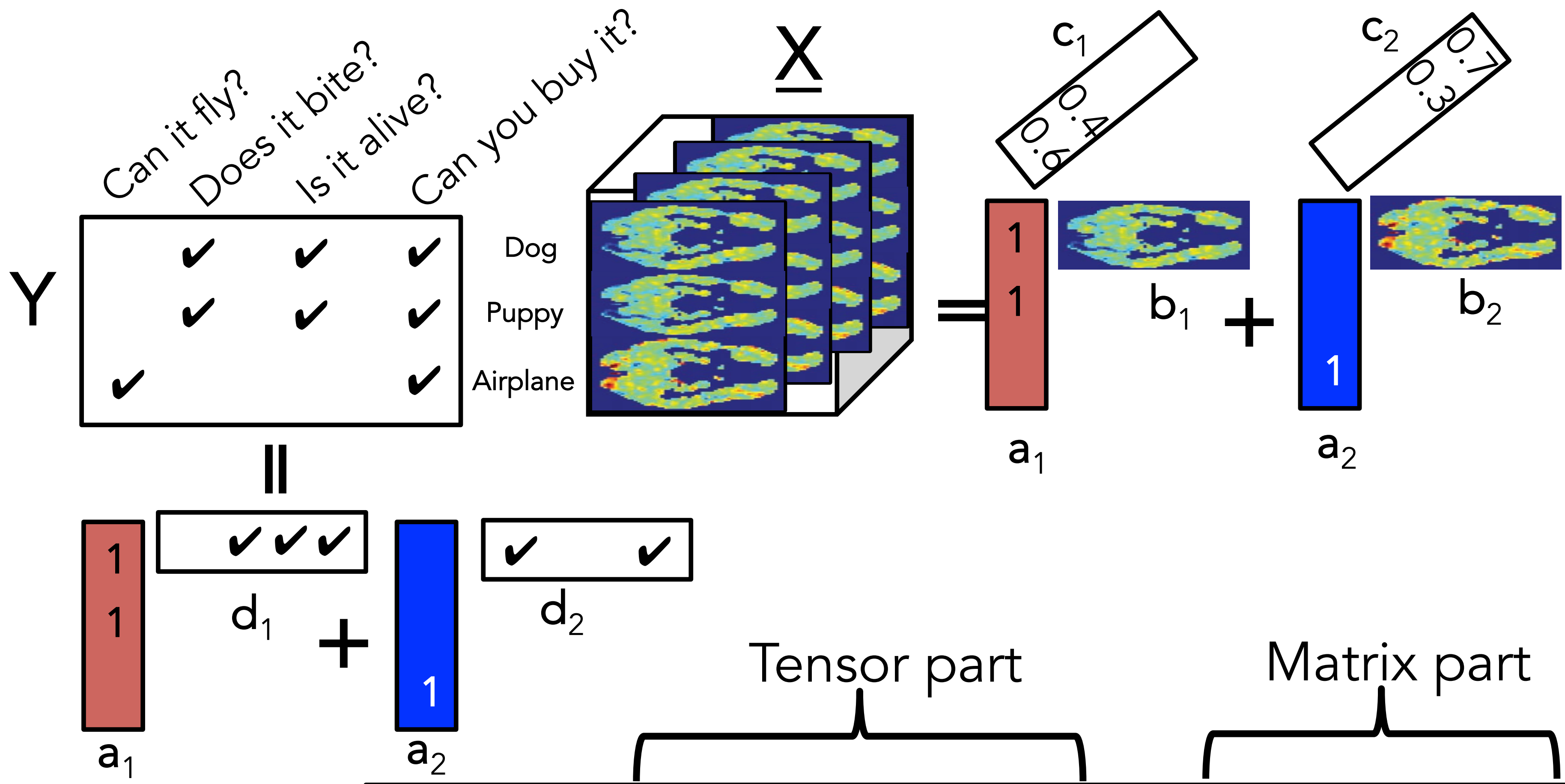


[Papalexakis et al. SDM 2014]

Proposed Modeling: Coupled Matrix-Tensor Factorization



Proposed Modeling: Coupled Matrix-Tensor Factorization



$$\overbrace{\min_{\mathbf{a}_r,\mathbf{b}_r,\mathbf{c}_r,\mathbf{d}_r} \|\underline{\mathbf{X}} - \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r\|_F^2}^{\text{Tensor part}} + \overbrace{\|\mathbf{Y} - \sum_{r=1}^{R} \mathbf{a}_r \mathbf{d}_r^T\|_F^2}^{\text{Matrix part}}$$

Pre-motor Cortex

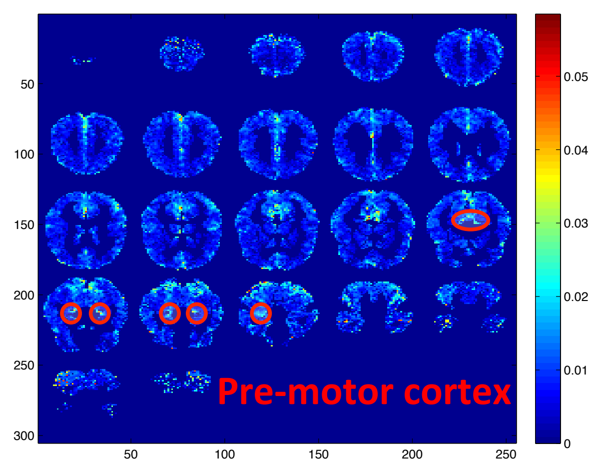
Nouns
glass
tomato

✔ Unsupervised

✔ Agrees with Neuroscience

can you hold it in one hand?

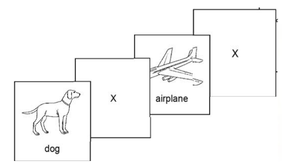
is it smaller than a golfball?'

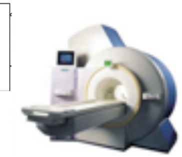


Multi-Aspect Data Everywhere!

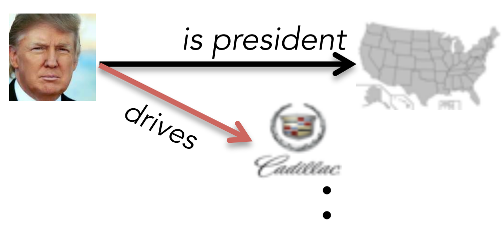
Social Networks & Urban Comp.

Neurosemantics

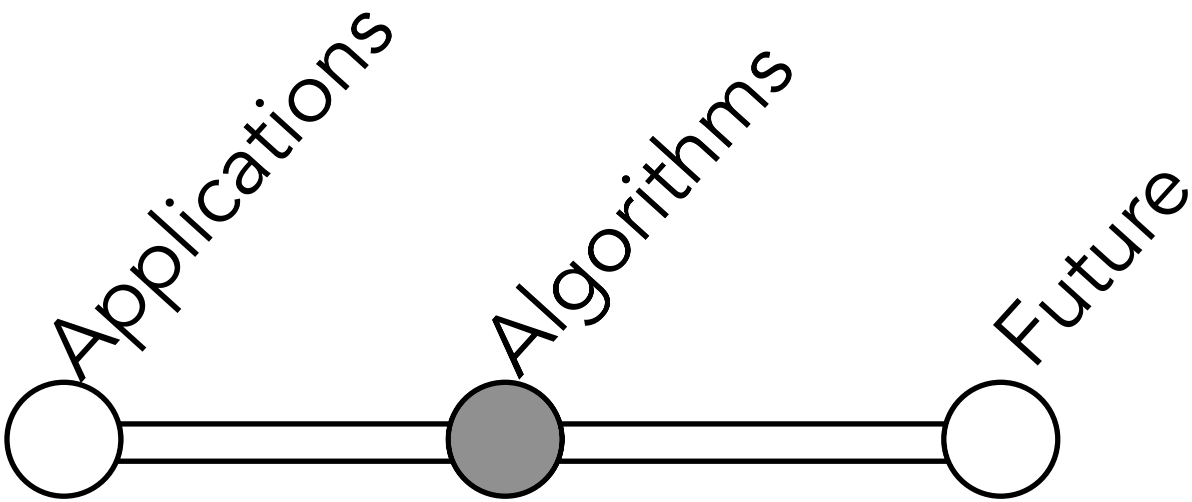




Web Knowledge

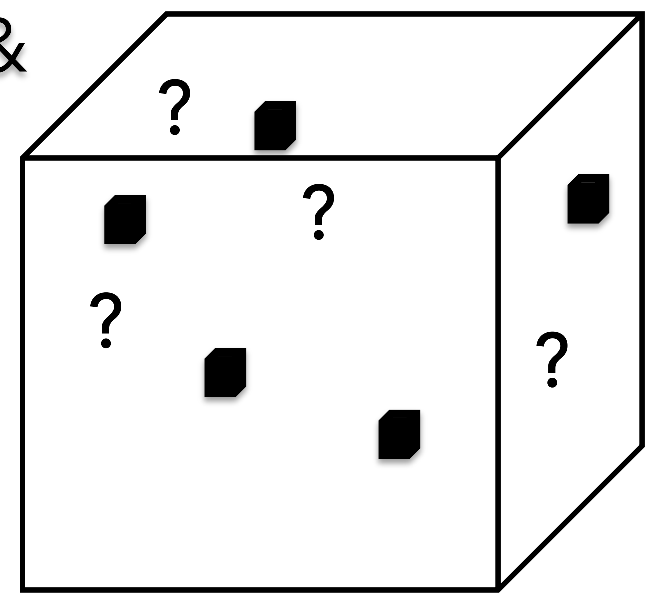


Roadmap



Tensors in Data Science

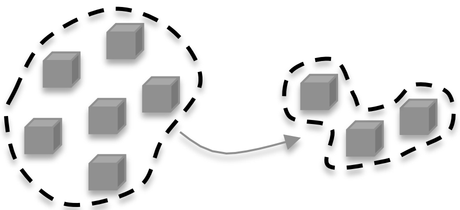
- Naturally model multi-aspect data
- Very powerful modeling tools
- **Big Challenges**
  - ✧ **C1**: Data Size & Scalability
  - ✧ **C2**: Model Selection, Quality & Interpretability

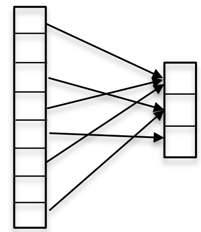


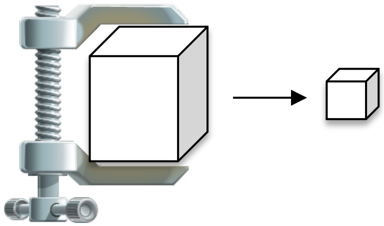
Fast and Scalable Tensor Decompositions

- Exploiting Sparsity
  - ✧ Tensor Toolbox for Matlab [Kolda et al.]
  - ✧ GigaTensor [Kang et al. 2012]
  - ✧ FlexiFaCT [Beutel et al. 2014]
  - ✧ DFacto [Choi et al. 2014]
  - ✧ SPLATT [Smith et al. 2015]
- All above methods are exact
  - ✧ Most of them focus on the "MTTKRP" operation
- Can we do something by **approximating**?

Approximate "Sketching" Methods







Sampling

Tensor CUR [Mahoney et al. 2008]

MACH [Tsourakakis 2010]

ParCube [Papalexakis et al. 2012]

Walk'n'Merge [Erdos et al 2013]

SPALS [Cheng et al 2016]

CPRAND[Battaglino et al 2017]

Hashing

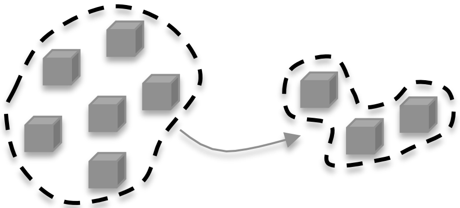
[Wang et al. 2015]

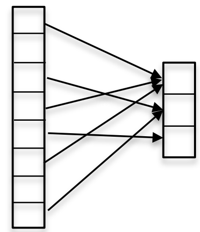
Compression

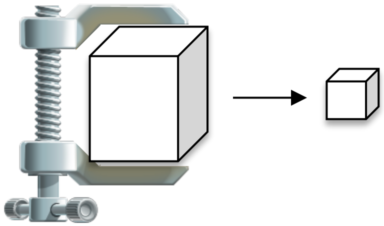
Tucker Compression [Bro et al. 1998]

PARACOMP [Sidiropoulos et al. 2014]

Approximate "Sketching" Methods







Sampling

Tensor CUR [Mahoney et al. 2008]

MACH [Tsourakakis 2010]

ParCube [Papalexakis et al. 2012]

Walk'n'Merge [Erdos et al 2013]

SPALS [Cheng et al 2016]

CPRAND[Battaglino et al 2017]

Hashing

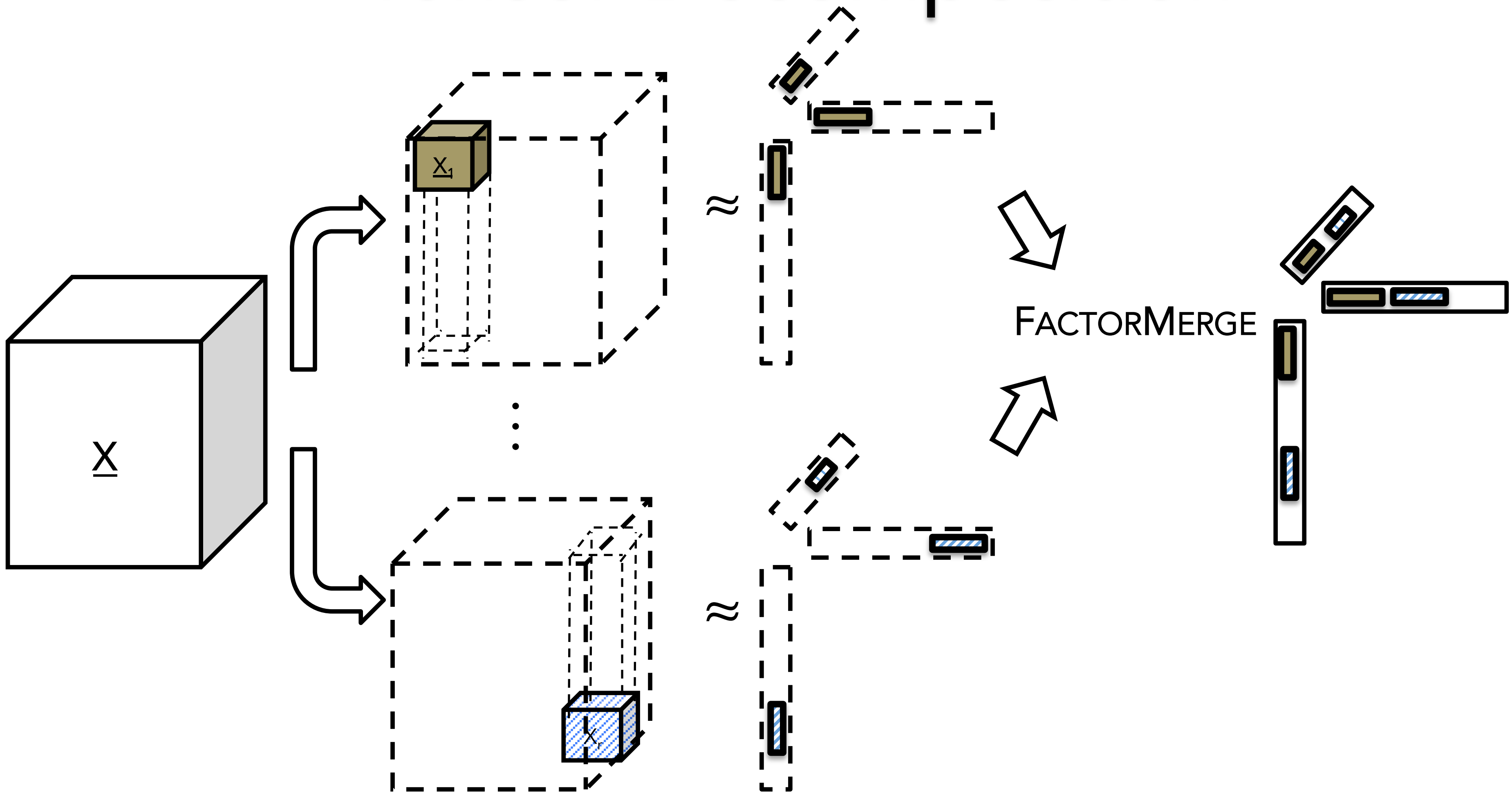
[Wang et al. 2015]

Compression

Tucker Compression
[Bro et al. 1998]

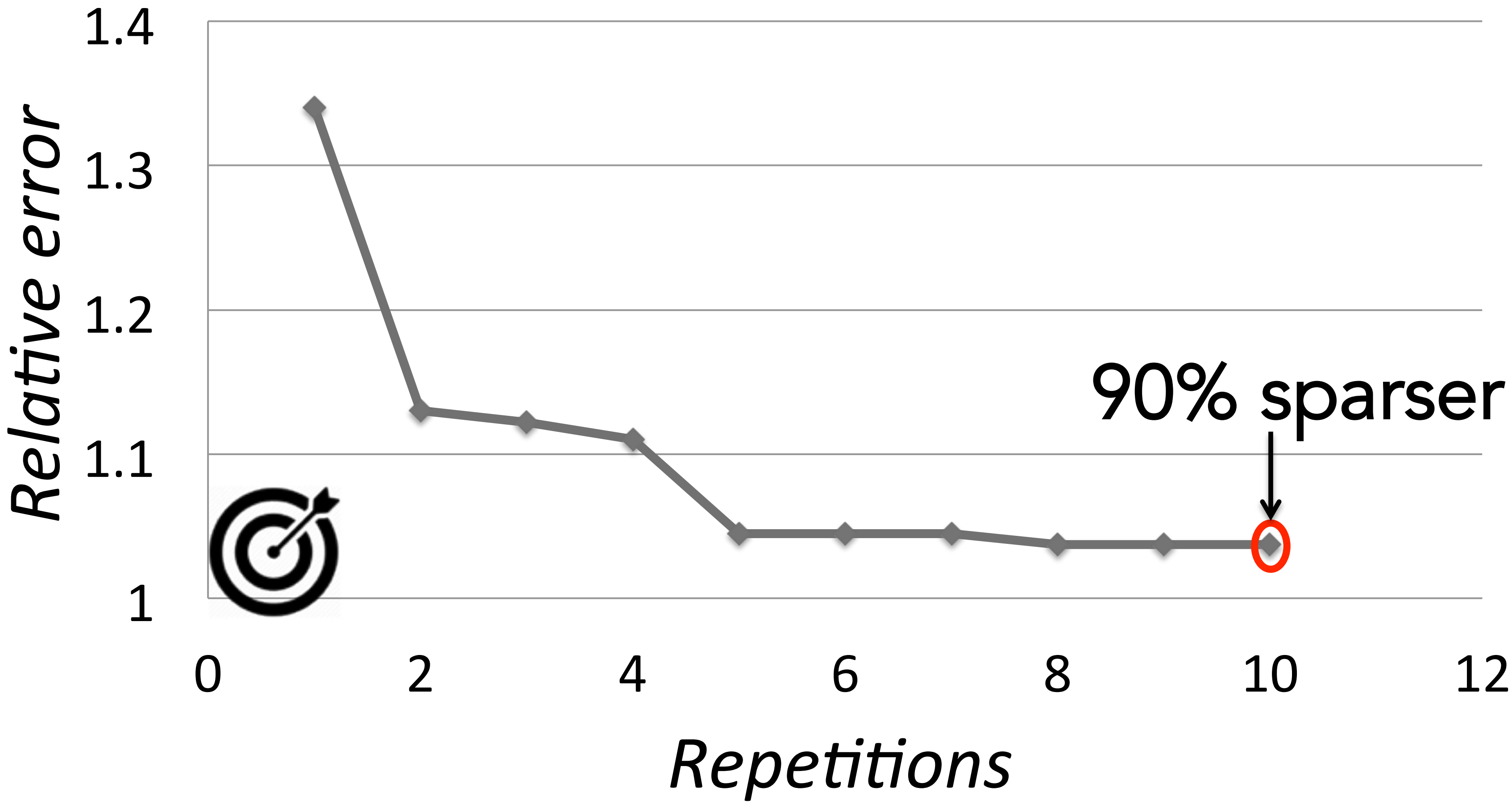
PARACOMP [Sidiropoulos et al. 2014]

ParCube: Sampling-based Parallel Tensor Decomposition



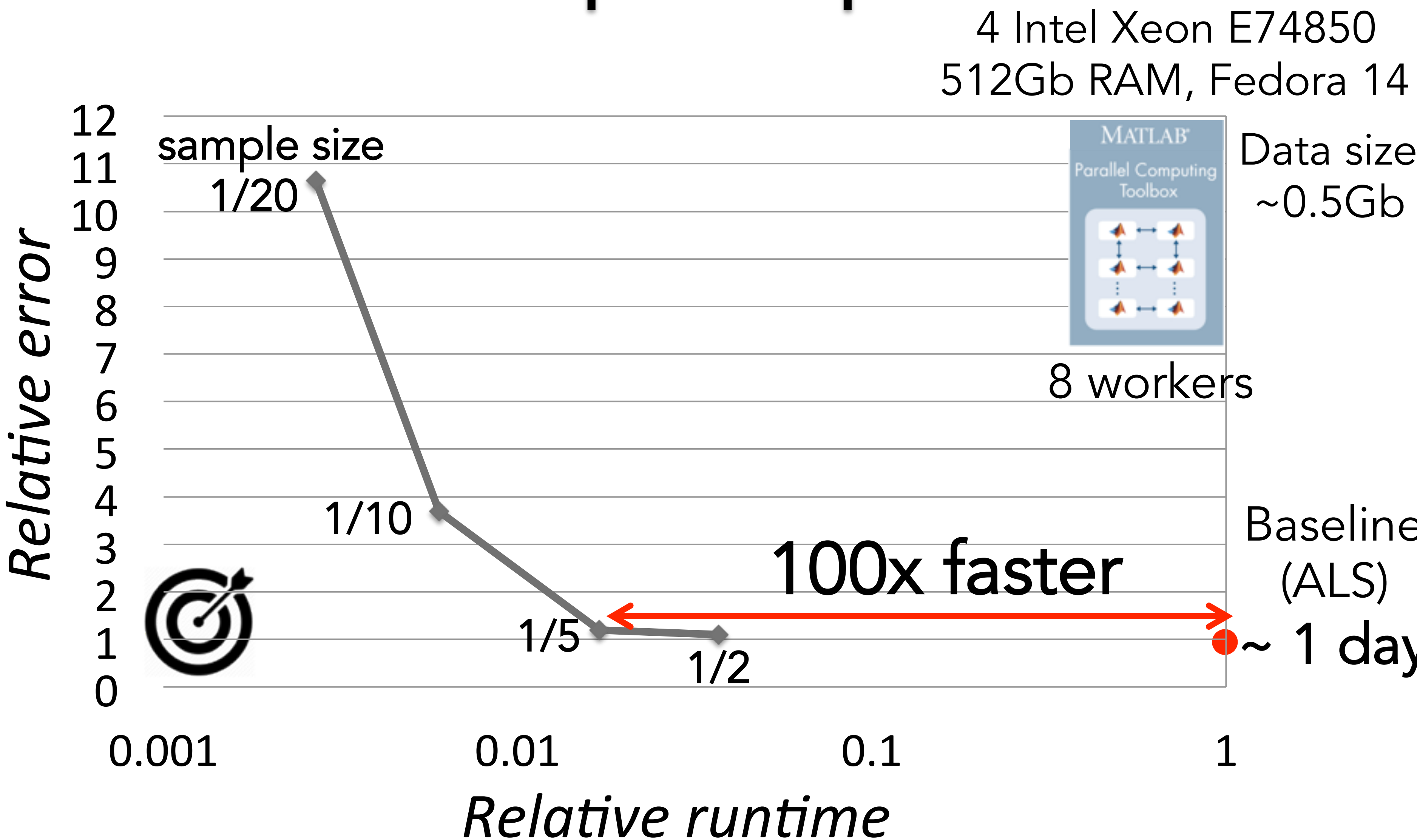
Papalexakis et al. ECML-PKDD 2012

Does it work?

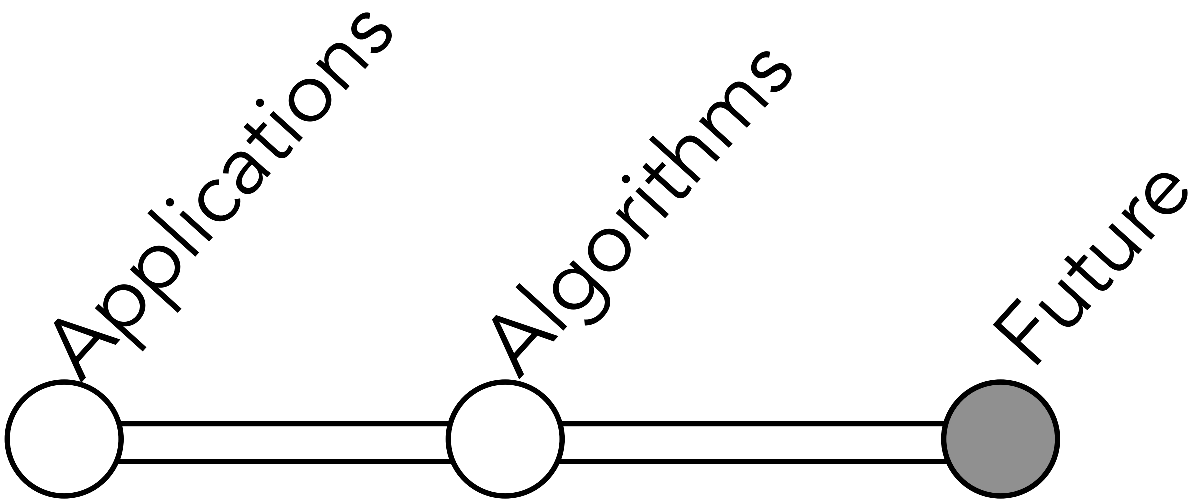


Achieves comparable accuracy to exact algorithm

Speedup


4 Intel Xeon E74850
512Gb RAM, Fedora 14
MATLAB
Parallel Computing Toolbox
8 workers
Data size ~0.5Gb
Baseline (ALS)
~ 1 day
100x faster
sample size
1/20
1/10
1/5
1/2
Relative error
12
11
10
9
8
7
6
5
4
3
2
1
0
0.001
0.01
0.1
1
Relative runtime

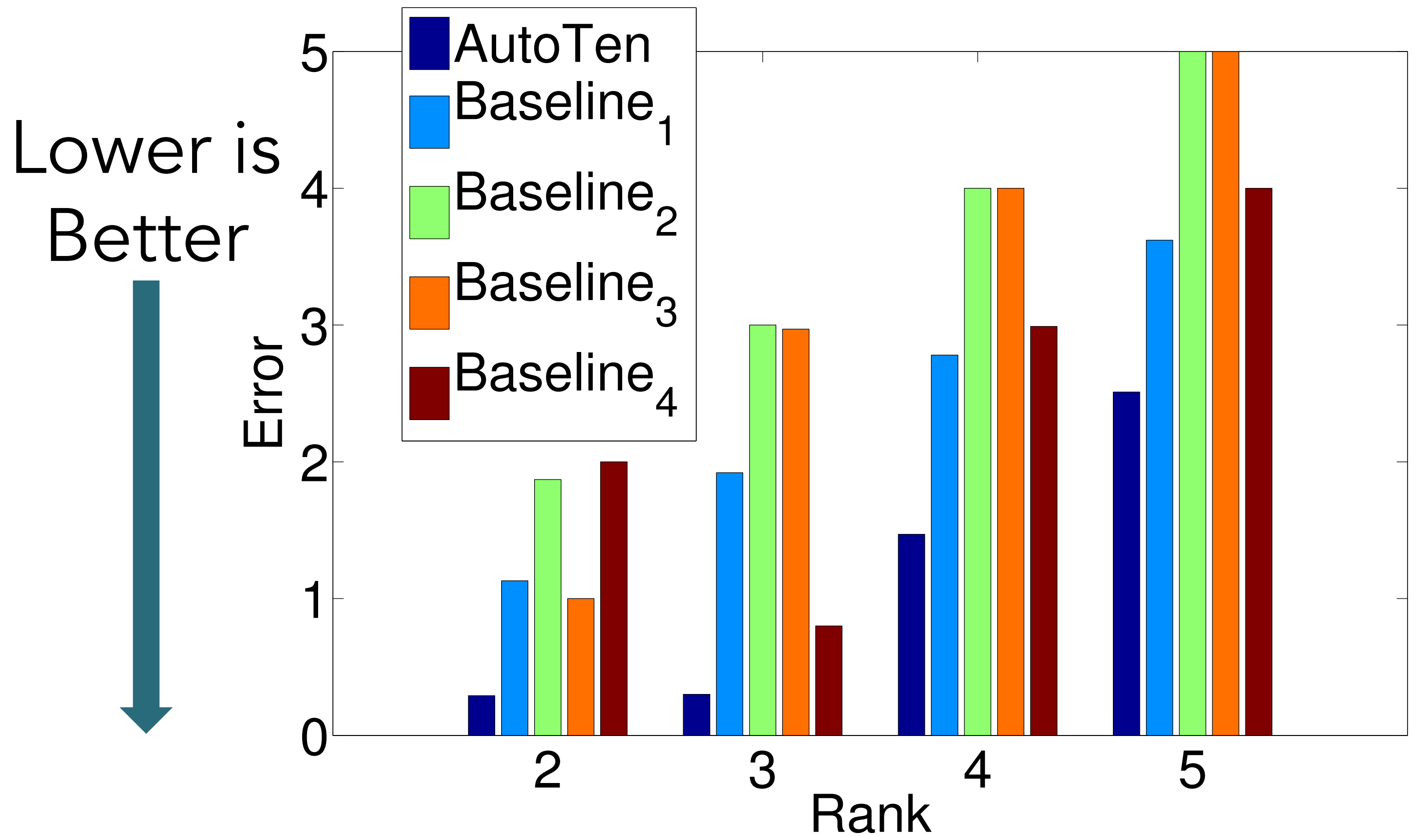
Roadmap



Model Selection & Quality

- Problem 1: **Rank Estimation**
  - Given a model (e.g. PARAFAC), choose the right number of components
  - Do this without any ground truth

Rank Estimation for CP/PARAFAC



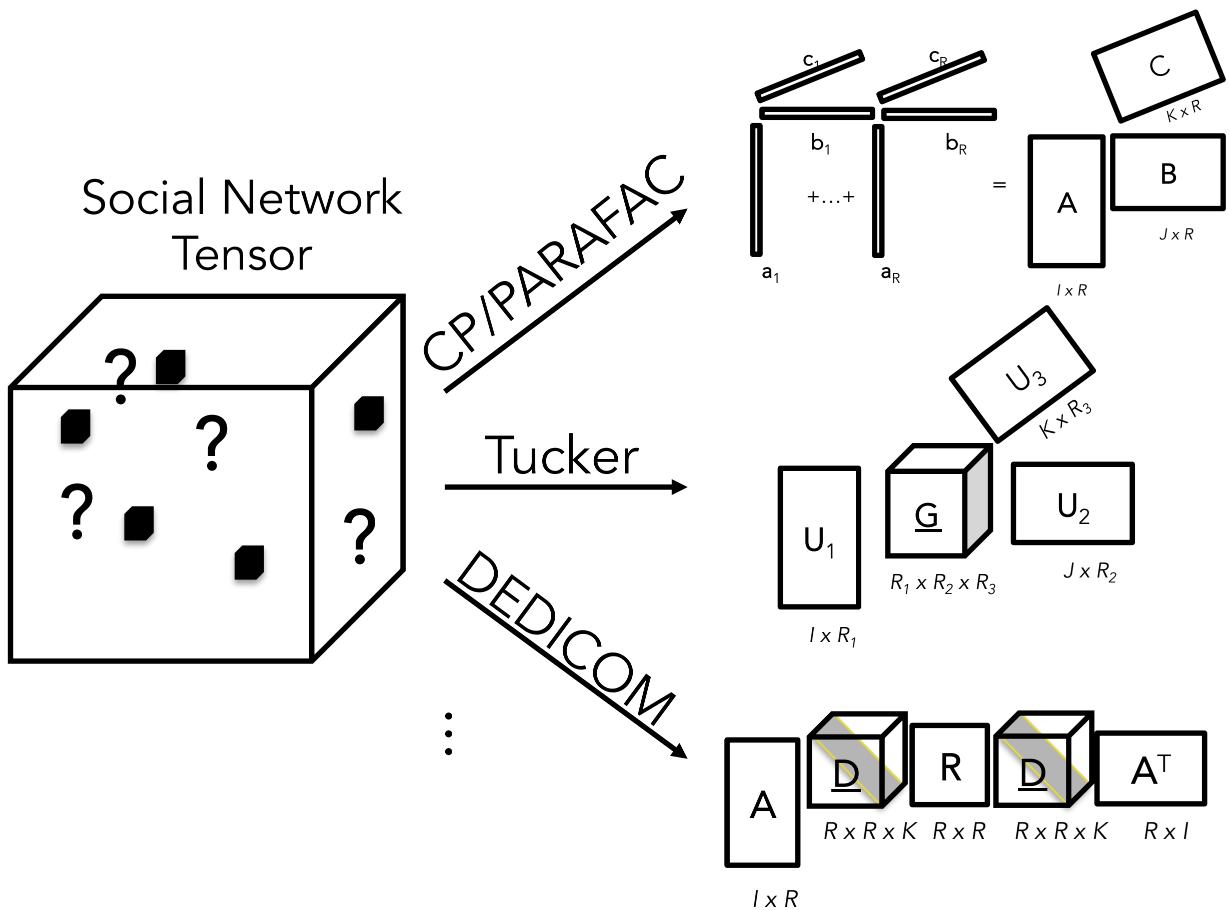
- Maximize both #components and "quality" of decomposition
- Quality is defined through Core Consistency [Bro et al. 2003]

Papalexakis SDM'16 [Best Student Paper Award]

Model Selection & Quality

- Problem 2: **Choosing the right model**
  - ✧ Given an application/dataset, choose the most appropriate tensor decomposition(s)
  - ✧ Again, assume no ground truth!

Choosing the right model



Easy to interpret
Captures trilinear structure
Used for community detection
[Papalexakis et al. Fusion 13]

Captures non-trilinear structure
Used for role discovery
[Gilpin et al. Arxiv.org 16]

Works on undirected graphs
Used for role discovery
[Kolda et al. ICDM 07]

Quick and effective diagnostics?

The End!

- Thank you! Questions??
- How to reach me:

web: http://www.cs.ucr.edu/~epapalex/

e-mail: epapalex@cs.ucr.edu