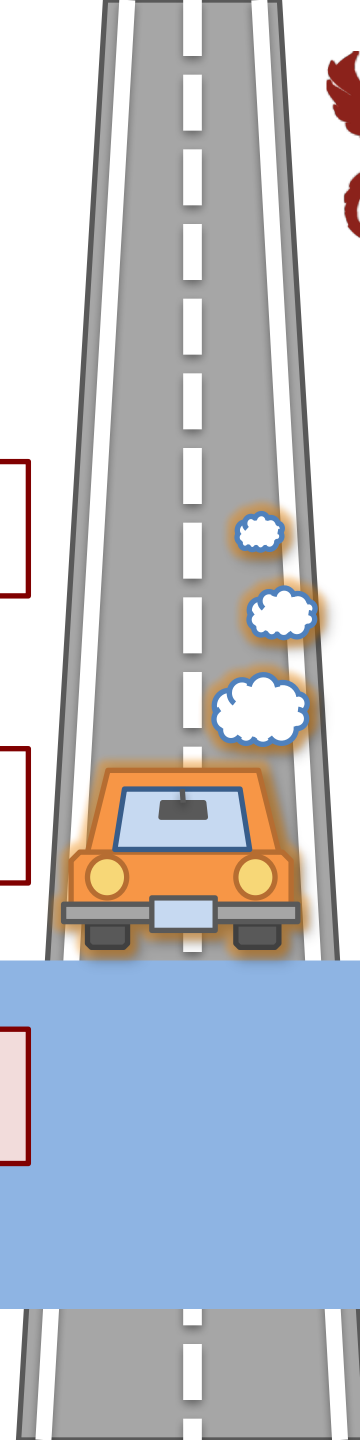
Smart Analytics for Big Time-series Data

Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)

Roadmap

- Motivation
- Similarity search, pattern discovery and summarization — **Part 1**
- Non-linear modeling and forecasting — **Part 2**
- **Extension of time-series data: tensor analysis** — **Part 3**



Part 3

Extension of time-series: tensor analysis

Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)

Outline

- Tensor decomposition
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

Outline

- Tensor decomposition
  - Motivation
  - Basic approaches
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

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-dim 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	...
t=1	13	11	22	55	...
t=2	5	4	6	7	...
t=3	...	...	...	...	...
t=4	...	...	...	...	...
...	...	...	...	...	...

Motivation 2: Why tensors?

- Q: what is a tensor?
- A: N-D generalization of matrix:

kdd' 17	network	search	graph	mining	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

Motivation 2: Why tensors?

- Q: what is a tensor?
- A: N-D generalization of matrix:

kdd’15

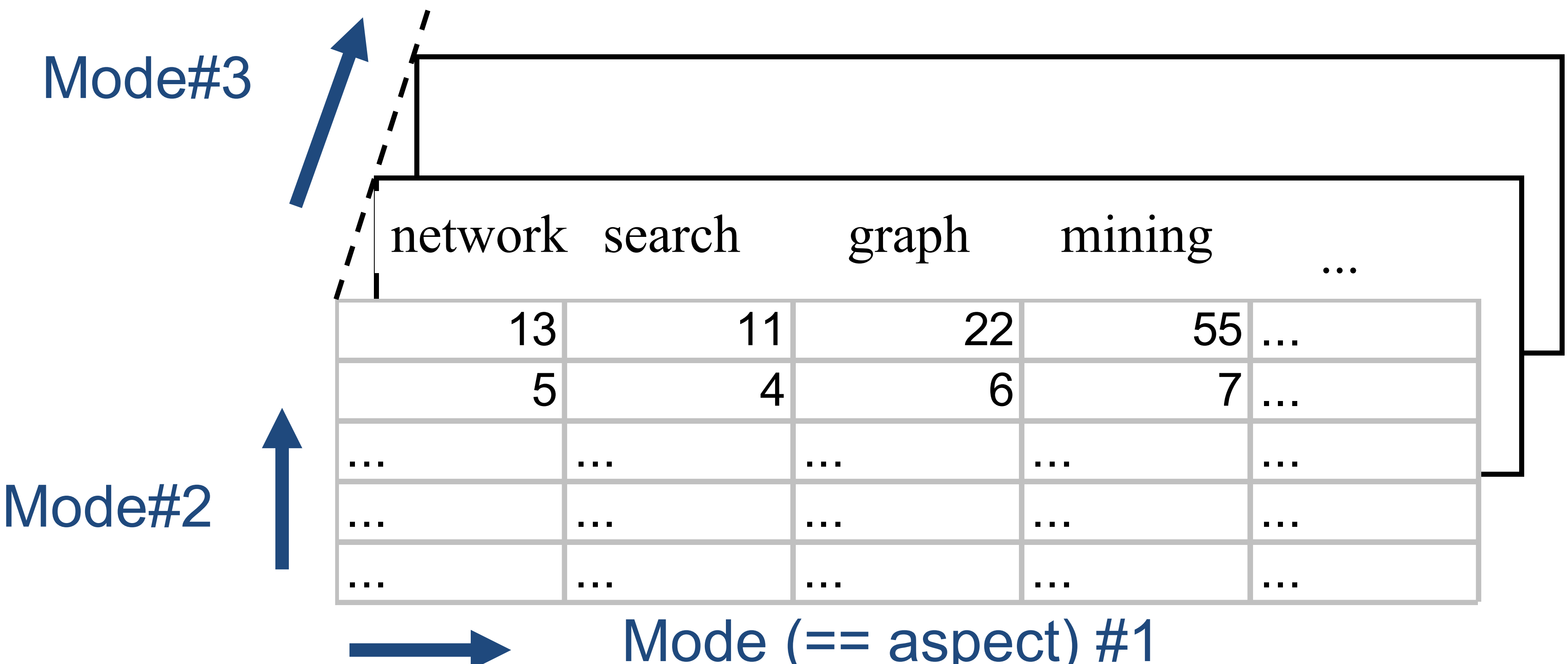
kdd’16

kdd’17

	network	search	graph	mining	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

Tensors are useful for 3 or more modes

Terminology: 'mode' (or 'aspect'):



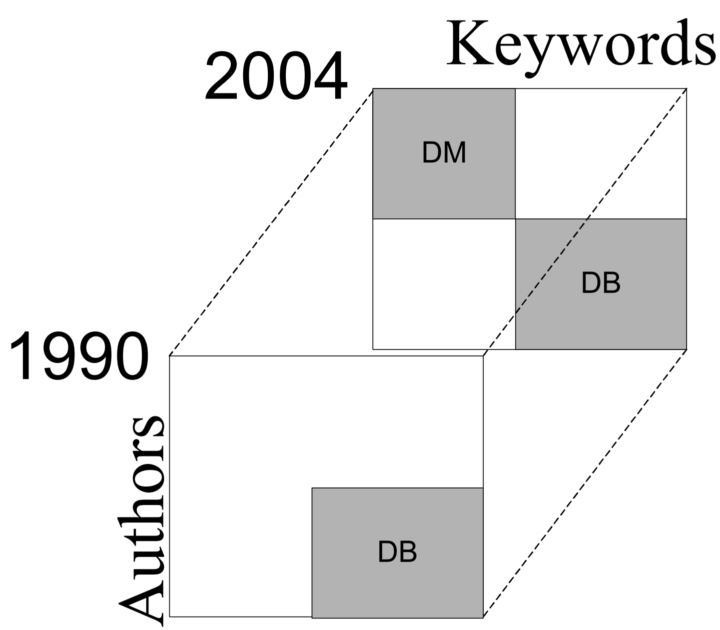
13	11	22	55	...
5	4	6	7	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Motivating Applications

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

P1: Social network analysis

- Monitoring networks and community structures over time



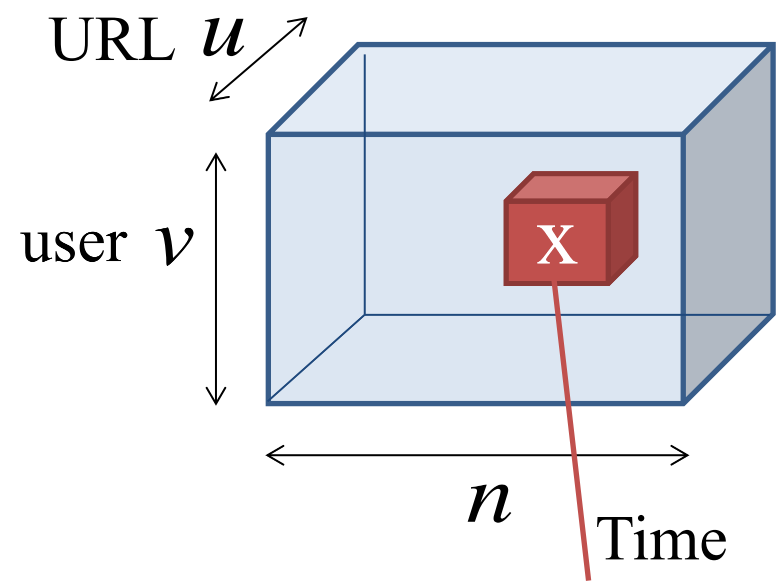
P2: Web graph mining

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

Tensor analysis for time-series data

- Time-stamped events
  - *e.g., web clicks*

Time	URL	User
08-01-12:00	CNN.com	Smith
08-02-15:00	YouTube.com	Brown
08-02-19:00	CNET.com	Smith
08-03-11:00	CNN.com	Johnson
…	…	…



Represent as
M^{th} order tensor (M=3)

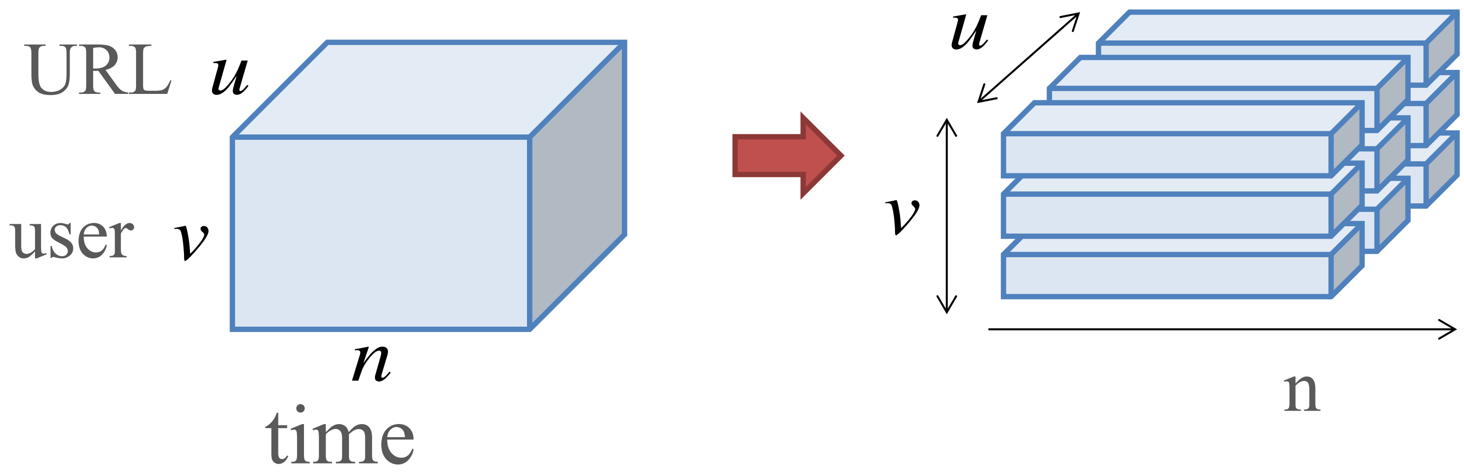
$$\mathcal{X} \in \mathbb{N}^{u \times v \times n}$$

Element x: # of events

e.g., ‘Smith’, ‘CNN.com’, ‘Aug 1, 10pm’; 21 times

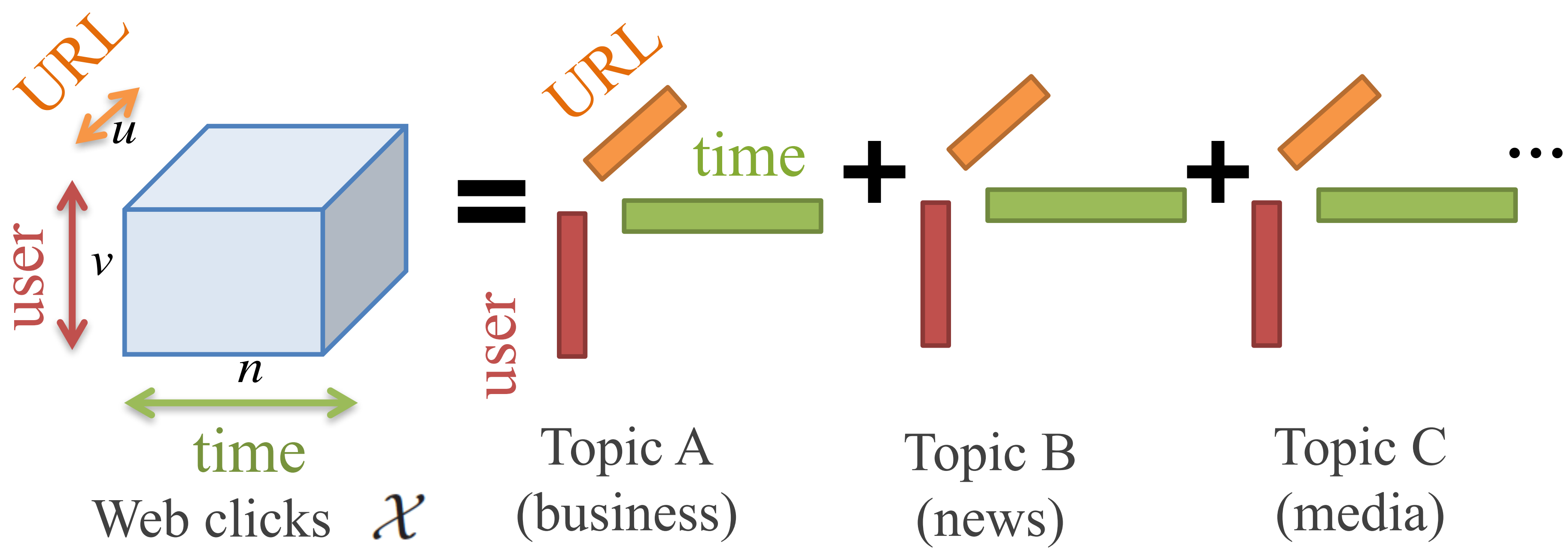
Tensor analysis for time-series data

- Individual-sequence mining
  - Create a set of (u * v) sequences of length (n)
  - Apply the mining algorithm for each sequence



Tensor analysis for time-series data

- Multi-aspect time-series analysis



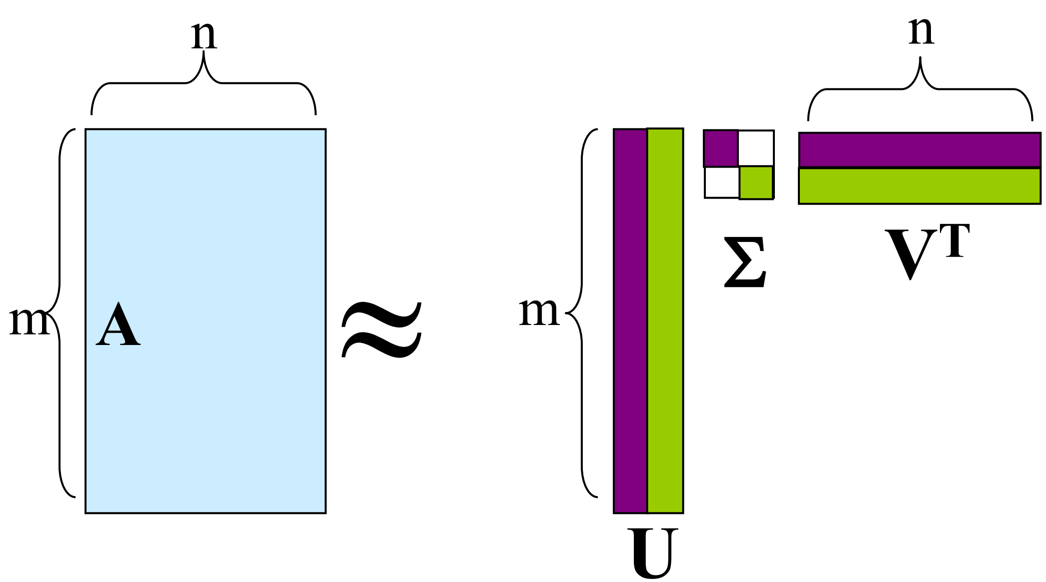
Outline

- Tensor decomposition
  - Motivation
  - Basic approaches
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

Reminder: SVD

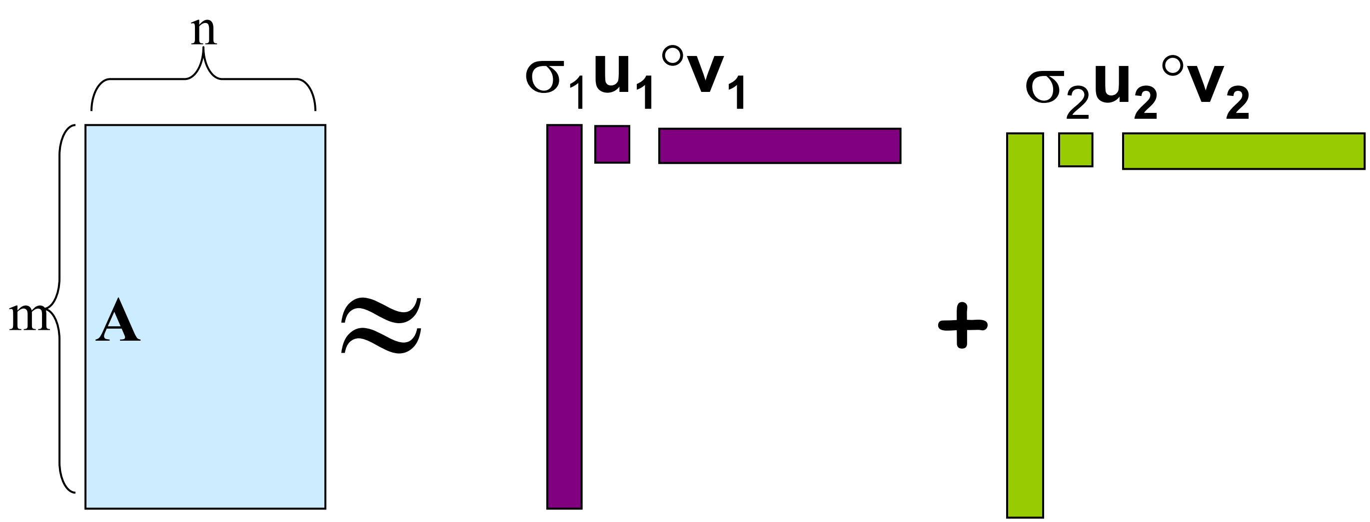
$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \textstyle\sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



– Best rank-k approximation in L2

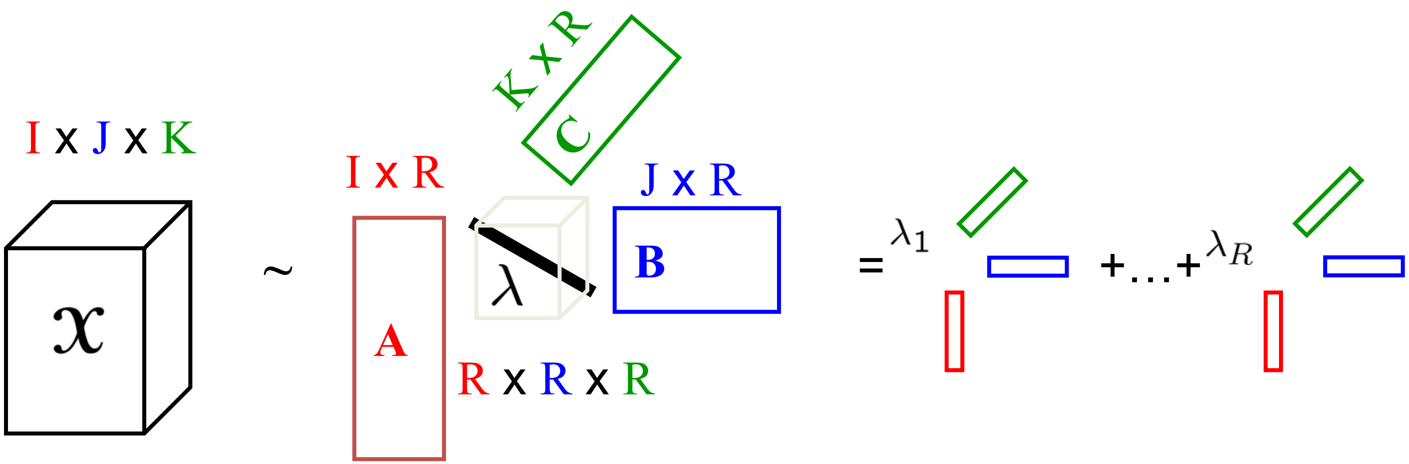
Reminder: SVD

$$\mathbf{A} \approx \mathbf{U}\Sigma\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



– Best rank-k approximation in L2

Goal: extension to >=3 modes



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

Main points:

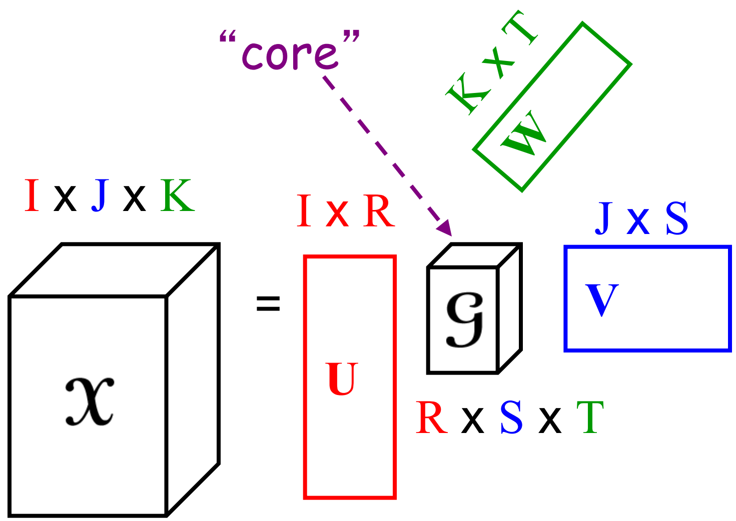
- 2 major types of tensor decompositions: PARAFAC and Tucker
- both can be solved with ``alternating least squares'' (ALS)

Specially Structured Tensors

- Tucker Tensor

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$
$$= \sum_r \sum_s \sum_t g_{rst}\, \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t$$
$$\equiv [\![\mathcal{G}\,;\mathbf{U},\mathbf{V},\mathbf{W}]\!]$$

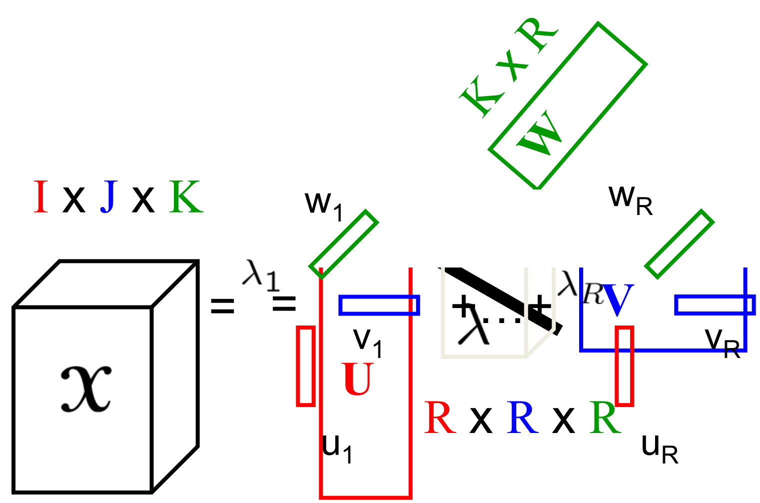
Our Notation



- Kruskal Tensor

$$\mathcal{X} = \sum_r \lambda_r\, \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$
$$\equiv [\![\lambda\,;\mathbf{U},\mathbf{V},\mathbf{W}]\!]$$

Our Notation



Details

Specially Structured Tensors

CMU CS

- Tucker Tensor

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$
$$= \sum_r \sum_s \sum_t g_{rst}\, \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t$$
$$\equiv [\![\mathcal{G} \,;\, \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

In matrix form:

$$\mathbf{X}_{(1)} = \mathbf{U}\mathbf{G}_{(1)}(\mathbf{W} \otimes \mathbf{V})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{V}\mathbf{G}_{(2)}(\mathbf{W} \otimes \mathbf{U})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{W}\mathbf{G}_{(3)}(\mathbf{V} \otimes \mathbf{U})^{\mathsf{T}}$$

$$\mathrm{vec}(\mathcal{X}) = (\mathbf{W} \otimes \mathbf{V} \otimes \mathbf{U})\mathrm{vec}(\mathcal{G})$$

- Kruskal Tensor

$$\mathcal{X} = \sum_r \lambda_r\, \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$
$$\equiv [\![\lambda \,;\, \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$$

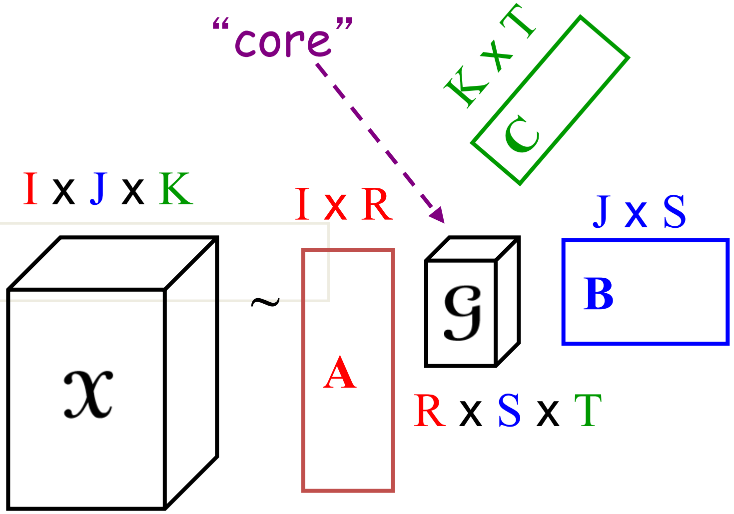
In matrix form:

Let $\mathbf{\Lambda} = \mathrm{diag}(\lambda)$

$$\mathbf{X}_{(1)} = \mathbf{U}\,\mathbf{\Lambda}\,(\mathbf{W} \odot \mathbf{V})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{V}\,\mathbf{\Lambda}\,(\mathbf{W} \odot \mathbf{U})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{W}\,\mathbf{\Lambda}\,(\mathbf{V} \odot \mathbf{U})^{\mathsf{T}}$$

$$\mathrm{vec}(\mathcal{X}) = (\mathbf{W} \odot \mathbf{V} \odot \mathbf{U})\,\lambda$$

Tucker Decomposition - intuition



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

Intuition behind core tensor

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

eg, terms x documents

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

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

med. doc

cs doc

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

med. terms

cs terms

common terms

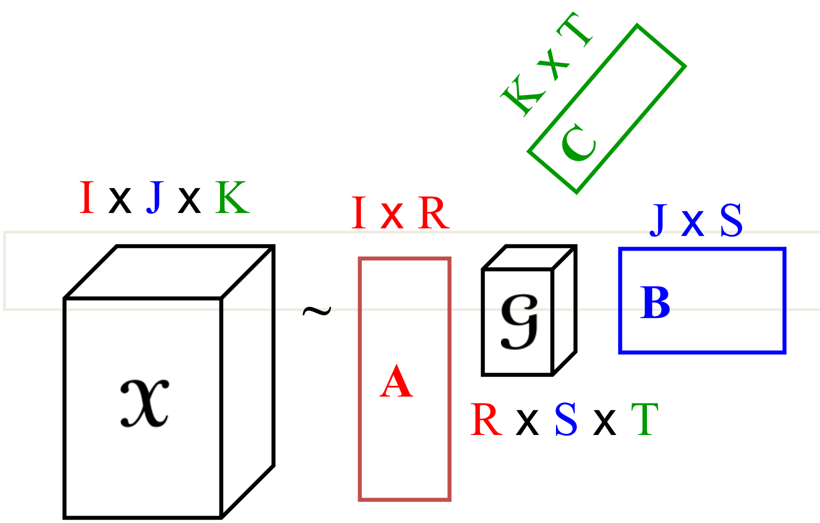
term group x doc. group

$$\begin{bmatrix} .5 & 0 & 0 \\ .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & .5 \end{bmatrix} \begin{bmatrix} .3 & 0 \\ 0 & .3 \\ .2 & .2 \end{bmatrix} \begin{bmatrix} .36 & .36 & .28 & 0 & 0 & 0 \\ 0 & 0 & 0 & .28 & .36 & .36 \end{bmatrix} = \begin{bmatrix} .054 & .054 & .042 & 0 & 0 & 0 \\ .054 & .054 & .042 & 0 & 0 & 0 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ .036 & .036 & 028 & .028 & .036 & .036 \\ .036 & .036 & .028 & .028 & .036 & .036 \end{bmatrix}$$

doc x doc group

term x term-group

Tucker Decomposition



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

Given A, B, 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
- **A**, **B**, and **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

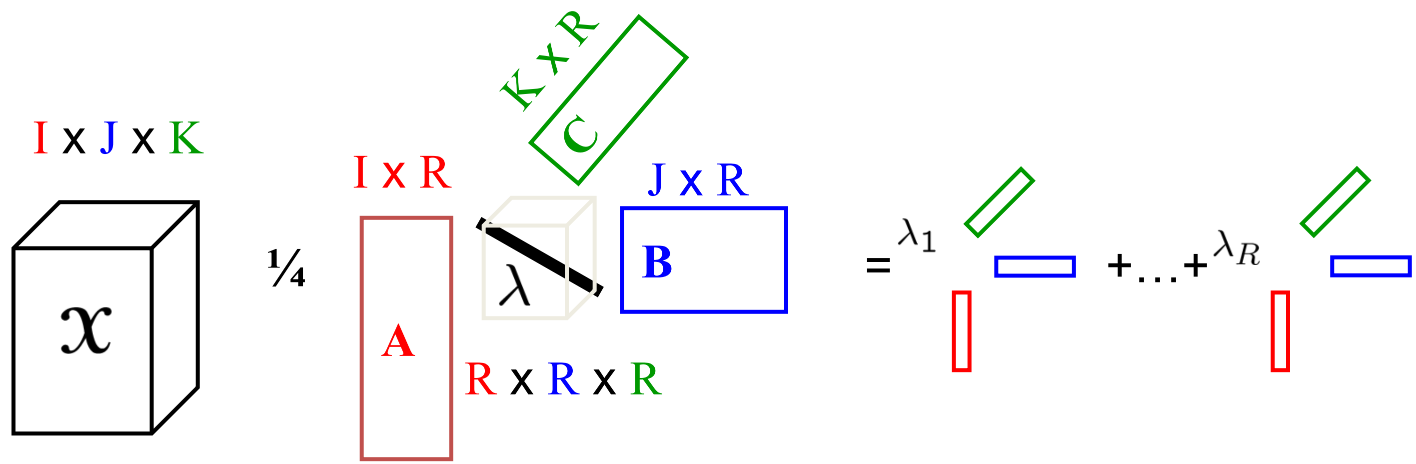
$$\mathbf{X}_{(1)} = \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}}$$

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

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

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

CANDECOMP/PARAFAC Decomposition



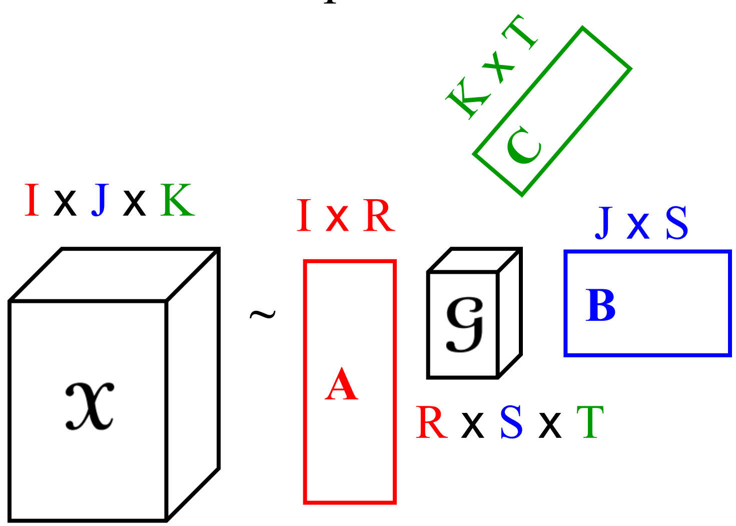
$$\mathcal{X} \approx [\![\lambda\,;\mathbf{A},\mathbf{B},\mathbf{C}]\!] = \sum_r \lambda_r\, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

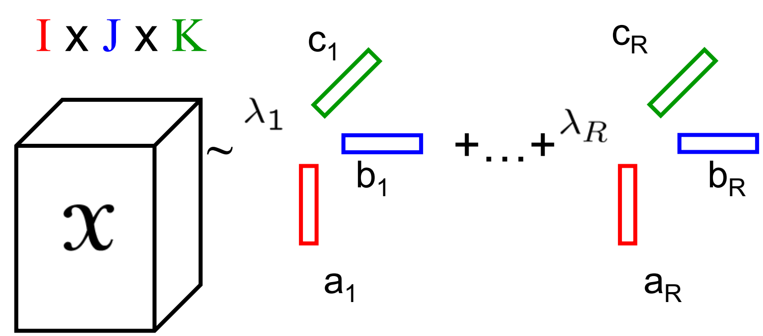
- CANDECOMP = Canonical Decomposition (Carroll & Chang, 1970)
- PARAFAC = Parallel Factors (Harshman, 1970)
- Core is <u>diagonal</u> (specified by the vector λ)
- Columns of **A**, **B**, and **C** are <u>not</u> orthonormal
- If R is <u>*minimal*</u>, then R is called the **rank** of the tensor (Kruskal 1977)
- Can have rank( ) > 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:

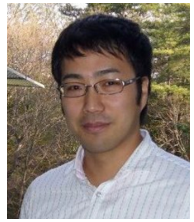
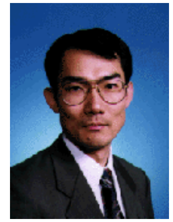
http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/

Outline

- Tensor decomposition
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

M A N T

[Matsubara+ KDD’12]

Fast Mining and Forecasting of Complex Time-Stamped Events

Yasuko Matsubara (Kyoto University)

Yasushi Sakurai (NTT)

Christos Faloutsos (CMU)

Tomoharu Iwata (NTT)

Masatoshi Yoshikawa (Kyoto University)

Motivation

- Complex time-stamped events

{timestamp + multiple attributes}

e.g., web click events:

{timestamp, URL, user ID, access devices, http referrer,…}

Timestamp	URL	User	Device
2012-08-01-12:00	CNN.com	Smith	iphone
2012-08-02-15:00	YouTube.com	Brown	iphone
2012-08-02-19:00	CNET.com	Smith	mac
2012-08-03-11:00	CNN.com	Johnson	ipad
…	…	…	…

Motivation

Q1. Are there any topics ?

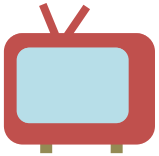
- news, tech, media, sports, etc...

Timestamp	URL	User	Device
2012-08-01-12:00	CNN.com	Smith	iphone
2012-08-02-15:00	YouTube.com	Brown	iphone
2012-08-02-19:00	CNET.com	Smith	mac
2012-08-03-11:00	CNN.com	Johnson	ipad
…	…	…	…

e.g.,

CNN.com, CNET.com -> news topic

YouTube.com -> media topic



Motivation

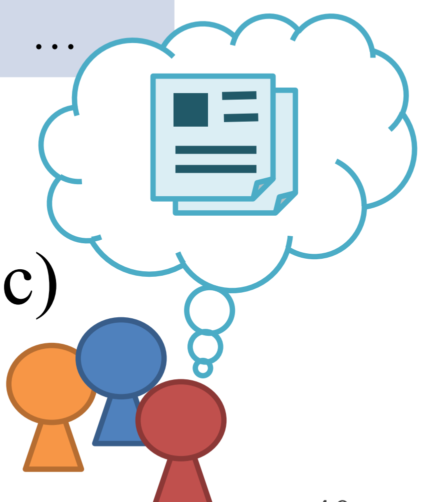
Q2. Can we group URLs/users accordingly?

Timestamp	URL	User	Device
2012-08-01-12:00	CNN.com	Smith	iphone
2012-08-02-15:00	YouTube.com	Brown	iphone
2012-08-02-19:00	CNET.com	Smith	mac
2012-08-03-11:00	CNN.com	Johnson	ipad
…	…	…	…

e.g.,

CNN.com & CNET.com (related to news topic)

Smith & Johnson (related to news topic)



Motivation

Q3. Can we forecast future events?

- How many clicks from 'Smith' tomorrow?
- How many clicks to 'CNN.com' over next 7 days?

future clicks?

Timestamp	URL	User	Device
2012-08-01-12:00	CNN.com	Smith	iphone
2012-08-02-15:00	YouTube.com	Brown	iphone
2012-08-02-19:00	CNET.com	Smith	mac
2012-08-03-11:00	CNN.com	Johnson	ipad
2012-08-05-12:00	CNN.com	Smith	iphone
2012-08-05-19:00	CNET.com	Smith	iphone

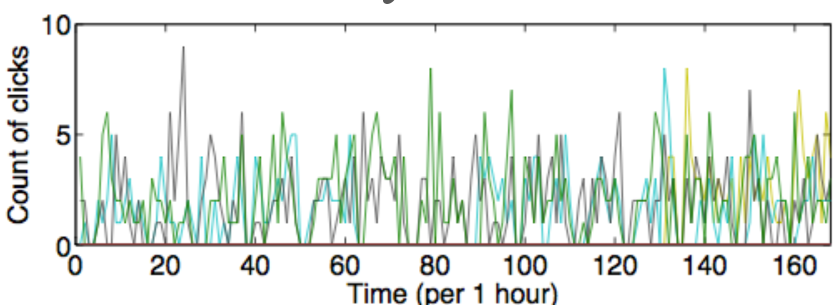
Motivation

Web click events – can we see any trends?

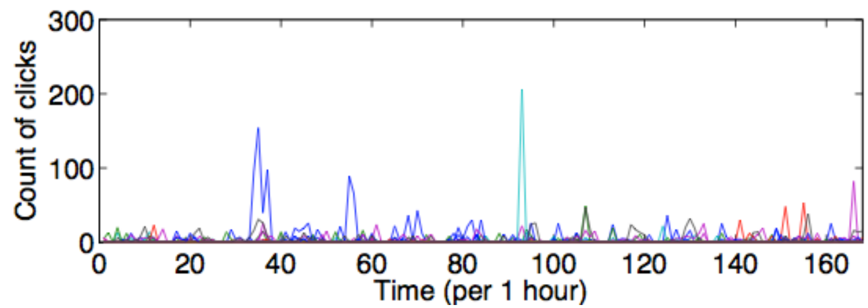
Original access counts of each URL

- 100 random users
- 1 week (window size = 1 hour)

URL: money site



URL: blog site



Motivation

Web click events – can we see any trends?

Original access counts of each URL

users

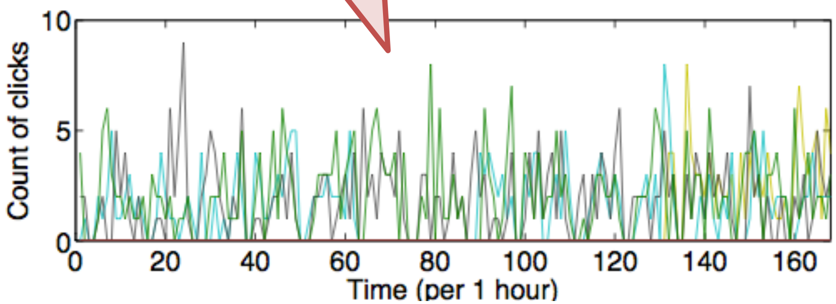
indow siz

Noisy ☹

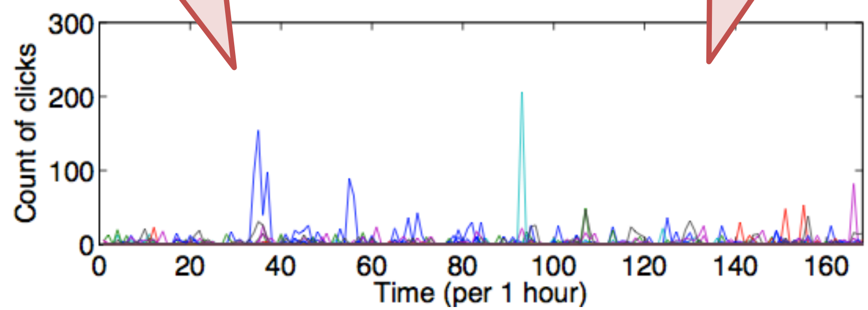
Sparse ☹

Bursty ☹

URL: mo y site



UR blog site



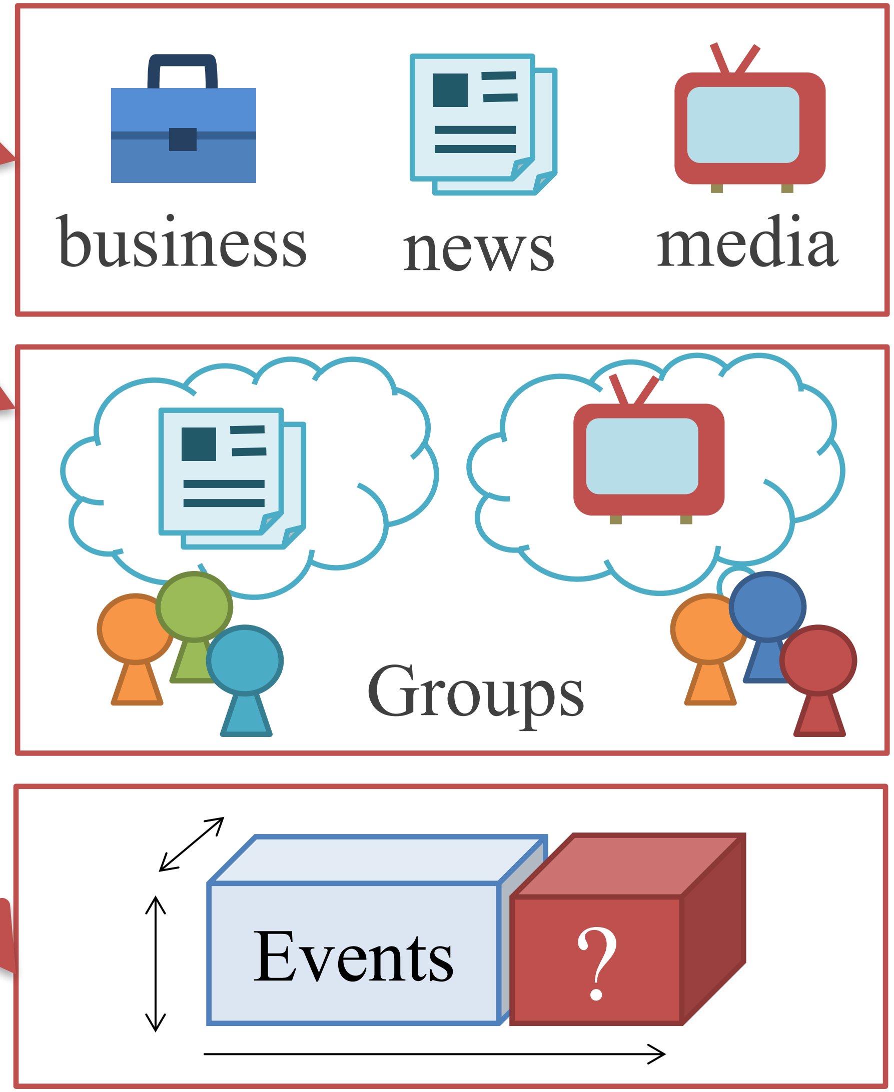
☹ We cannot see any trends !!

Our goals

Q1: Hidden topics

Q2: Groups

Q3: Forecasting



Kumamoto U

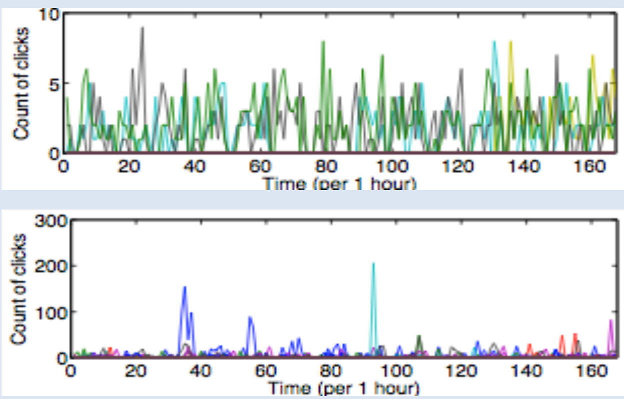
CMU CS

Problem definition

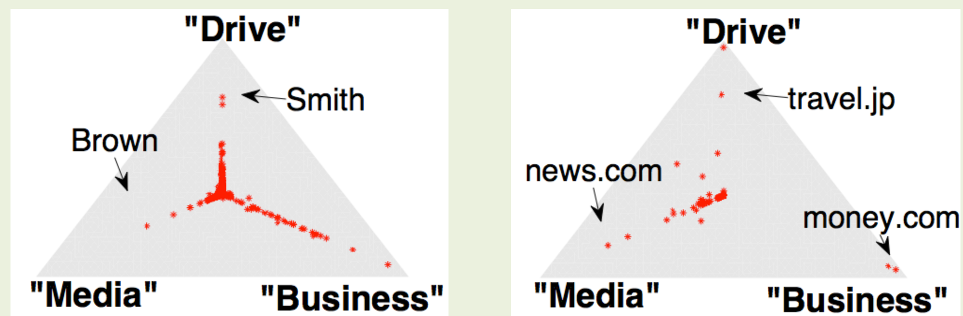
Given: a set of complex time-stamped events

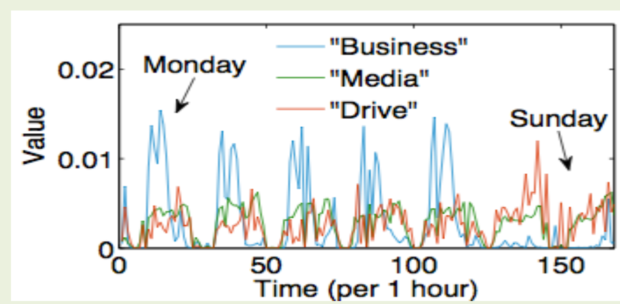
1. **Find**: major topics/trends
2. **Forecast**: future events

Original web-click events



URL in topic space | **User in topic space**





Time evolution

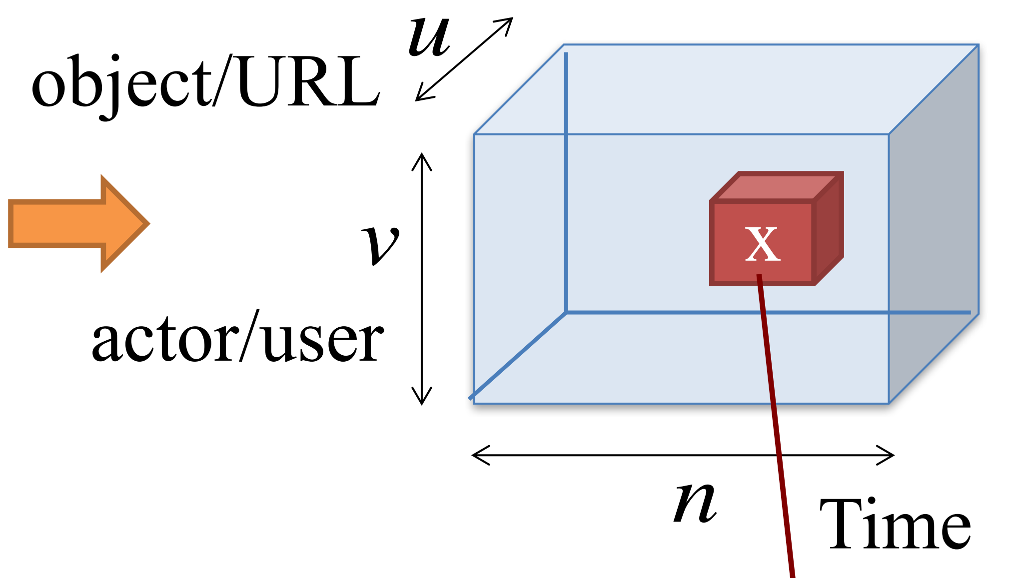
"Hidden topics" wrt each aspect *(URL, user, time)*

Main idea (1) : M-way analysis

Complex time-stamped events

e.g., web clicks

Time	URL	User
08-01-12:00	CNN.com	Smith
08-02-15:00	YouTube.com	Brown
08-02-19:00	CNET.com	Smith
08-03-11:00	CNN.com	Johnson
…	…	…



Represent as Mth order tensor (M=3)

$$\mathcal{X} \in \mathbb{N}^{u \times v \times n}$$

Element x: # of events

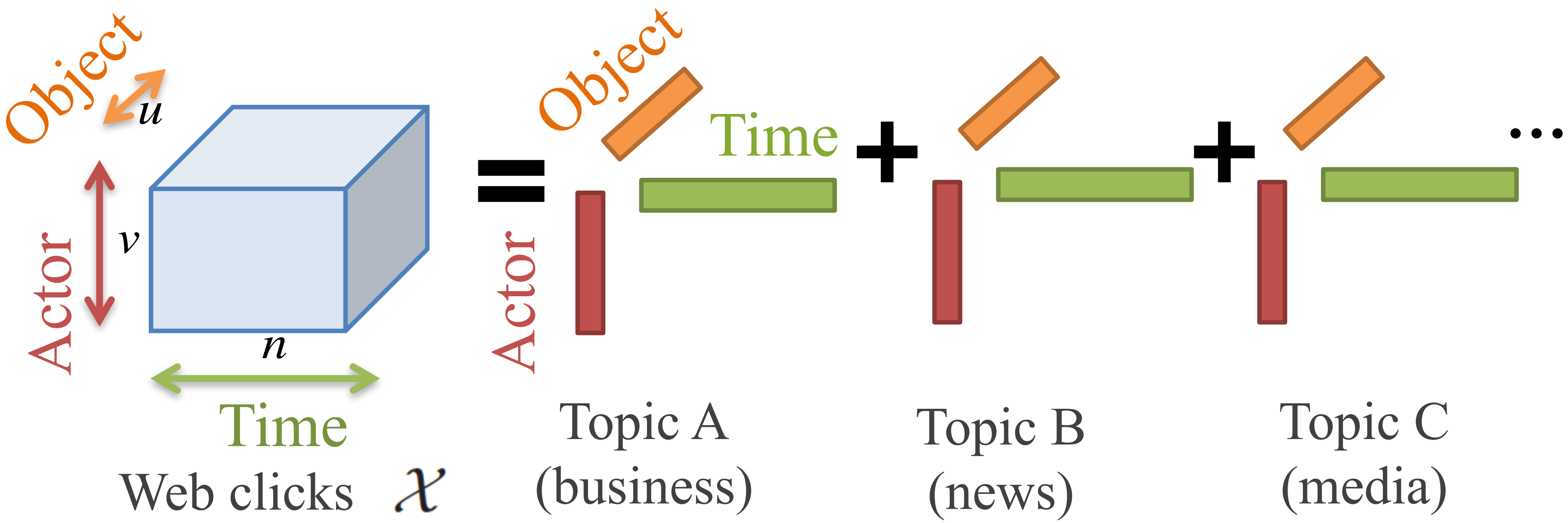
e.g., 'Smith', 'CNN.com', 'Aug 1, 10pm'; 21 times

Main idea (1) : M-way analysis

A. decompose to a set of **3 topic vectors:**

- **Object vector** **Actor vector** **Time vector**

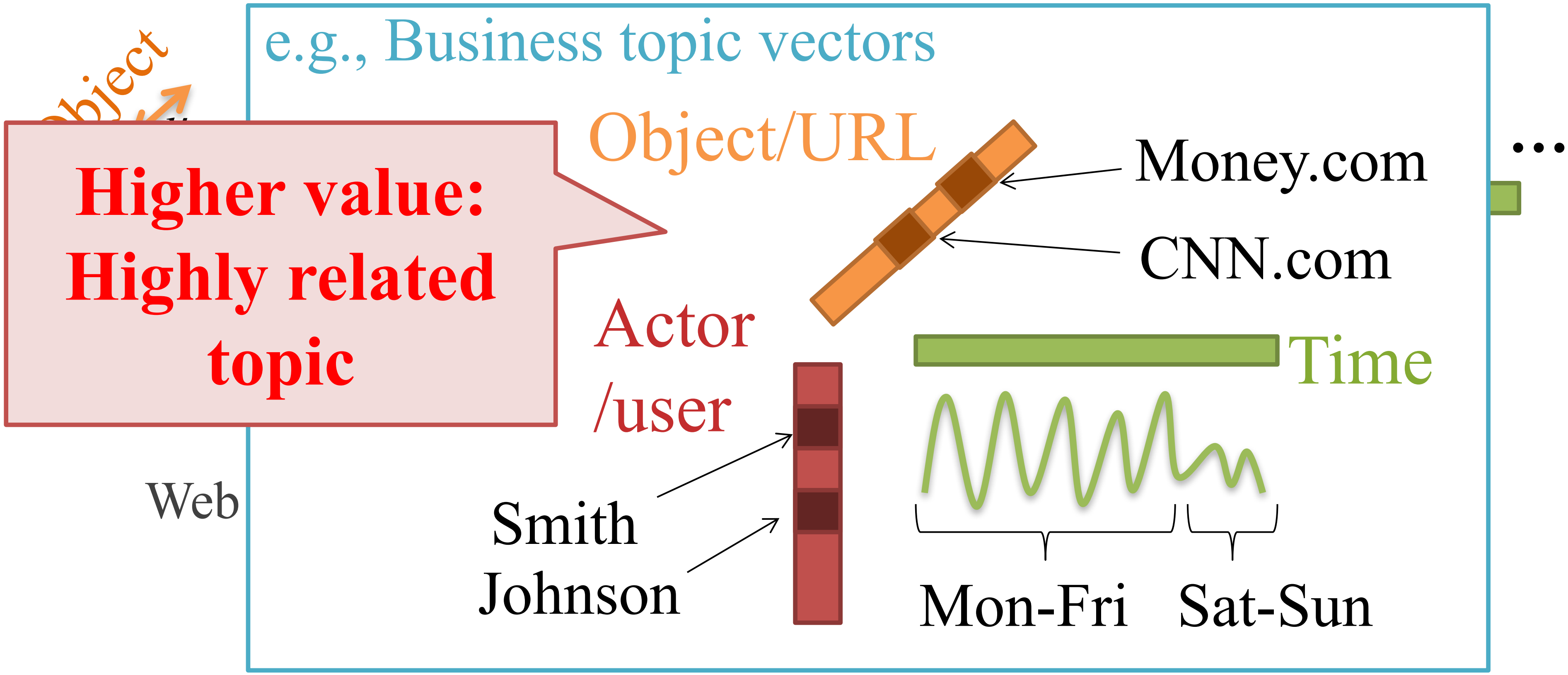
URL user Time



Main idea (1) : M-way analysis

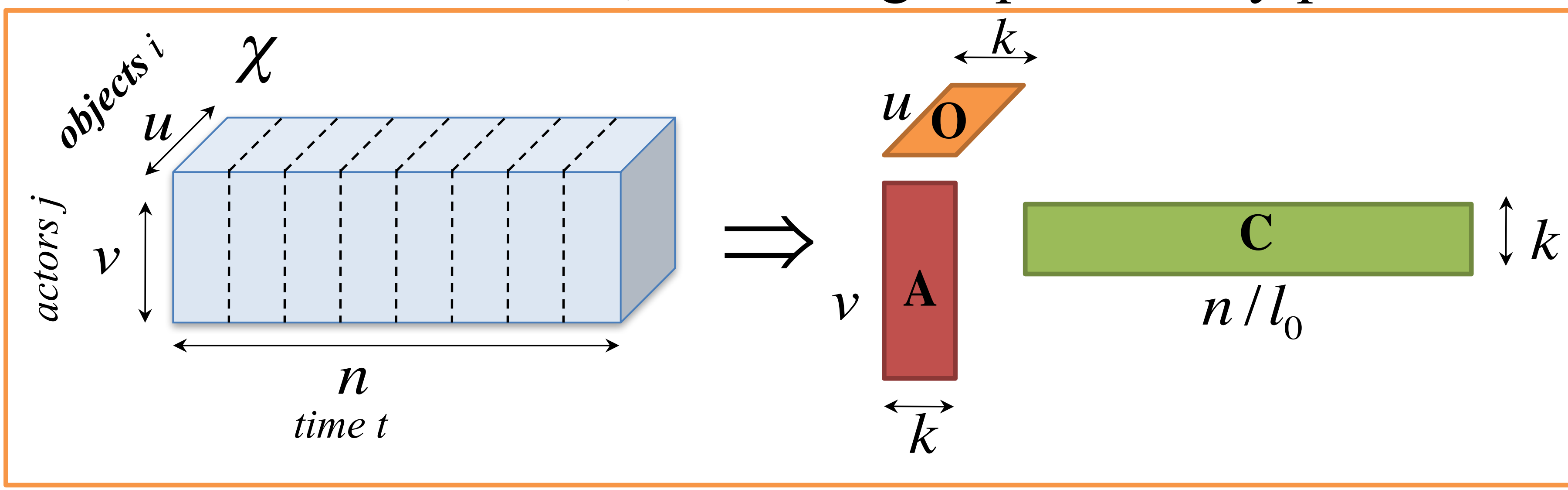
A. decompose to a set of **3 topic vectors:**

- **Object vector** **Actor vector** **Time vector**



Main idea (1) : M-way analysis (details)

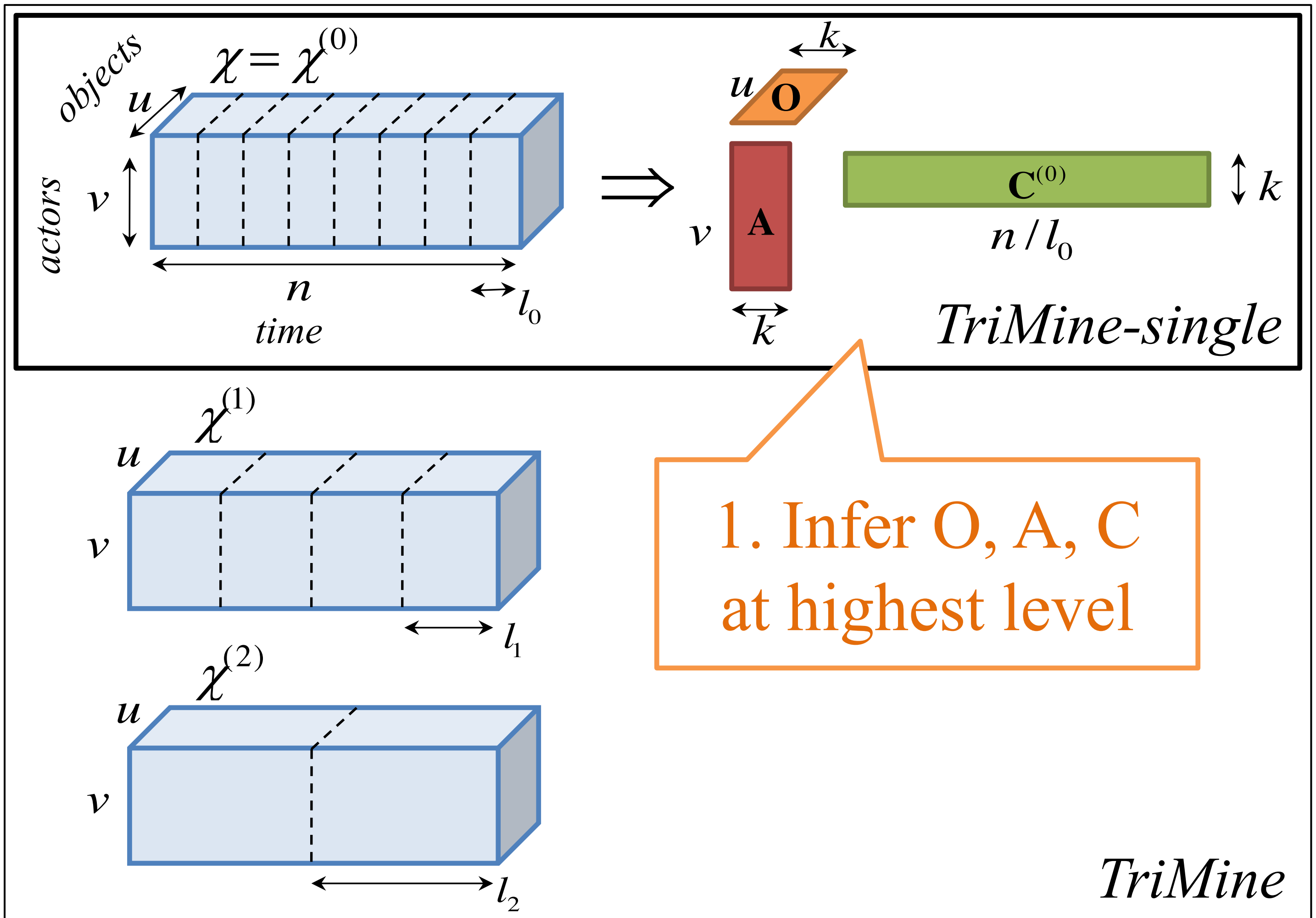
- M-way decomposition (M=3)
  - **[Gibbs sampling]** infer k hidden topics for each non-zero element of X, according to probability p



$$
\begin{aligned}
p(z_{i,j,t} &= r | \mathcal{X}, \mathbf{O}', \mathbf{A}', \mathbf{C}', \alpha, \beta, \gamma) \qquad (1) \\
&\propto \frac{o'_{i,r} + \alpha}{\sum_r o'_{i,r} + \alpha k} \cdot \frac{a'_{r,j} + \beta}{\sum_j a'_{r,j} + \beta v} \cdot \frac{c'_{r,t} + \gamma}{\sum_t c'_{r,t} + \gamma n}
\end{aligned}
$$

Main idea (2) : Multi-scale analysis (details)

- Tensors with multiple window sizes



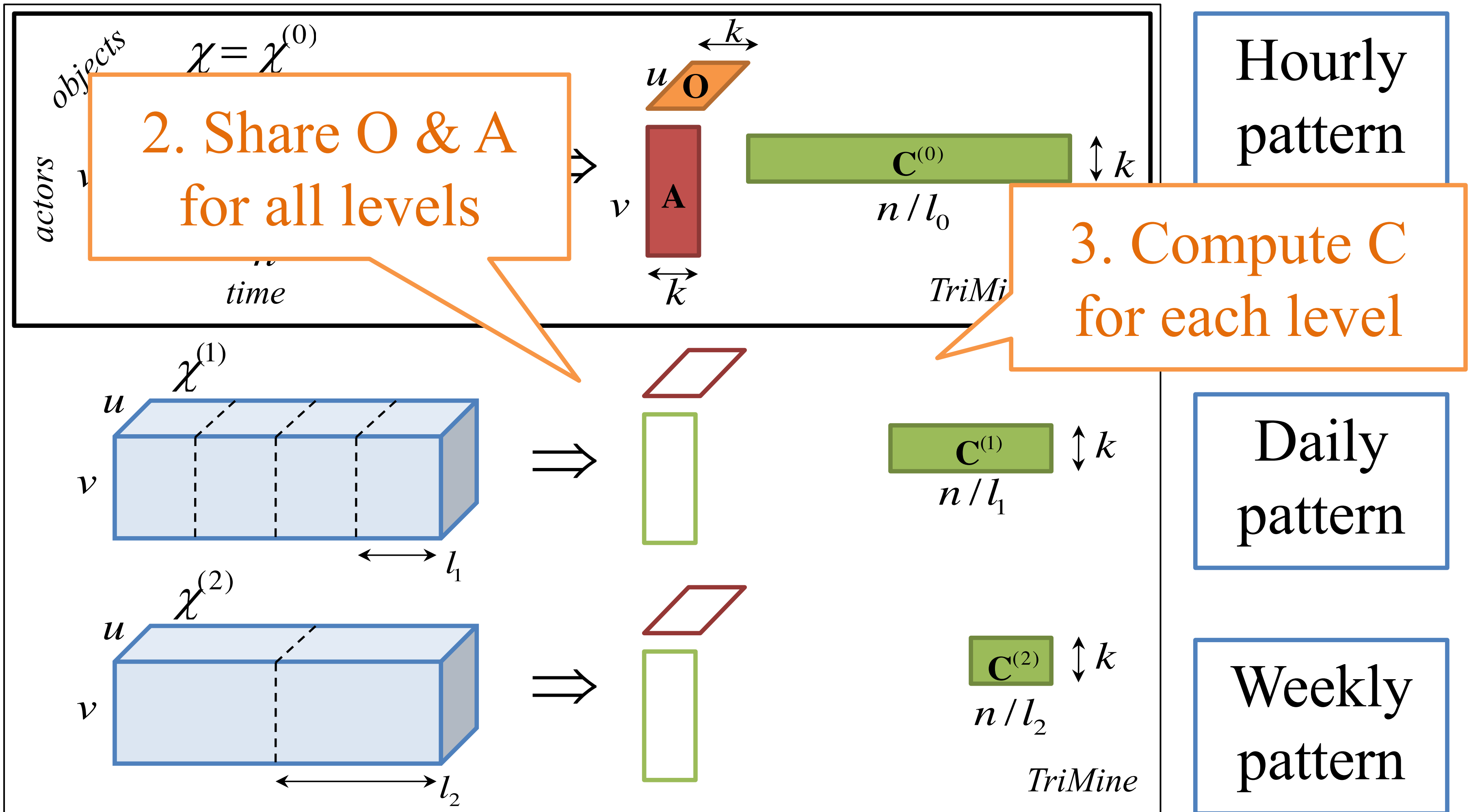
Hourly pattern

Daily pattern

Weekly pattern

Main idea (2) : Multi-scale analysis (details)

- Tensors with multiple window sizes



Hourly pattern

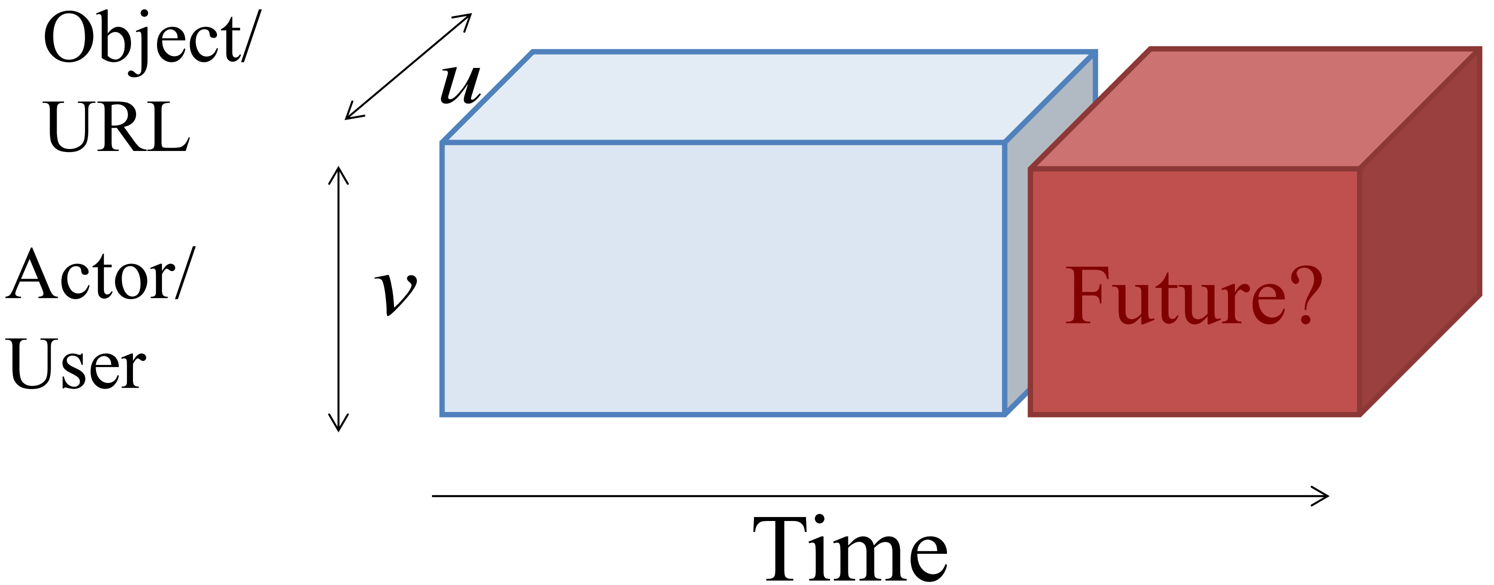
Daily pattern

Weekly pattern

TriMine-Forecasts

Our final goal: “forecast future events”!

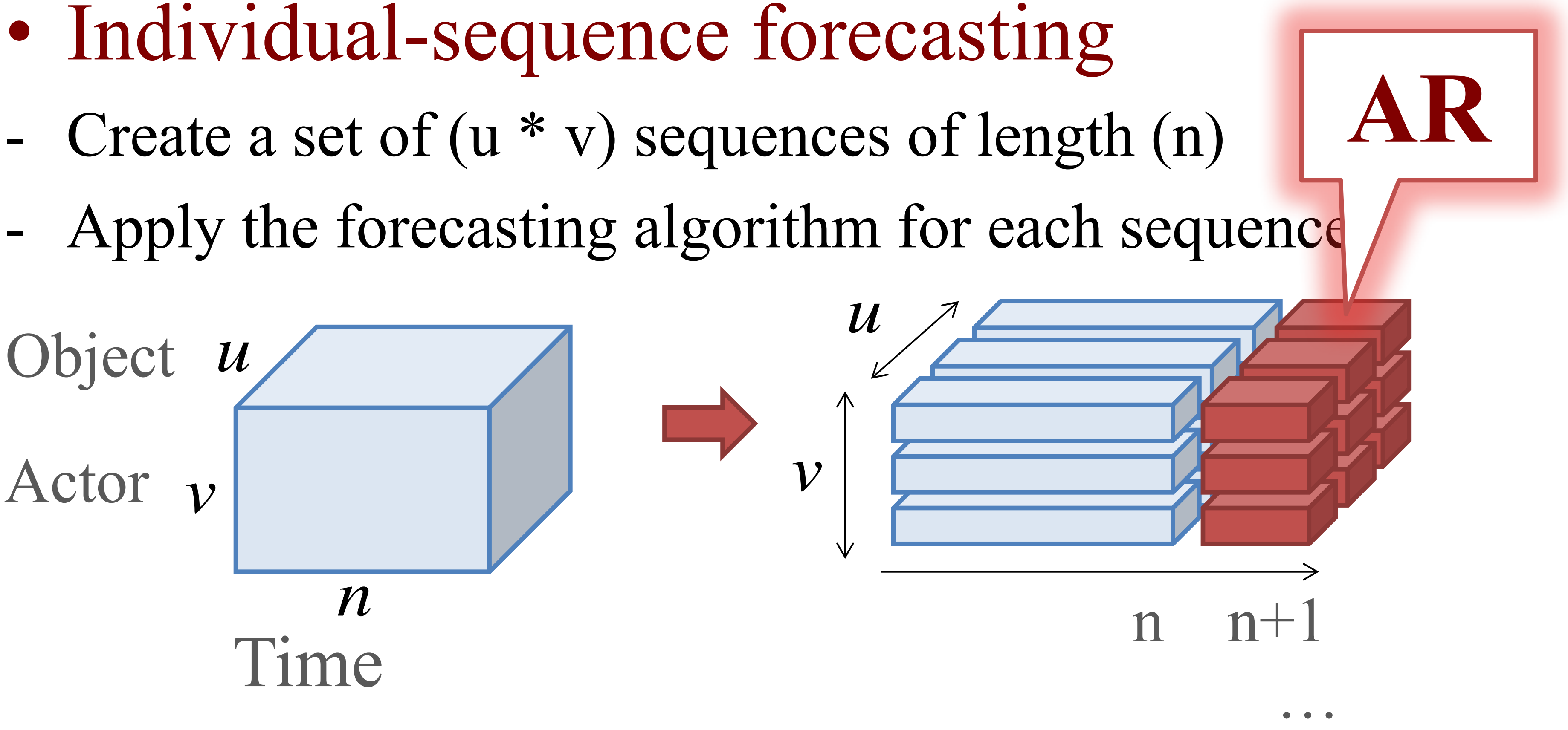
Q. How can we generate a realistic events?



e.g., estimate the number of clicks for user “smith”, to URL “CNN.com”, for next 10 days

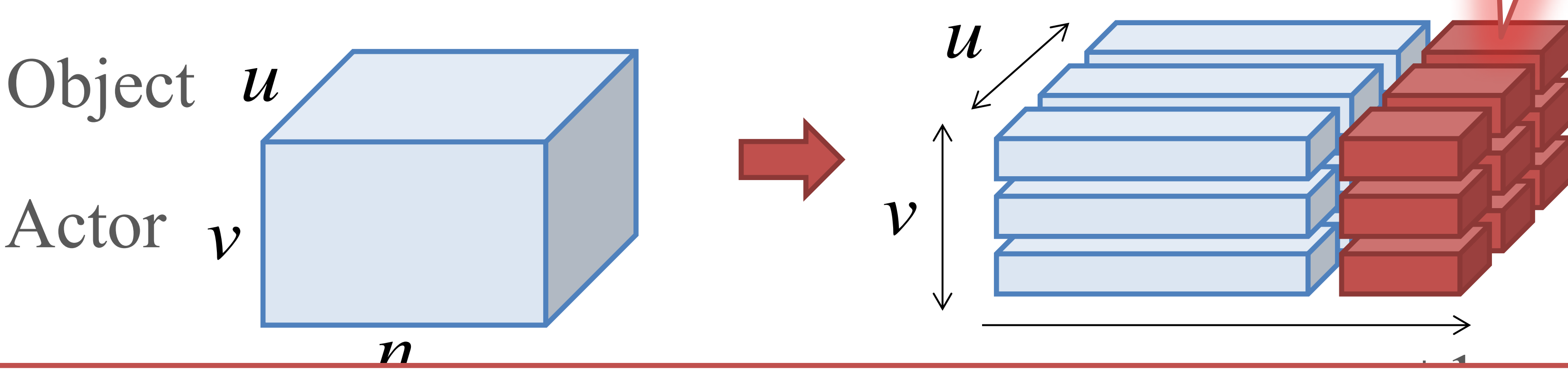
Why not naïve?

- Individual-sequence forecasting
  - Create a set of (u * v) sequences of length (n)
  - Apply the forecasting algorithm for each sequence



Why not naïve?

- Individual-sequence forecasting
  - Create a set of (u * v) sequences of length (n)
  - Apply the forecasting algorithm for each sequence

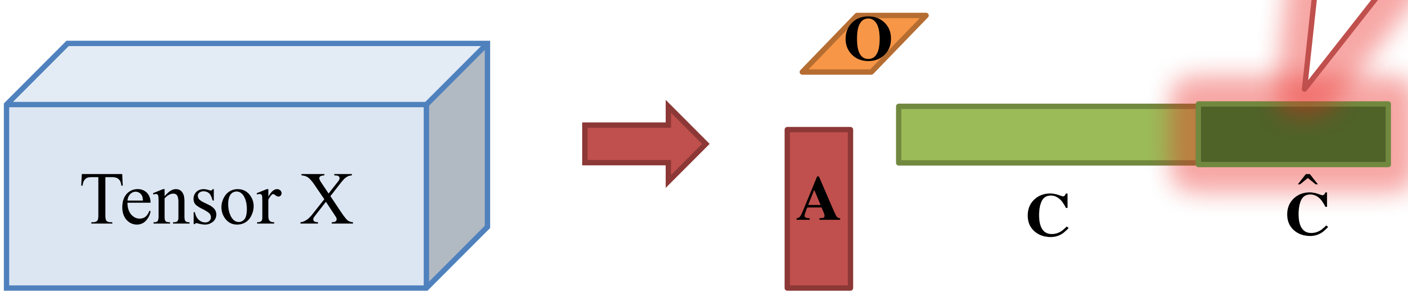


- ☹ **Scalability** : time complexity is at least $O(uvn)$
- ☹ **Accuracy** : each sequence "looks" like noise, (e.g., {0, 0, 0, 1, 0, 0, 2, 0, 0, ….}) -> hard to forecast

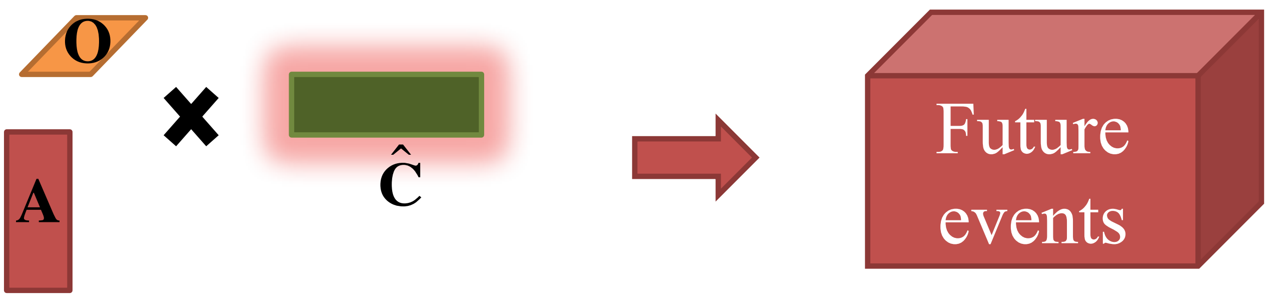
TriMine-F

Our approach:

– Step 1: Forecast time-topic matrix:



– Step 2: Generate events using 3 matrices



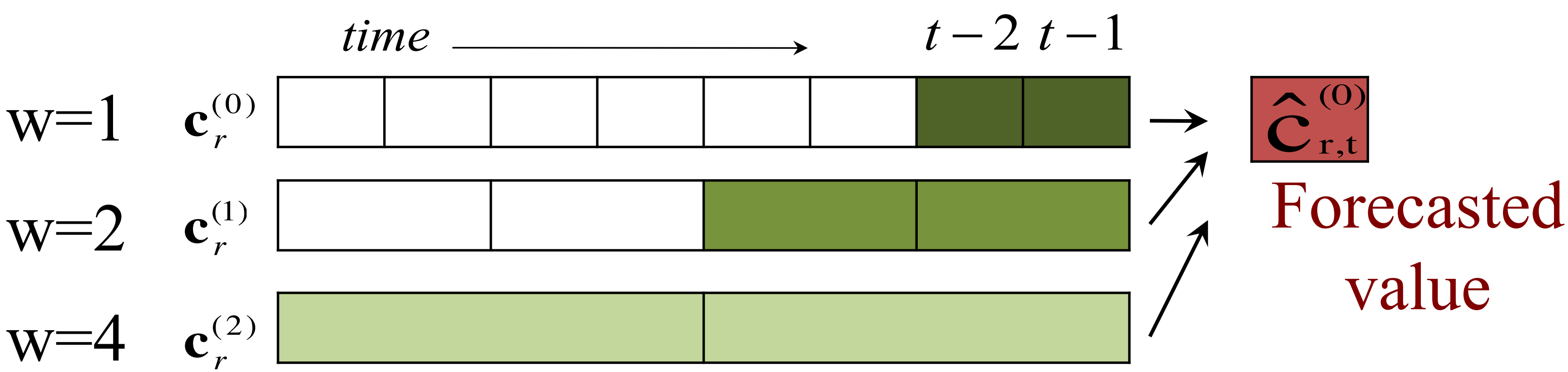
Forecast ‘time-topic matrix’ (details)

Q. How to capture multi-scale dynamics ?

e.g., bursty pattern, noise, multi-scale period

Multi-scale forecasting

Forecast $\hat{\mathbf{c}}^{(0)}_{r,t}$ using multiple levels of matrices

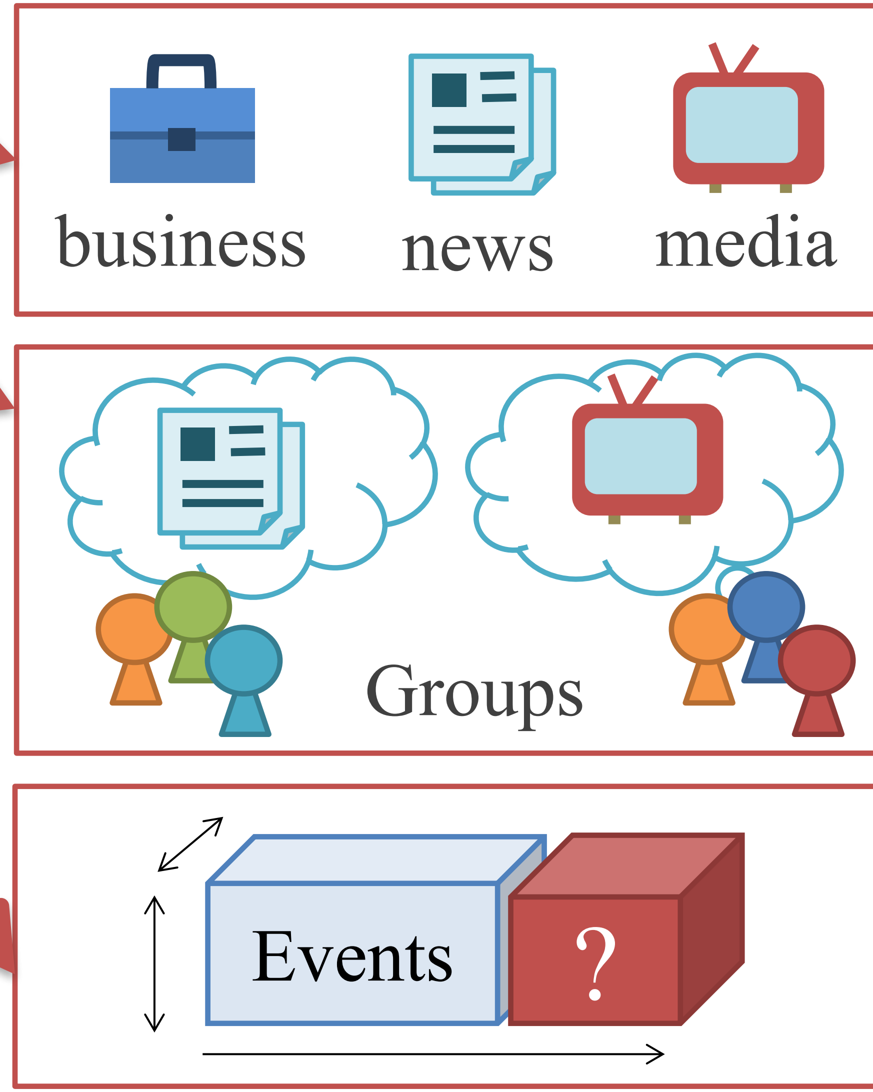


$$c^{(0)}_{r,t} = \sum_{h=0}^{\lceil \log n \rceil} \sum_{i=1}^{w} \lambda^{(h)}_{i,r} c^{(h)}_{r,t-i} + \epsilon_t .$$

(Details in paper)

Our goals

- Q1: Hidden topics
- Q2: Groups
- Q3: Forecasting

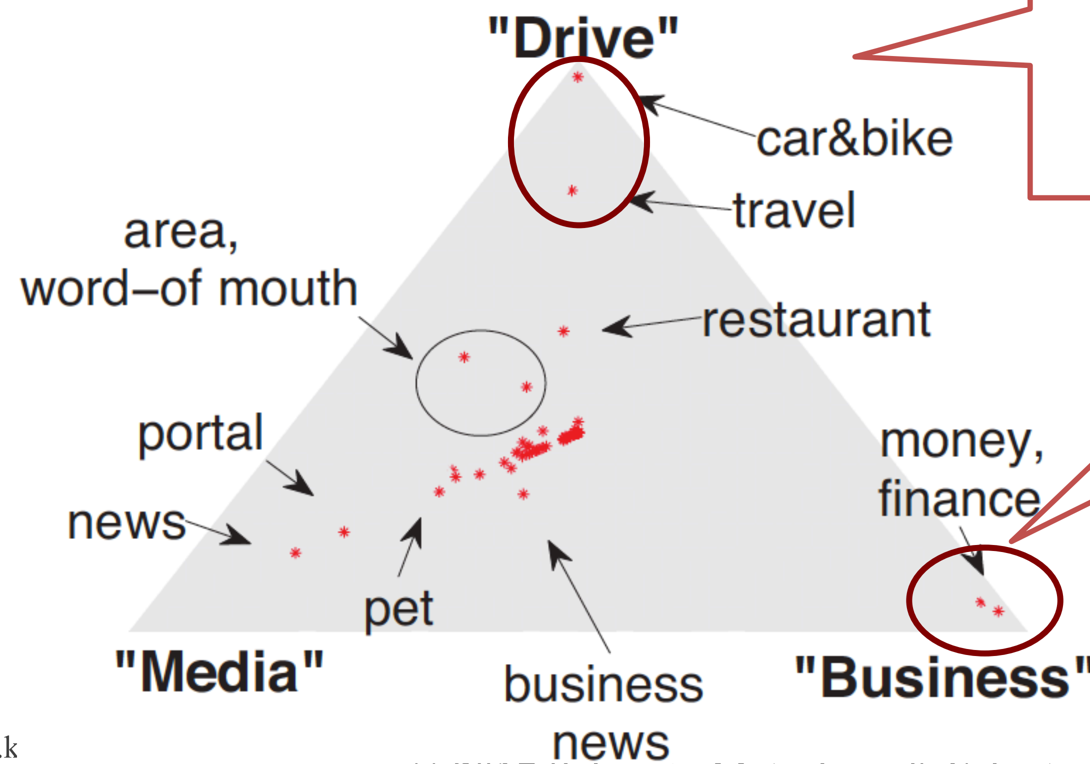


Q1&2. WebClick data

URL-topic matrix (O)

Three hidden topics: “drive”, “business”, “media”

* Red point : each web site



Car & bike site is related to travel site

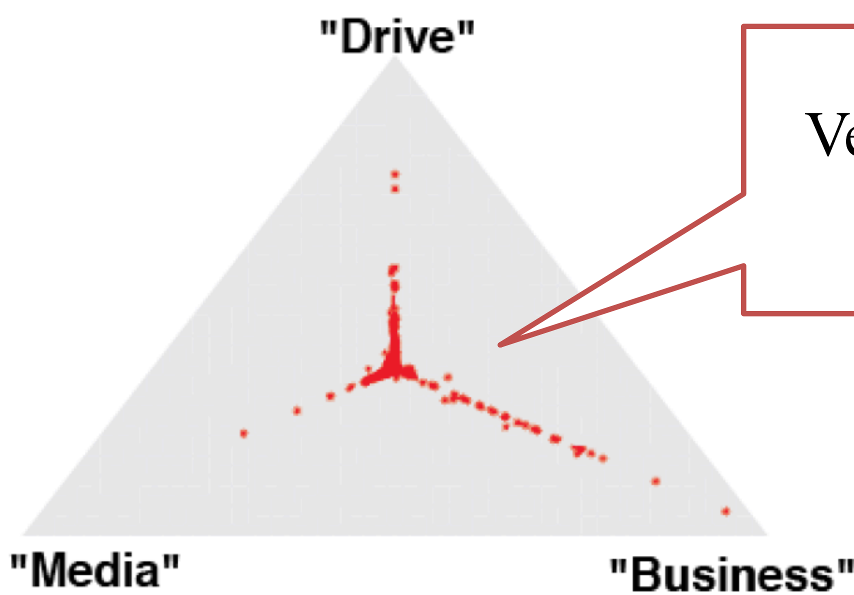
Money site& Finance site have similar trends

Q1&2. WebClick data

User-topic matrix (A)

Three hidden topics: “drive”, “business”, “media”

* Red point : each user



Very clear user groups along the spokes

Q1&2. WebClick data

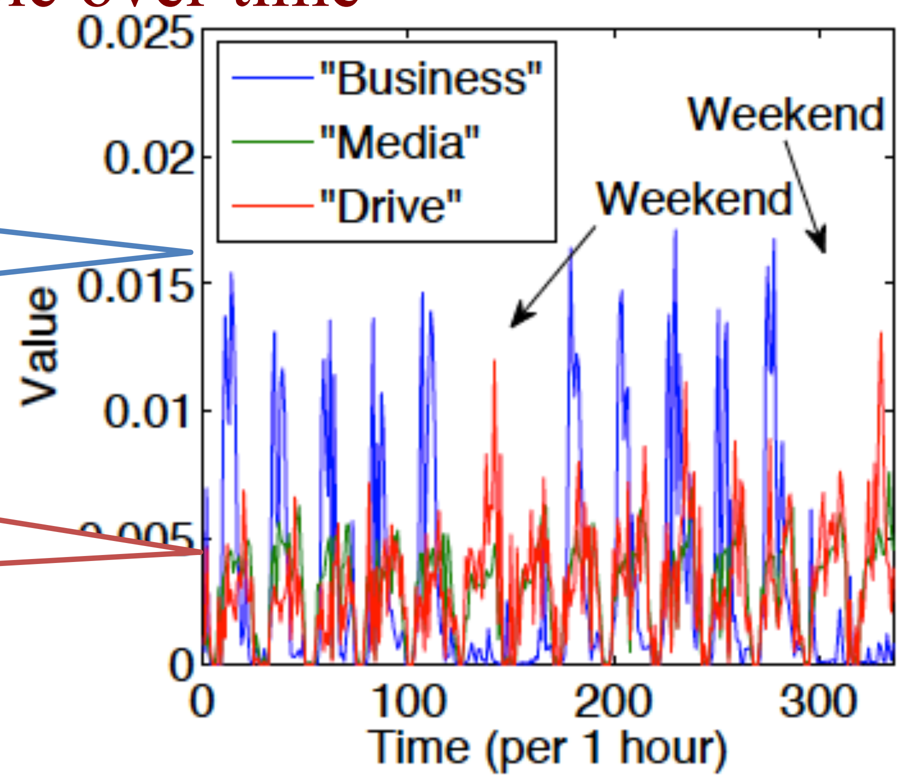
Time-topic matrix (C)

Three hidden topics: “drive”, “business”, “media”

* Each sequence: each topic over time

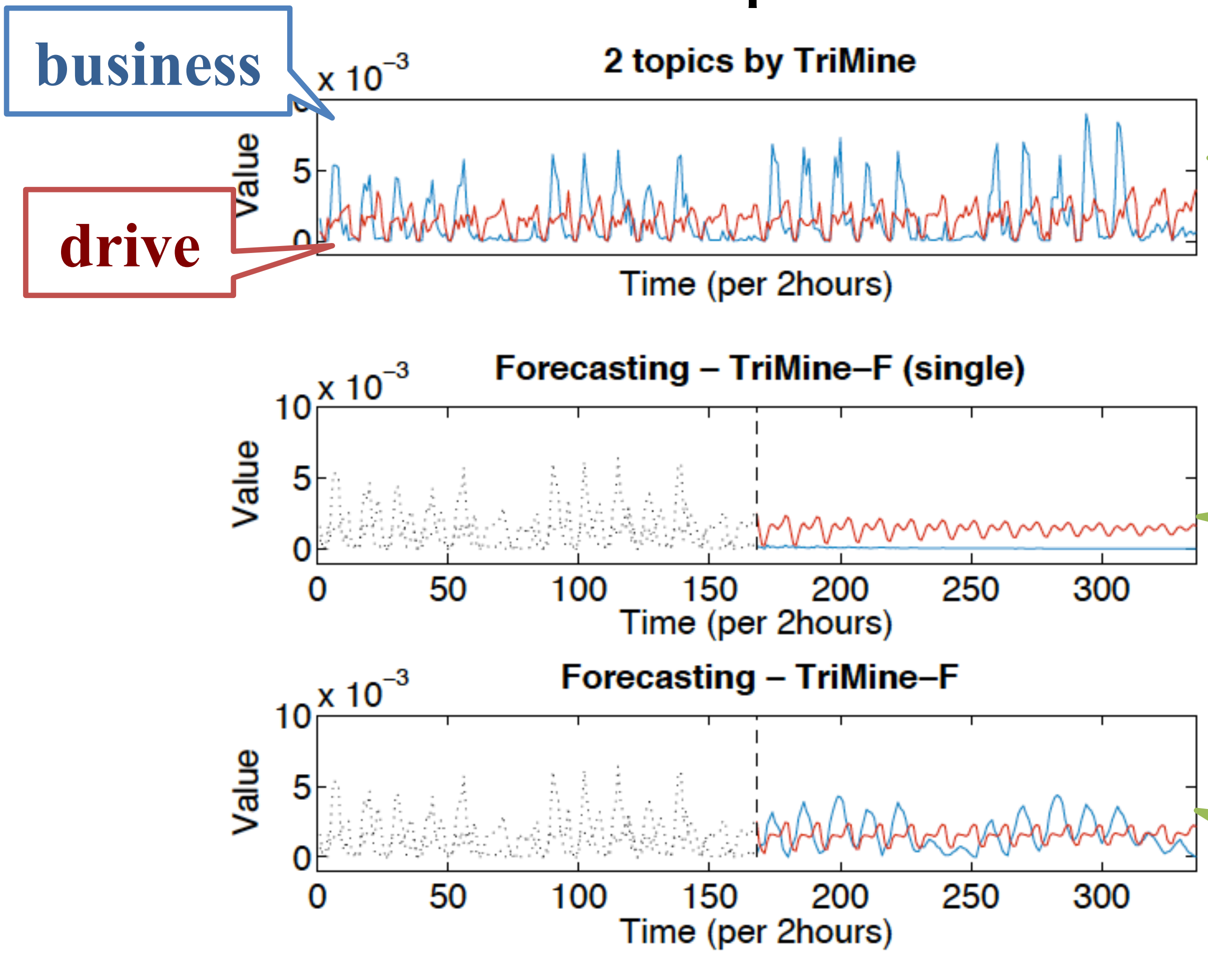
“Business” topic: Less access during weekend

“Drive” topic: Spikes during weekend



Q3. Forecasting accuracy

- Benefit of multiple time-scale forecasting



Original sequence of matrix (C)

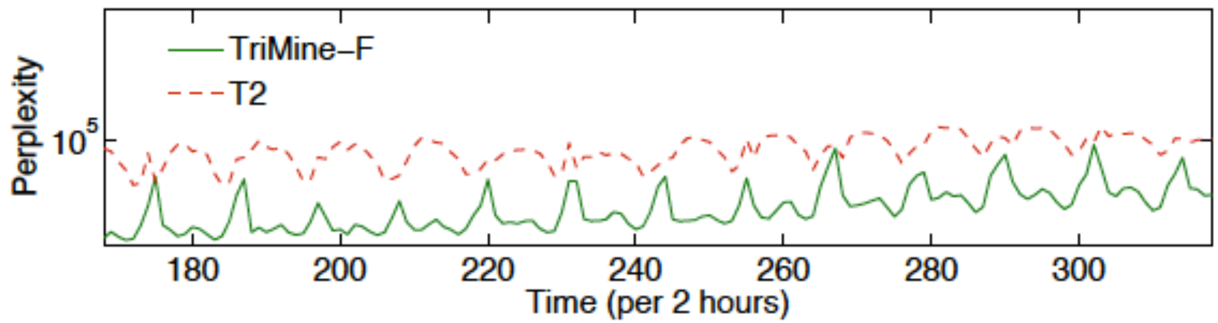
Forecast C' using single level -> failed

Multi-scale forecast -> captured cyclic patterns

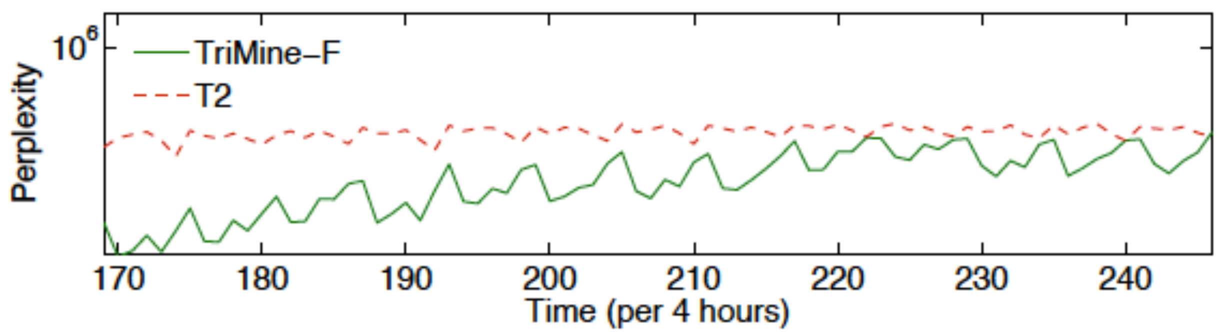
Q3. Forecasting accuracy

Temporal perplexity (entropy for each time-tick)

Lower perplexity: higher predictive accuracy



(a) *WebClick*



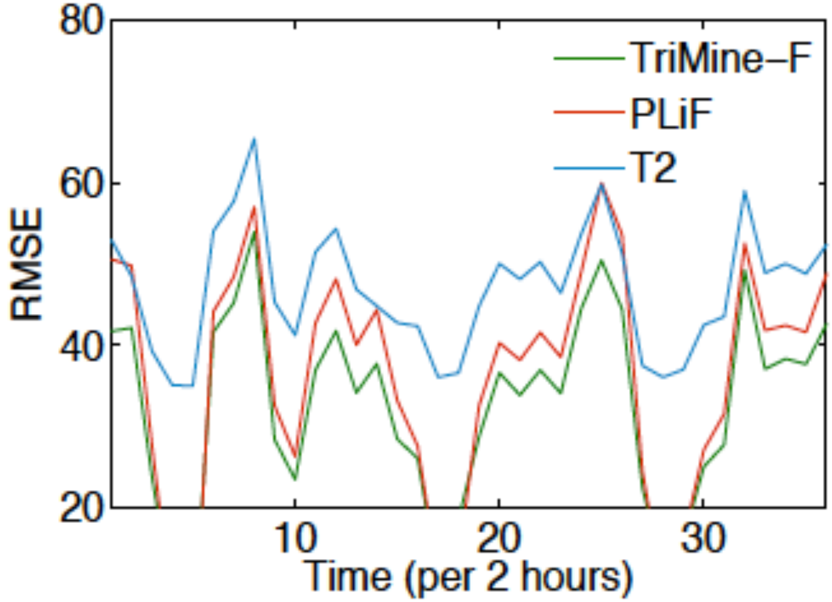
(b) *Ondemand TV*

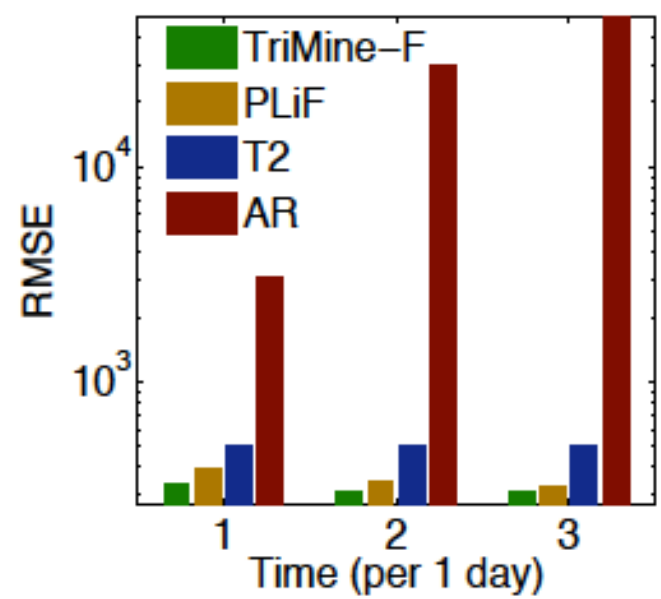
T2: [Hong et al. KDD'11]

Q3. Forecasting accuracy

Accuracy of event forecasting

RMSE between original and forecasted events
(lower is better)

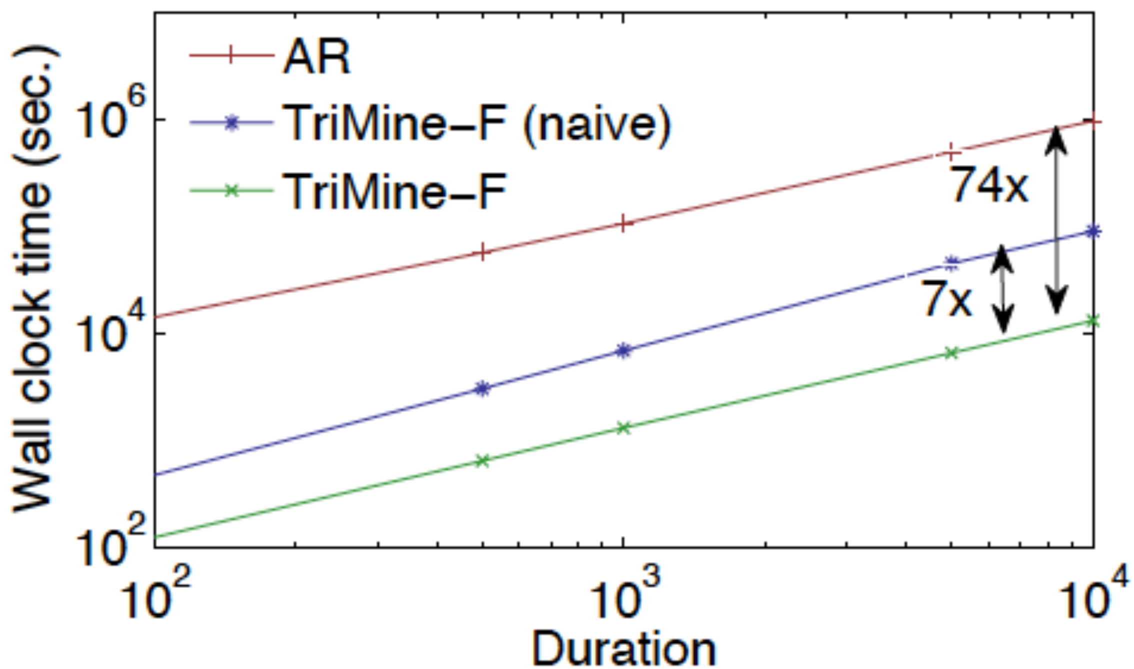




PLiF [Li et al.VLDB’10] , T2: [Hong et al.KDD’11]

Q3. Scalability

- Computation cost (vs. AR)

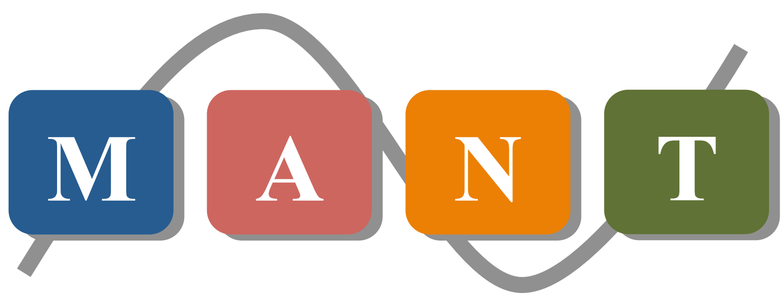


- **TriMine** provides a reduction in computation time (up to 74x)

Outline

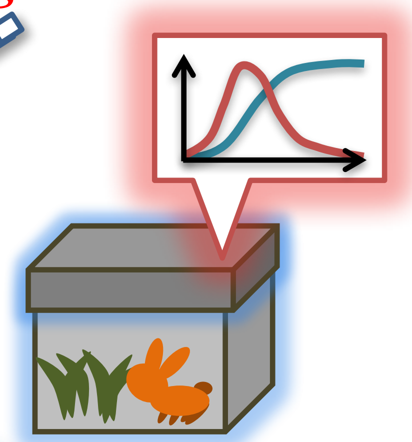
- Tensor decomposition
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series



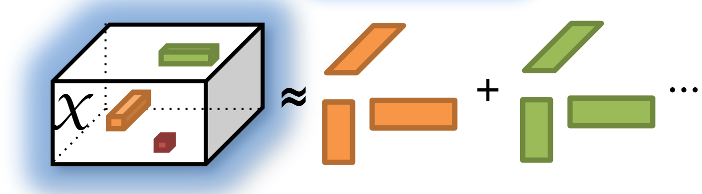
Non-linear tensor analysis

NO magic !



New research directions

1. Automatic mining
2. Non-linear (gray-box) modeling
3. Large-scale tensor analysis



Put all together

New challenge: MANT analysis

Multi-Aspect Non-linear Time-series




[Matsubara+ KDD'14]

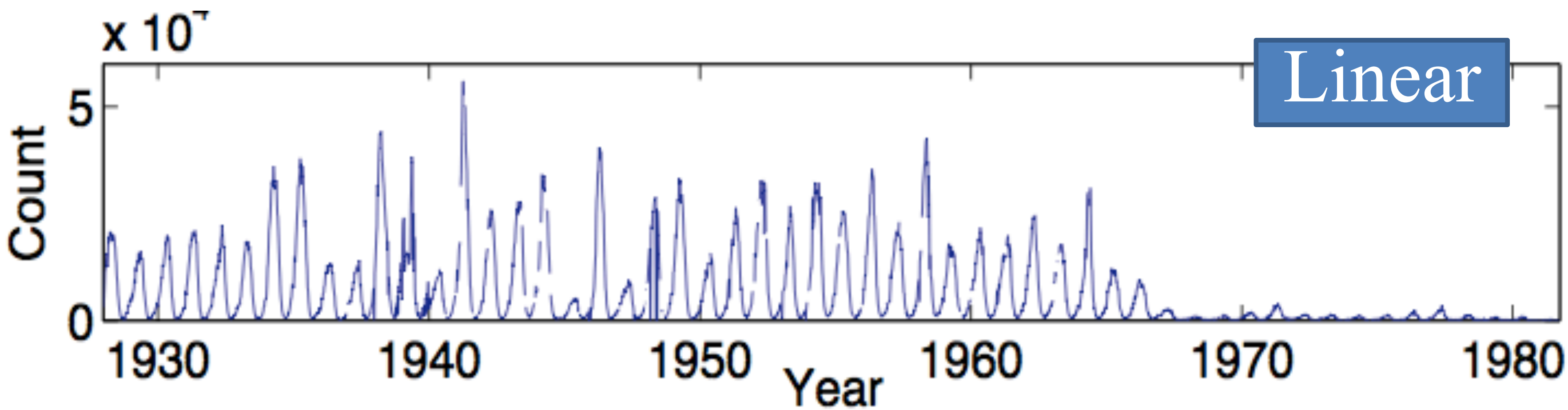
FUNNEL: Automatic Mining of Spatially Coevolving Epidemics

Yasuko Matsubara, Yasushi Sakurai (Kumamoto University)
Willem G. van Panhuis (University of Pittsburgh)
Christos Faloutsos (CMU)

Motivation

Given: Large set of epidemiological data

e.g., Measles cases in the U.S.

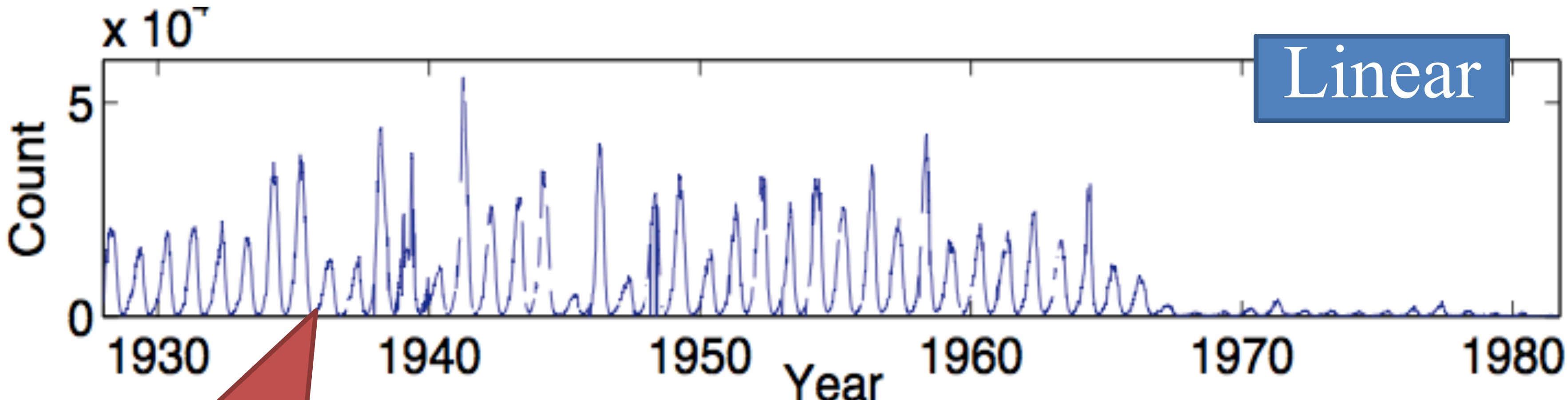


(Weekly)

Motivation

Given: Large set of epidemiological data

e.g., Measles cases in the U.S.



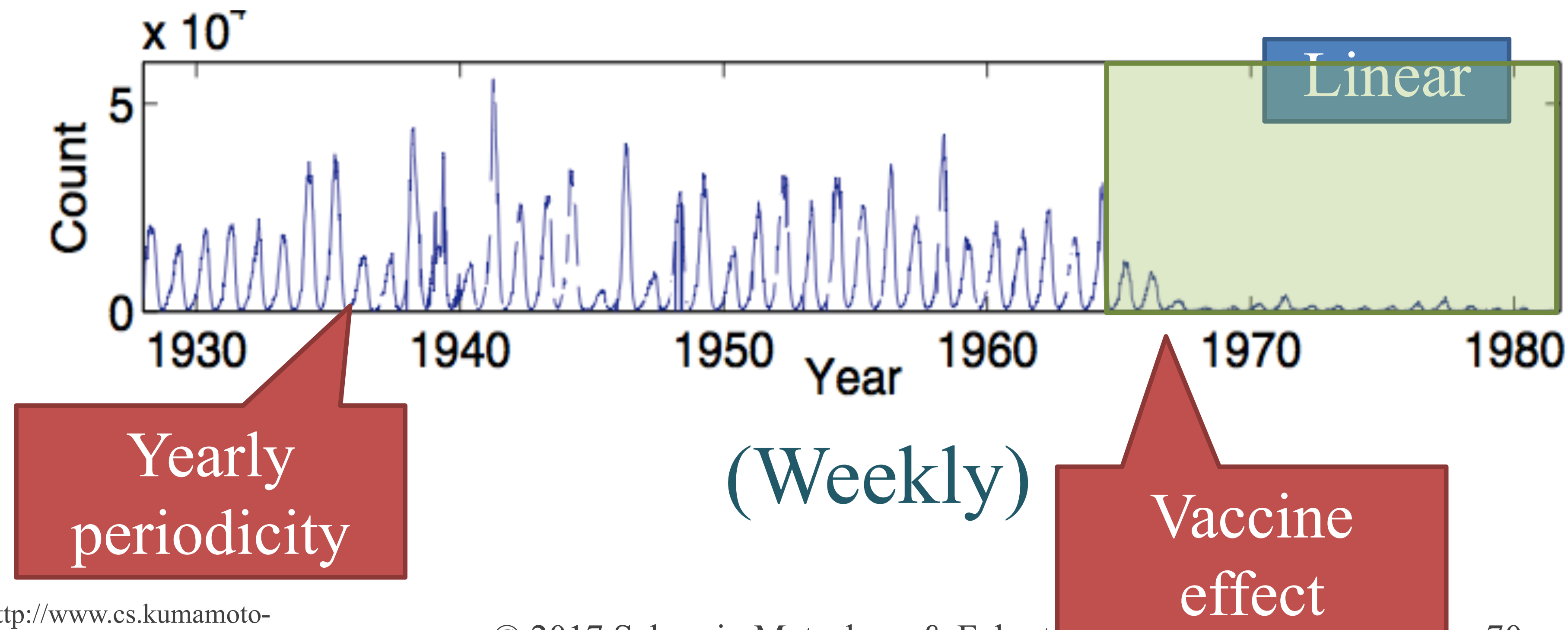
Yearly periodicity

(Weekly)

Motivation

Given: Large set of epidemiological data

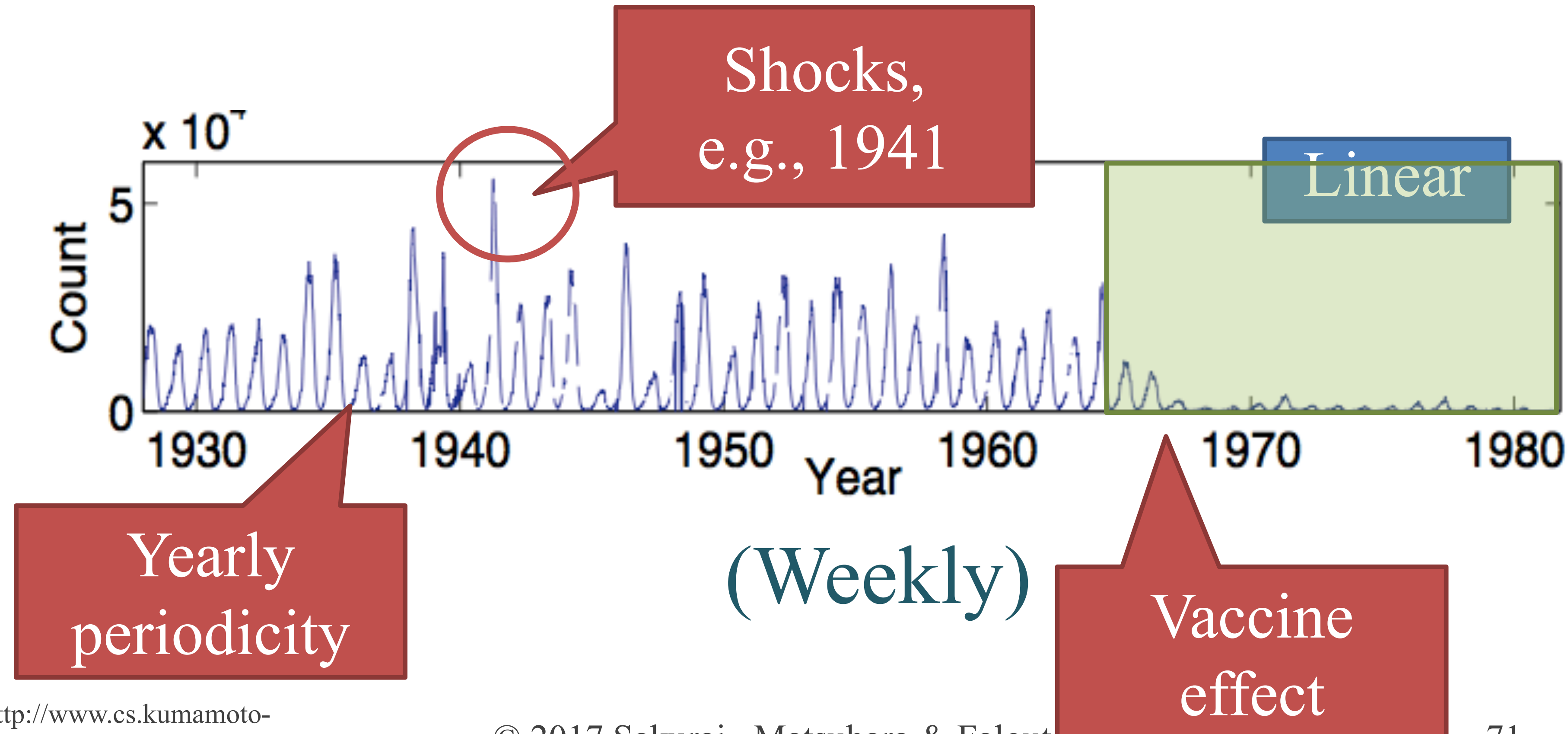
e.g., Measles cases in the U.S.



Motivation

Given: Large set of epidemiological data

e.g., Measles cases in the U.S.



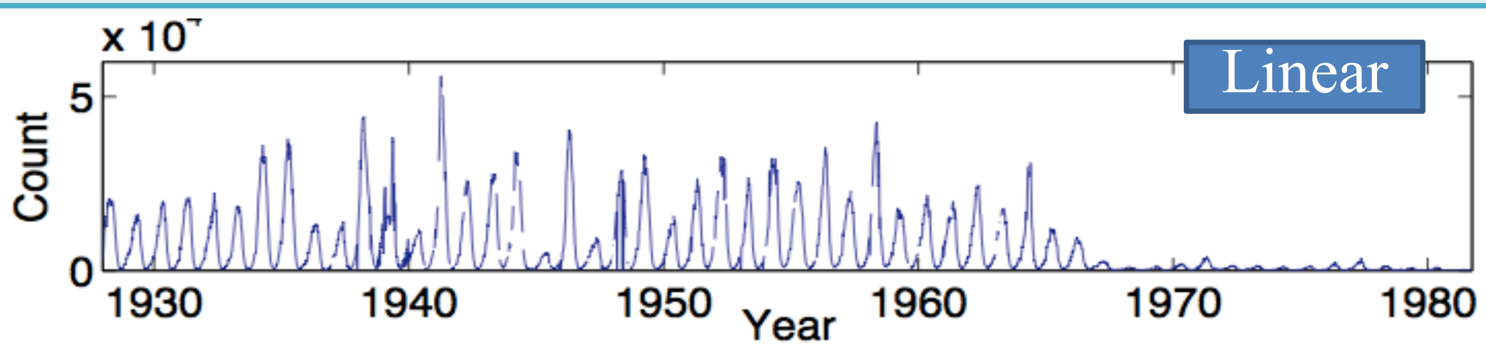
(Weekly)

Motivation

Given: Large set of epidemiological data

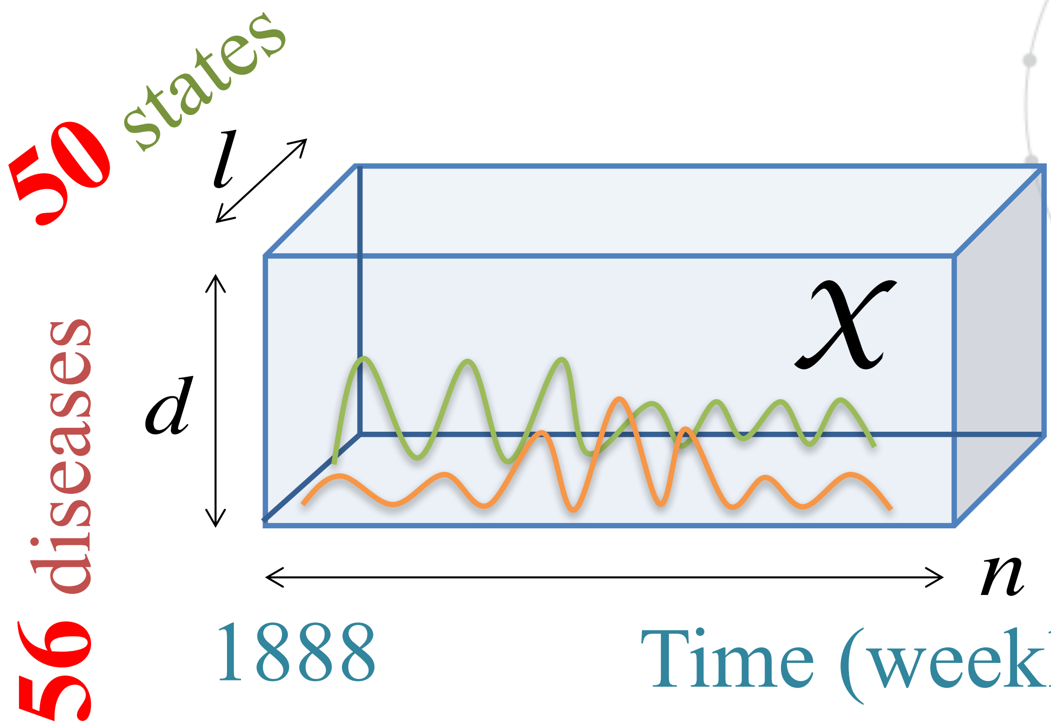
e.g., Measles cases in the U.S.

Goal: summarize all the epidemic time-series, **"fully-automatically"**



Data description

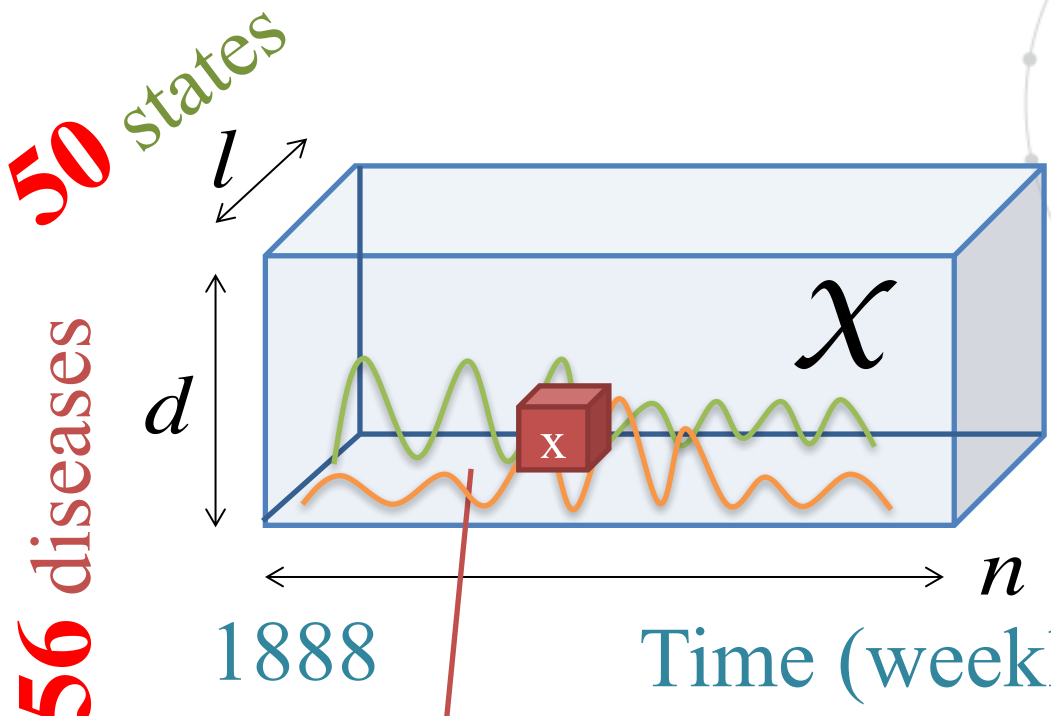
Project Tycho: infectious diseases in the U.S.



Time (weekly) (> **125** years)

Data description

Project Tycho: infectious diseases in the U.S.



Time (weekly) (> **125** years)

Element x : # of cases

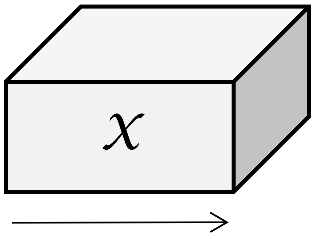
e.g., ‘measles’, ‘NY’, ‘April 1-7, 1931’, ‘4000’

http
u.ac

Problem definition

Given:

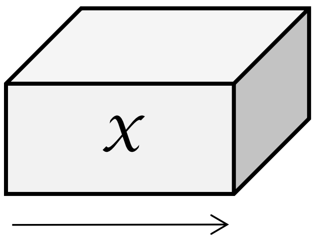
Tensor $\mathcal{X}$ (disease x state x time)



Problem definition

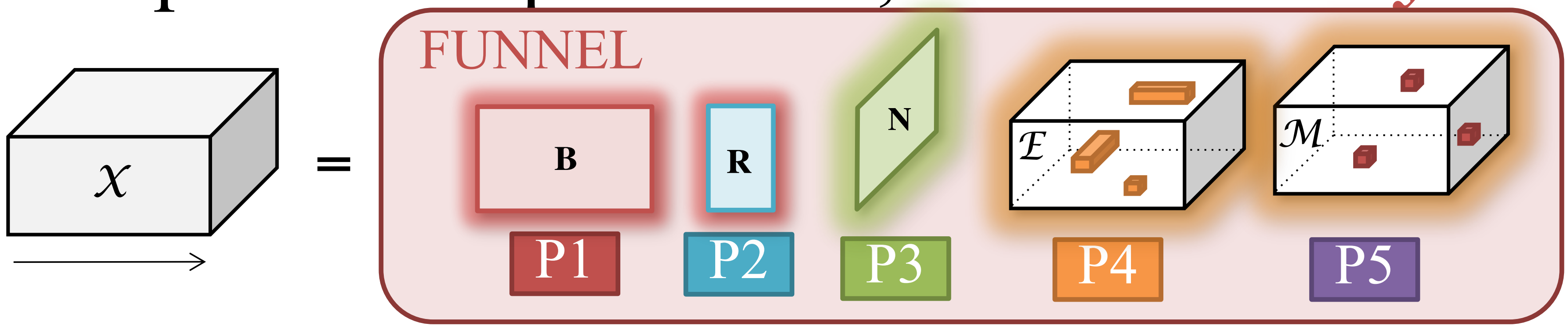
Given:

Tensor $\mathcal{X}$ (disease x state x time)



Find:

Compact description of $\mathcal{X}$, ***"automatically"***

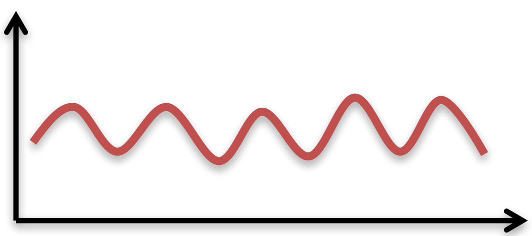


Problem definition

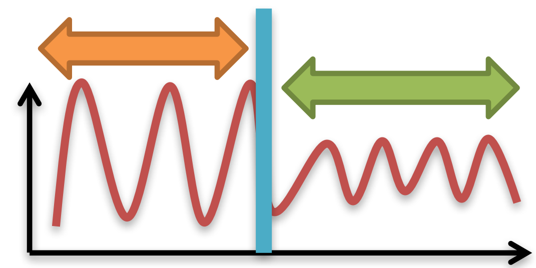
G

Te state

Seasonality

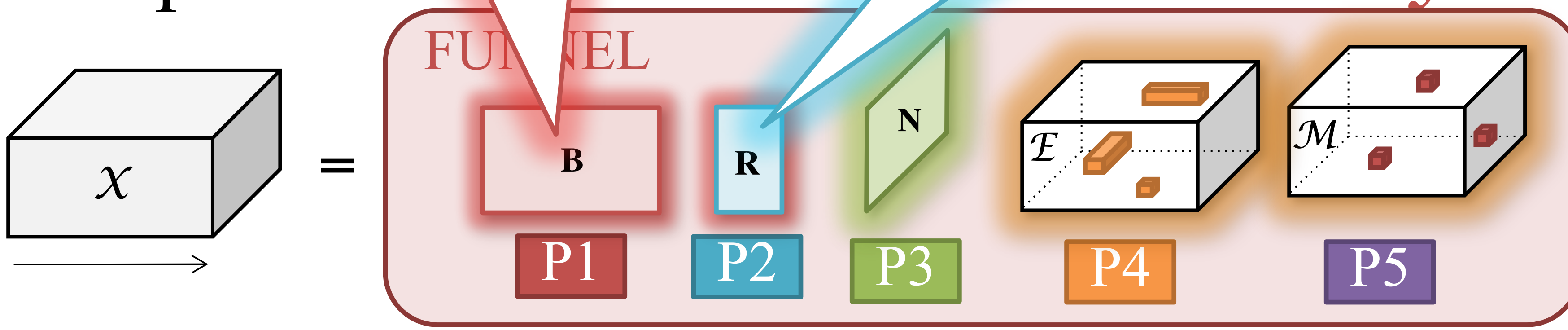


Discontinuities



Find:

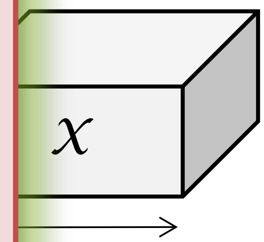
Compact desc tion of $\mathcal{X}$ *automatically"*



Problem definition

Given:

Tensor

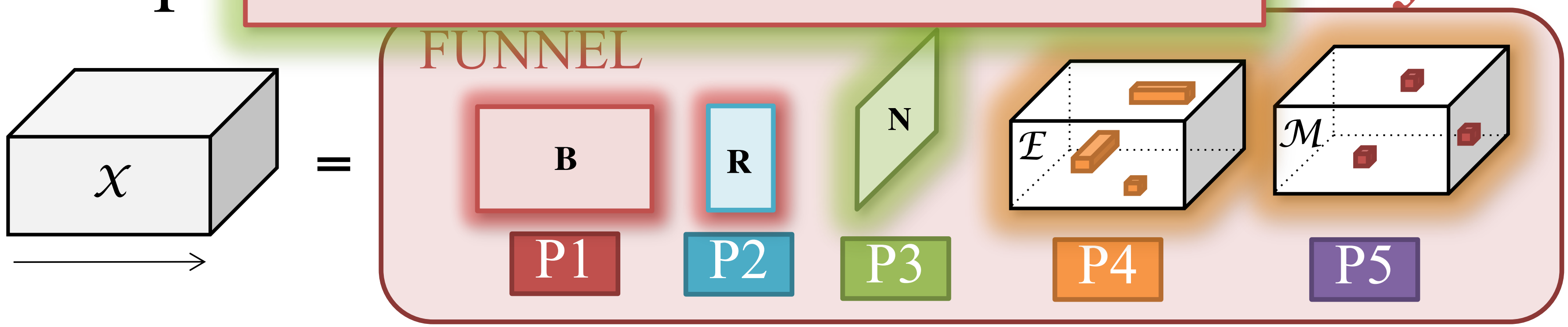


Find:

Compa ...ically”

NO magic numbers !

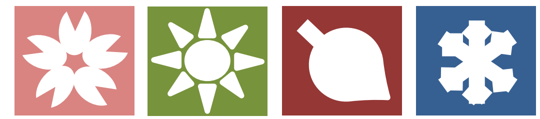
Parameter-free!



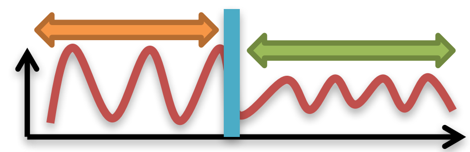
Modeling power of FUNNEL

Our model can capture 5 properties

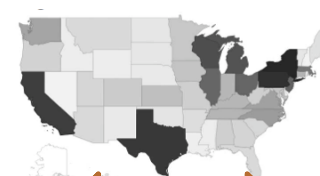
P1 Seasonality



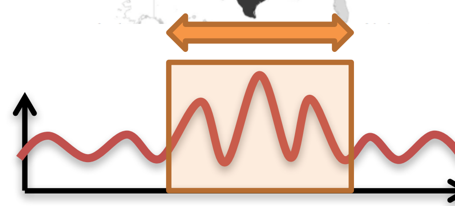
P2 Disease reductions



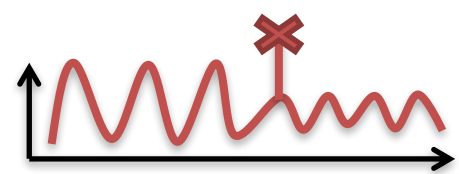
P3 Area sensitivity



P4 External events

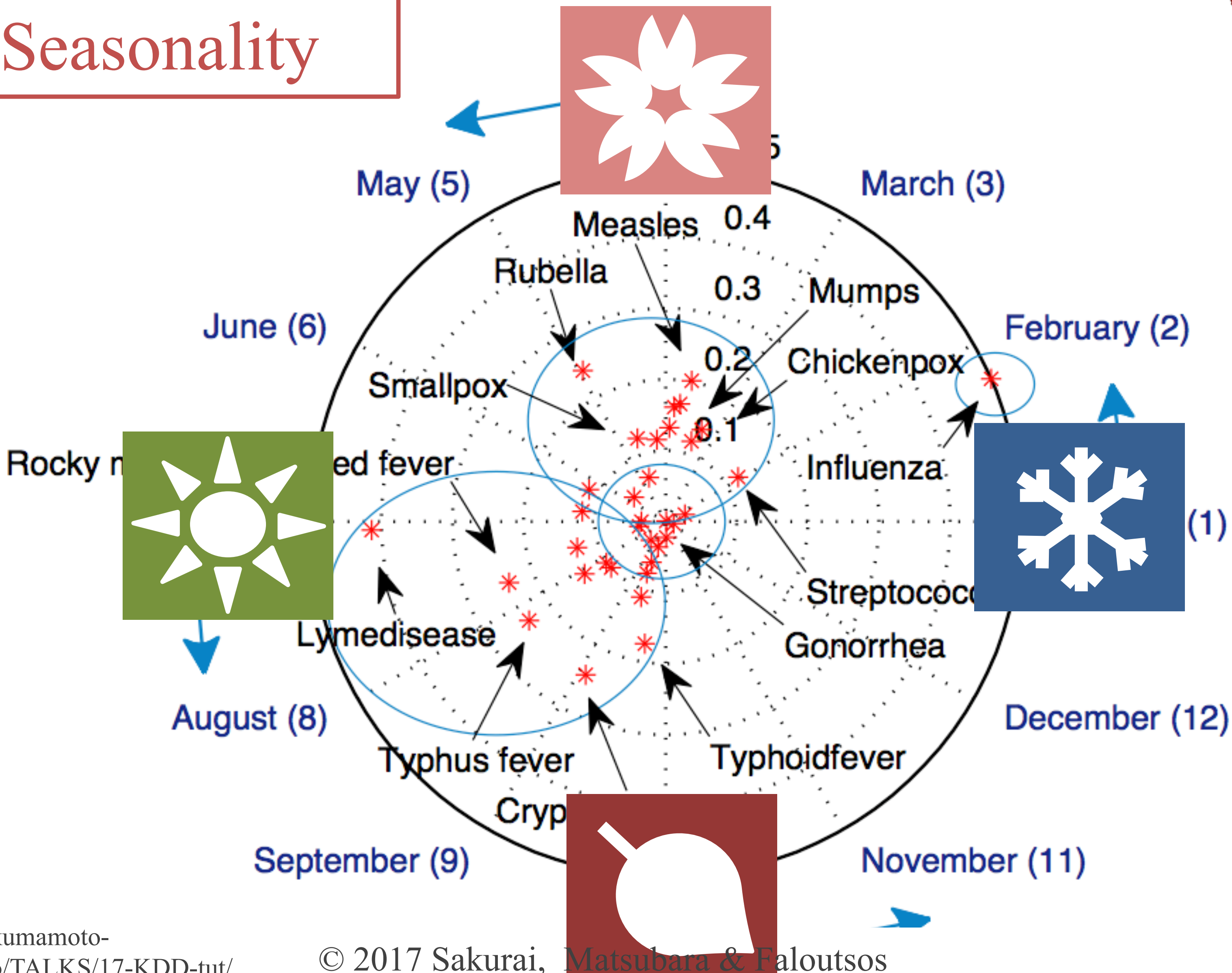


P5 Mistakes



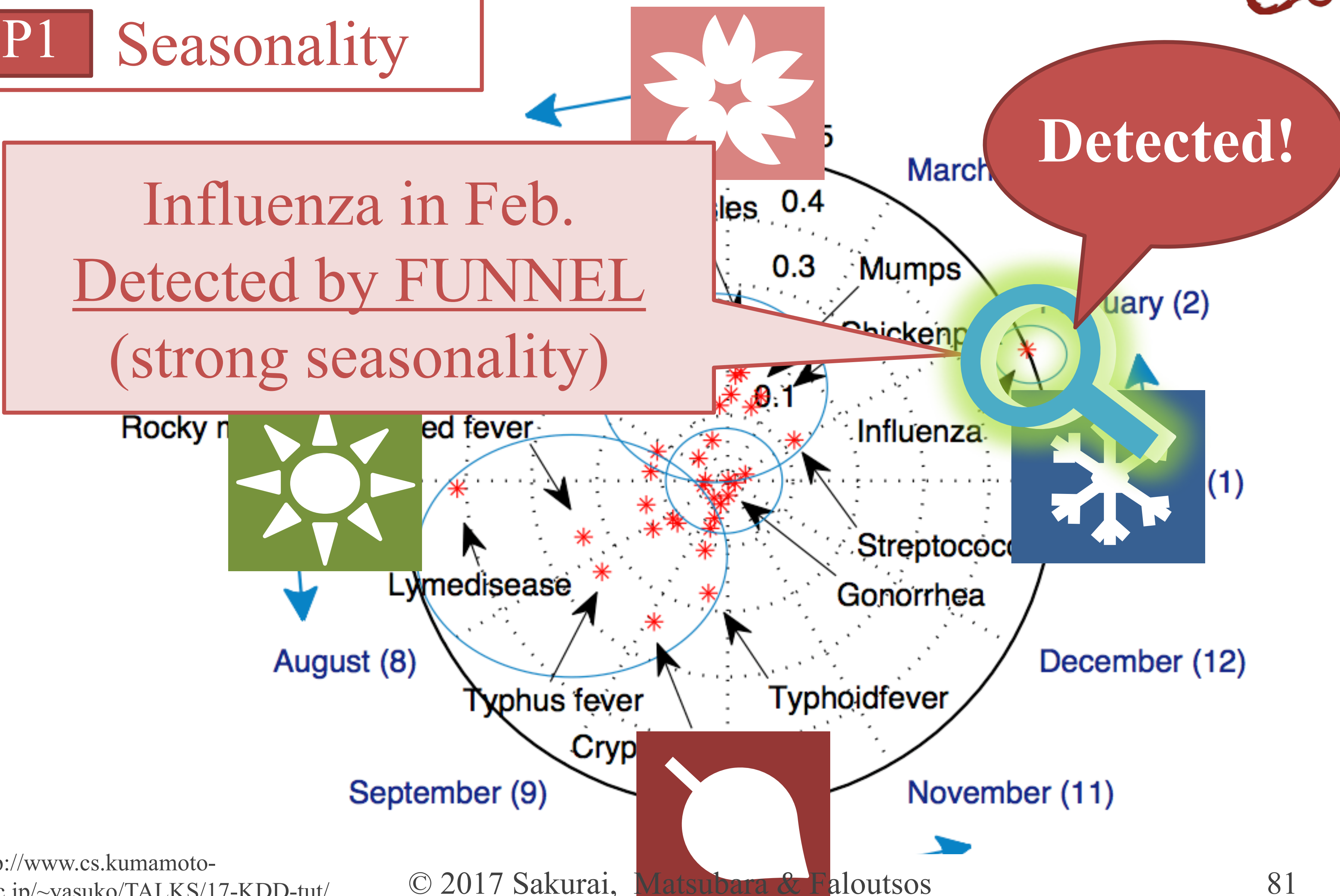
Modeling power of FUNNEL

P1 Seasonality



Modeling power of FUNNEL

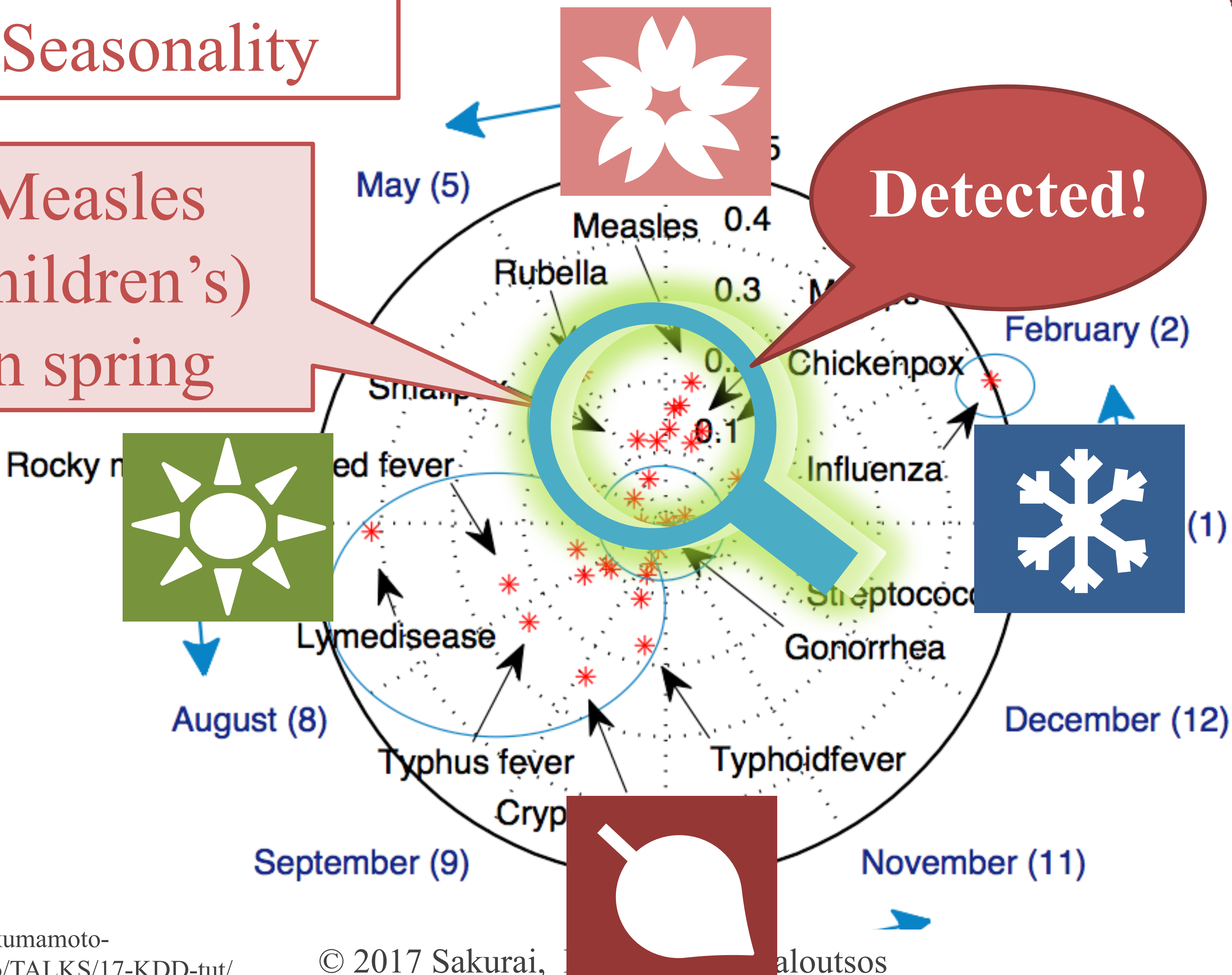
P1 Seasonality



Modeling power of FUNNEL

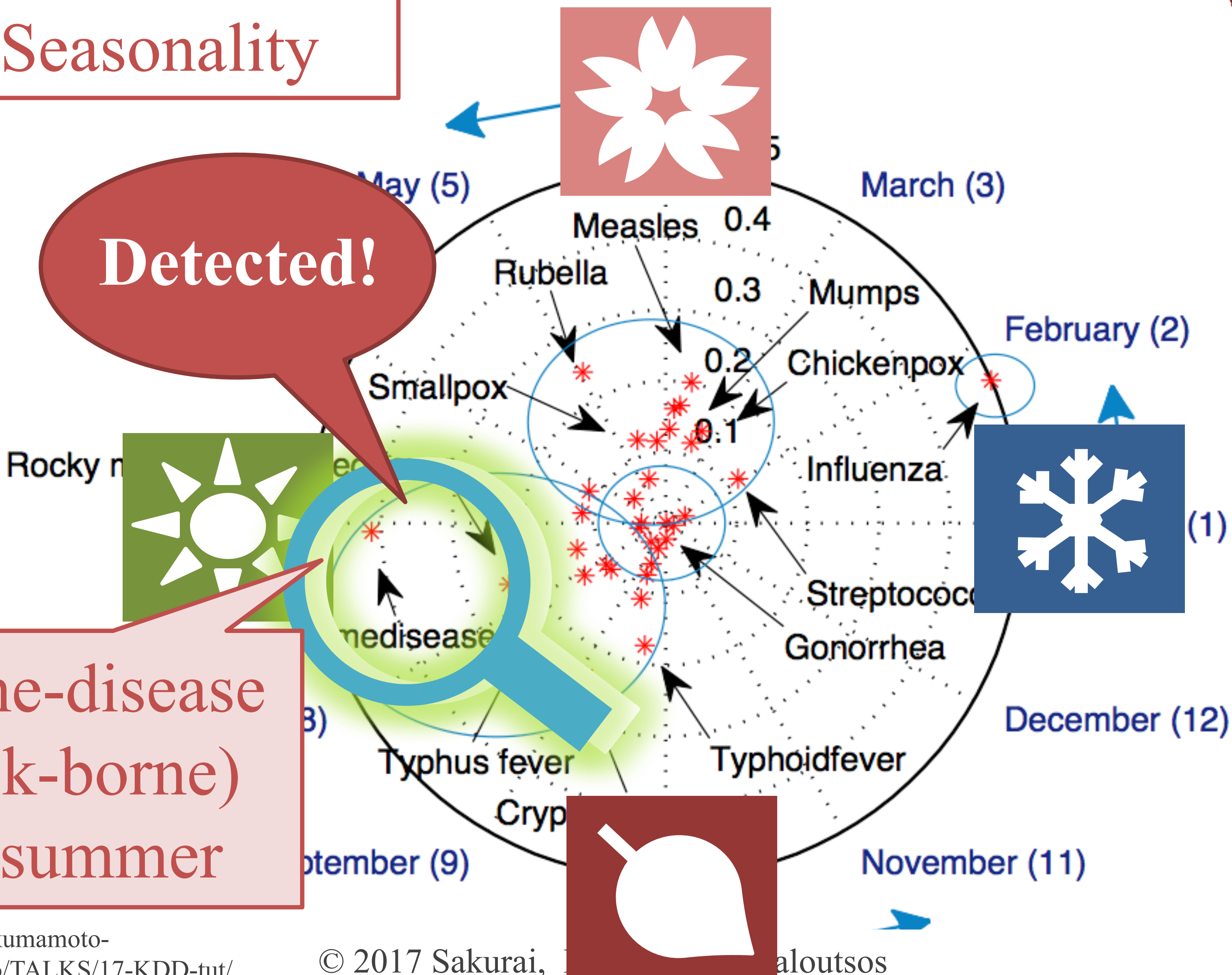
P1 Seasonality

Measles (children's) in spring



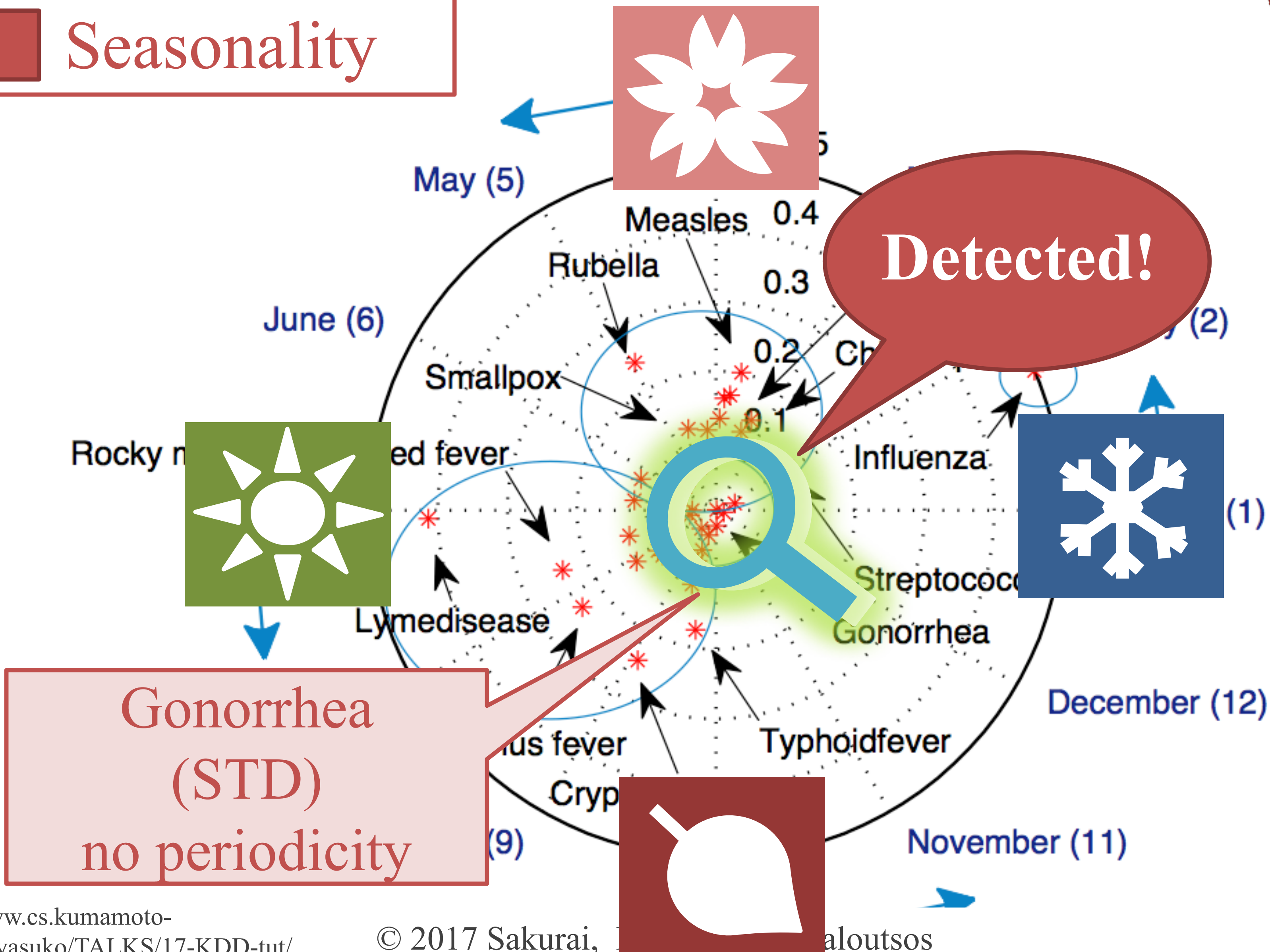
Modeling power of FUNNEL

P1 Seasonality



Modeling power of FUNNEL

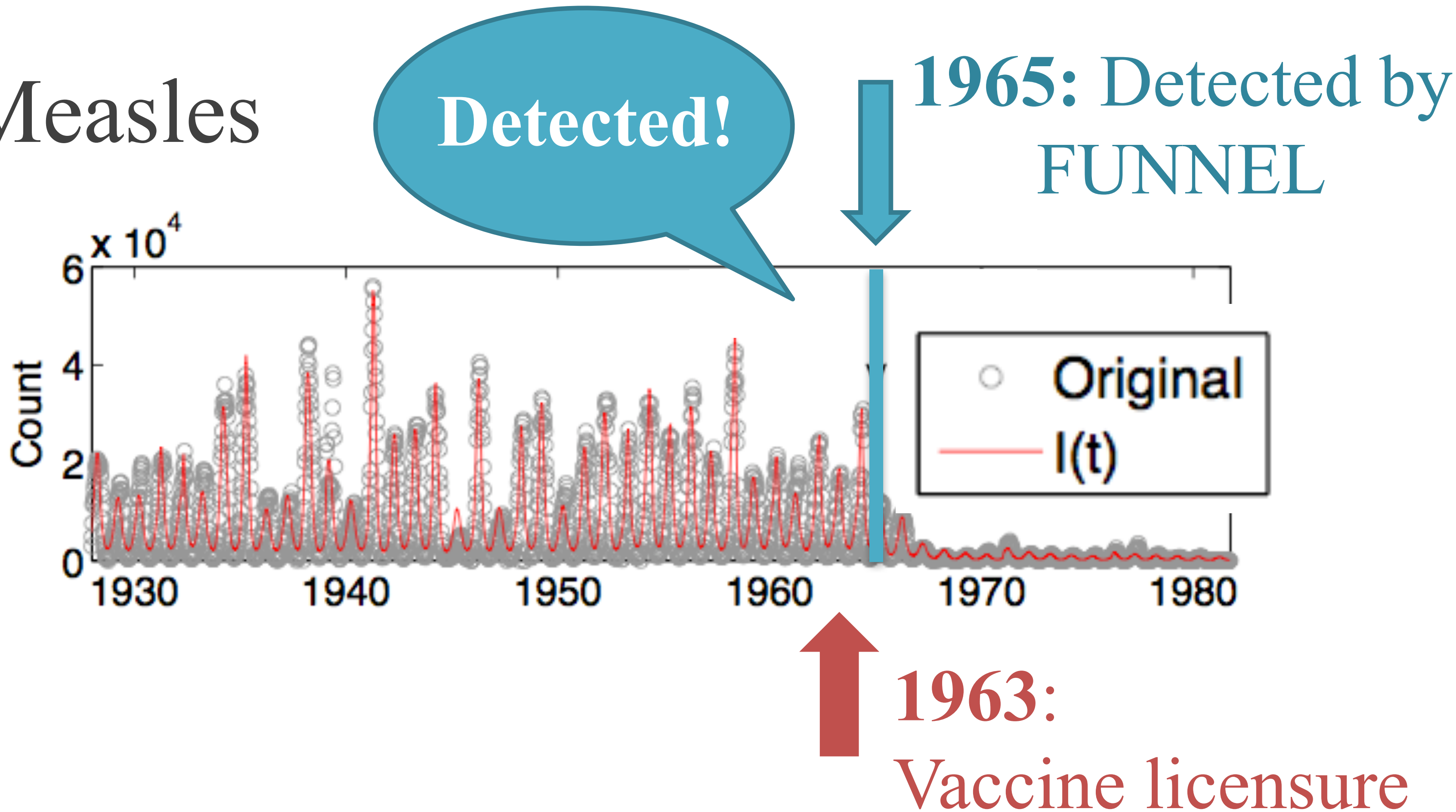
P1 Seasonality



Modeling power of FUNNEL

P2 Disease reduction effect (discontinuities)

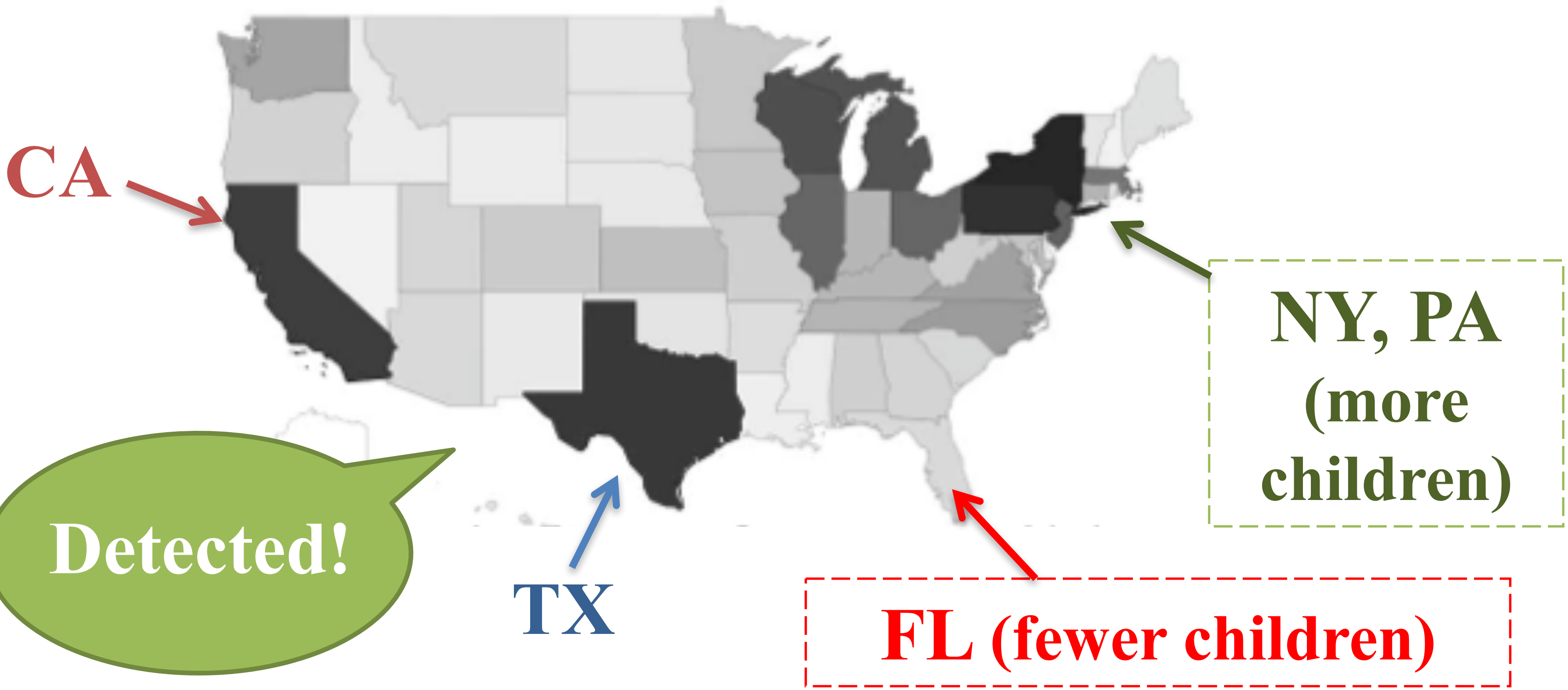
Measles



Modeling power of FUNNEL

P3 Area sensitivity

FUNNEL's guess of susceptibles (measles)



Detected!

Modeling power of FUNNEL

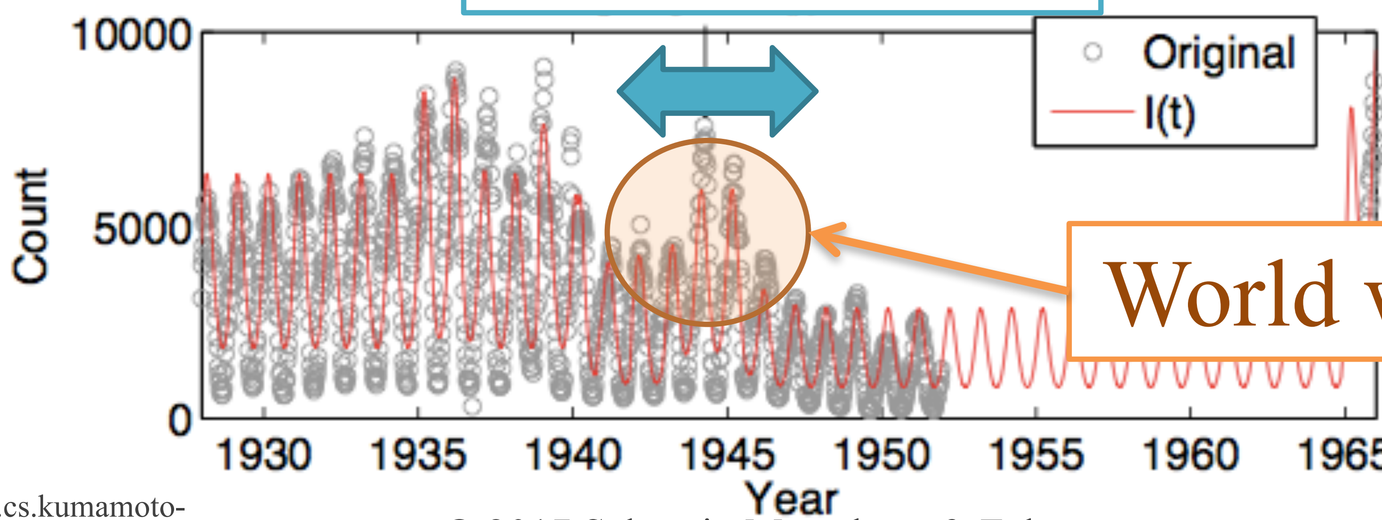
P4 External shock events

Funnel can detect external shocks "**fully-automatically**" !

Scarlet fever

Detected by FUNNEL

Detected!



World war II

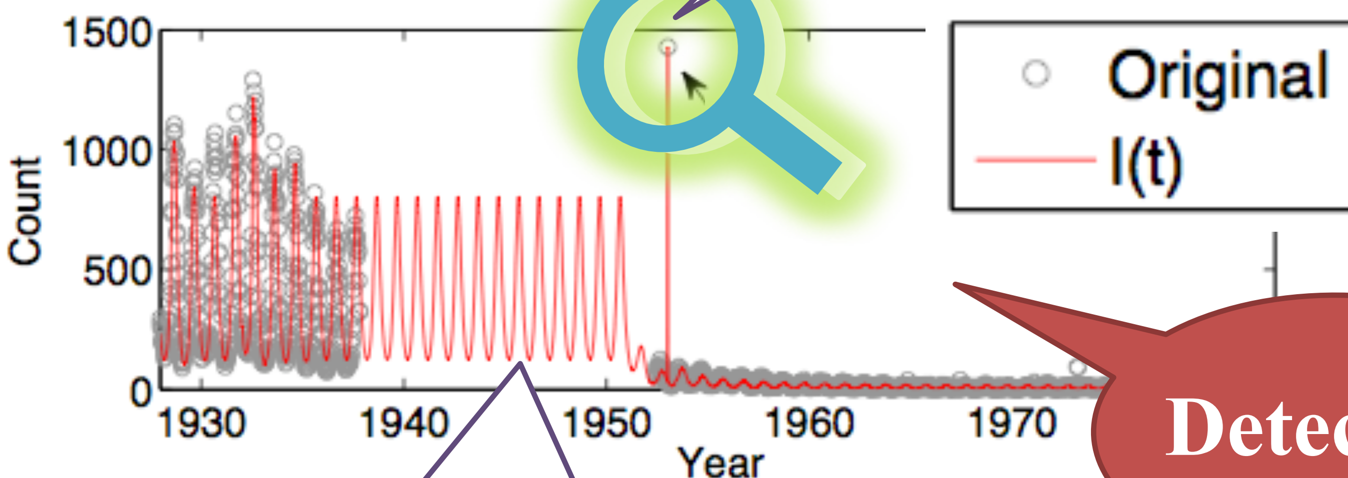
Modeling power of FUNNEL

P5 Mistakes, incorrect values

It can also detect typos, “**automatically**” !!

Typhoid fever cases

Mistake

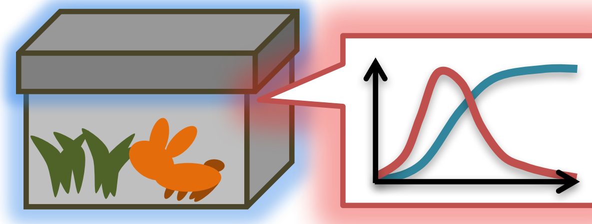


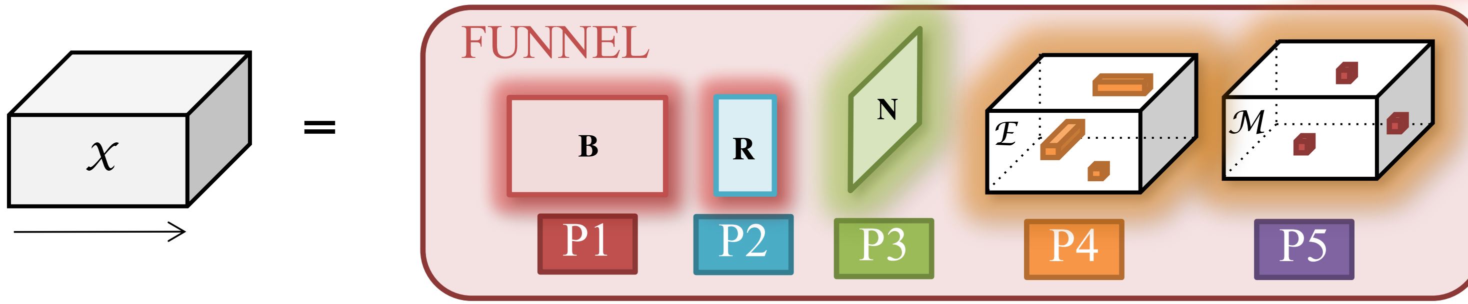
Detected!

Missing values

Two main ideas

Idea #1: Grey-box model



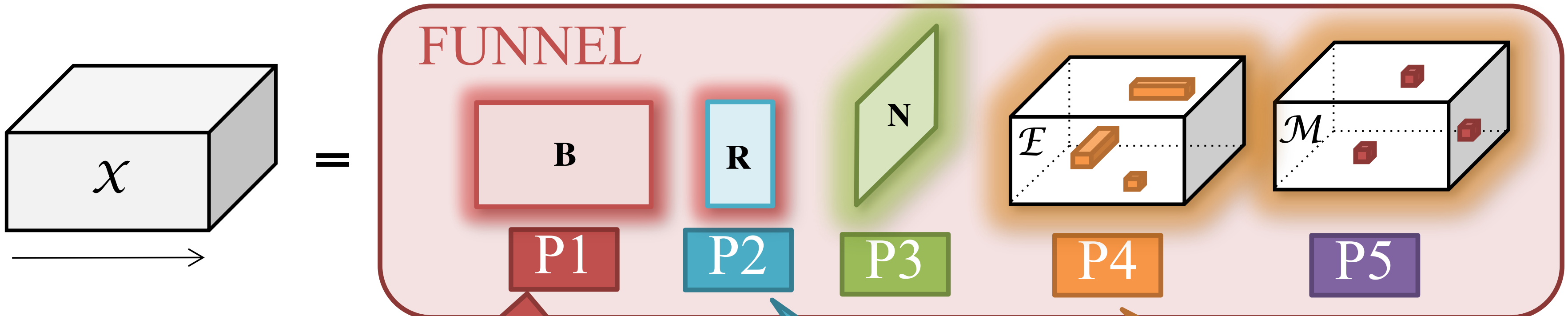


Idea #2: MDL for tensor analysis

NO magic numbers ! (parameter-free)

Two main ideas

Idea #1: Grey-box model - domain knowledge

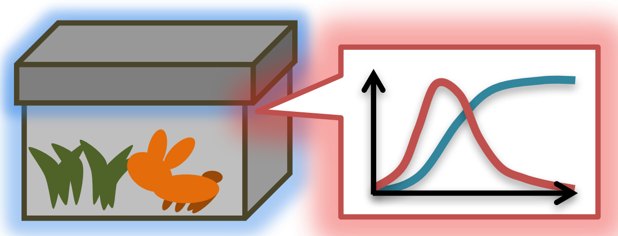


(SIRS+) : 6 parameters

$$\boxed{S(t+1)} = S(t) - \beta(t)\epsilon(t)S(t)I(t)$$
$$\boxed{I(t+1)} = I(t) + \beta(t)\epsilon(t)S(t)I(t)$$
$$\boxed{V(t+1)} = V(t) + \delta I(t) - \gamma V(t) +$$

Vaccine

Shocks

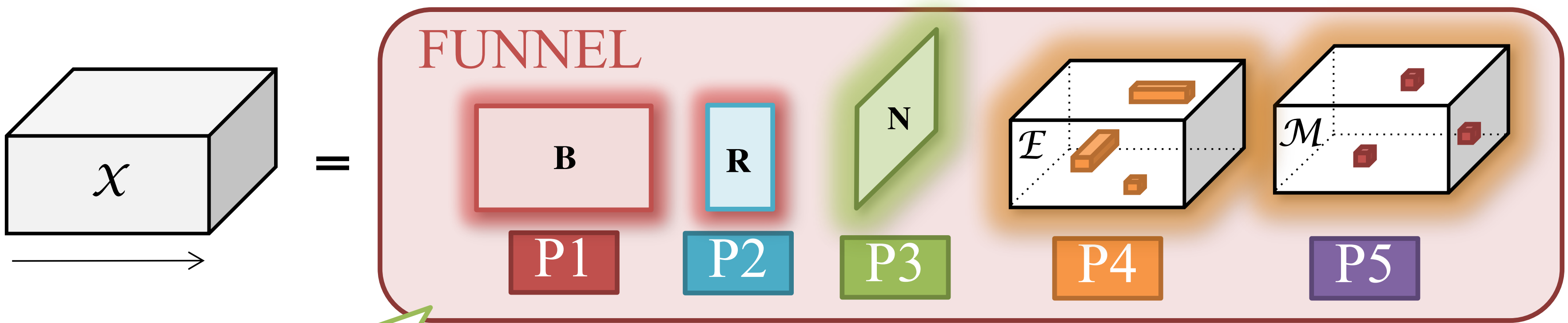


Kumamoto U

CMU CS

Two main ideas

Idea #2: Fitting with MDL -> automatic!



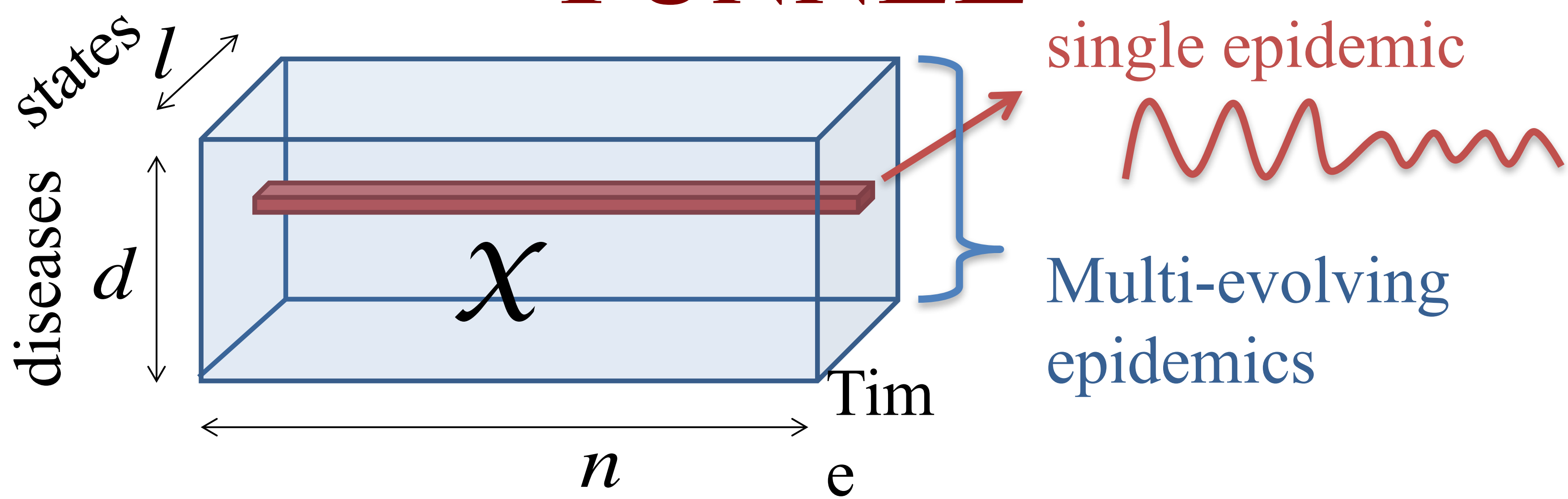
$$Cost_T(\mathcal{X};\mathcal{F}) = \log^*(d) + \log^*(l) + \log^*(n)$$
$$+Cost_M(\mathbf{B}) + Cost_M(\mathbf{R}) + Cost_M(\mathbf{N})$$
$$+Cost_M(\mathcal{E}) + Cost_M(\mathcal{M}) + Cost_C(\mathcal{X}|\mathcal{F})$$

Cost function

NO magic numbers

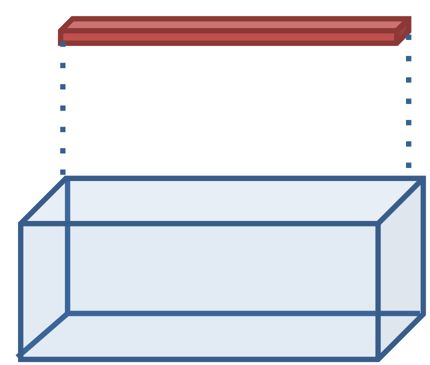
Parameter-free!

Proposed model: FUNNEL

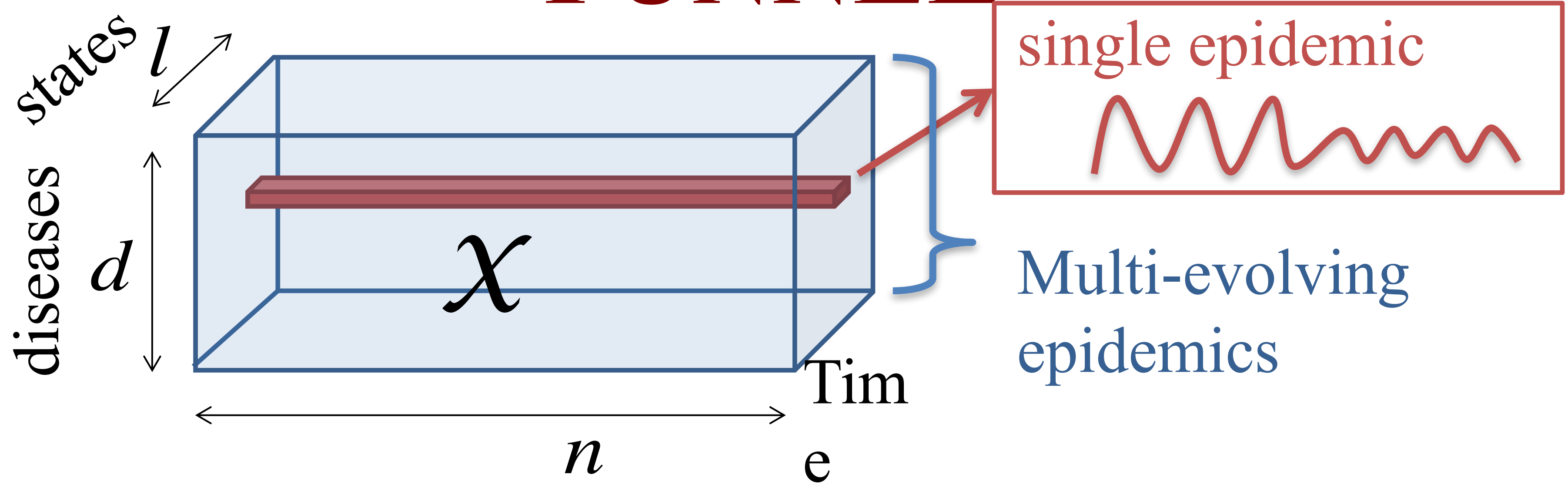


(a) FUNNEL-single

(b) FUNNEL-full (tensor)

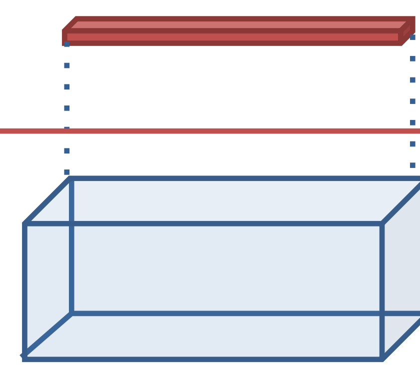


Proposed model: FUNNEL



(a) FUNNEL-single

(b) FUNNEL-full (tensor)

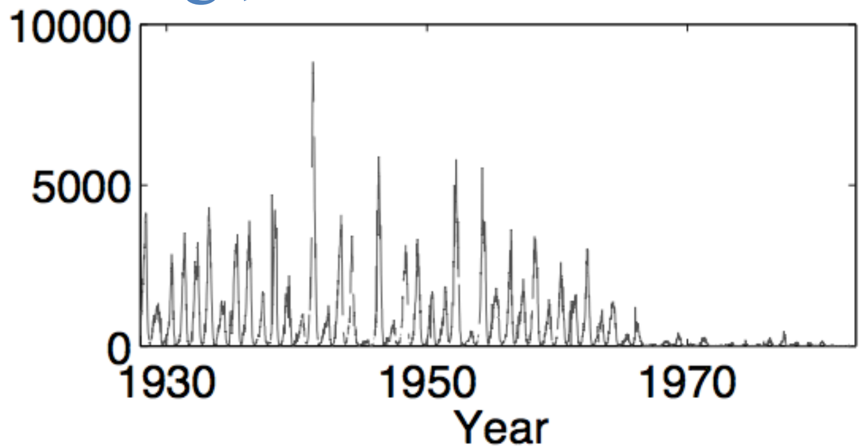


FUNNEL – with a single epidemic

Given:

"single" epidemic sequence

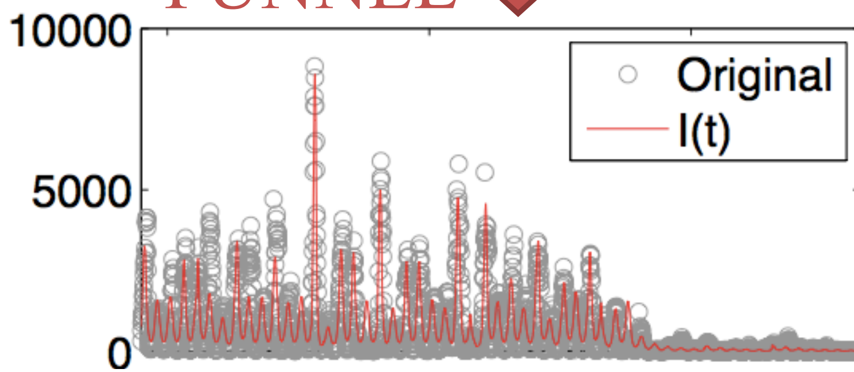
e.g., measles in NY



FUNNEL

Find:

nonlinear equation, model parameters



Kumamoto U

FUNNEL – with a single epidemic

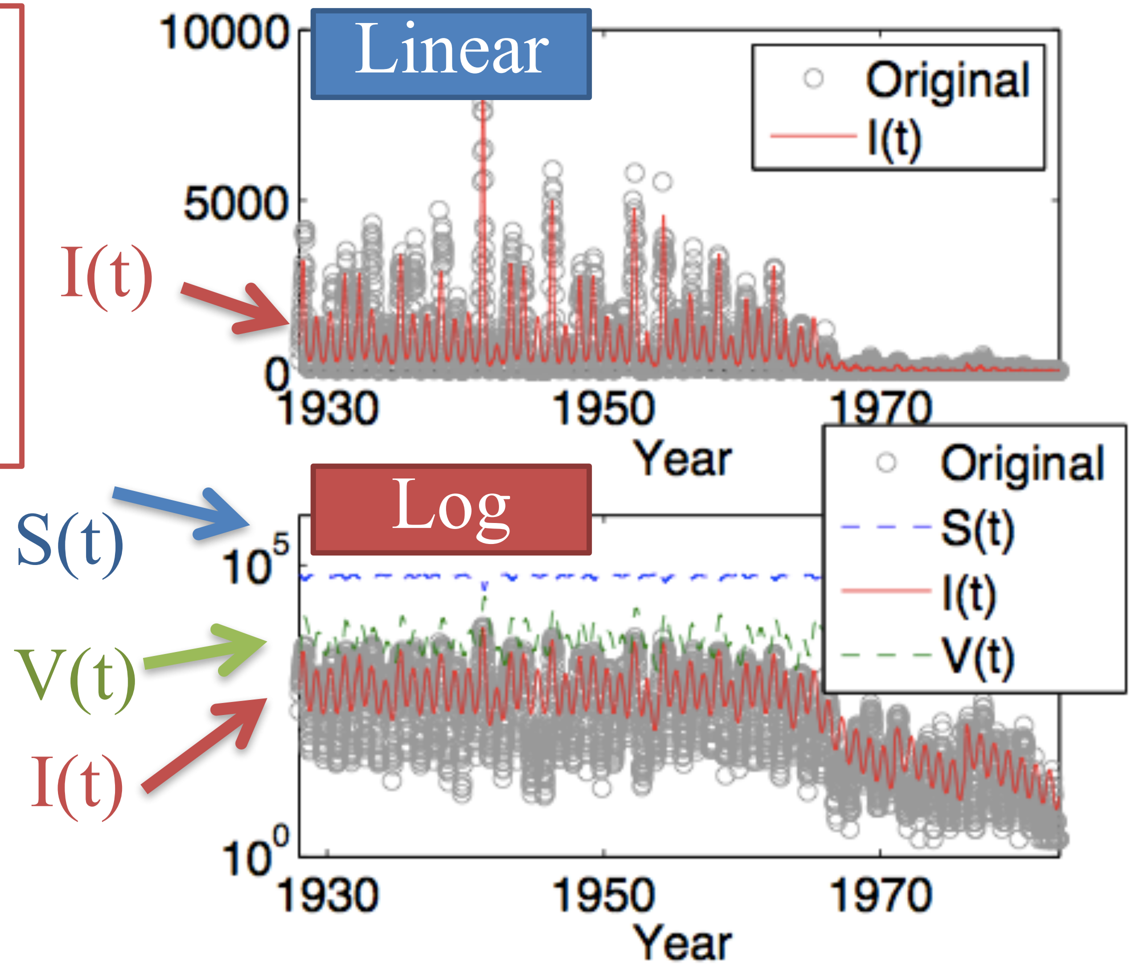
CMU CS

Details

With a single epidemic: Funnel-RE

People of 3 classes

- **S** : Susceptible
- **I** : Infected
- **V** : Vigilant/ vaccinated



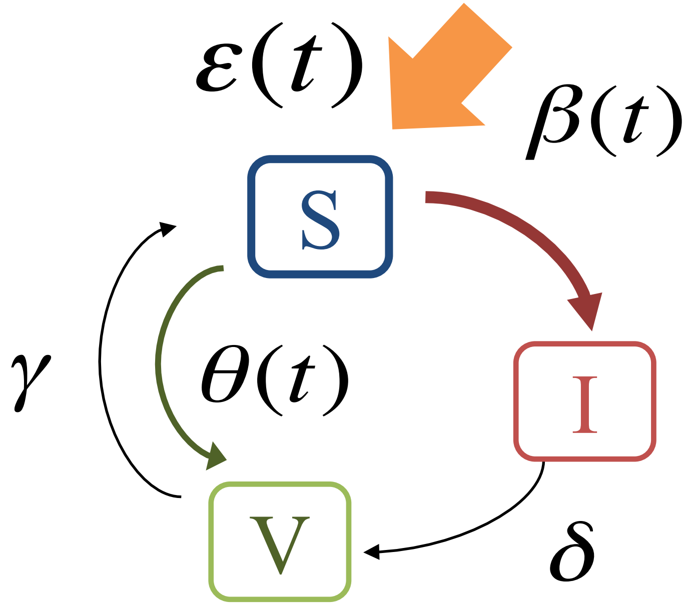
FUNNEL – with a single epidemic

Details

With a single epidemic: Funnel-RE

$$\begin{aligned} \boxed{S(t+1)} &= S(t) - \beta(t)\epsilon(t)S(t)I(t) + \gamma V(t) - \theta(t)S(t) \\ \boxed{I(t+1)} &= I(t) + \beta(t)\epsilon(t)S(t)I(t) - \delta I(t) \\ \boxed{V(t+1)} &= V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t) \qquad (3) \end{aligned}$$

S(t) : susceptible
I (t) : Infected
V(t) : Vigilant /Vaccinated



FUNNEL – with a single epidemic

Details

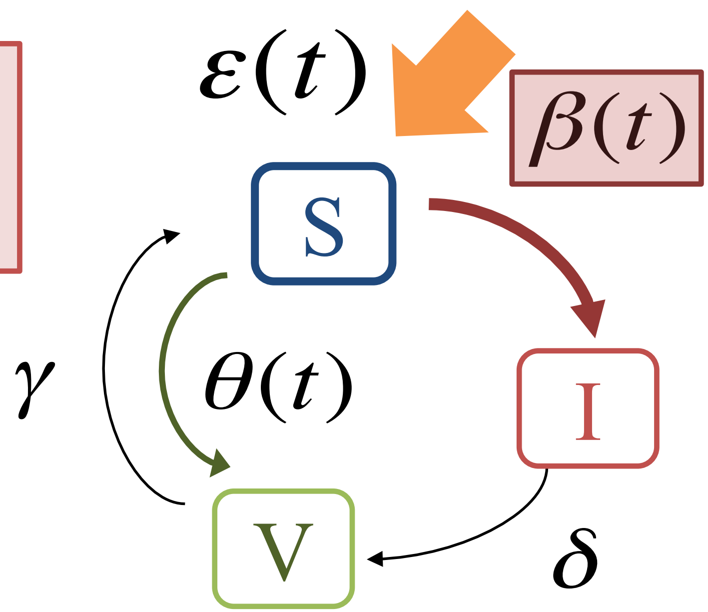
With a single epidemic: Funnel-RE

$$
\begin{aligned}
\boxed{S(t+1)} &= S(t) - \boxed{\beta(t)}\epsilon(t)S(t)I(t) + \gamma V(t) - \theta(t)S(t) \\
\boxed{I(t+1)} &= I(t) + \boxed{\beta(t)}\epsilon(t)S(t)I(t) - \delta I(t) \\
\boxed{V(t+1)} &= V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t) \qquad (3)
\end{aligned}
$$

$\beta(t)$: strength of infection
(yearly periodic func)

$$\beta(t) = \beta_0 \cdot \left(1 + P_a \cdot cos\left(\frac{2\pi}{P_p}(t + P_s)\right)\right)$$

$$P_p = 52$$



Kumamoto U

CMU CS

FUNNEL – with a single epidemic

Details

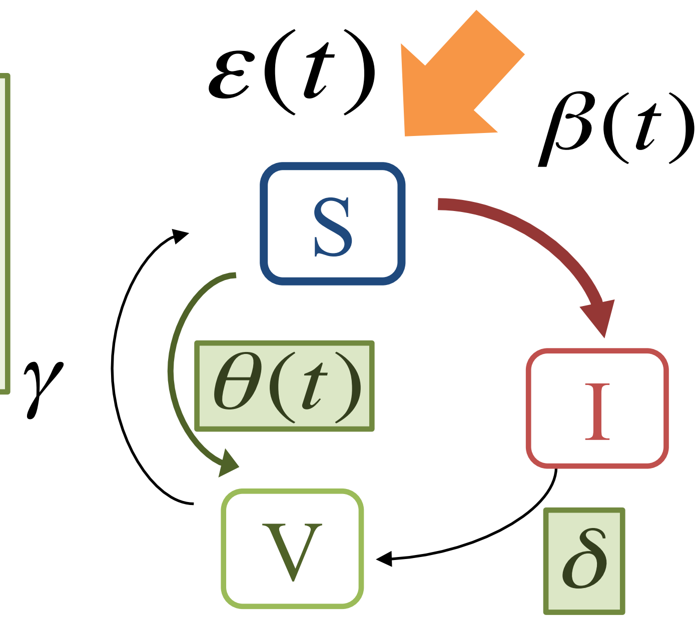
With a single epidemic: Funnel-RE

$$
\begin{aligned}
\boxed{S(t+1)} &= S(t) - \beta(t)\epsilon(t)S(t)I(t) + \gamma V(t) - \boxed{\theta(t)}S(t) \\
\boxed{I(t+1)} &= I(t) + \beta(t)\epsilon(t)S(t)I(t) - \boxed{\delta}I(t) \\
\boxed{V(t+1)} &= V(t) + \boxed{\delta}I(t) - \gamma V(t) + \boxed{\theta(t)}S(t) \qquad (3)
\end{aligned}
$$

δ : healing rate

$\theta(t)$: disease reduction effect

$$\theta(t) = \begin{cases} 0 & (t < t_\theta) \\ \theta_0 & (t \geq t_\theta) \end{cases}$$



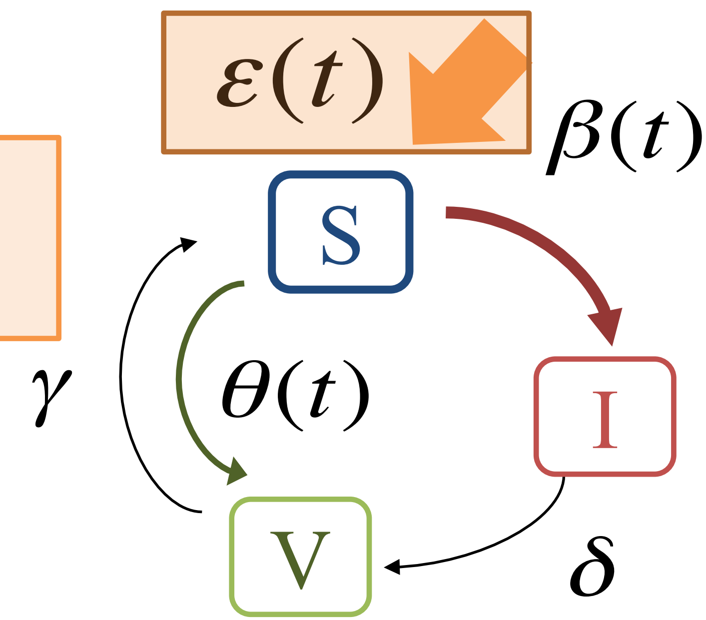
FUNNEL – with a single epidemic

Details

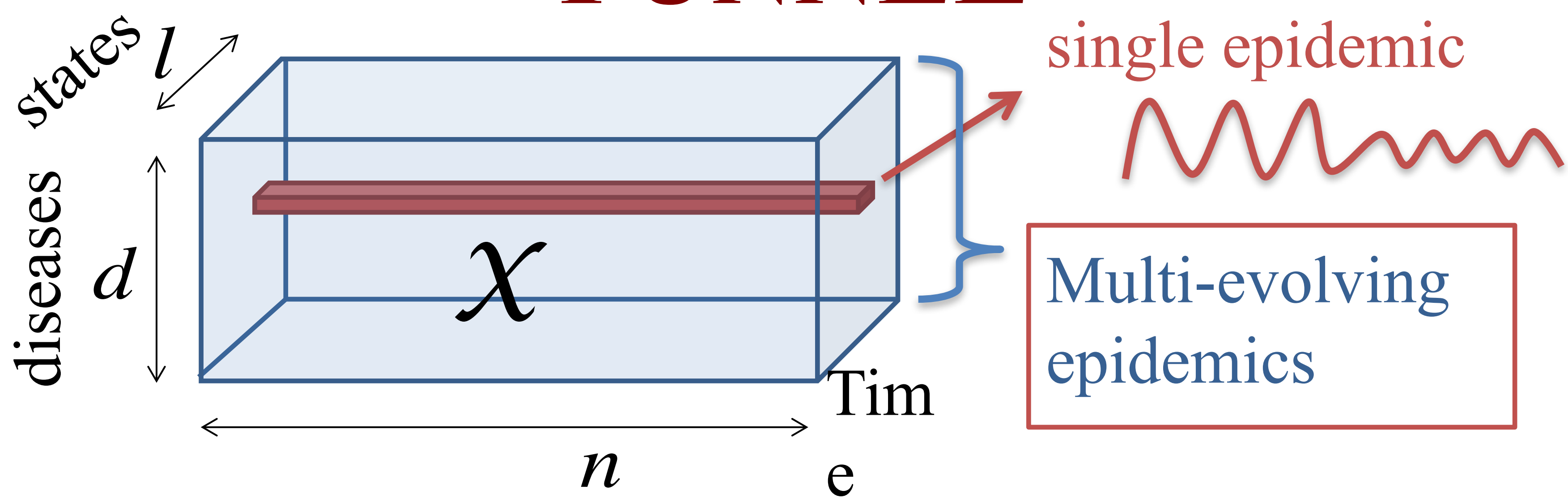
With a single epidemic: Funnel-RE

$$
\begin{aligned}
\boxed{S(t+1)} &= S(t) - \beta(t)\boxed{\epsilon(t)}S(t)I(t) + \gamma V(t) - \theta(t)S(t) \\
\boxed{I(t+1)} &= I(t) + \beta(t)\boxed{\epsilon(t)}S(t)I(t) - \delta I(t) \\
\boxed{V(t+1)} &= V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t) \qquad (3)
\end{aligned}
$$

$\varepsilon(t)$: temporal susceptible rate

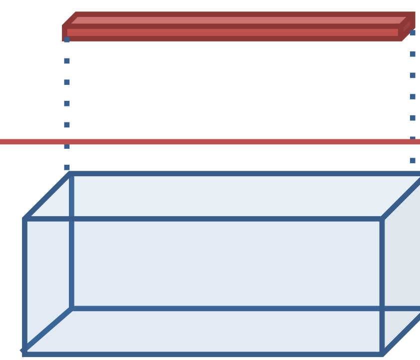


Proposed model: FUNNEL

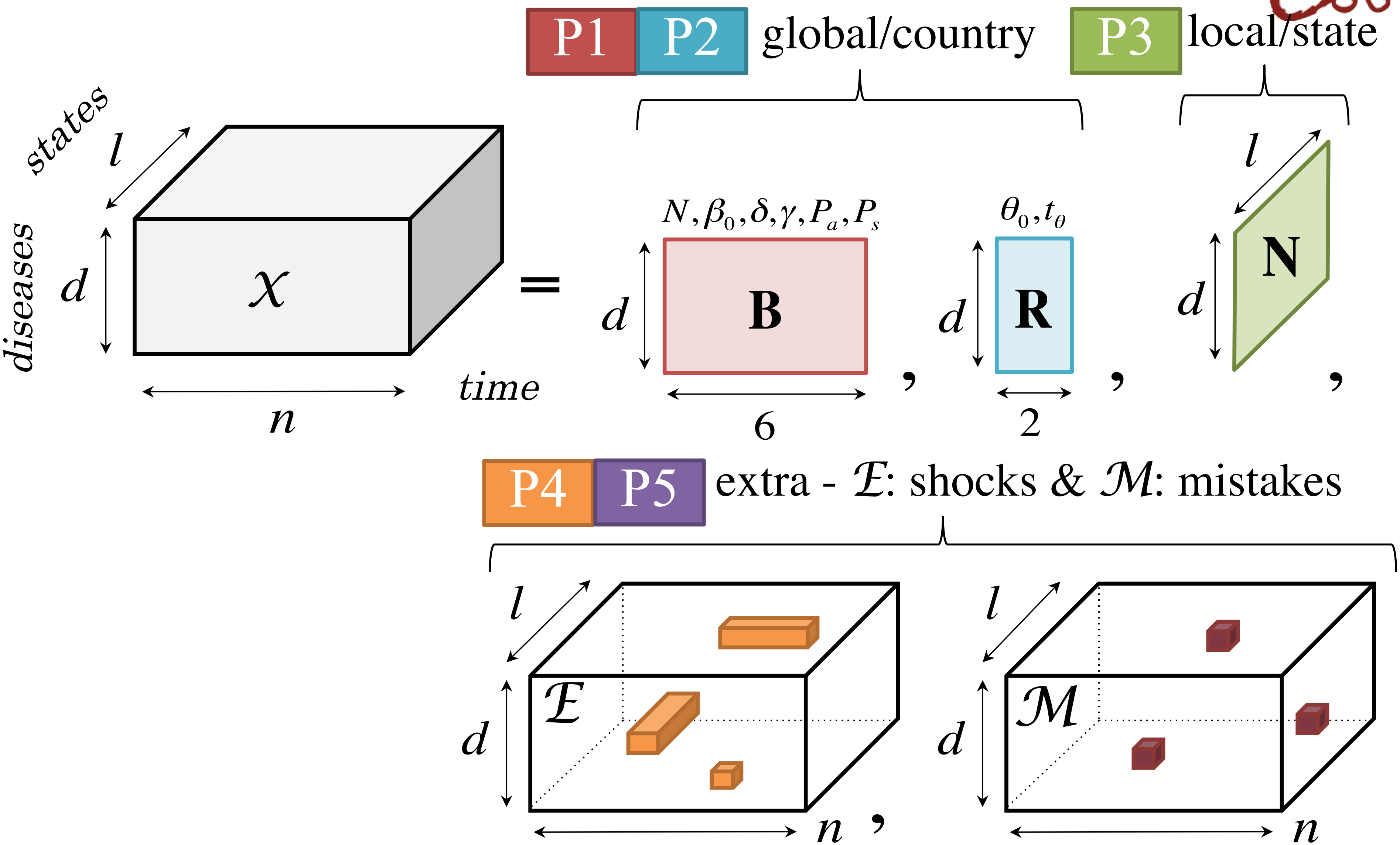


(a) FUNNEL-single

(b) FUNNEL-full (tensor)

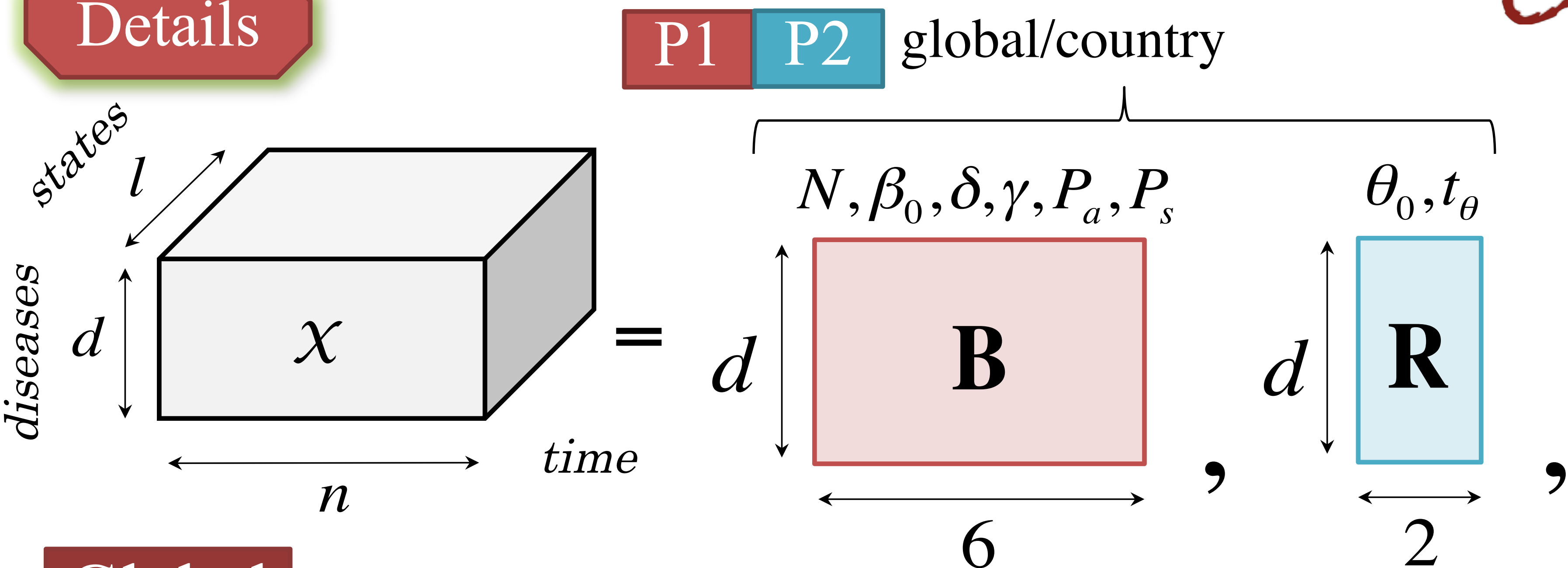


FUNNEL-full


P1 P2 global/country
P3 local/state
states
l
d
diseases
X
n
time
=
$N, \beta_0, \delta, \gamma, P_a, P_s$
B
6
θ_0, t_θ
R
2
N
P4 P5 extra - $\mathcal{E}$: shocks & $\mathcal{M}$: mistakes
$\mathcal{E}$
$\mathcal{M}$

FUNNEL-full

Details



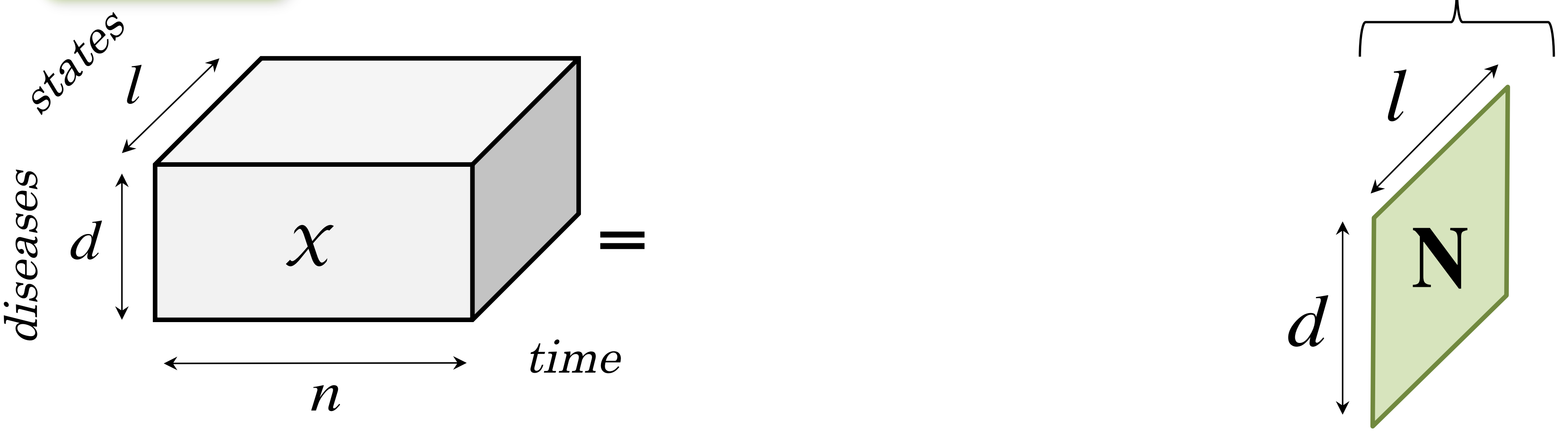
Global

P1 **Base matrix B (d x 6)**

P2 **Disease reduction matrix R (d x 2)**

FUNNEL-full

Details



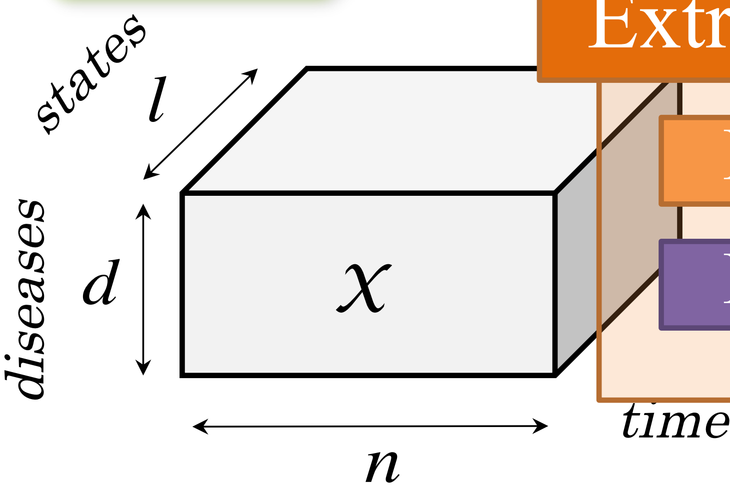
Local

P3 **Geo-disease matrix N (d x l)**

$$\mathbf{N} = \{N_{ij}\}_{i,j=1}^{d,l}$$: potential population of disease i in state j

FUNNEL-full

Details

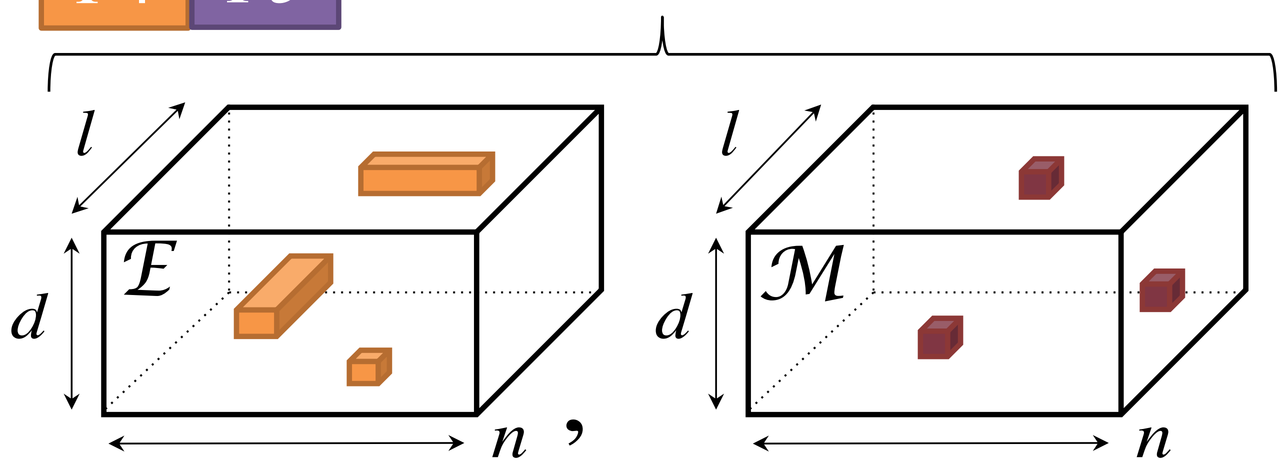


Extra

P4 **External shock tensor $\mathcal{E}$**

P5 **Mistake tensor $\mathcal{M}$**

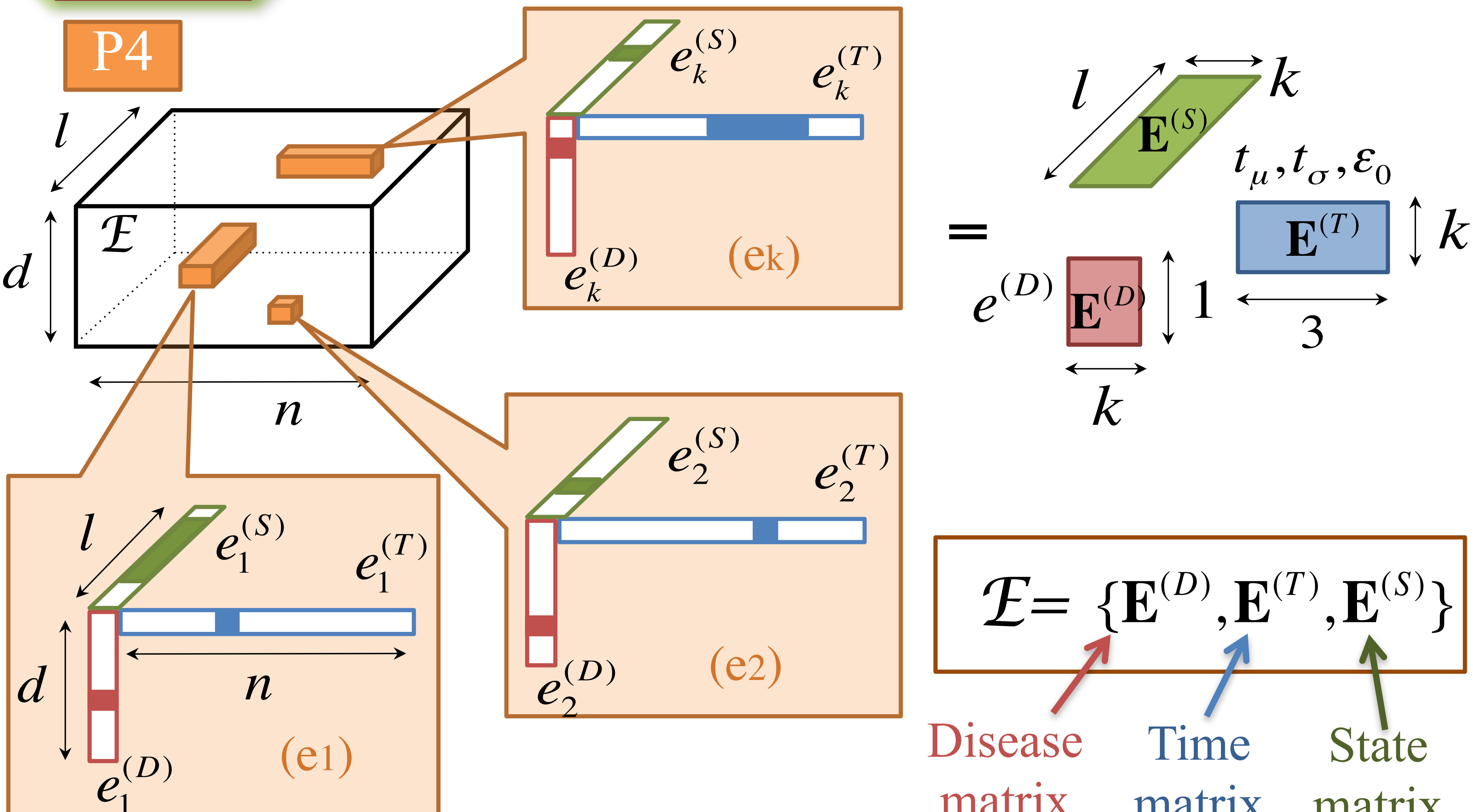
P4 P5 extra - $\mathcal{E}$: shocks & $\mathcal{M}$: mistakes



FUNNEL-full

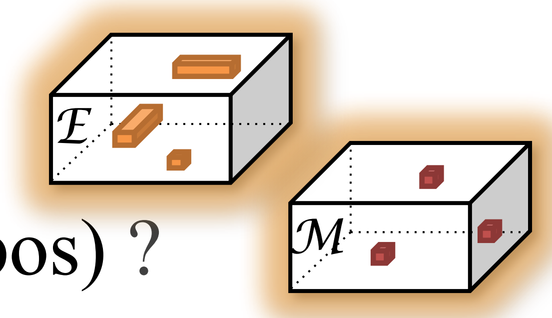
Details

P4

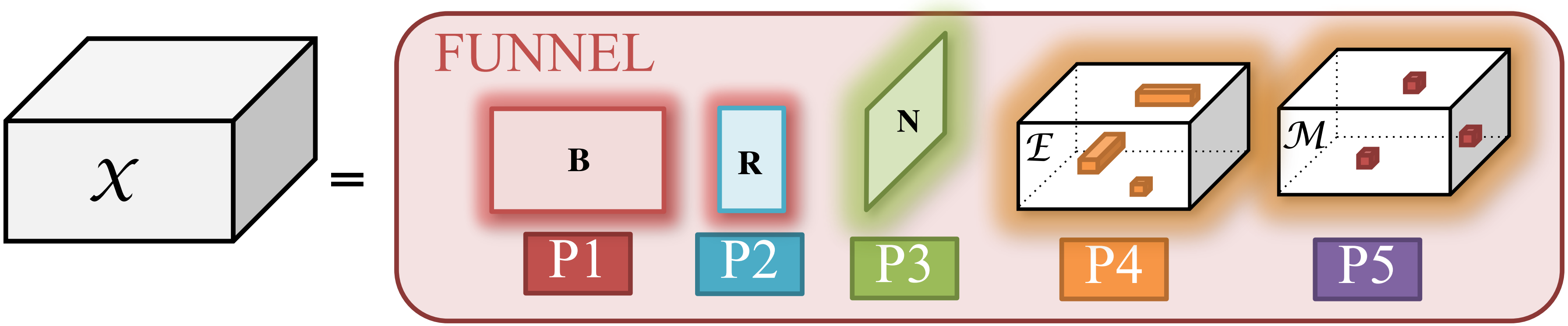


Challenges

Q1. How to automatically
- find "external shocks" ?
- ignore "mistakes" (i.e., typos) ?

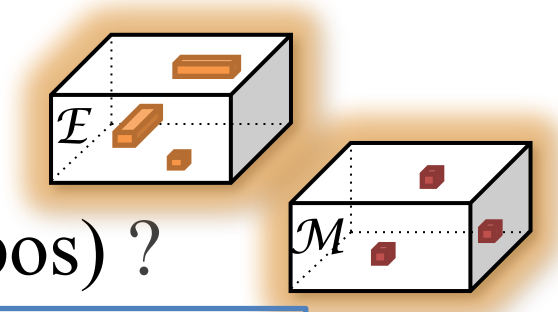


Q2. How to efficiently estimate model parameters ?



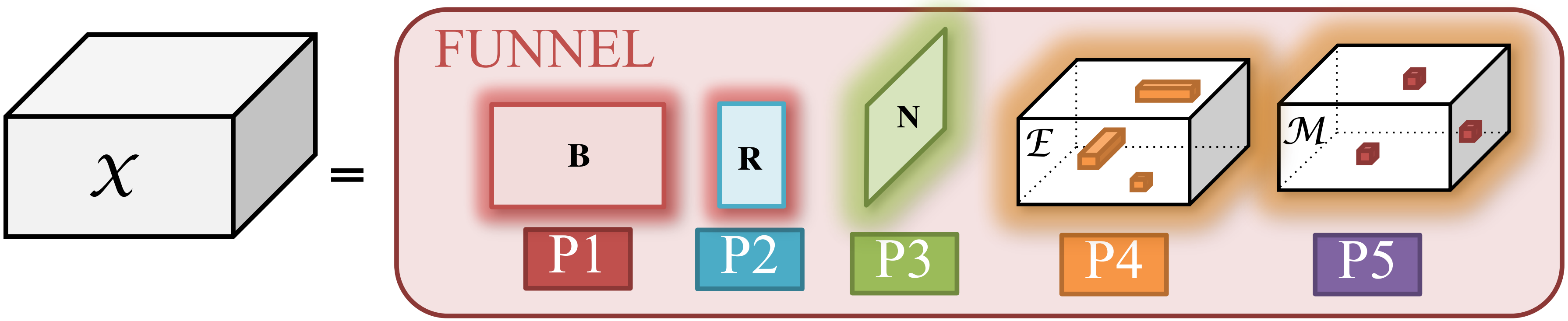
Challenges

Q1. How to automatically
- find "external shocks" ?
- ignore "mistakes" (i.e., typos) ?



Idea (1) : Model description cost

Q2. How to efficiently estimate model parameters ?



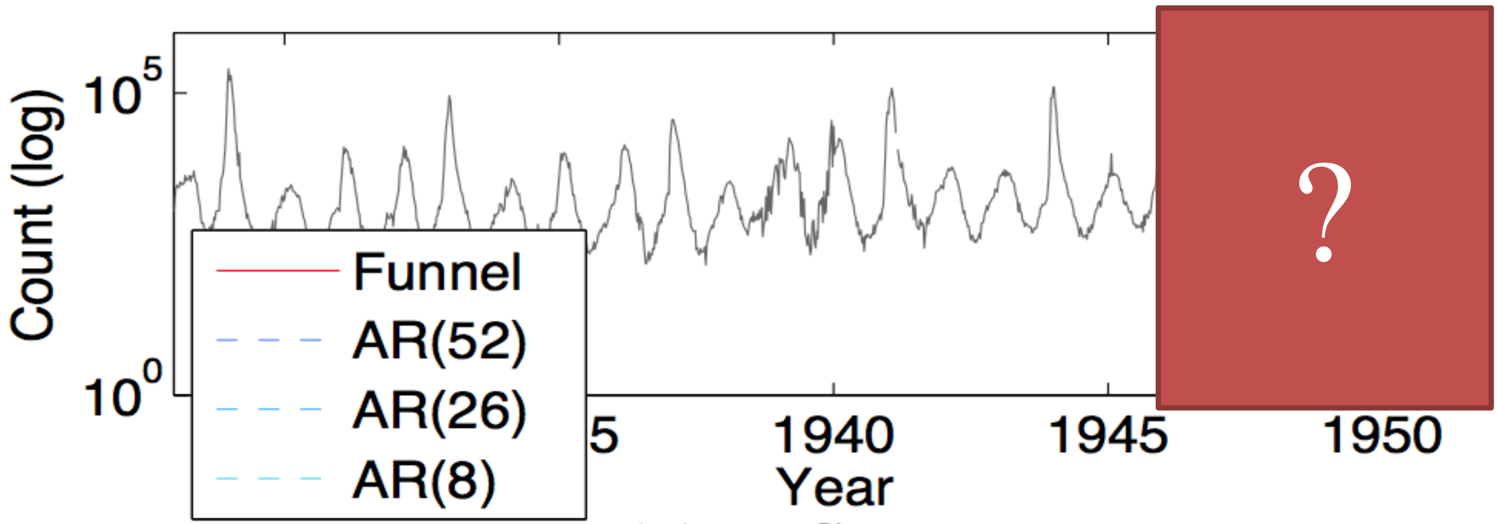
Idea (2): Multi-layer optimization - *O(dln)*

FUNNEL at work - forecasting

Forecasting future epidemics

Train:
2/3 sequences

Forecast:
1/3 following years



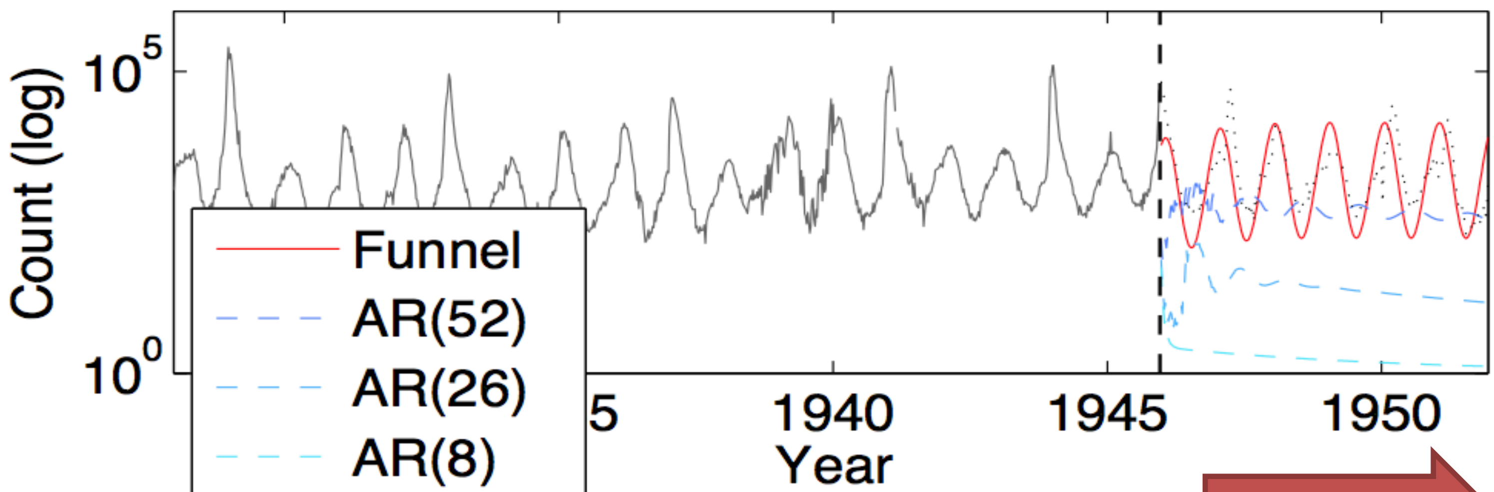
(a) Influenza

FUNNEL at work - forecasting

Forecasting future epidemics

Train:
2/3 sequences

Forecast:
1/3 following years



(a) Influenza

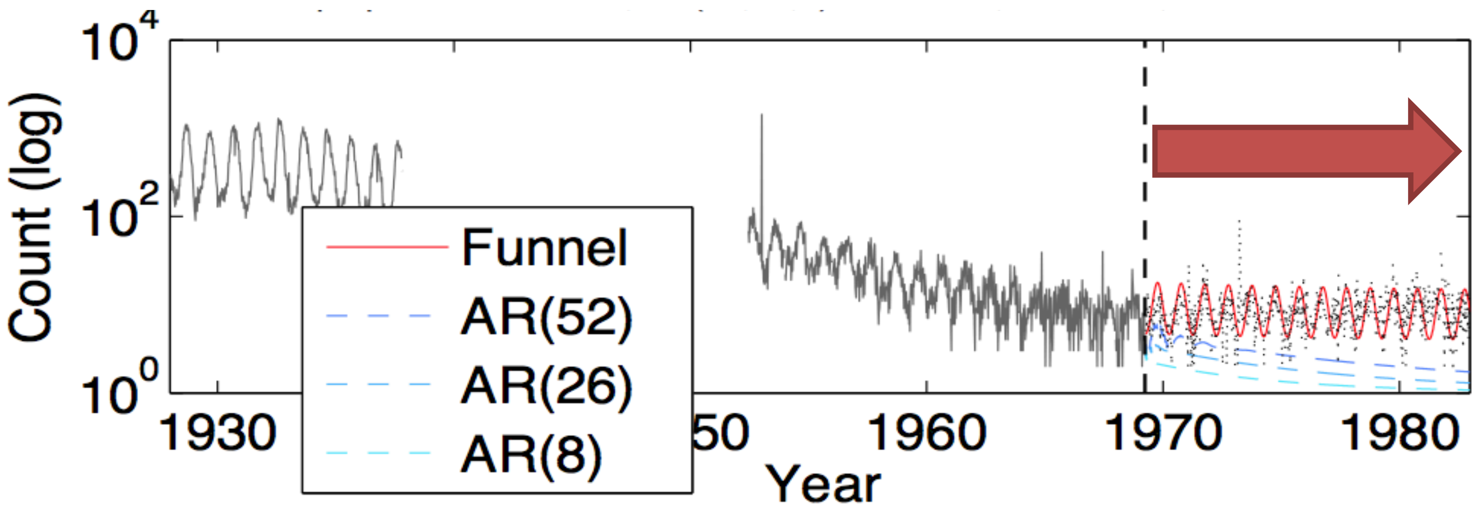
Funnel can capture future epidemics (AR: fail)

FUNNEL at work - forecasting

Forecasting future epidemics

Train:
2/3 sequences

Forecast:
1/3 following years



(c) Typhoid fever

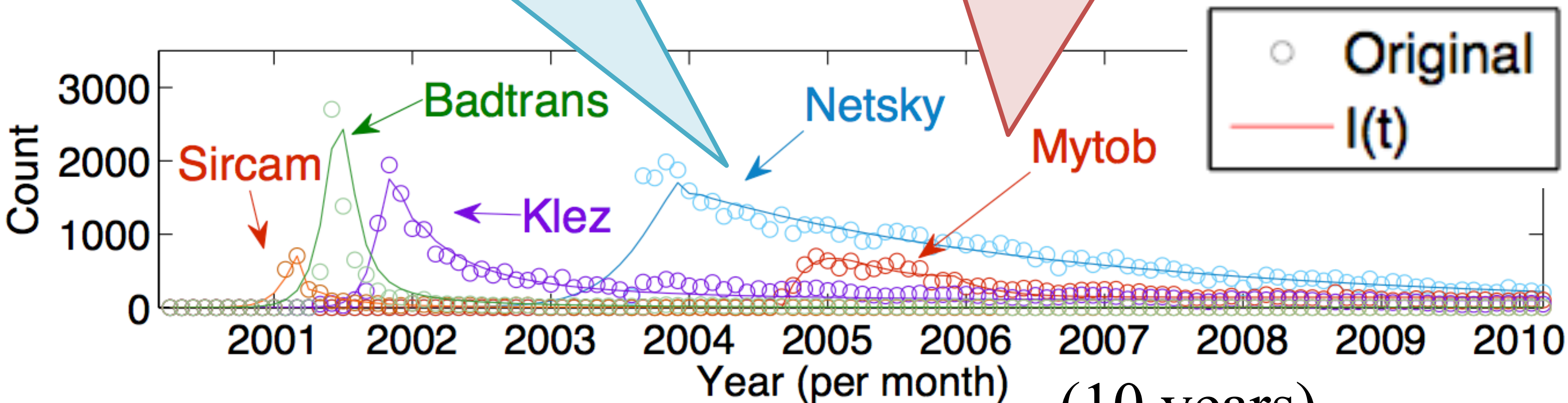
Funnel can capture future epidemics (AR: fail)

Generality of FUNNEL

Epidemics on computer networks

Spread via email attachment

Spread through corporate networks



(10 years)

Funnel is general: it fits computer virus very well!

[Matsubara+ WWW'16]

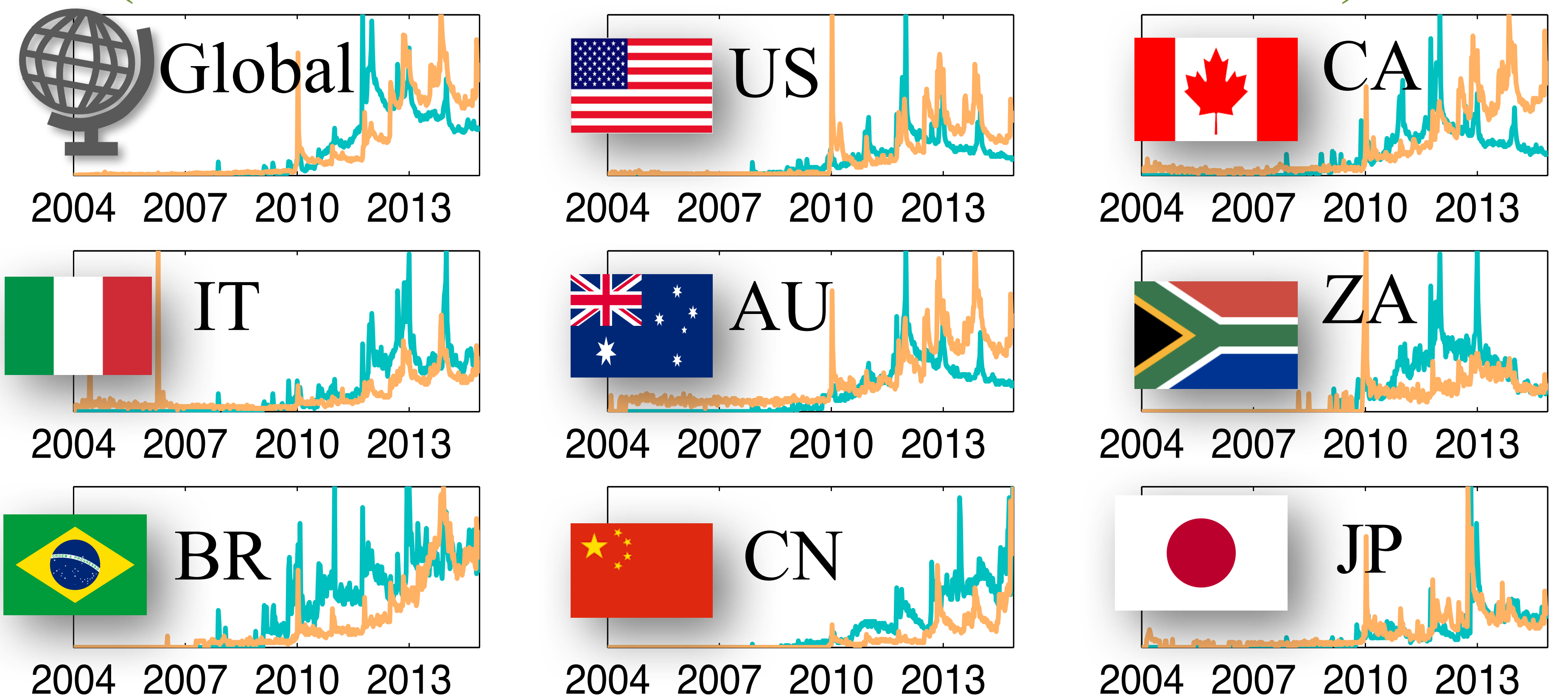
Non-linear Mining of Competing Local Activities

Yasuko Matsubara (Kumamoto University)
Yasushi Sakurai (Kumamoto University)
Christos Faloutsos (CMU)

Given: local user activities

e.g., *Google* search volumes for **Kindle, Nexus**

(for 236 countries, from 2004 to 2015)



Given: local user activities

e.g., *Google* search volumes for **Kindle, Nexus**

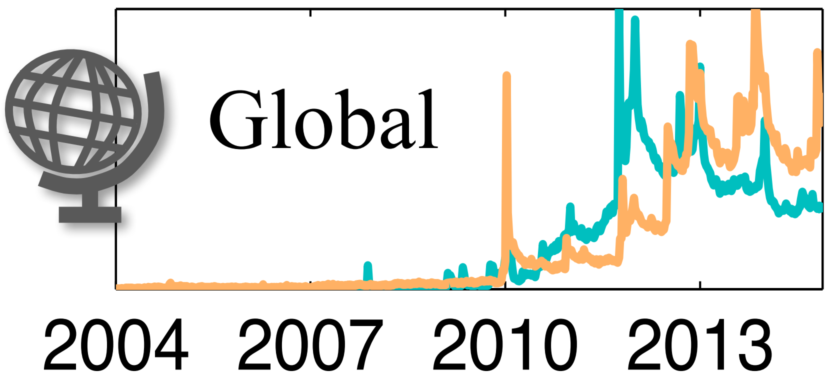
(for 236 countries, from 2004 to 2015)

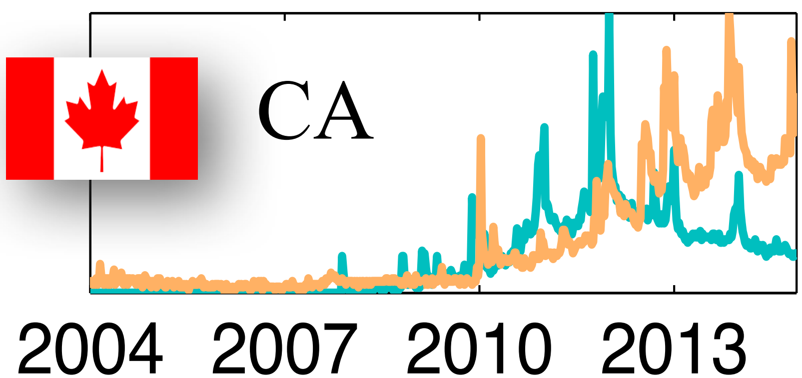
Q. Any global/local trends?

Given: local user activities

e.g., *Google* search volumes for **Kindle**, **Nexus**

(for 236 countries, from 2004 to 2015)

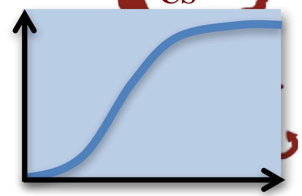




Nexus

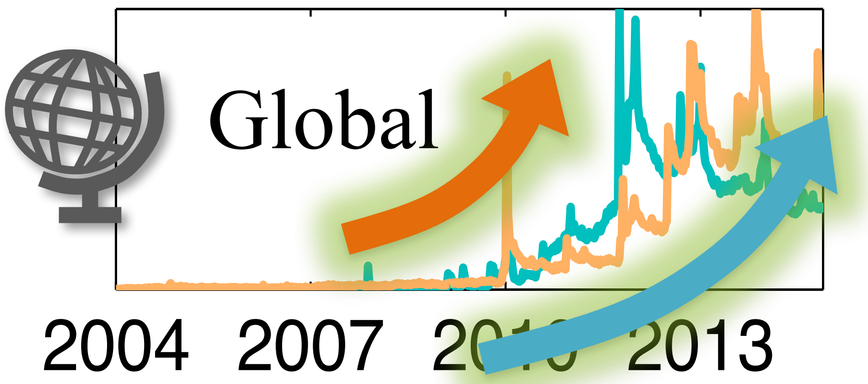
Kindle

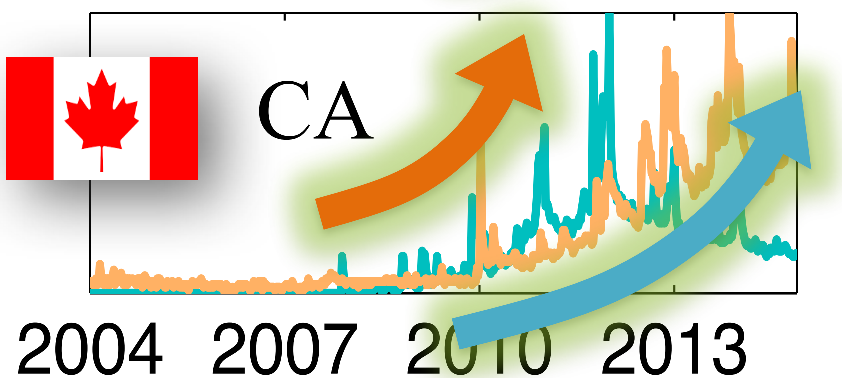
Given: local user activities



e.g., *Google* search volumes for **Kindle**, **Nexus**

(for 236 countries, from 2004 to 2015)





1. Exponential growth

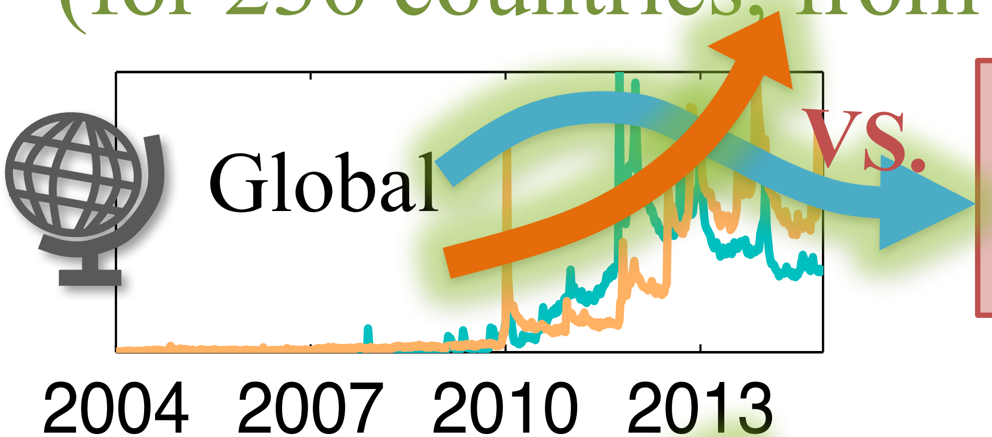
Nexus

Kindle

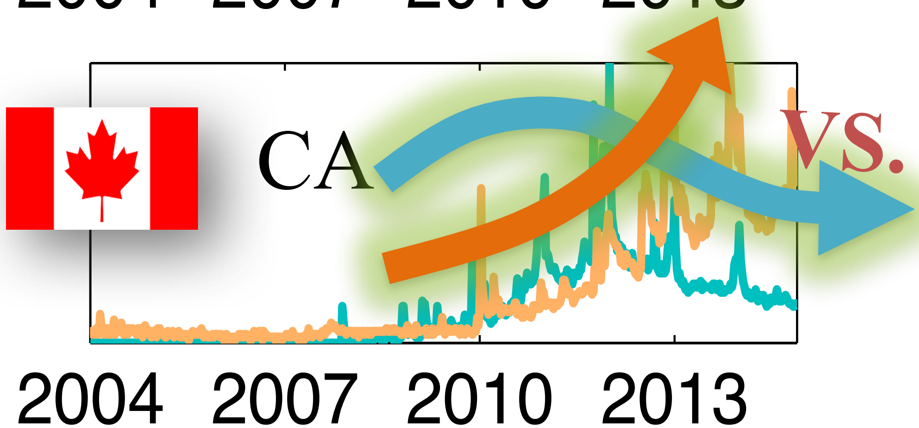
Given: local user activities

e.g., *Google* search volumes for **Kindle**, **Nexus**

(for 236 countries, from 2004 to 2015)



2. Competition



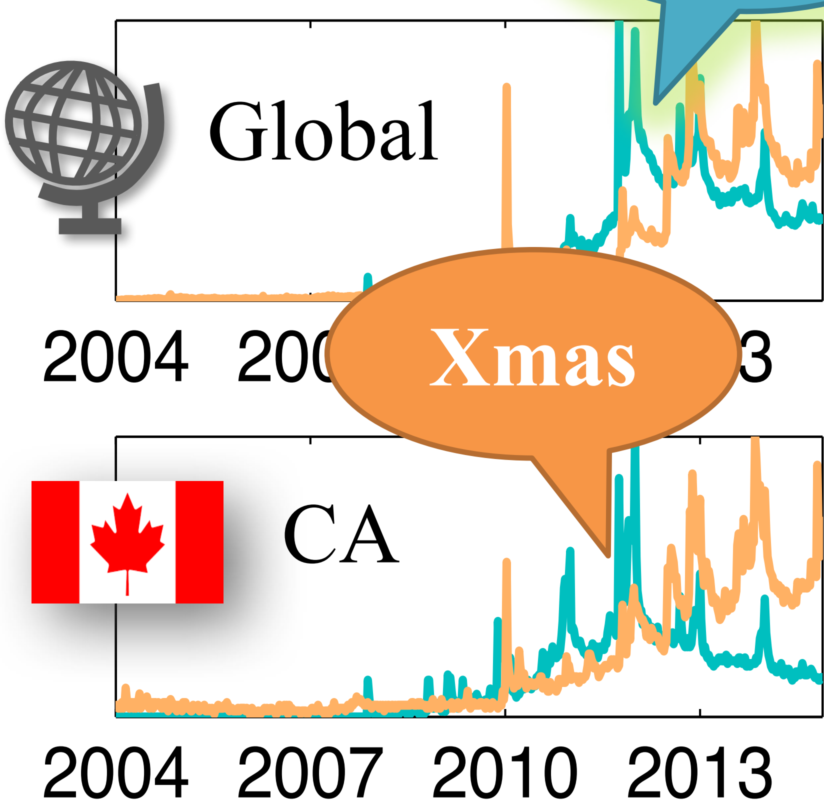
Nexus

Kindle

Given: local user activities

e.g., *Google* search volumes for **Kindle**, **Nexus**

(for 236 cou... 2004 to 2015)



3. Seasonality

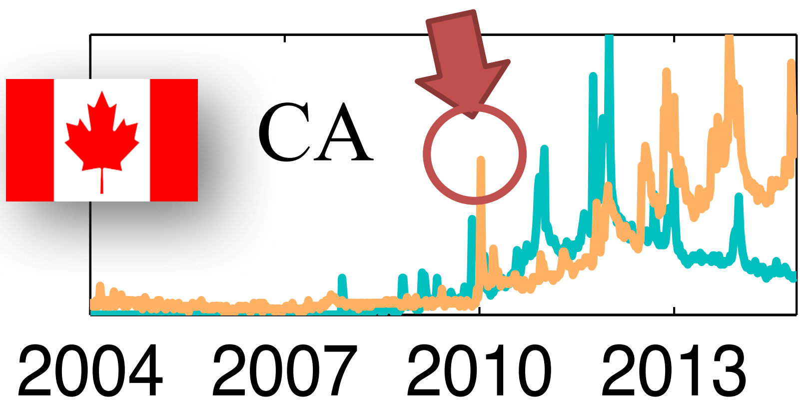
Nexus

Kindle

Given: local user activities

e.g., *Google* search volumes for **Kindle**, **Nexus**

(for 236 countries, from 2004 to 2015)

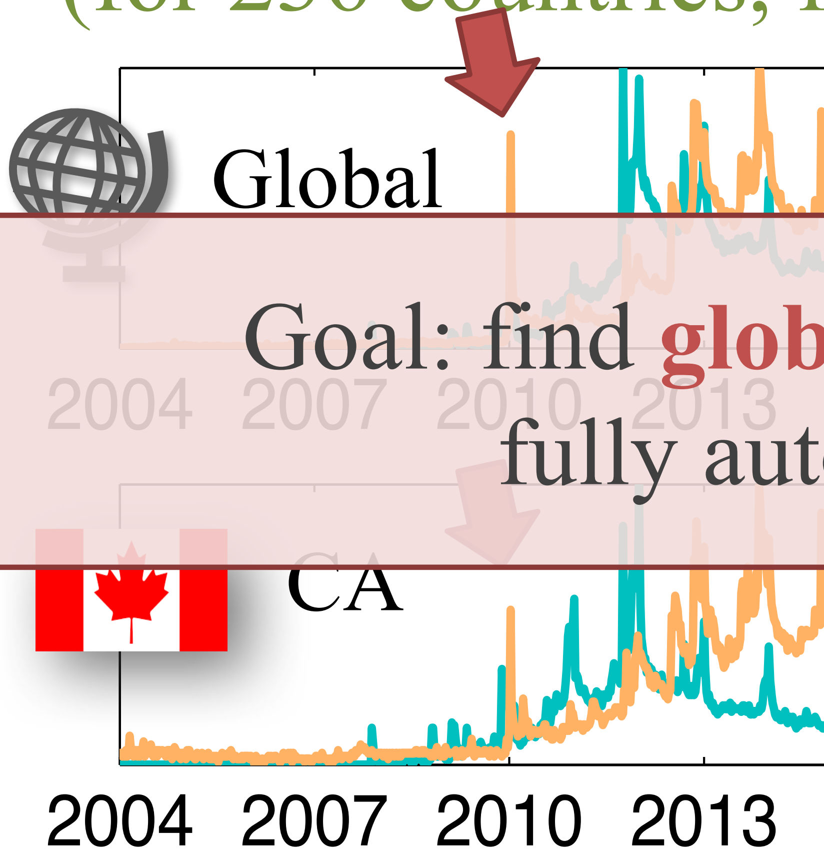


4. Deltas (outliers)

Given: local user activities

e.g., *Google* search volumes for **Kindle**, **Nexus**

(for 236 countries, from 2004 to 2015)



4. Deltas
(outliers)

Goal: find **global/local** patterns,
fully automatically

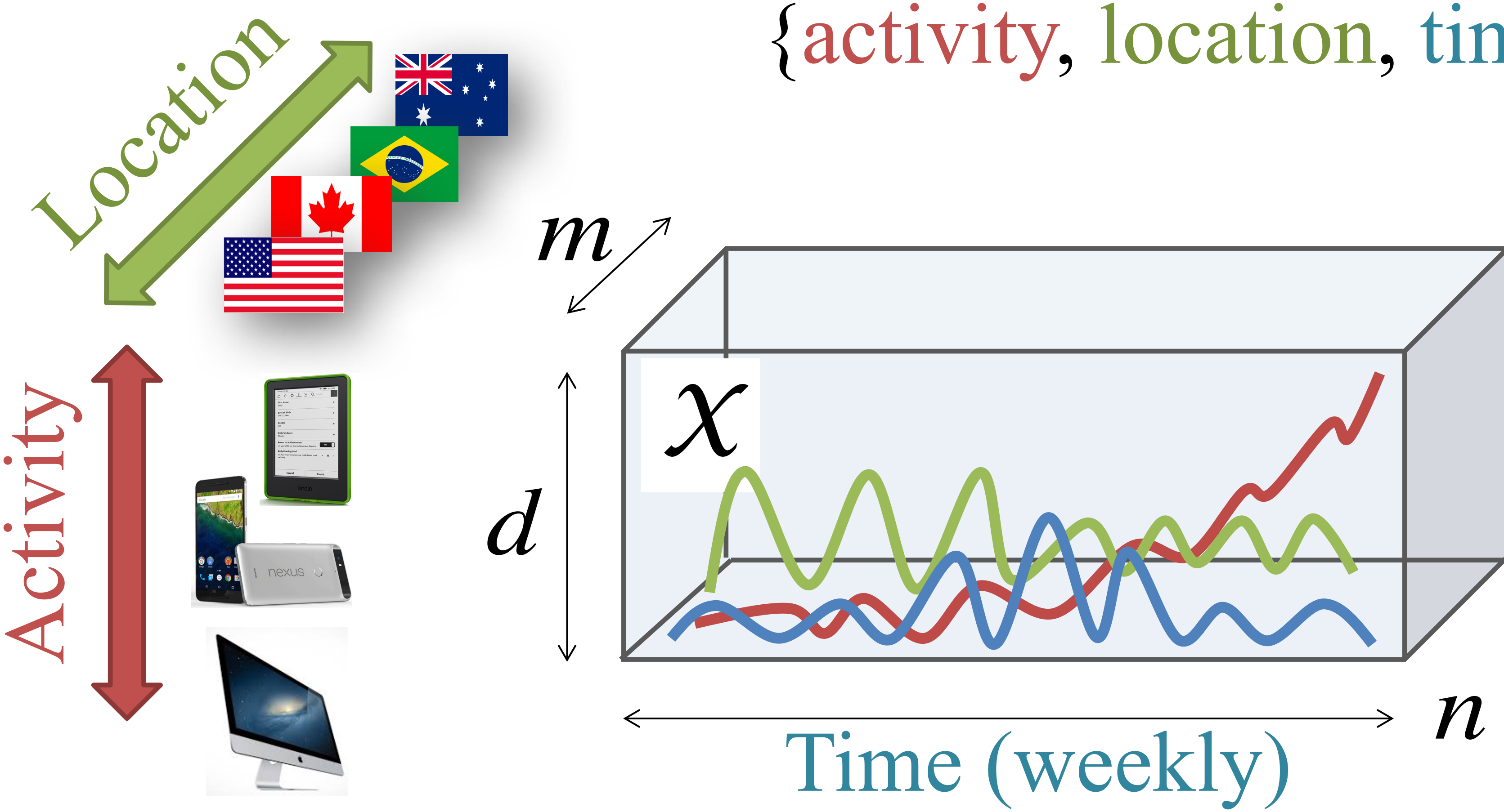
Nexus

Kindle

Data description

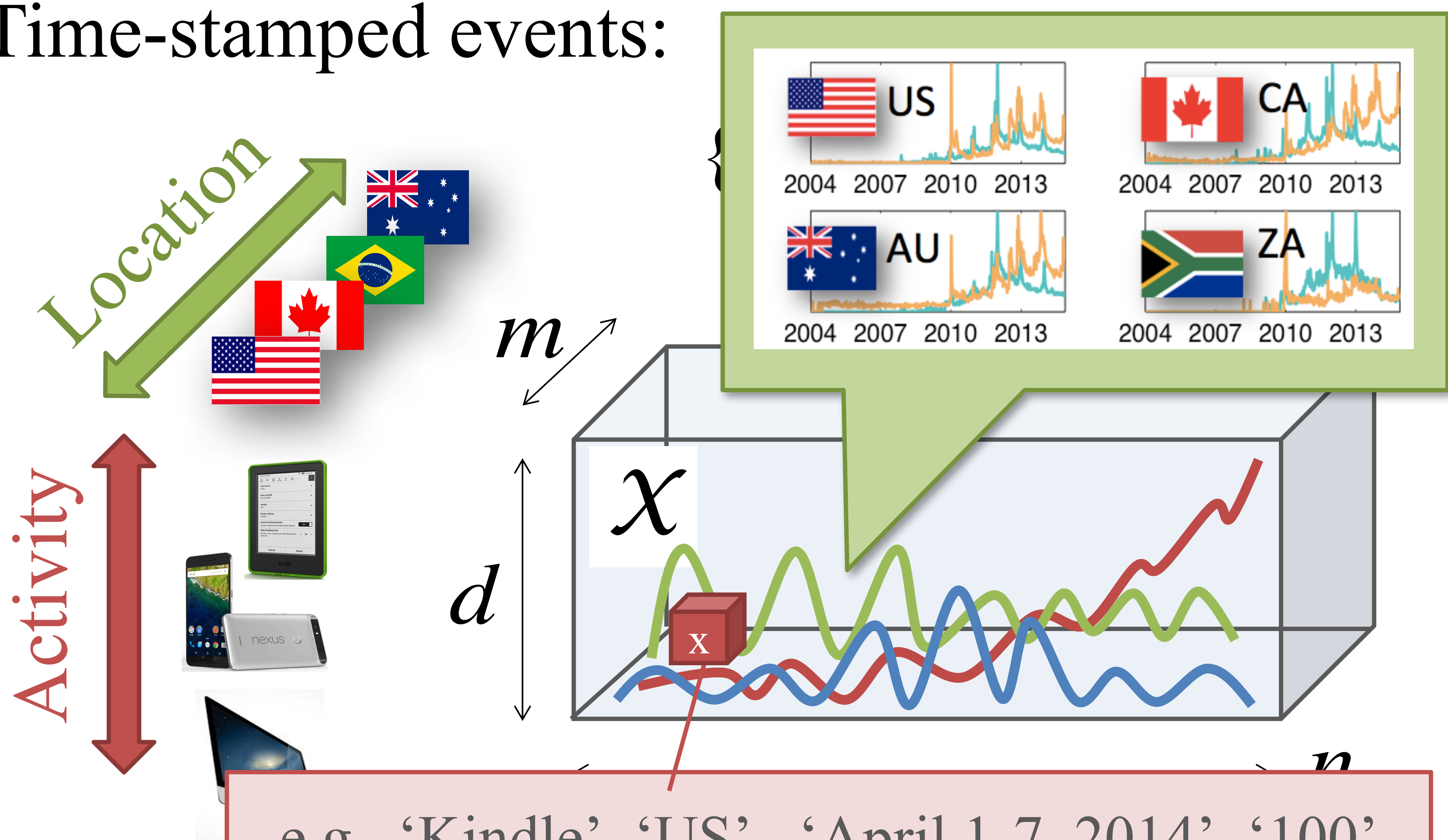
Time-stamped events:

{activity, location, time}



Data description

Time-stamped events:

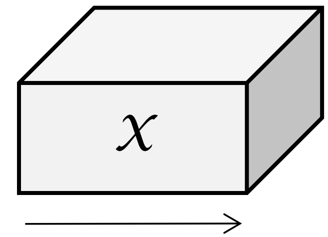


e.g., 'Kindle', 'US', 'April 1-7, 2014', '100'

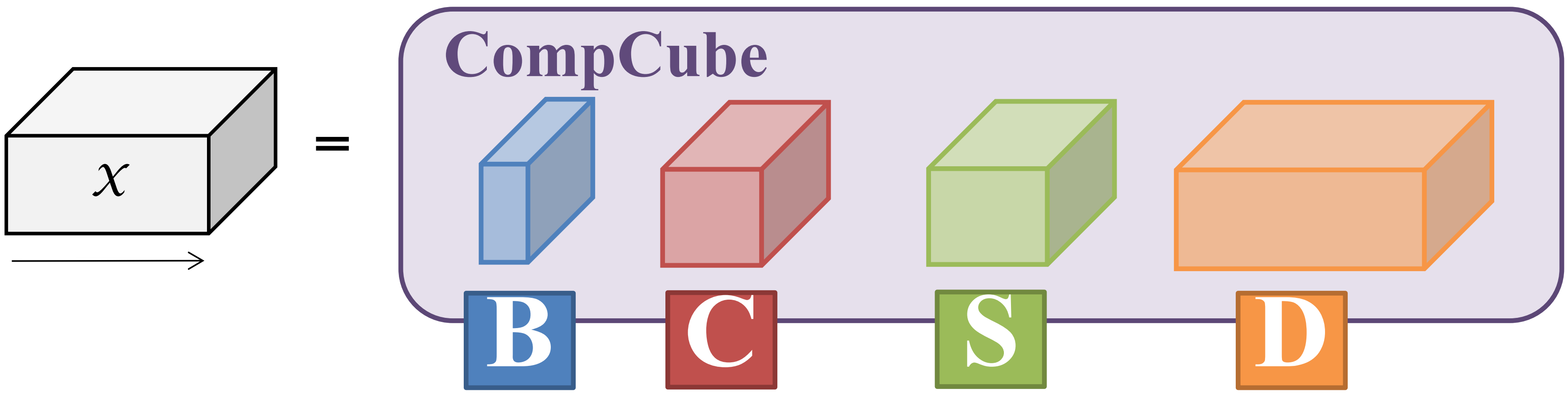
Problem definition

Given: Tensor $\mathcal{X}$

(activity x location x time)

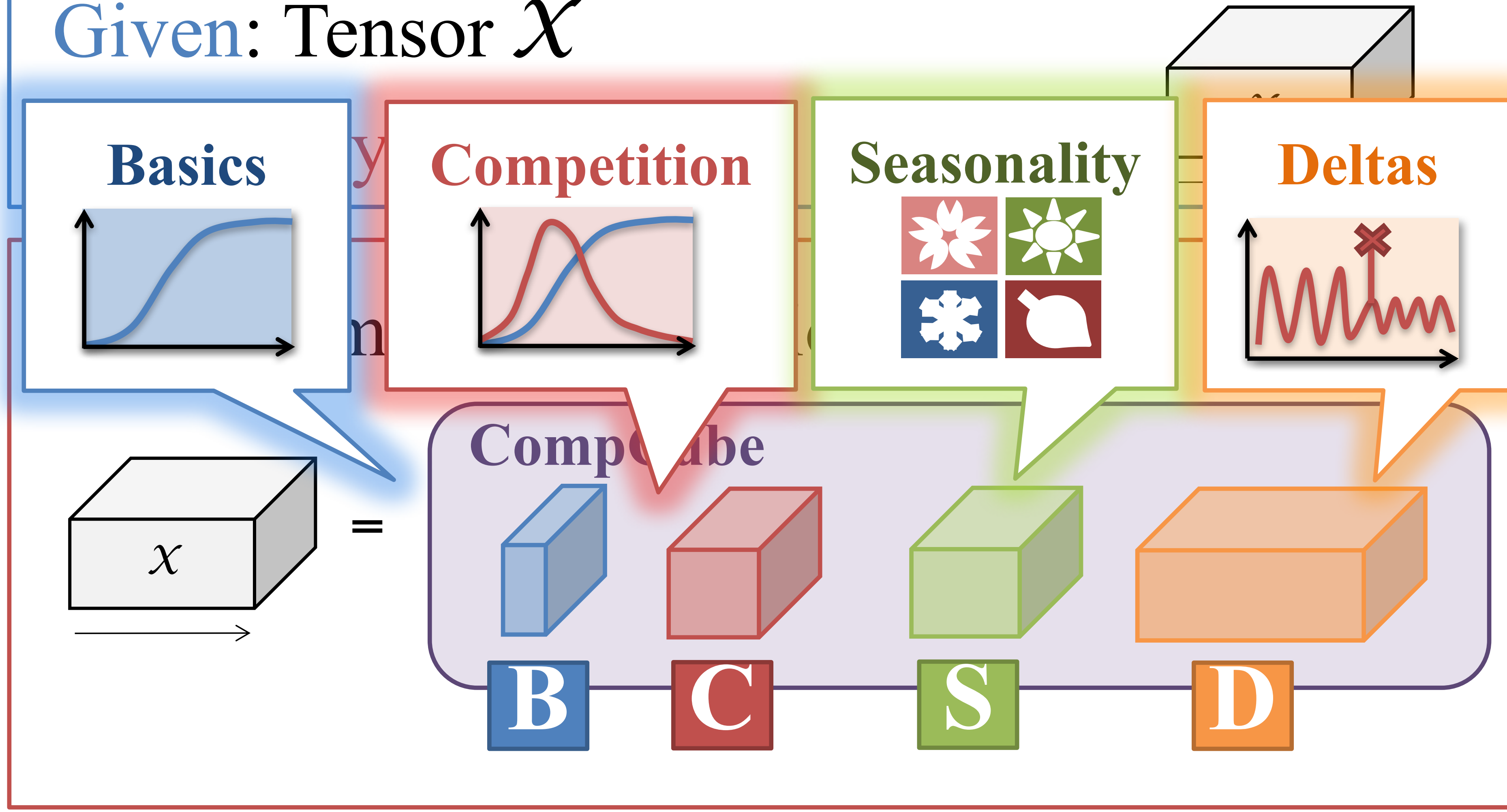


Find: Compact description of $\mathcal{X}$



Problem definition

Given: Tensor $\mathcal{X}$

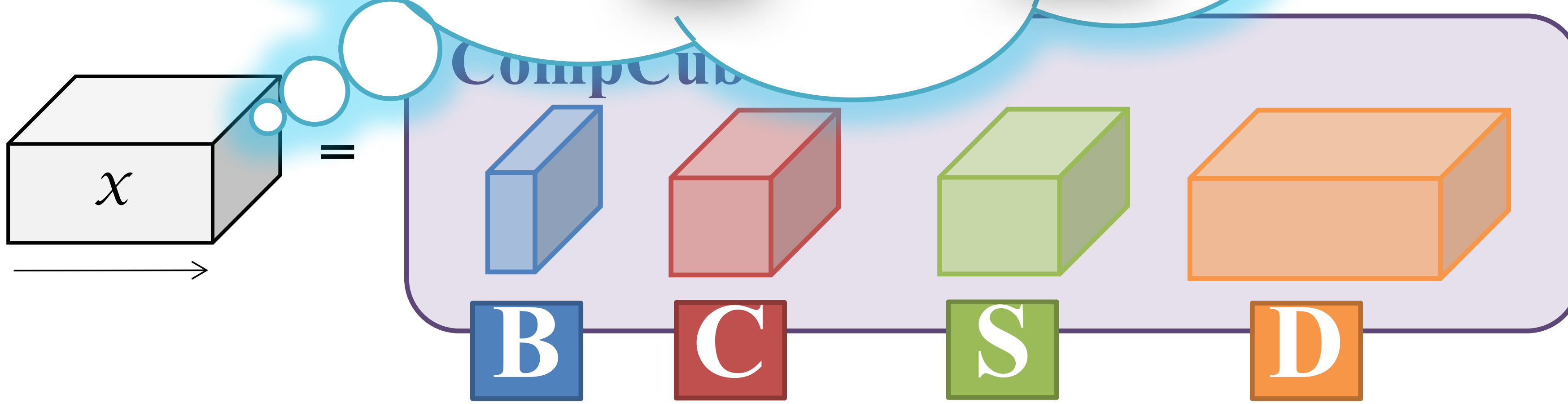


Problem definition

Given: Ten

(activi

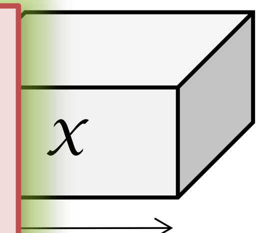
Find: C



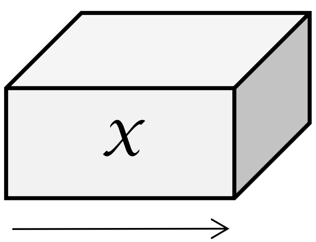
Problem definition

Given: Te

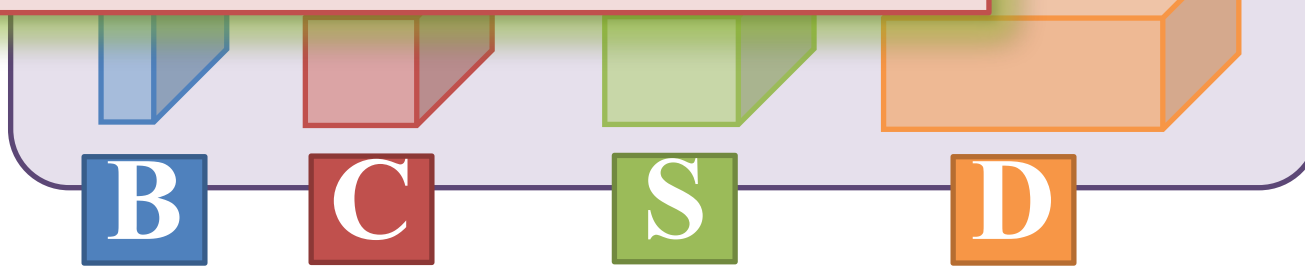
(activity



Find: Com



=



NO magic numbers !

Parameter-free!

Modeling power of CompCube

Products

News sources

Microsoft

AP

CNN

YAHOO! NEWS

TIME

H

BBC NEWS

Modeling power of CompCube

Products

News sources

Microsoft
AP
CNN
YAHOO! NEWS
TIME
H
BBC NEWS

Modeling power of CompCube

Produc

Q. Any global/local competition?

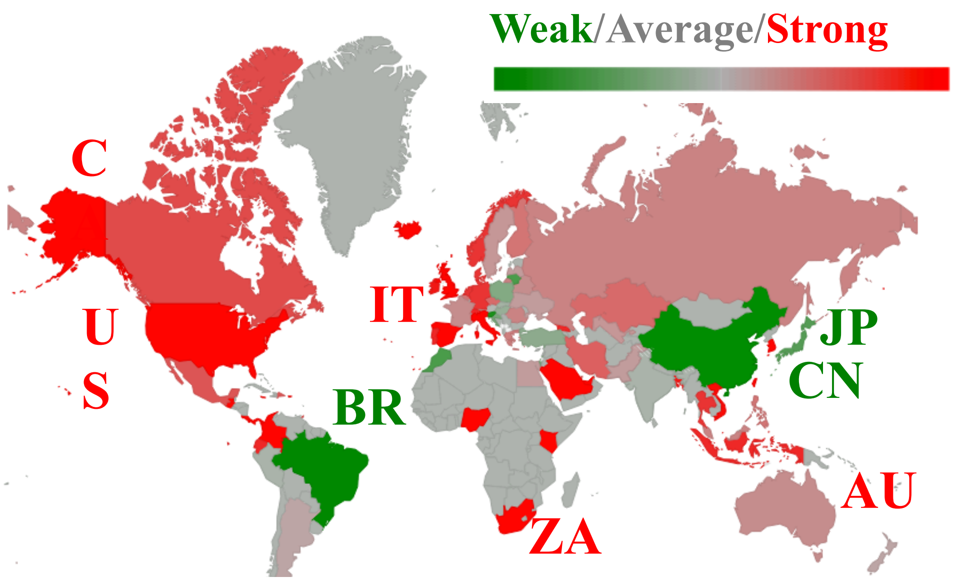
Nexus Kindle

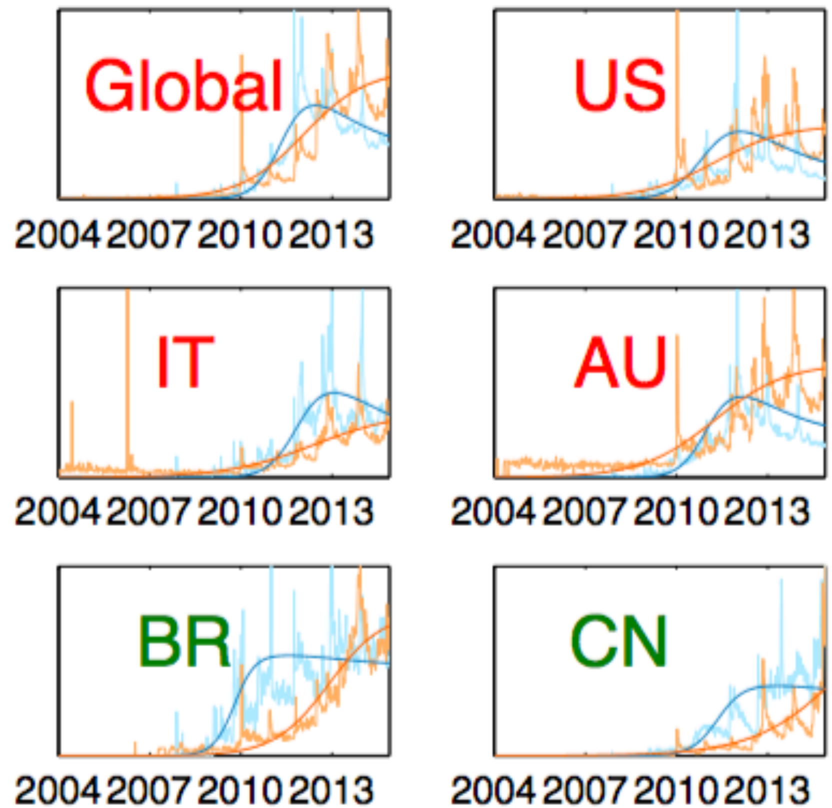
VS.

e.g., in

Modeling power of CompCube

e.g., *Google* search volumes for **Kindle, Nexus**

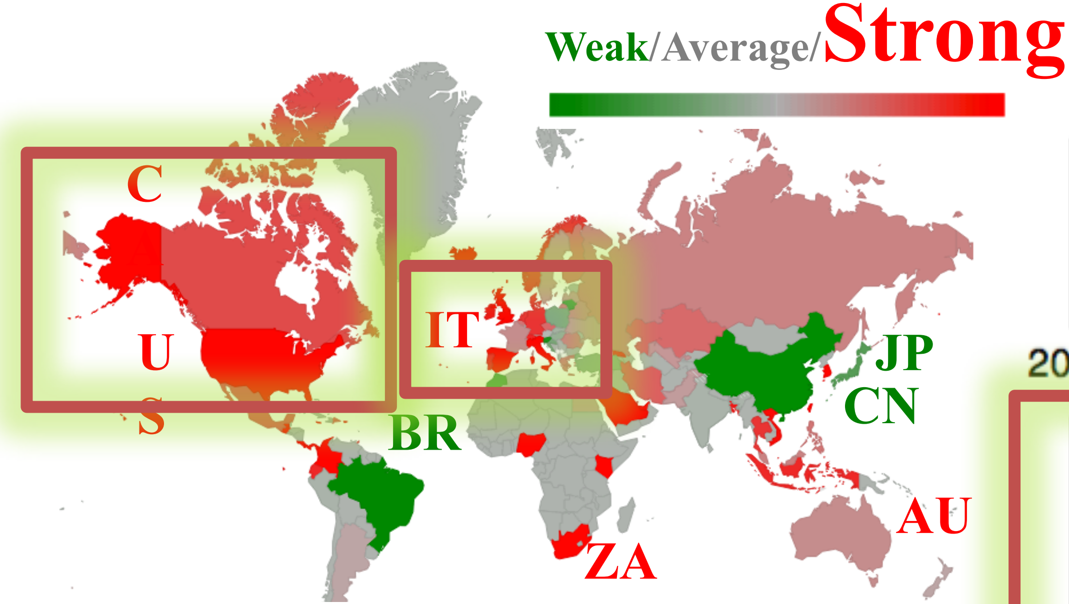


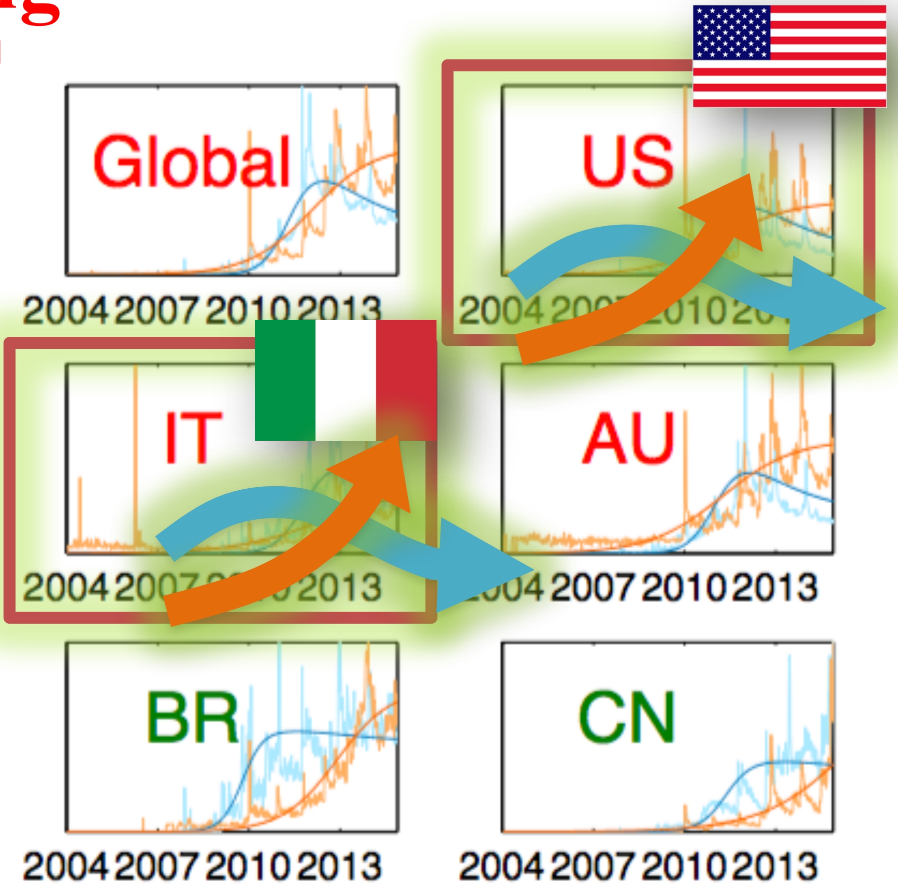


Local Competition strength

Modeling power of CompCube

e.g., *Google* search volumes for **Kindle**, **Nexus**



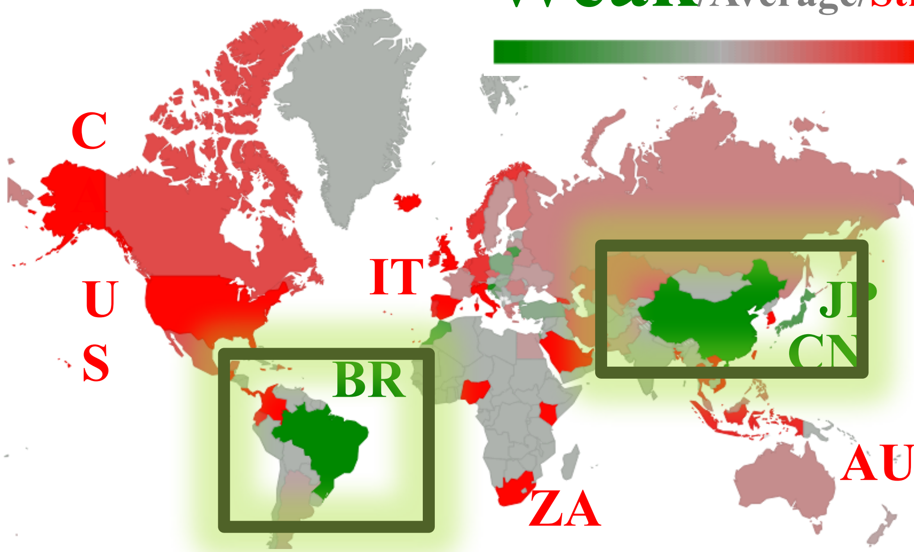


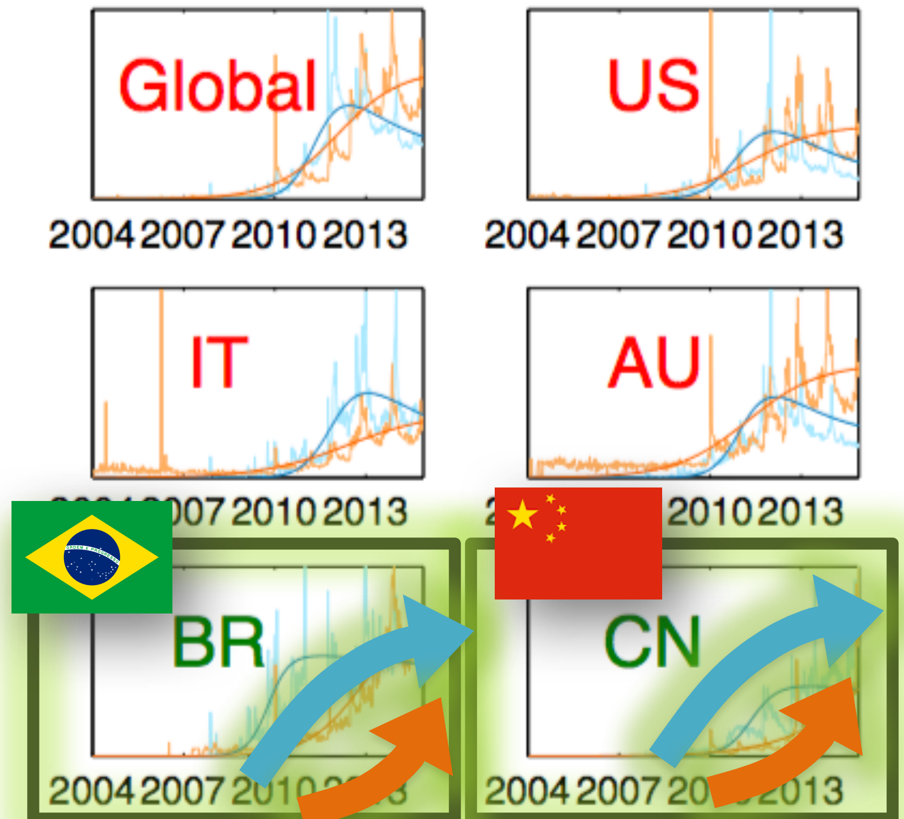
Local Competition strength

Modeling power of CompCube

e.g., *Google* search volumes for **Kindle**, **Nexus**

Weak/Average/**Strong**





Local Competition strength

Modeling power of CompCube

Products

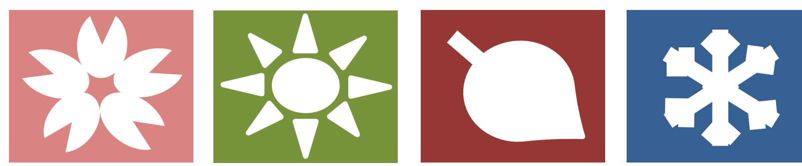
News sources

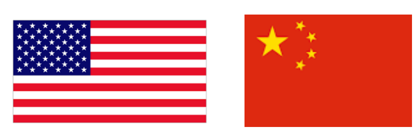
Modeling power of CompCube

Products

News sources

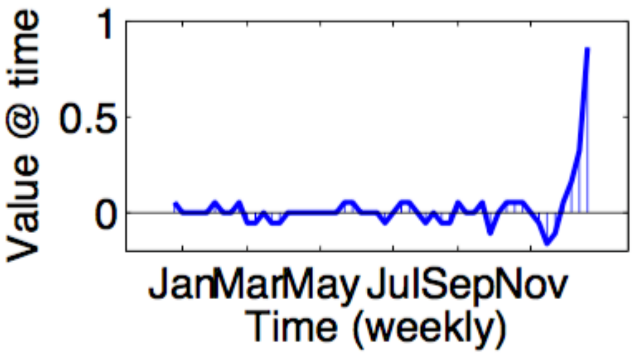
Q. Any seasonality?



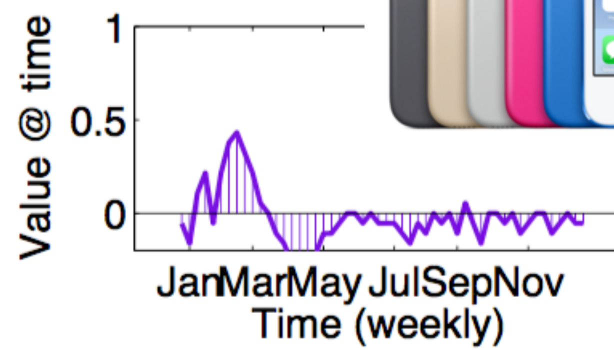
in 

Modeling power of CompCube

e.g., Local seasonality for **iPod**



Component #1

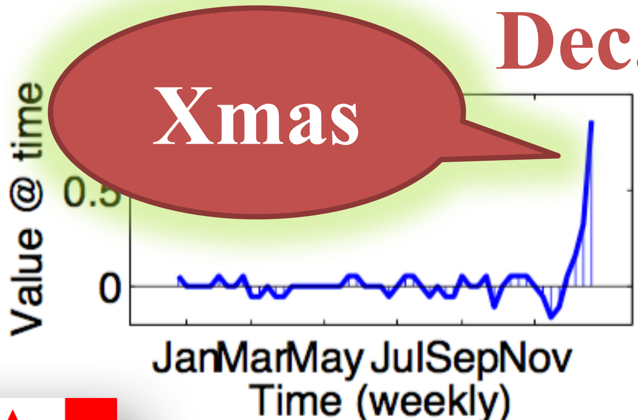


Component #2

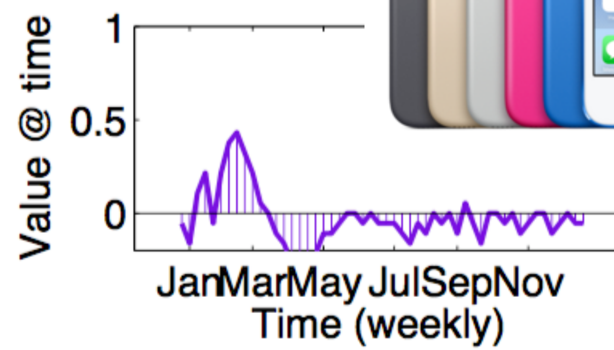
Modeling power of CompCube

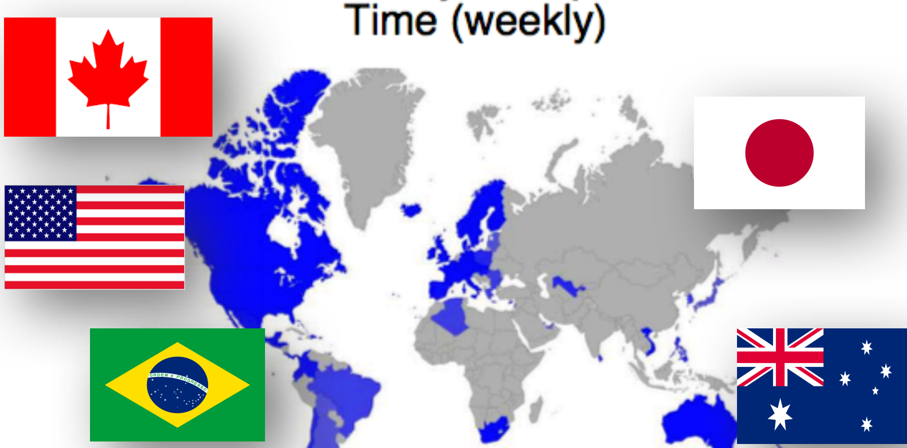
e.g., Local seasonality for

Dec.



iPod



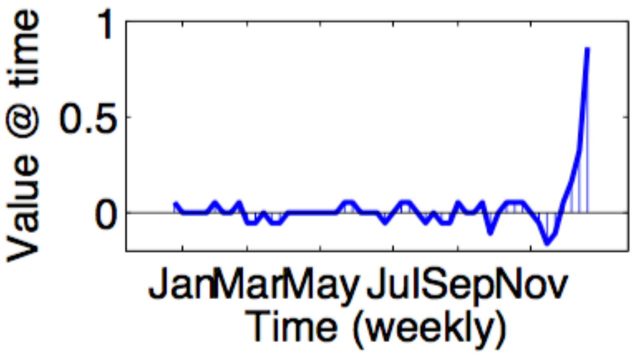


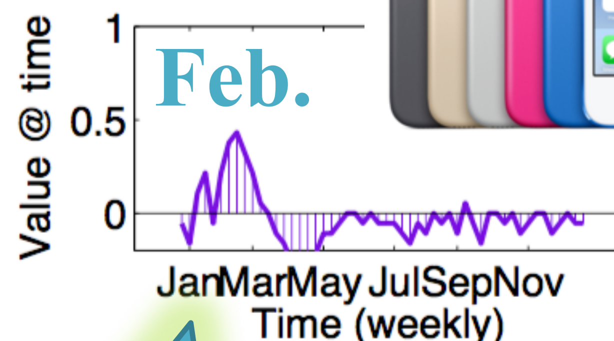
Component #1

Component #2

Modeling power of CompCube

e.g., Local seasonality for iPod





Component #1 **Component #2**

Modeling power of CompCube

Products

News sources

Modeling power of CompCube

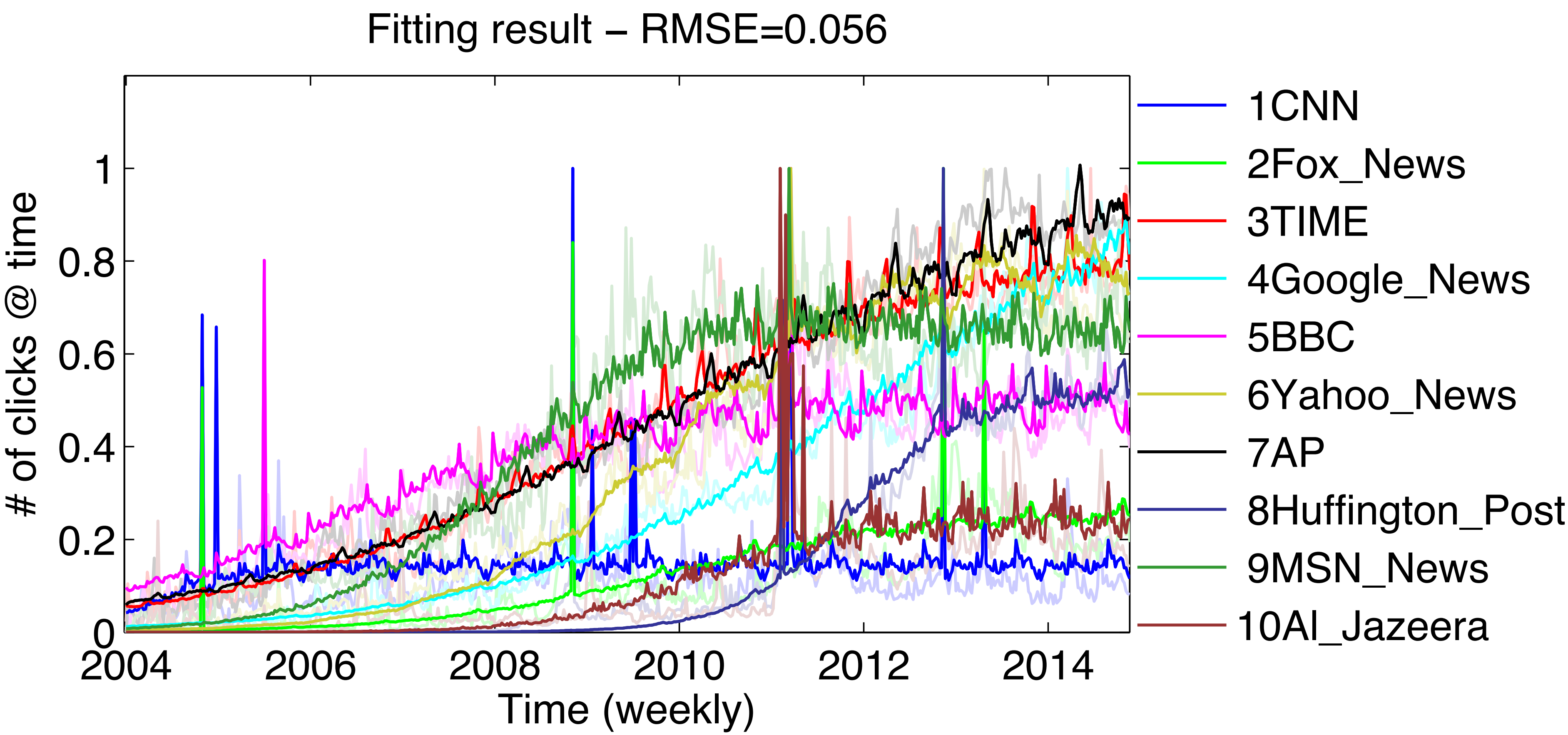
Products

News sources

Q. Any world-wide events?

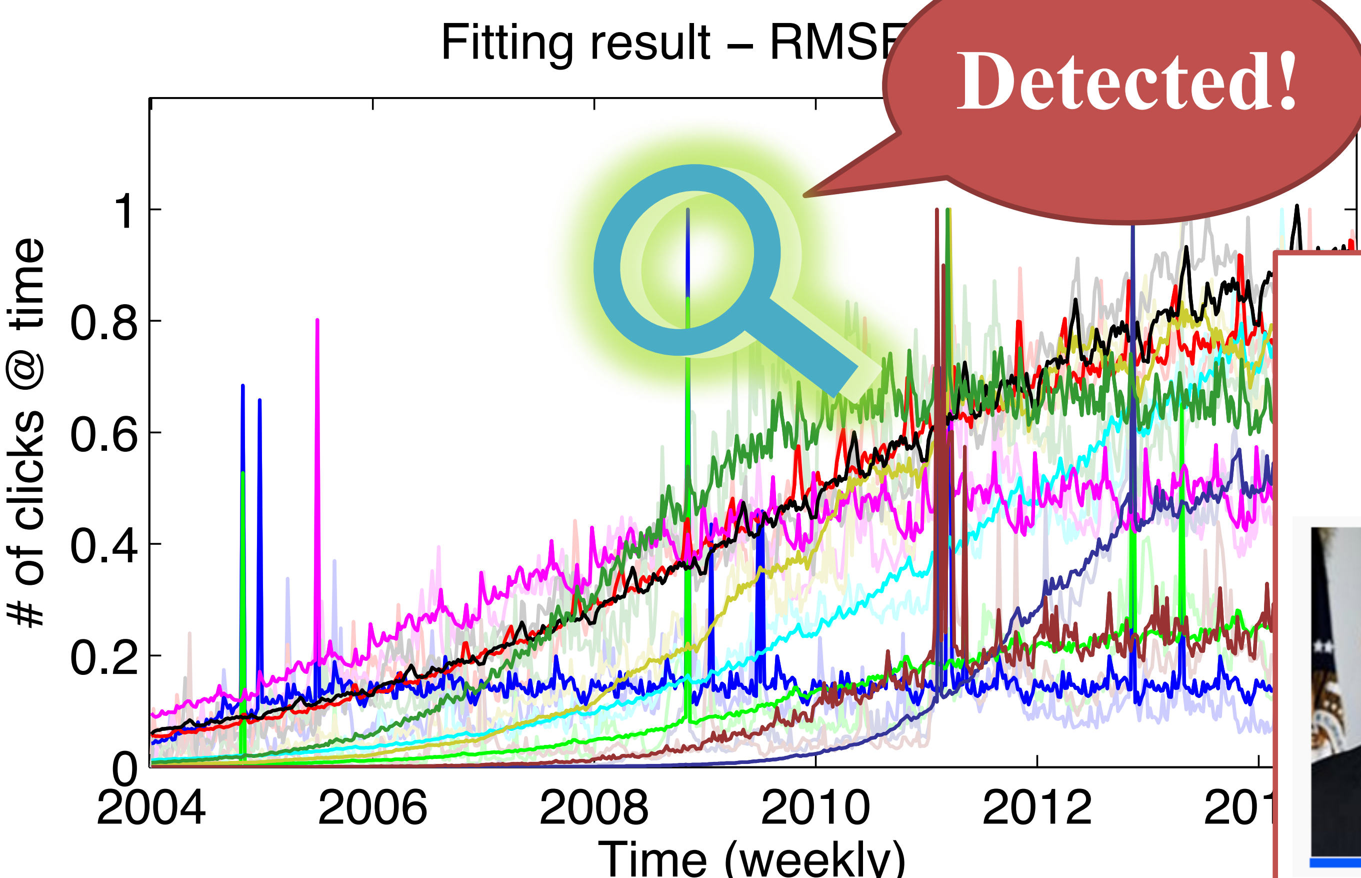
Modeling power of CompCube

Fitting result for **News resources**



Modeling power of CompCube

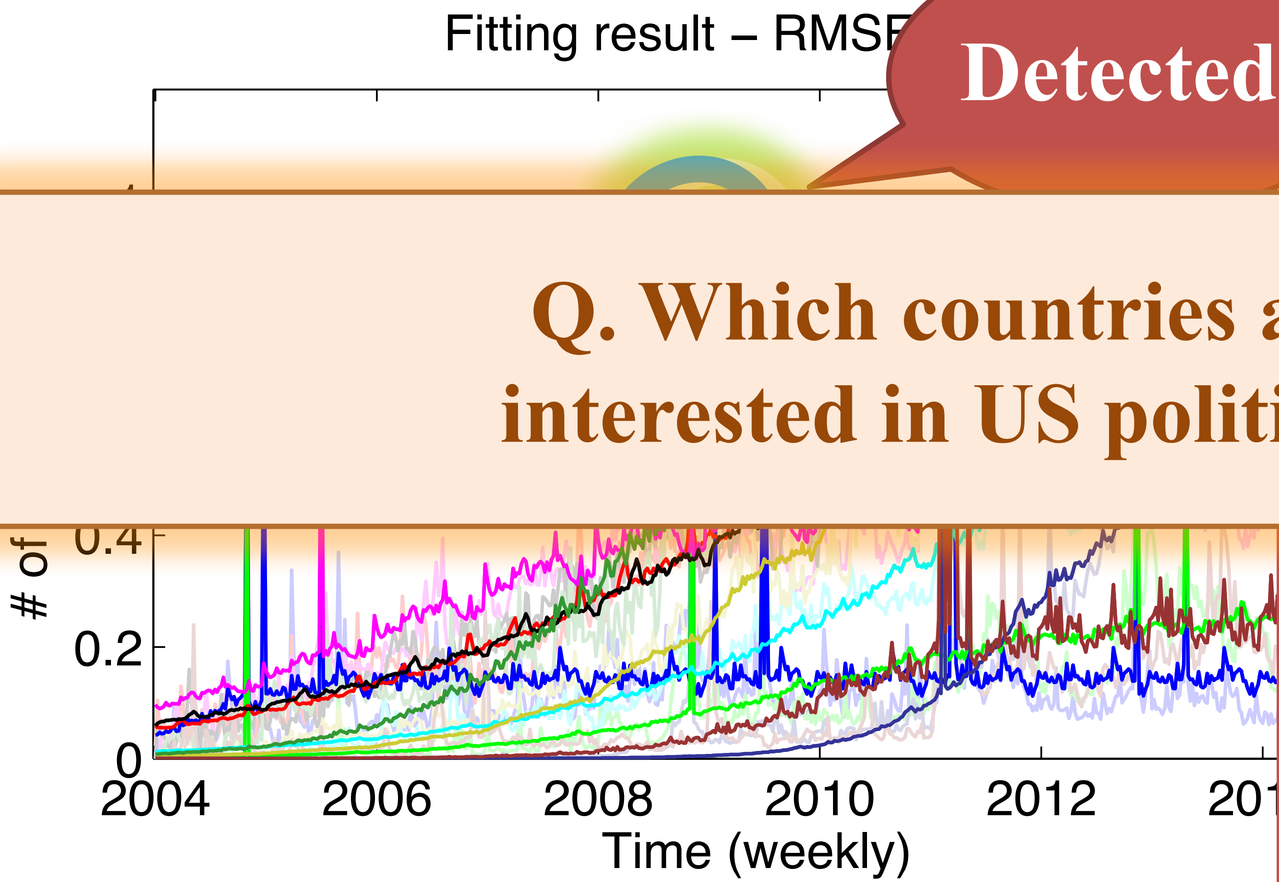
Fitting result for News resources



US election Nov. 2008

Modeling power of CompCube

Fitting result for **News resources**



Q. Which countries are interested in US politics?


Wikipedia

Modeling power of CompCube

Fitting result for **News resources**

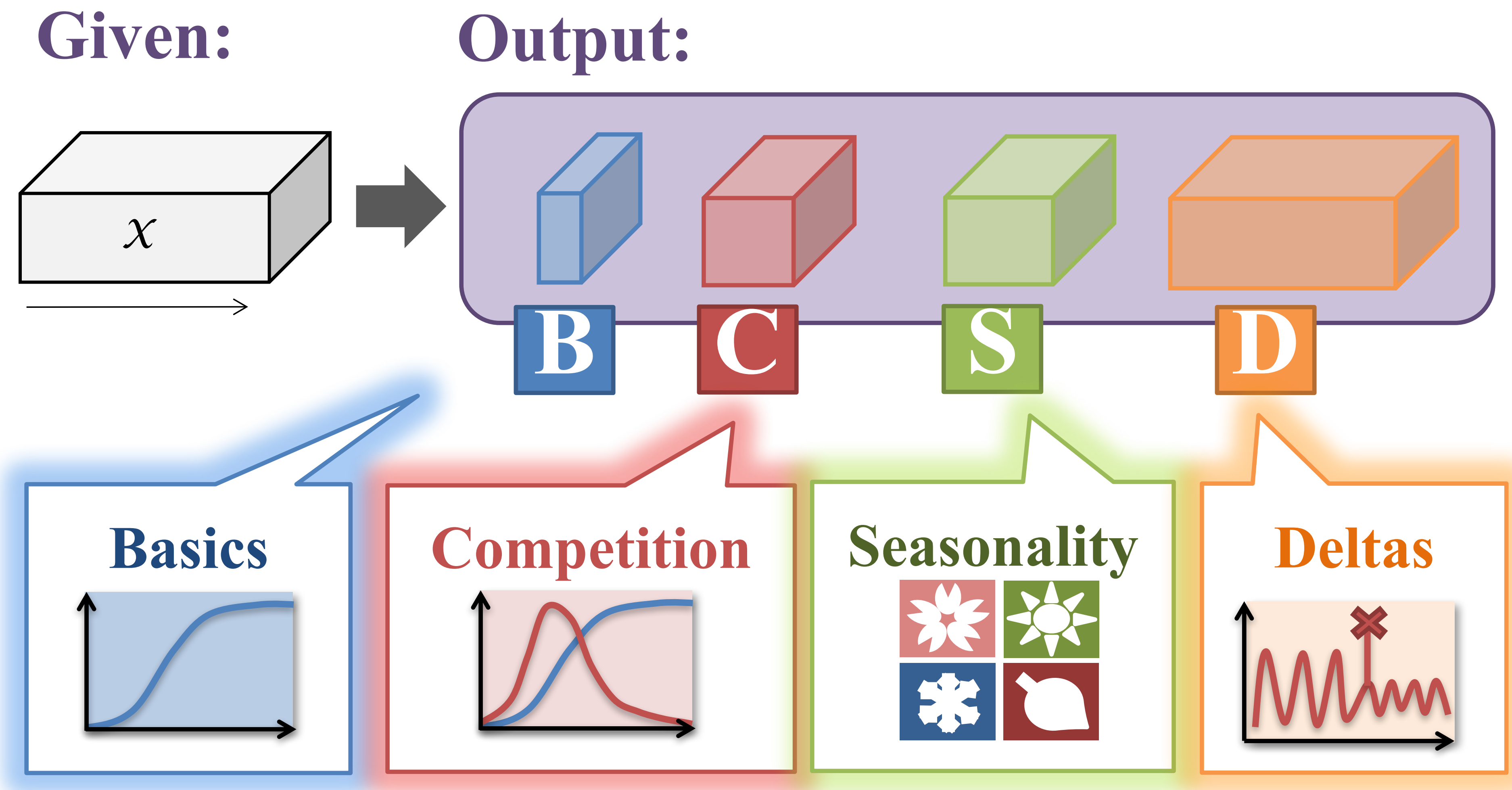
Weak/Strong

Local attention to US election

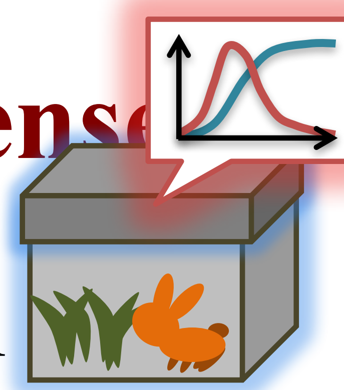
US election Nov. 2008

Wikipedia

CompCube-dense



CompCube-dens



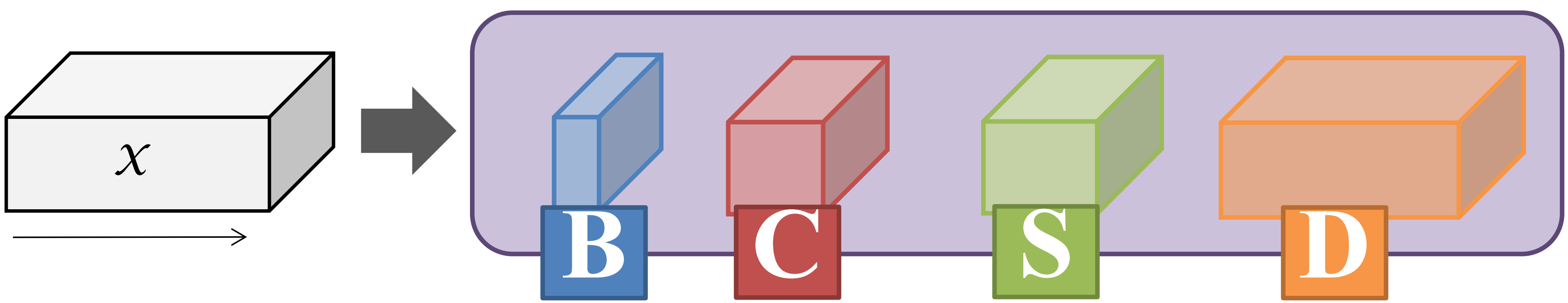
Details

Non-linear dynamical system

$$P_{il}(t) = P_{il}(t-1)\left[1 + r_{il}\left(1 - \frac{\sum_{j=1}^{d} c_{ijl}\cdot P_{jl}(t-1)}{K_{il}}\right)\right]$$

$$V_{il}(t) = P_{il}(t)\left[1 + s_{il}(t \mod n_p)\right] + \delta_{il}(t)$$

$$(i = 1, \cdots, d;\ l = 1, \cdots, m;\ t = 1, \cdots, n) \quad P_{il}(0) = p_{il}$$

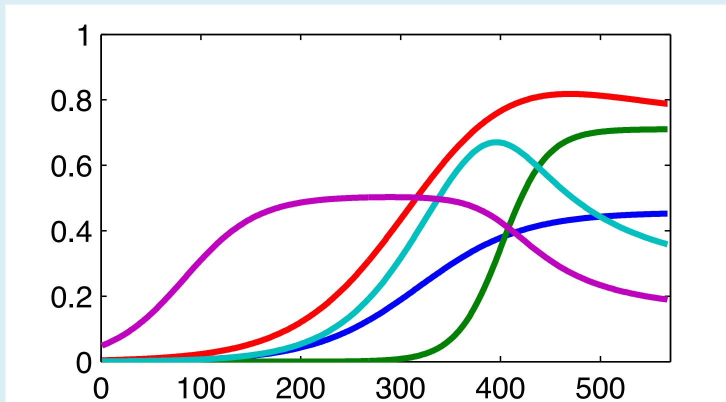


CompCube-dens

Details

Non-linear

P: latent popularity

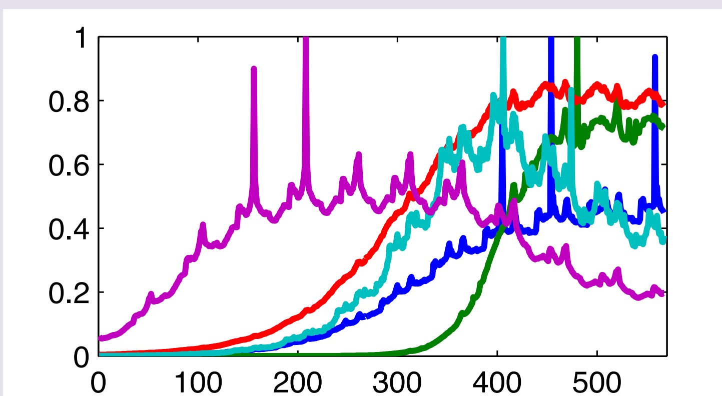


$P_{il}(t) = P_{il}(t$ … $_{jl} \cdot P_{jl}(t-1)$ … K_{il}

$V_{il}(t) = P_{il}(t)$ … (t)

$(i =$ … $= 1, \cdots, n)\ P_{il}(0) = p_{il}$

V: estimated volume

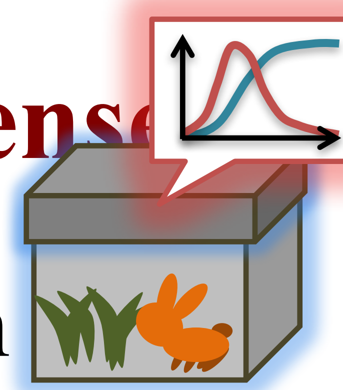


X

S

D

CompCube-dense



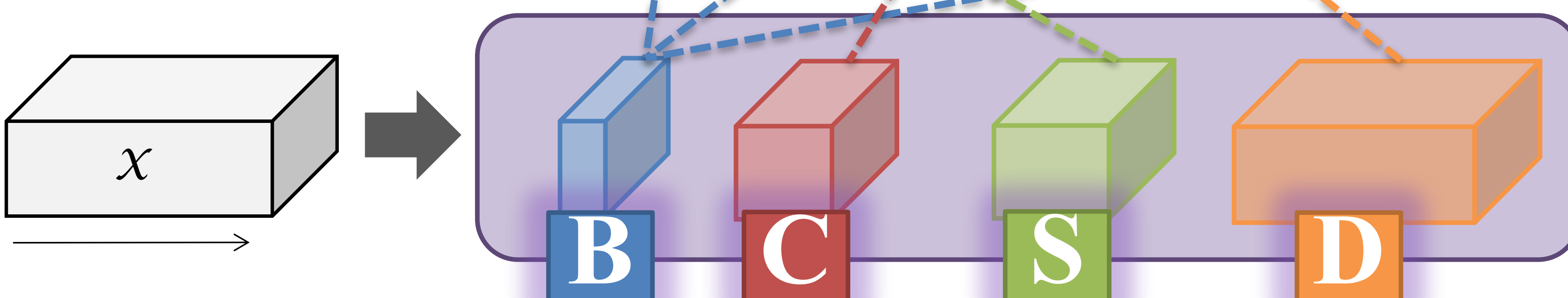
Details

Non-linear dynamical system

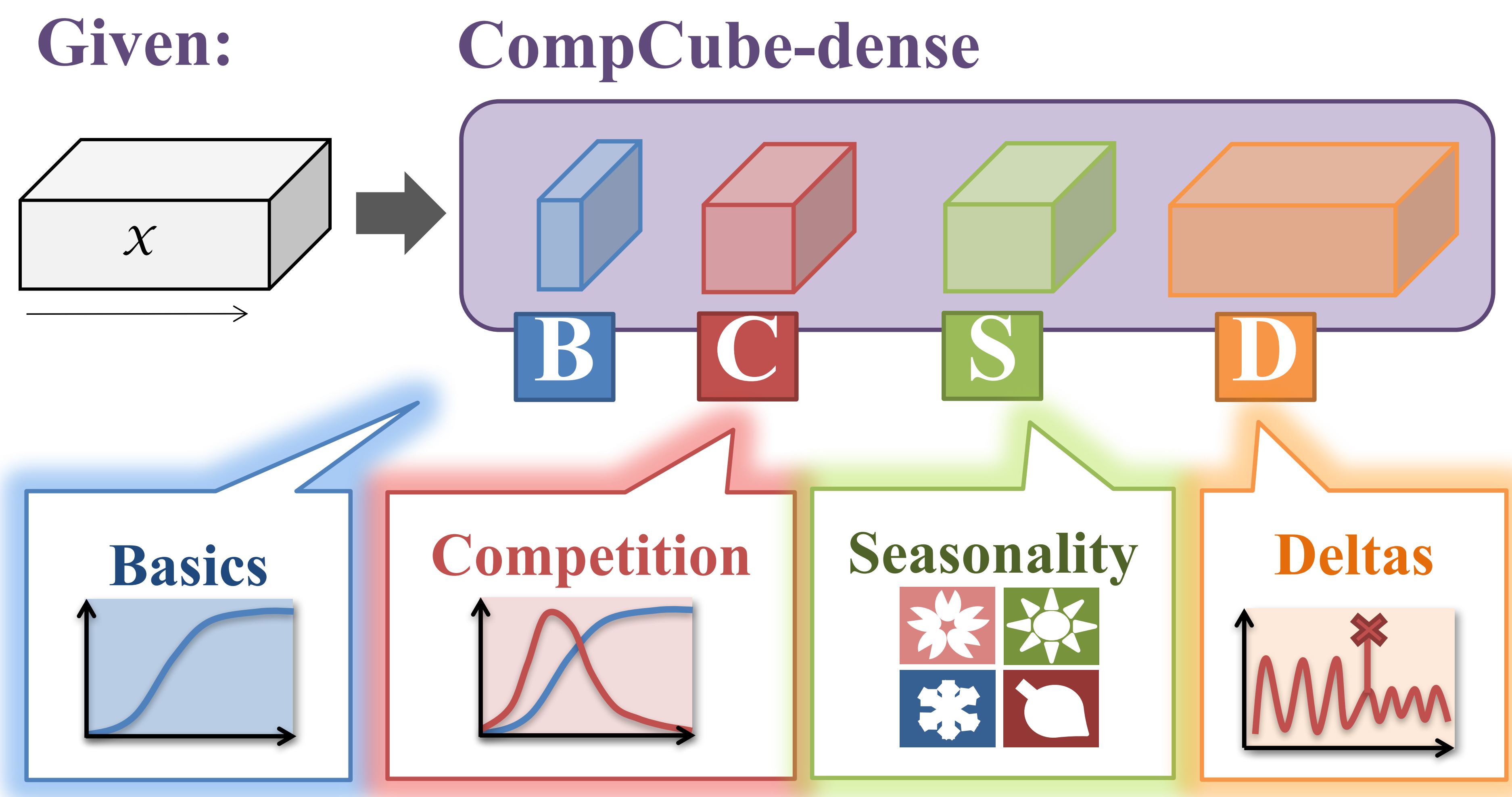
$$P_{il}(t) = P_{il}(t-1)\left[1 + \boxed{r_{il}}\left(1 - \frac{\sum_{j=1}^{d} \boxed{c_{ijl}} \cdot P_{jl}(t-1)}{\boxed{K_{il}}}\right)\right]$$

$$V_{il}(t) = P_{il}(t)\left[1 + \boxed{s_{il}}(t \mod n_p)\right] + \boxed{\delta_{il}(t)}$$

$$(i = 1, \cdots, d;\ l = 1, \cdots, m;\ t = 1, \cdots, n)\quad P_{il}(0) = \boxed{p_{il}}$$

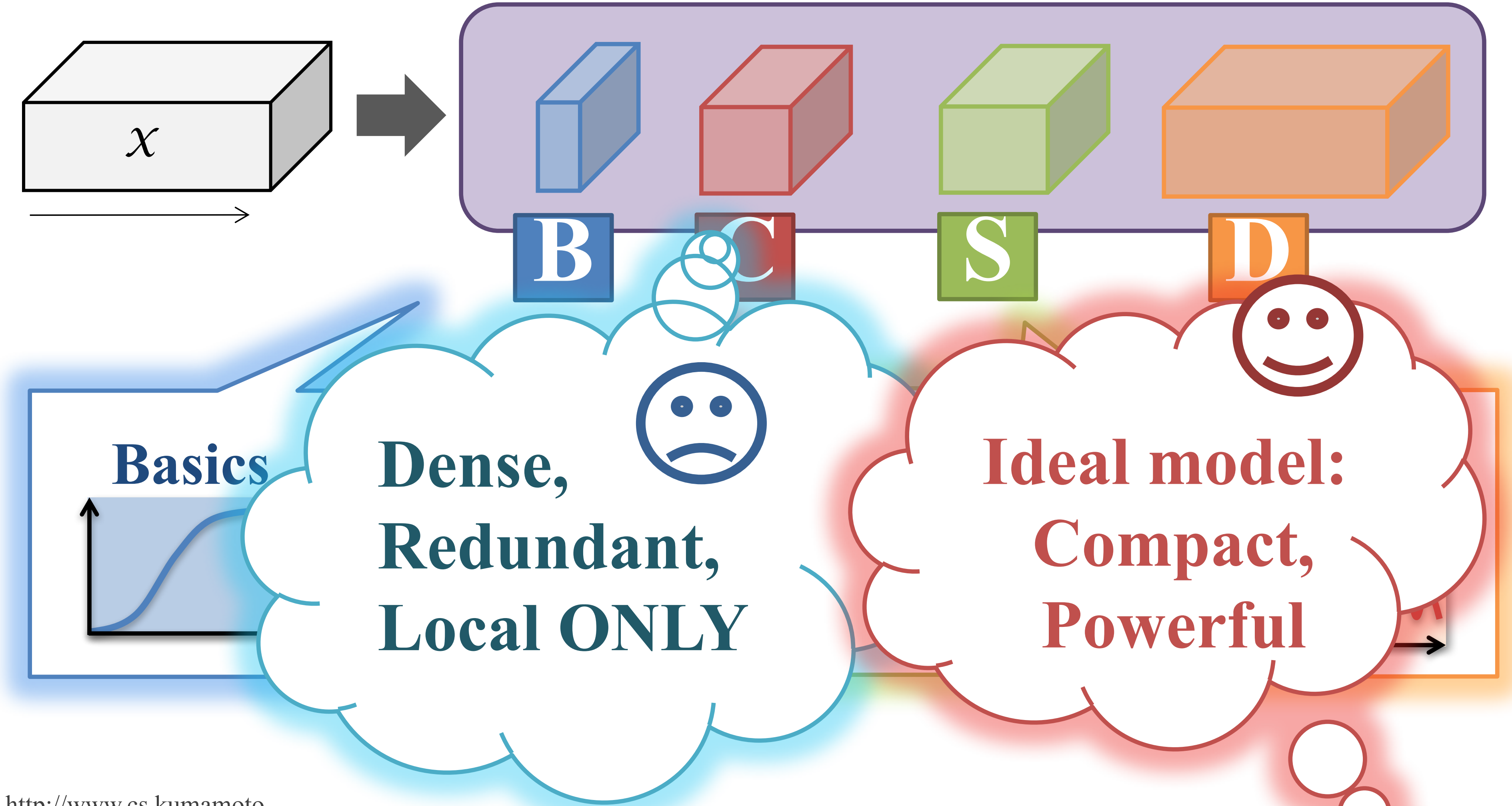


Initial attempt: CompCube-dense

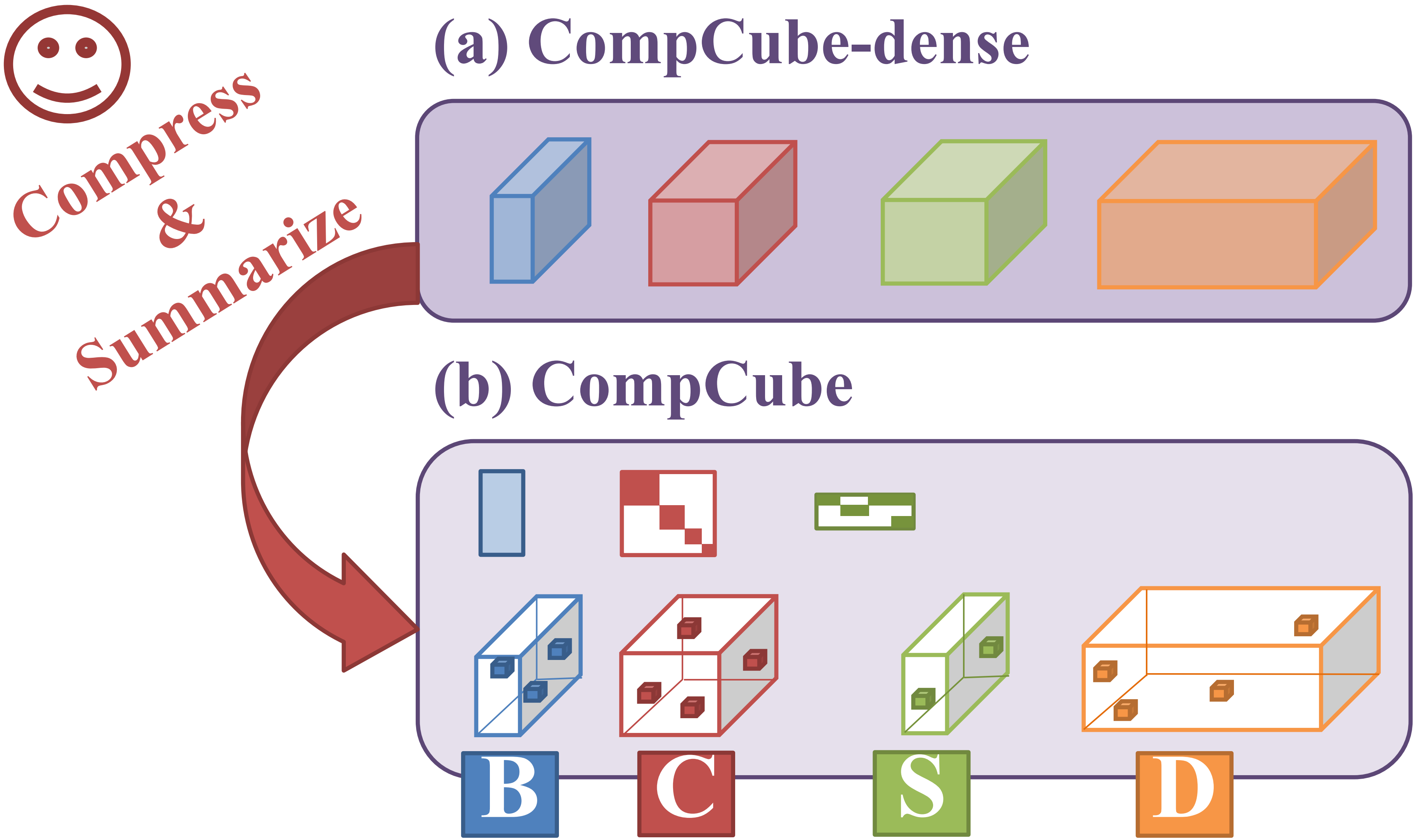


Initial attempt: CompCube-dense

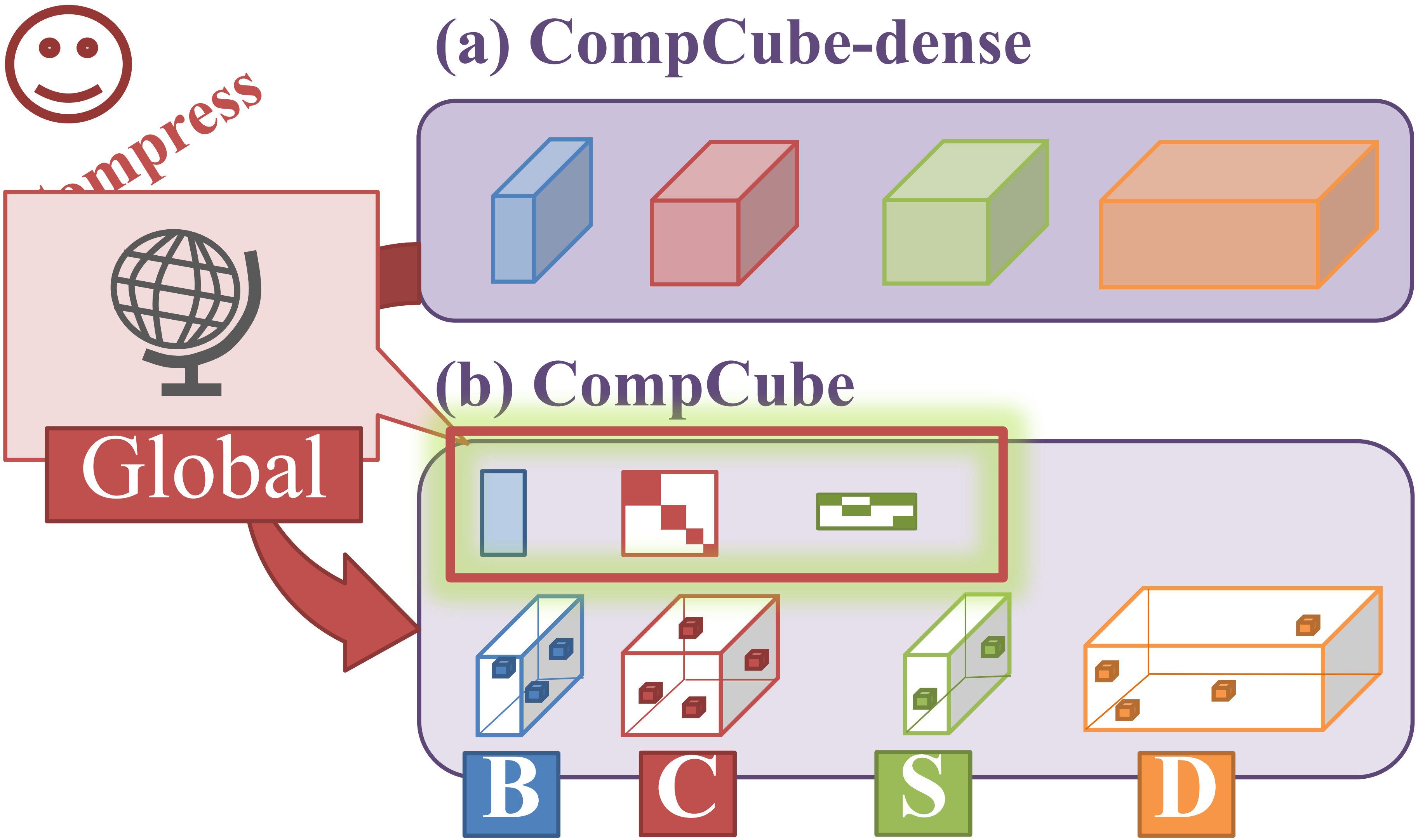
Given: **CompCube-dense**



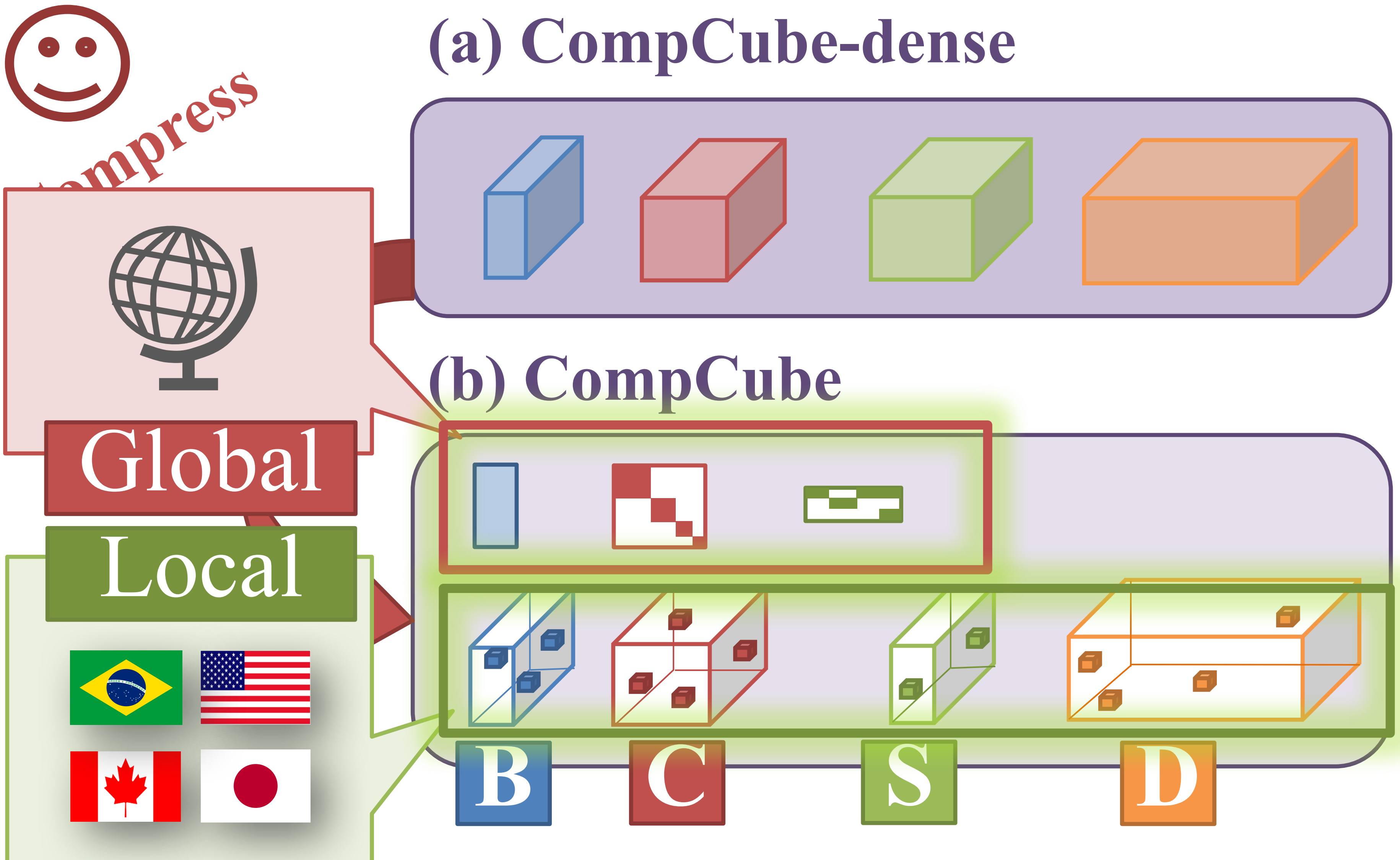
Final model: CompCube



Final model: CompCube

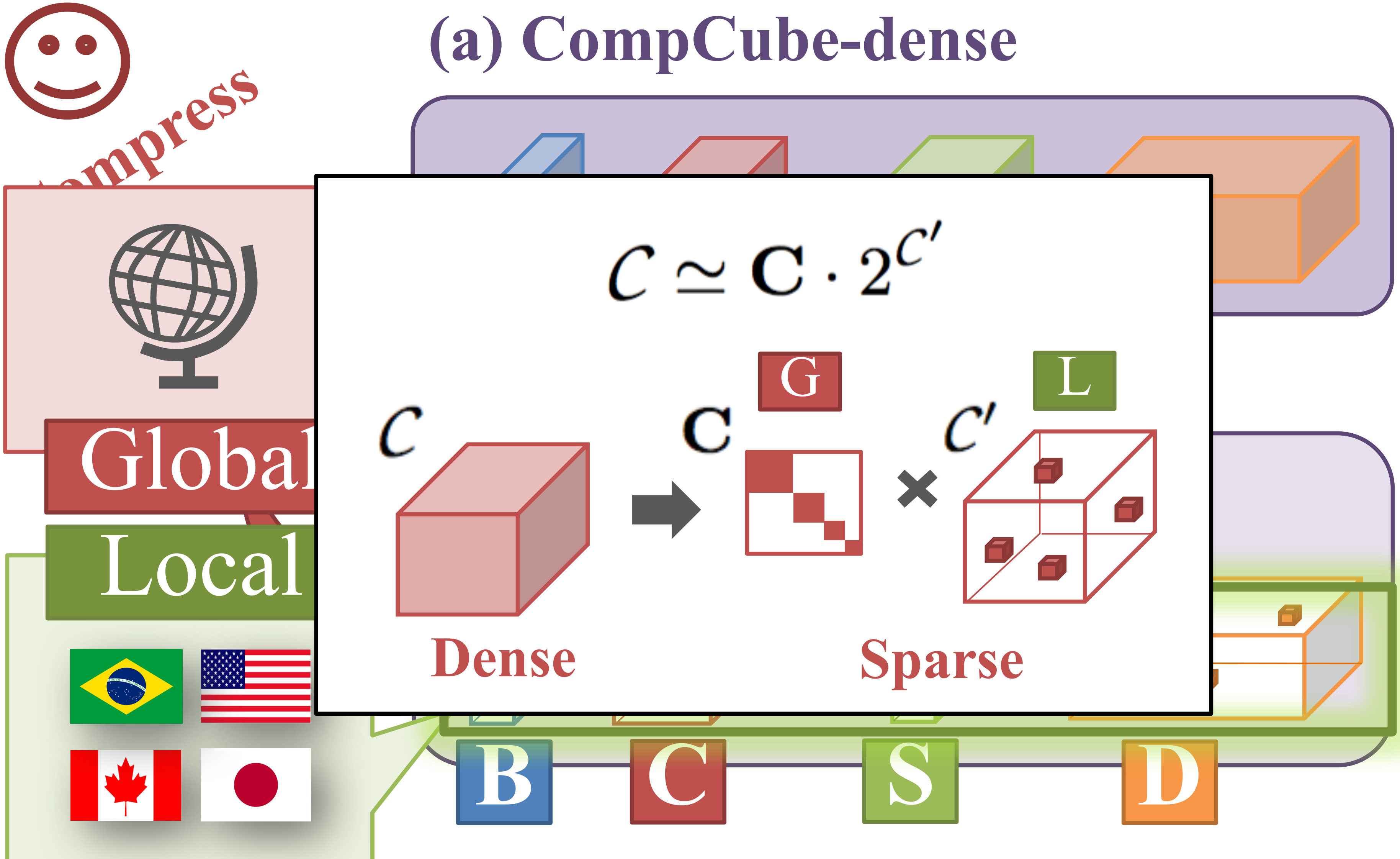


Final model: CompCube



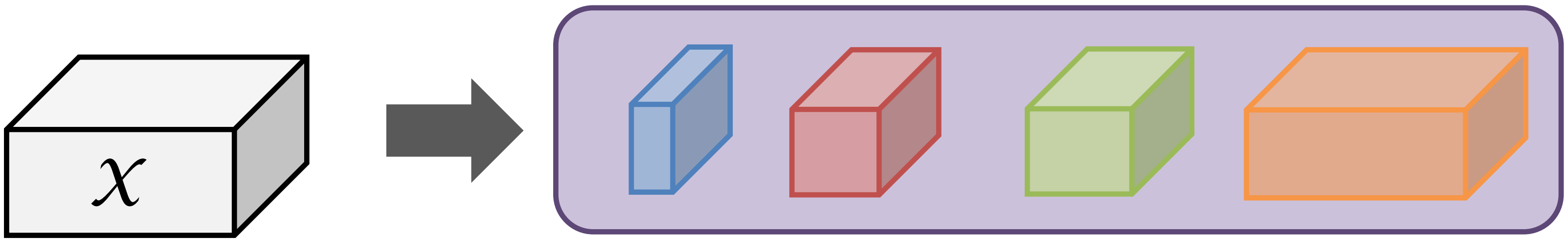
Final model: CompCube

(a) CompCube-dense

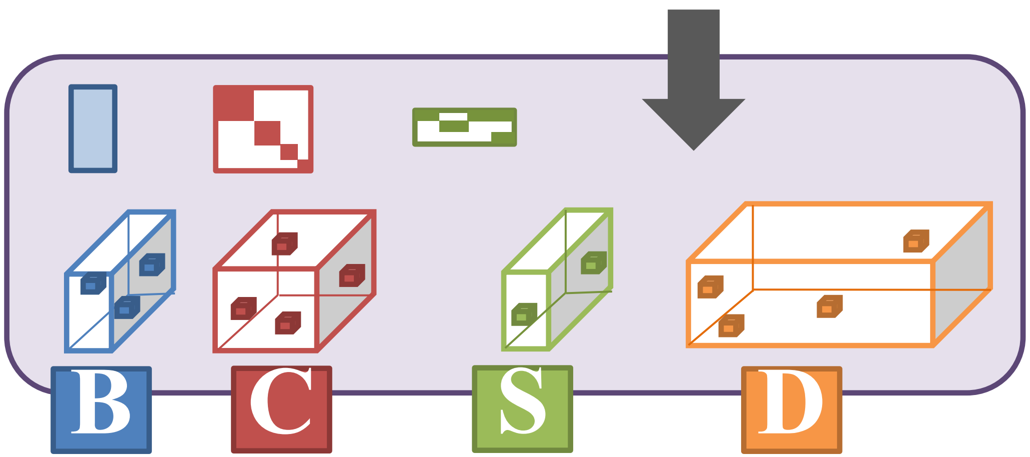


Algorithms

Q1. How can we efficiently estimate parameters?



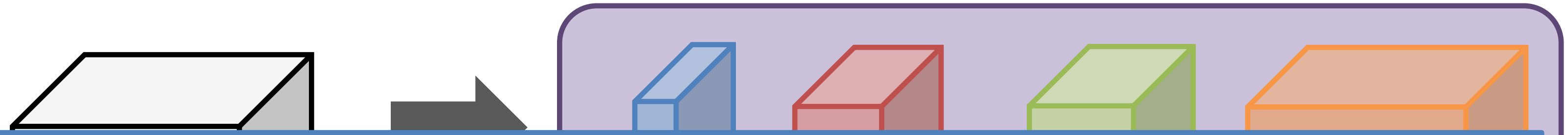
Q2. How can we automatically find best parameter sets?



Algorithms

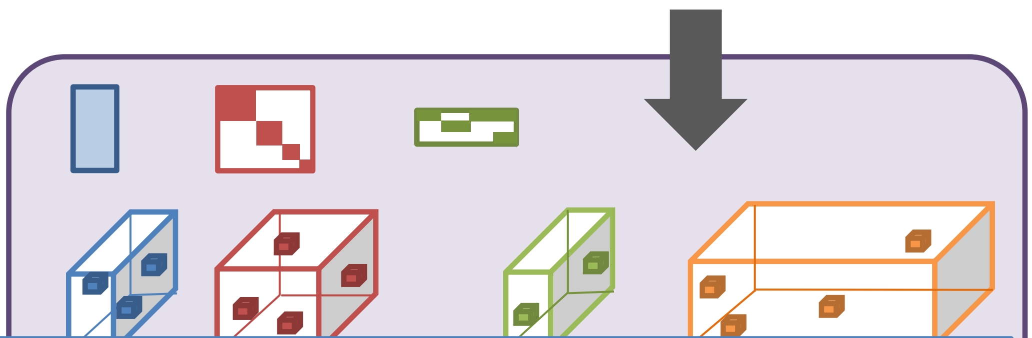
(Details in paper)

Q1. How can we efficiently estimate parameters?



Idea (1) : TetraFit algorithm

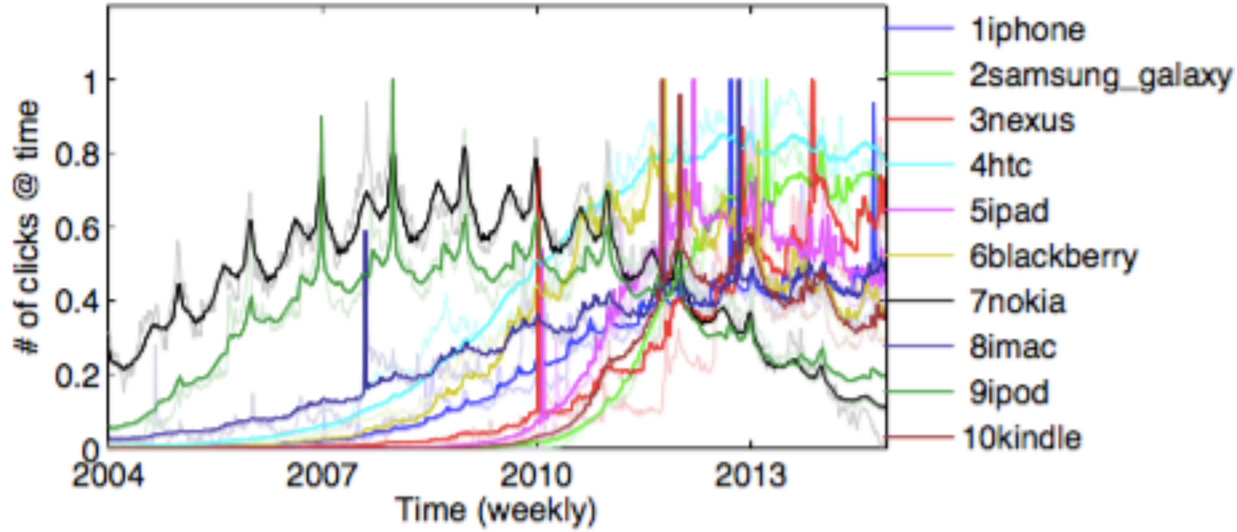
Q2. How can we automatically find best parameter sets?



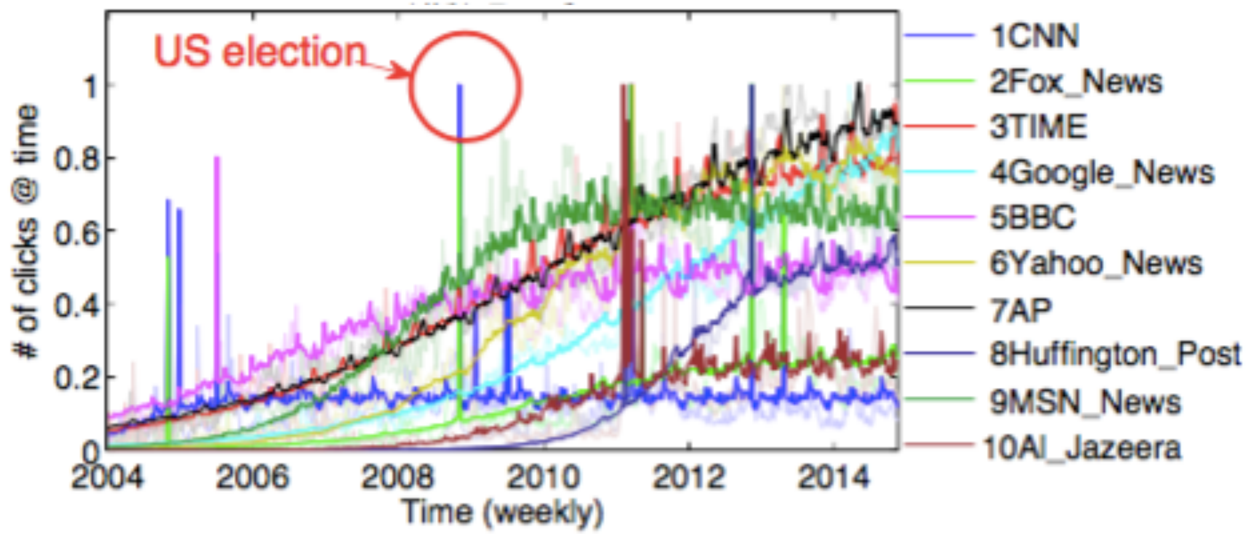
Idea (2): Model description cost

Effectiveness

1. Products



2. News



Effectiveness

1. Products

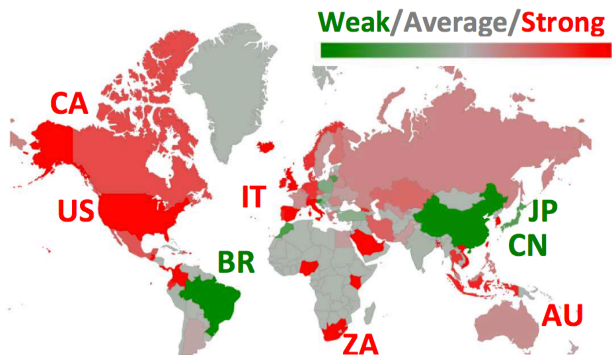
2. News

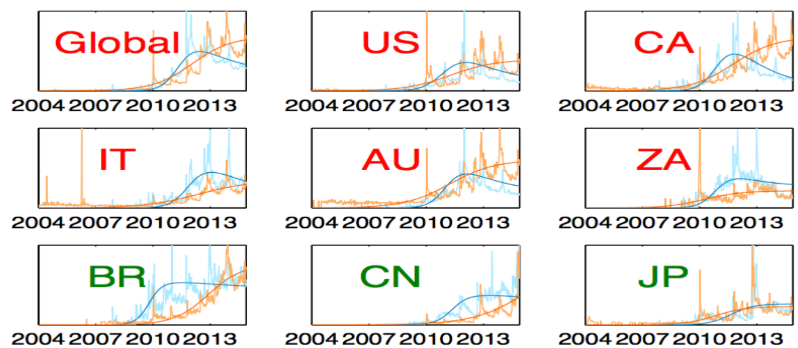
Microsoft

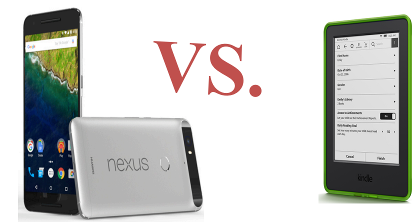
VS.

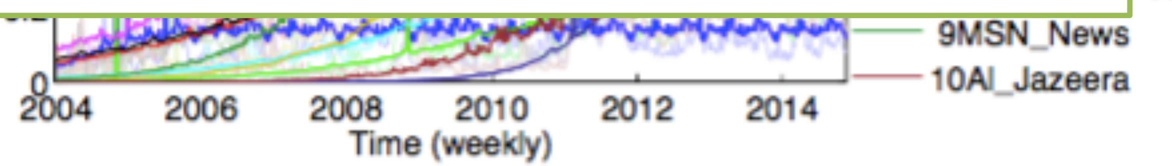
YAHOO! NEWS

Local competition









Effectiveness

1. Products

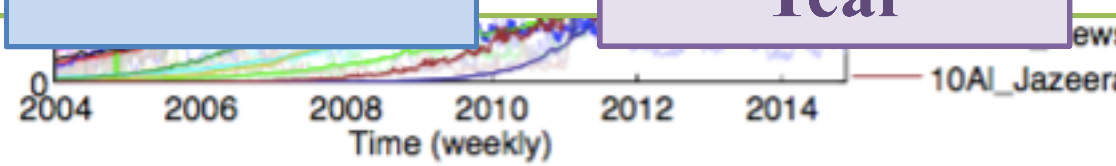
VS.

2. News

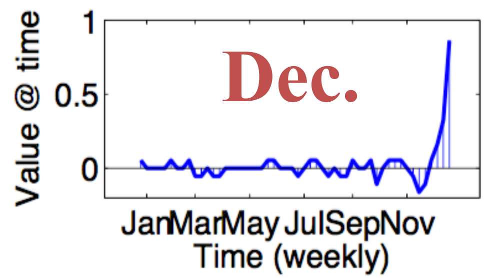
Microsoft

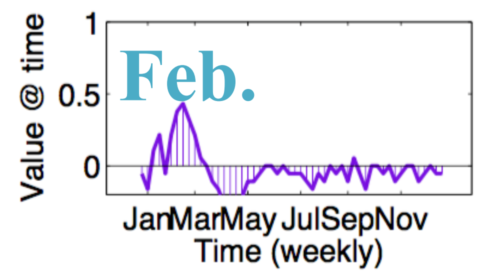
VS.

YAHOO! NEWS



Local seasonality





Christmas

Chinese New Year

Effectiveness

1. Products

VS.

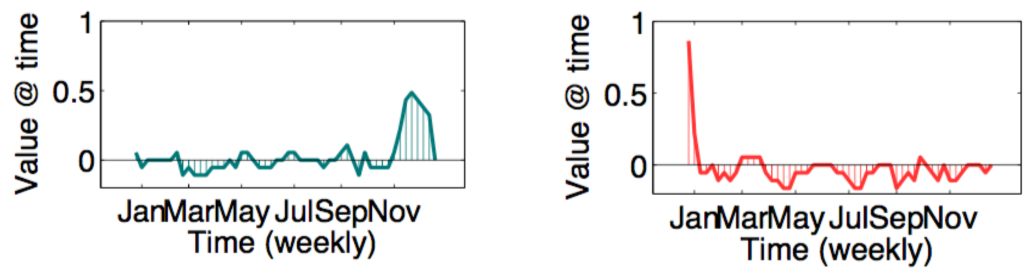
2. News

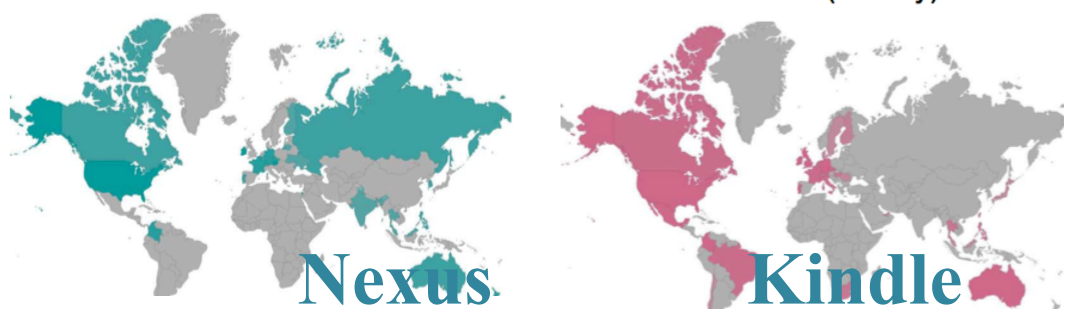
Microsoft

VS.

YAHOO! NEWS

Local seasonality





Nexus release

New Year sale

Effectiveness

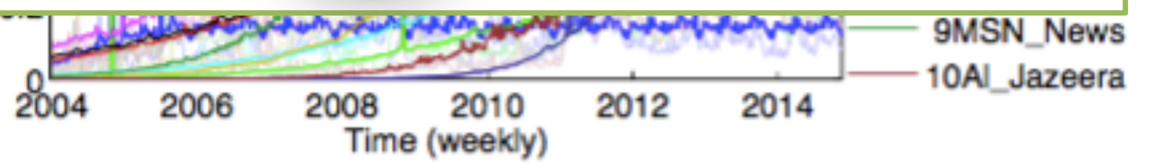
1. Products

2. News

Microsoft

VS.

YAHOO!
NEWS



Deltas

Weak/Strong

CNN

US election Nov. 2008

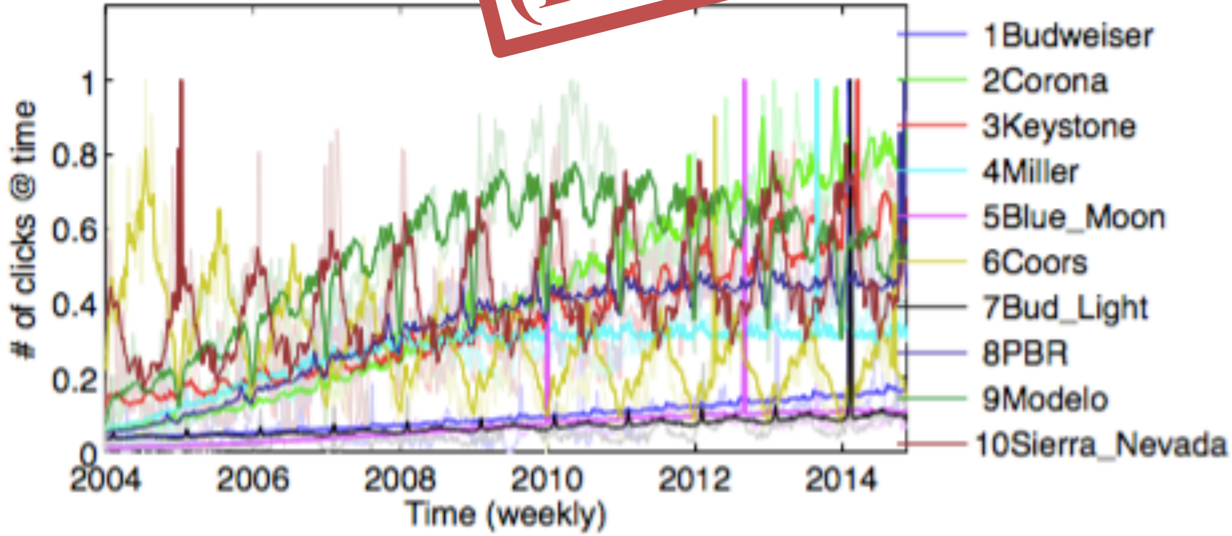
Wikipedia

Effectiveness

(Details in paper)

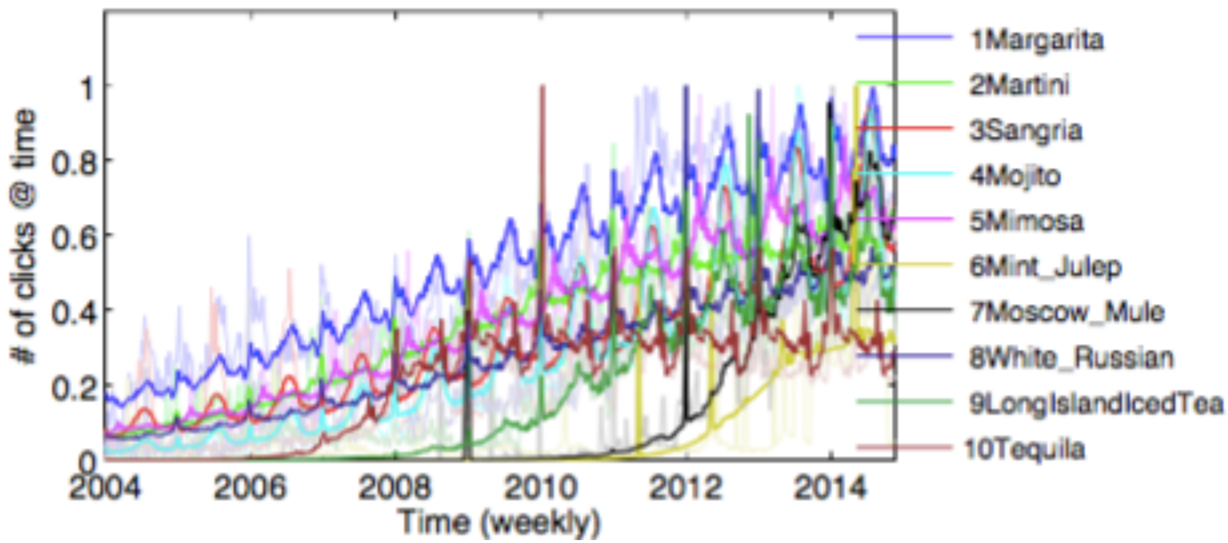
3. Beers

VS.



4. Cocktails

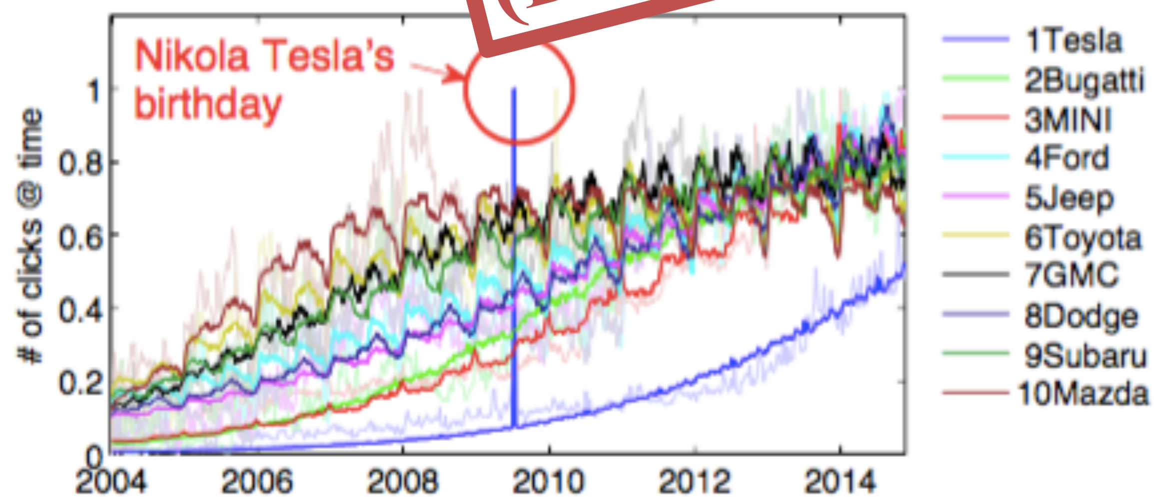
VS.



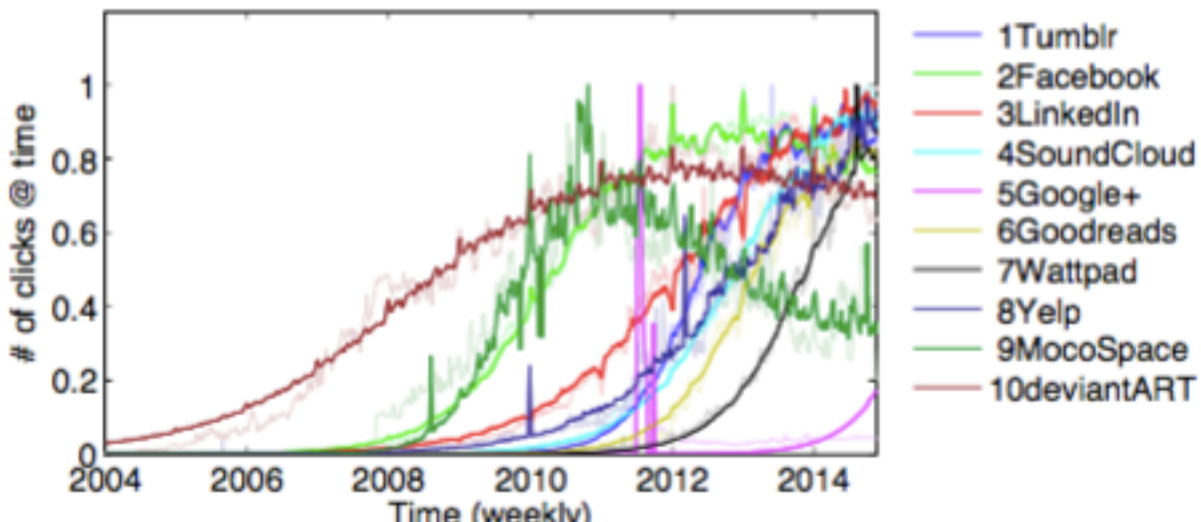
Effectiveness

(Details in paper)

5. Cars



6. SNS

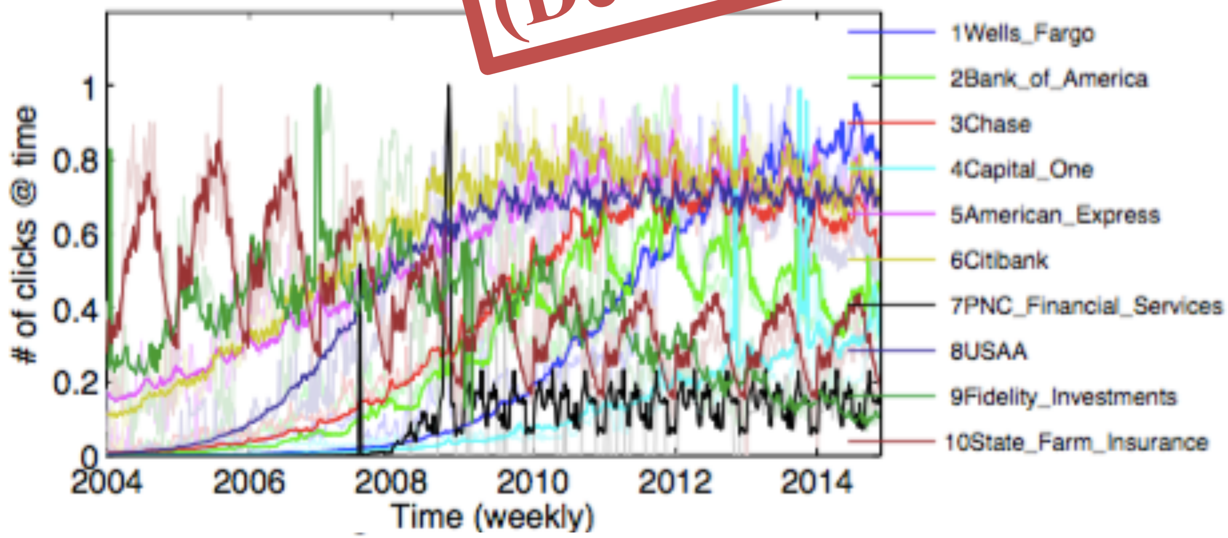


Effectiveness

(Details in paper)

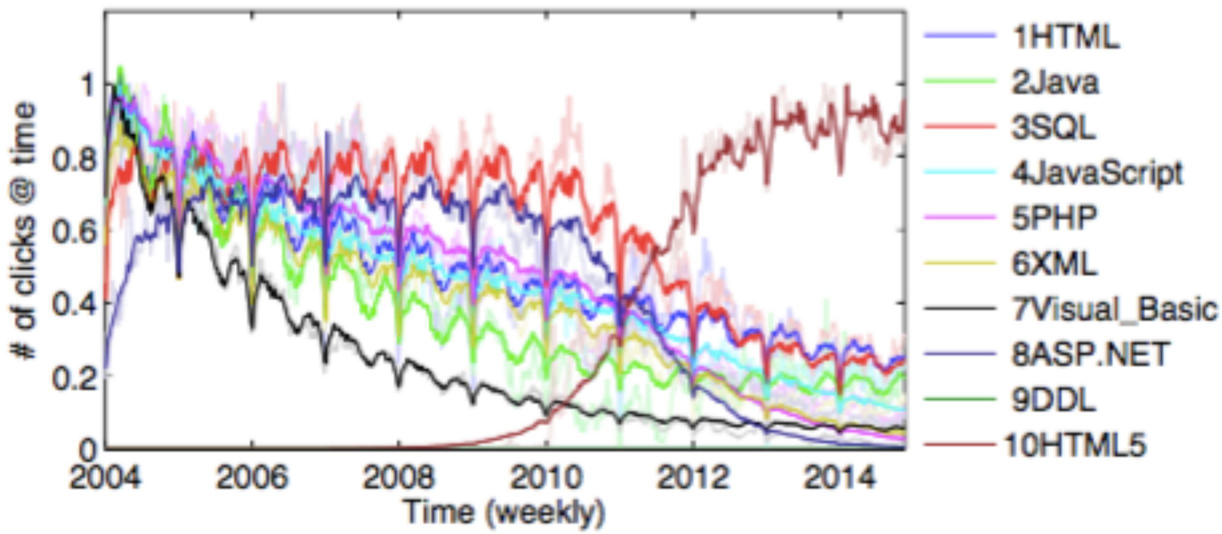
7. Finance

WELLS FARGO
citi CITIBANK



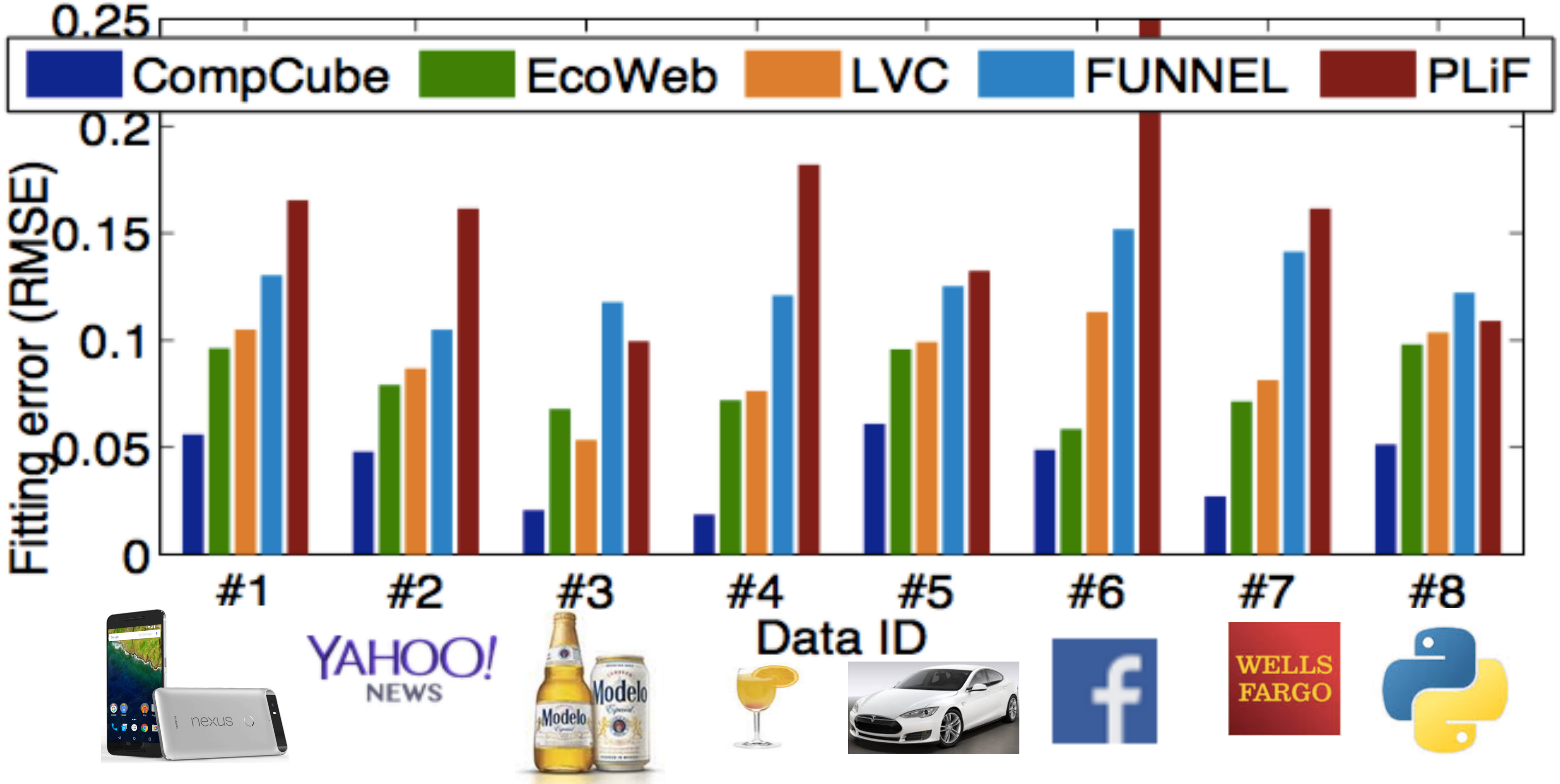
8. Software

HTML



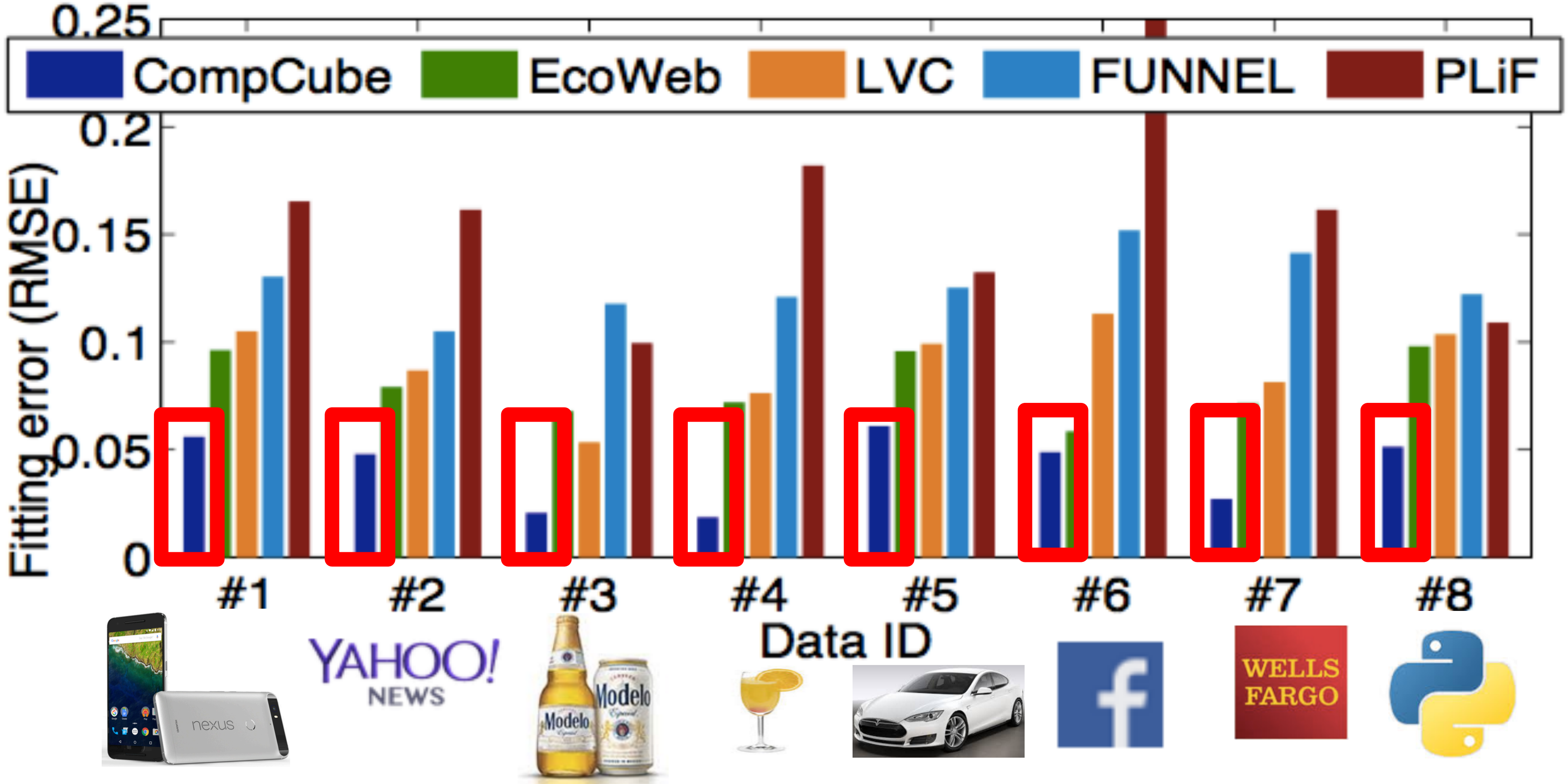
Accuracy

RMSE between original and fitted volume



Accuracy

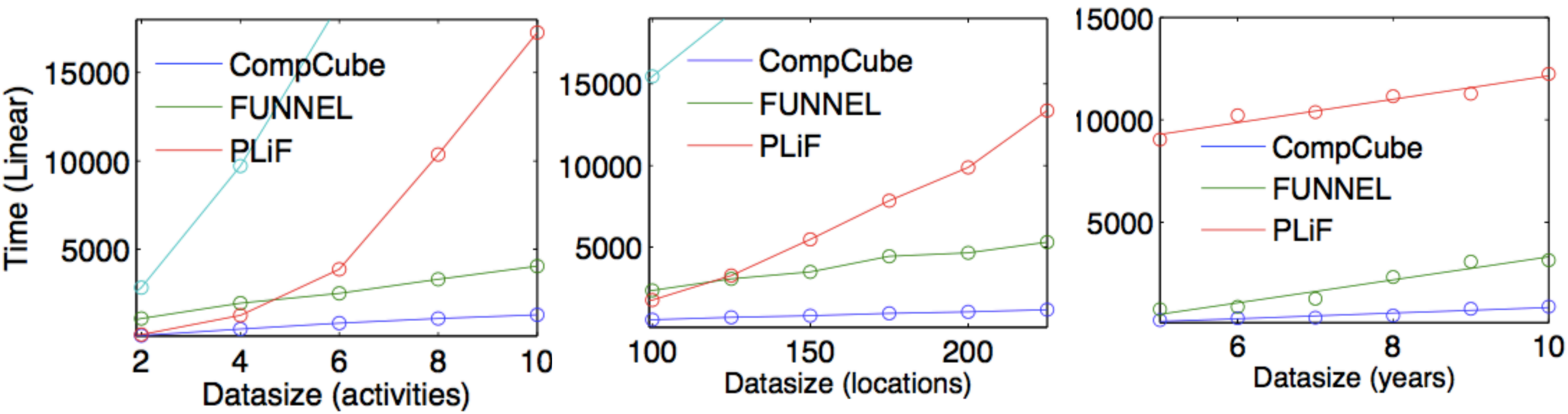
RMSE between original and fitted volume

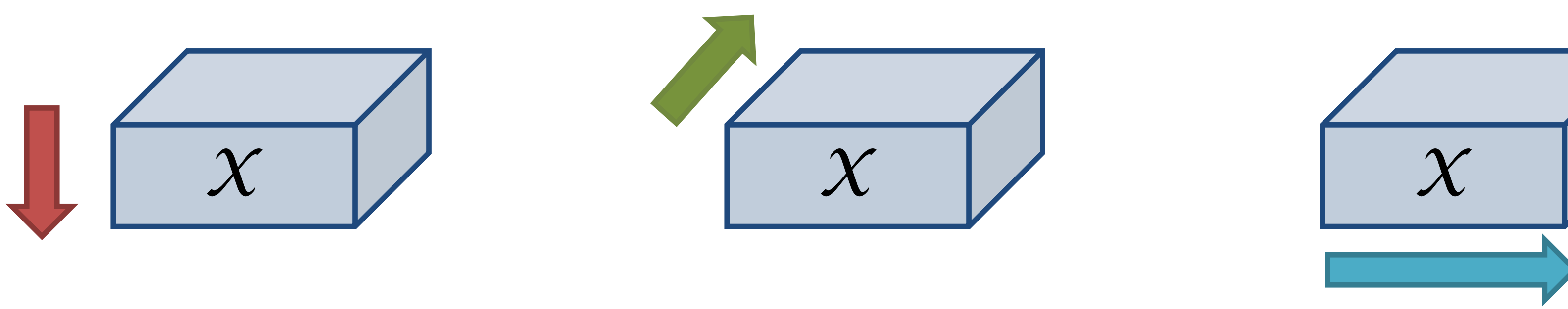


CompCube consistently wins!

Scalability

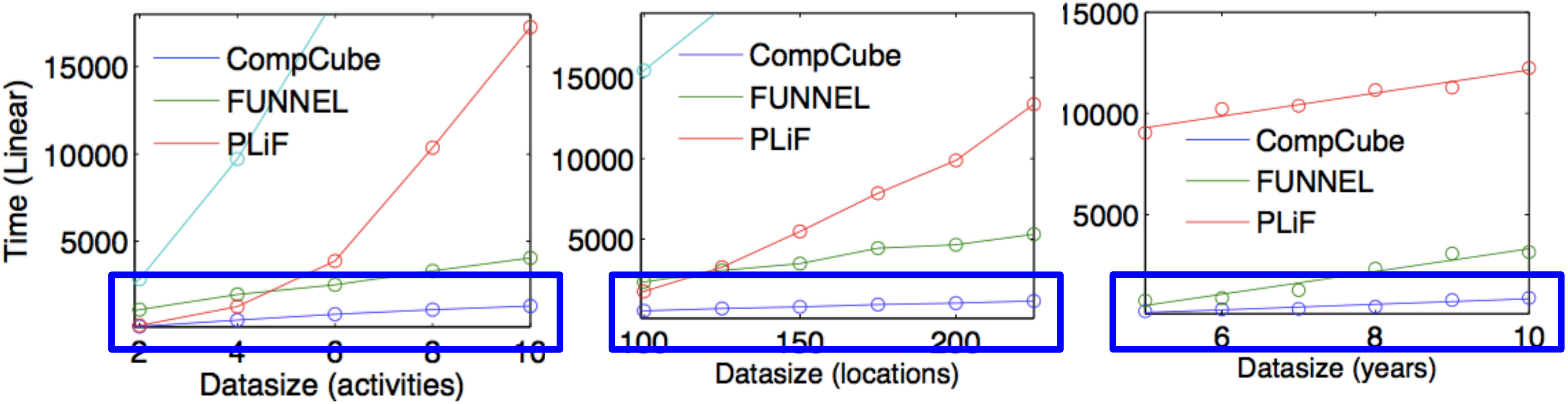
Wall clock time vs. activity , location , Time

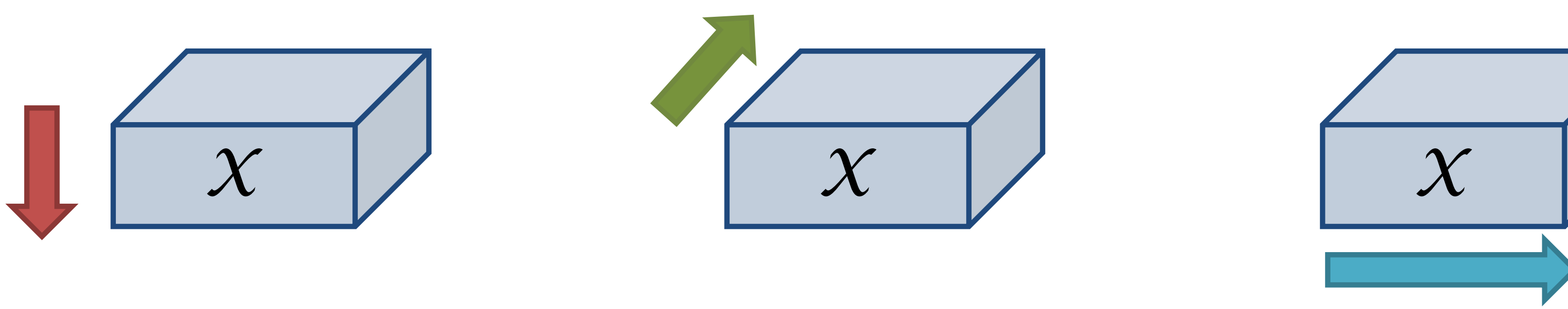




Scalability

Wall clock time vs. activity , location , Time

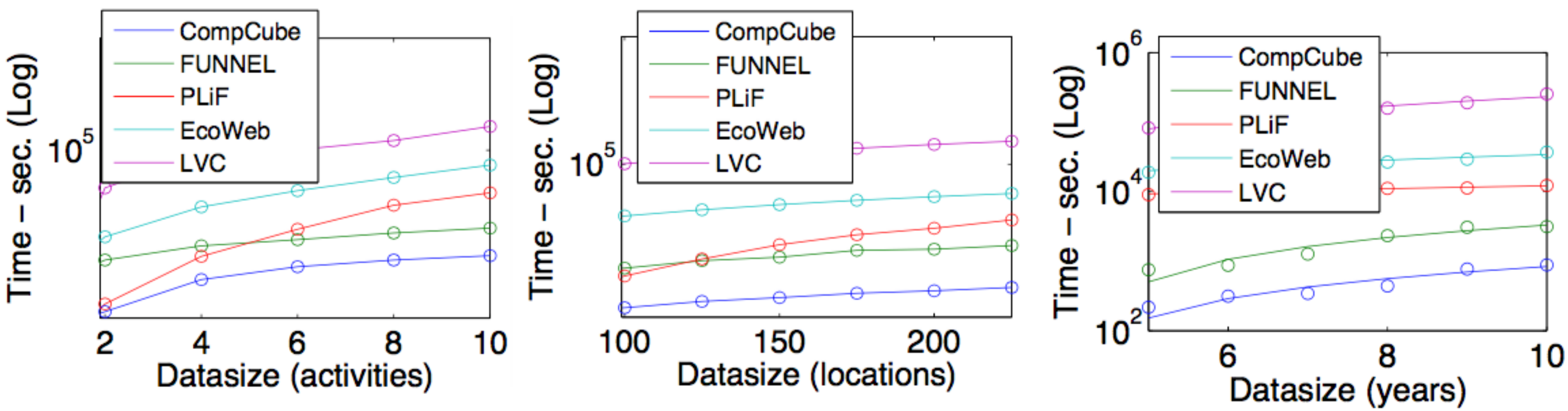


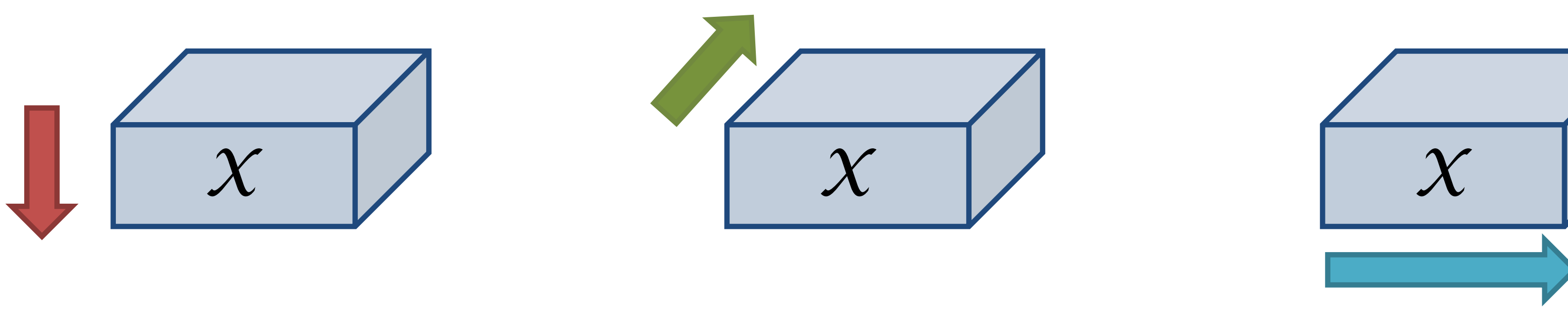


CompCube is linear w.r.t. data size : O(dmn)

Scalability

Wall clock time vs. activity , location , Time





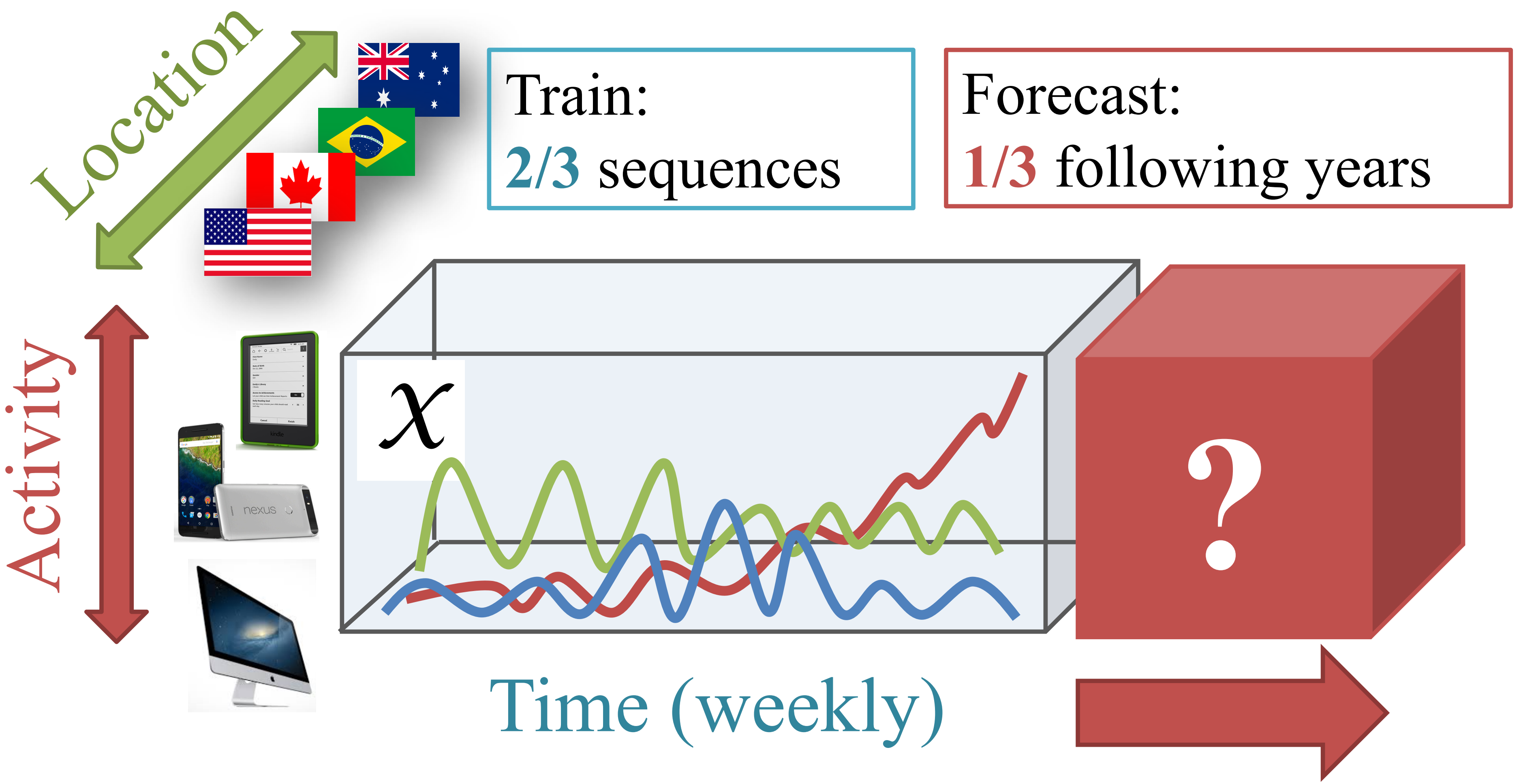
CompCube is linear w.r.t. data size : O(dmn)

CompCube at work - forecasting

Forecasting future local activities

Train:
2/3 sequences

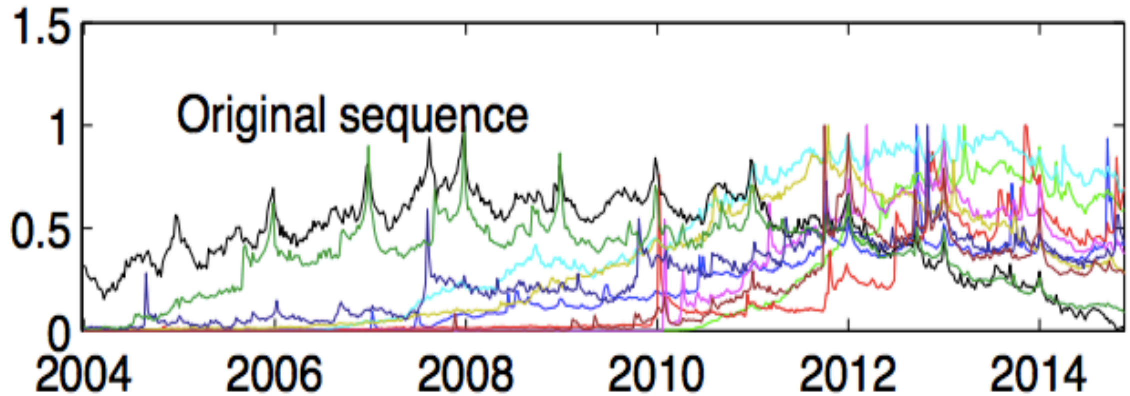
Forecast:
1/3 following years

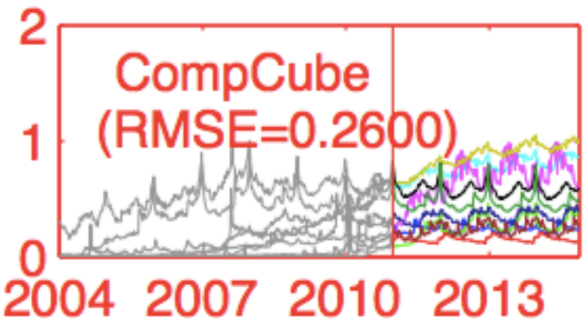


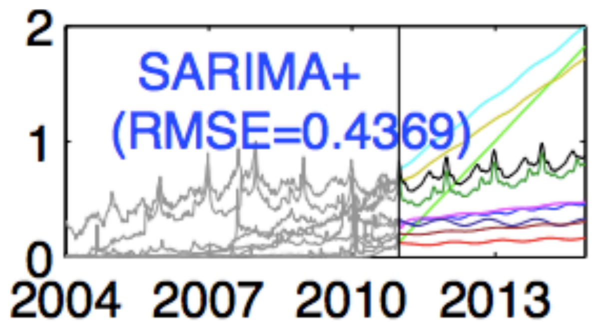
CompCube at work - forecasting

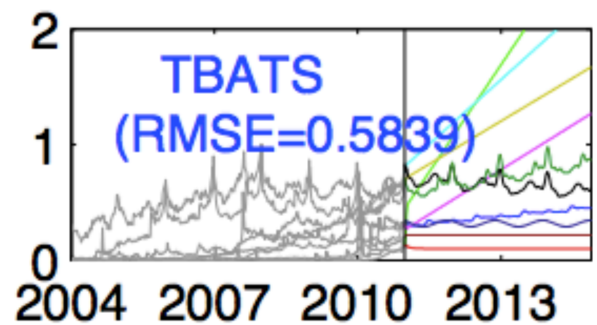
Forecasting results for #1 Products

1. Products





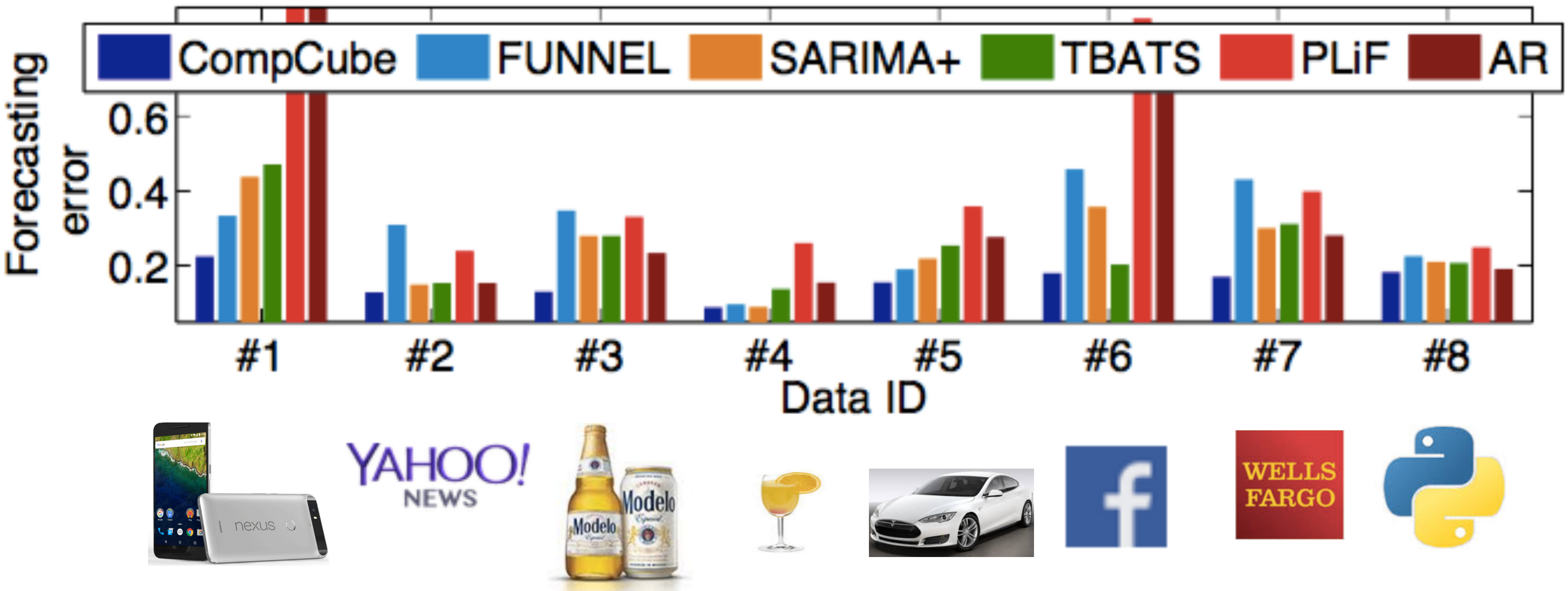




CompCube captures future activities very well

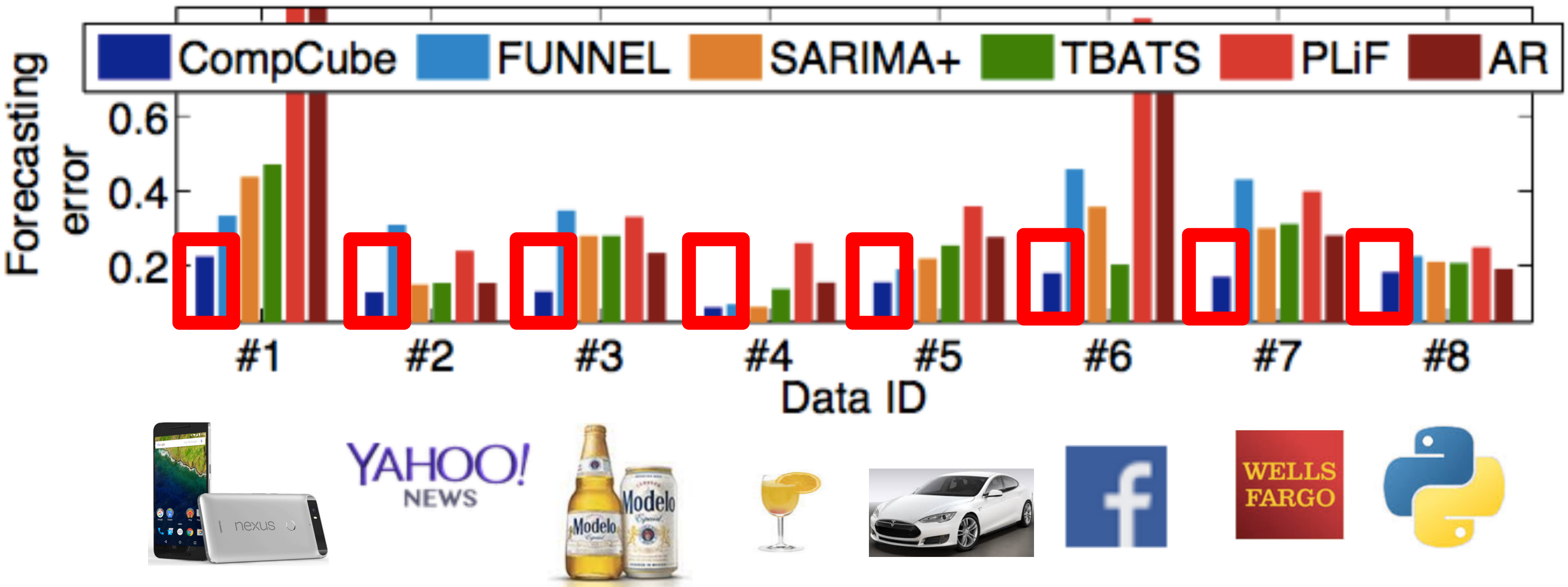
CompCube at work - forecasting

Forecasting error (original vs. forecasts)



CompCube at work - forecasting

Forecasting error (original vs. forecasts)

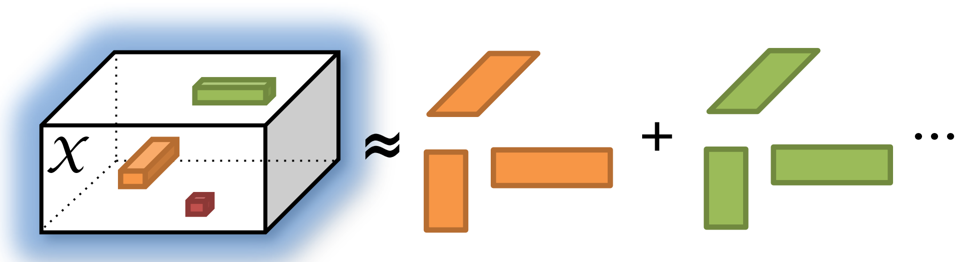


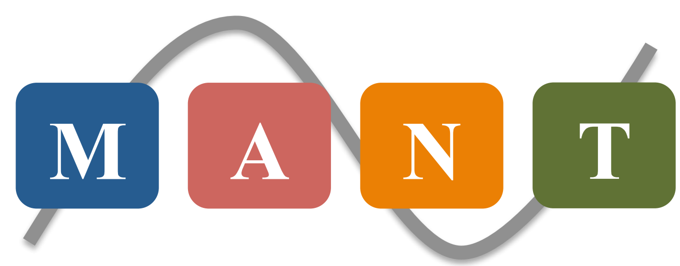
CompCube consistently wins!

Part 3 Conclusions

- Real data are often in high dimensions with multiple aspects (modes)
- Matrices and tensors provide elegant theory and algorithms
- MANT analysis

Multi-Aspect Non-linear Time-series





References

- I. S. Dhillon, S. Mallela, and D. S. Modha. *Information-theoretic co-clustering*. In *KDD*, pages 89-98, 2003.
- T. G. Kolda, B. W. Bader and J. P. Kenny. *Higher-Order Web Link Analysis Using Multilinear Algebra*. In: *ICDM*, Pages 242-249, 2005.
- J. Sun, S. Papadimitriou, and P. Yu. *Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams*, In *ICDM*, 2006.
- Y. Matsubara, Y. Sakurai, C. Faloutsos, T. Iwata, and M. Yoshikawa. *Fast Mining and Forecasting of Complex Time-Stamped Events*. In *KDD*, pages 271-279, 2012.
- Y. Matsubara, Y. Sakurai, W. G. van Panhuis, and C. Faloutsos. *FUNNEL: automatic mining of spatially coevolving epidemics*. In *KDD*, pages 105–114, 2014.
- Y. Matsubara, Y. Sakurai, and C. Faloutsos. *Non-linear Mining of Competing Local Activities*. In *WWW*, pages 737-747, 2016.

Part 3

Extension of time-series: tensor analysis

Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)