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

M A N T

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	...	...	...	...	...
...	...	...	...	...	...

Kumamoto U

Examples of Matrices: Market basket

CMU CS

- **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	...	...	...	...	...
...	...	...	...	...	...

Kumamoto U

Examples of Matrices: Authors and terms

CMU CS

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

Kumamoto U

CMU CS

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?

Motivation 2: Why tensors?

- A: N-D generalization of matrix:

sigmod' 07	data	mining	query	tree	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

Motivation 2: Why tensors?

- A: N-D generalization of matrix:

sigmod' 05

sigmod' 06

sigmod' 07

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

Kumamoto U

Tensors are useful for 3 or more modes

CMU CS

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

Mode#3

data	mining	query	tree	...
13	11	22	55	...
5	4	6	7	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Mode#2

Mode (== aspect) #1

Motivating Applications

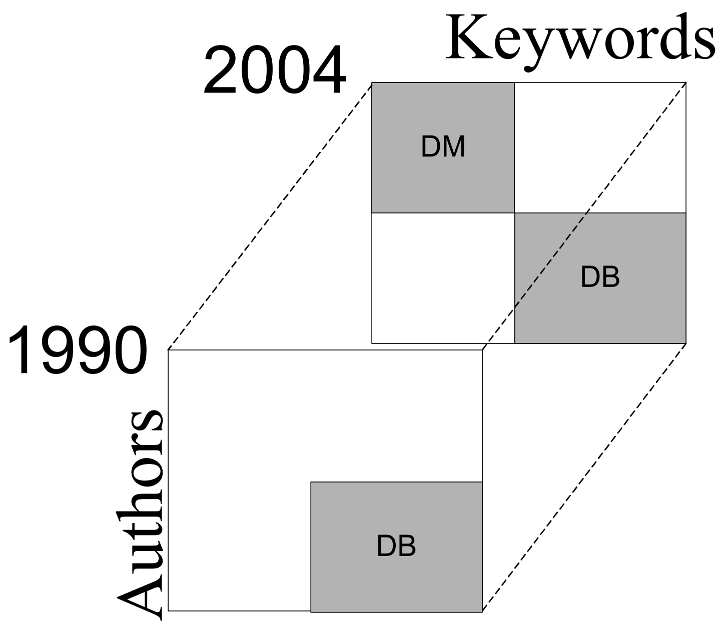
- Why matrices are important?
- Why tensors are useful?
  - P1: social networks
  - P2: web mining

Kumamoto U

P1: Social network analysis

CMU CS

- 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

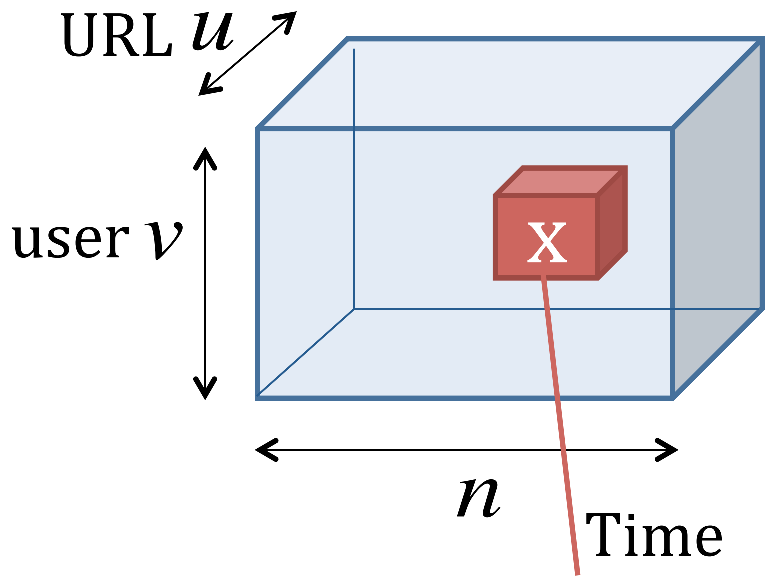
Kumamoto U

CMU CS

Tensors for time-series analysis

- 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)

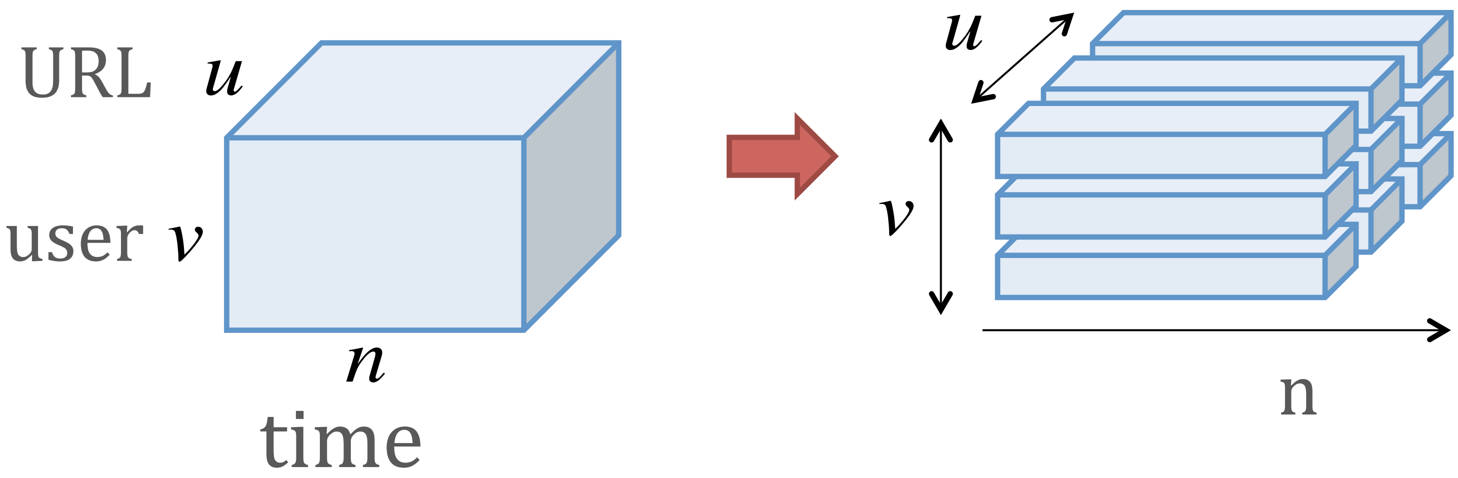
$$\mathcal{X} \in \mathbb{N}^{u \times v \times n}$$

Element x: # of events

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

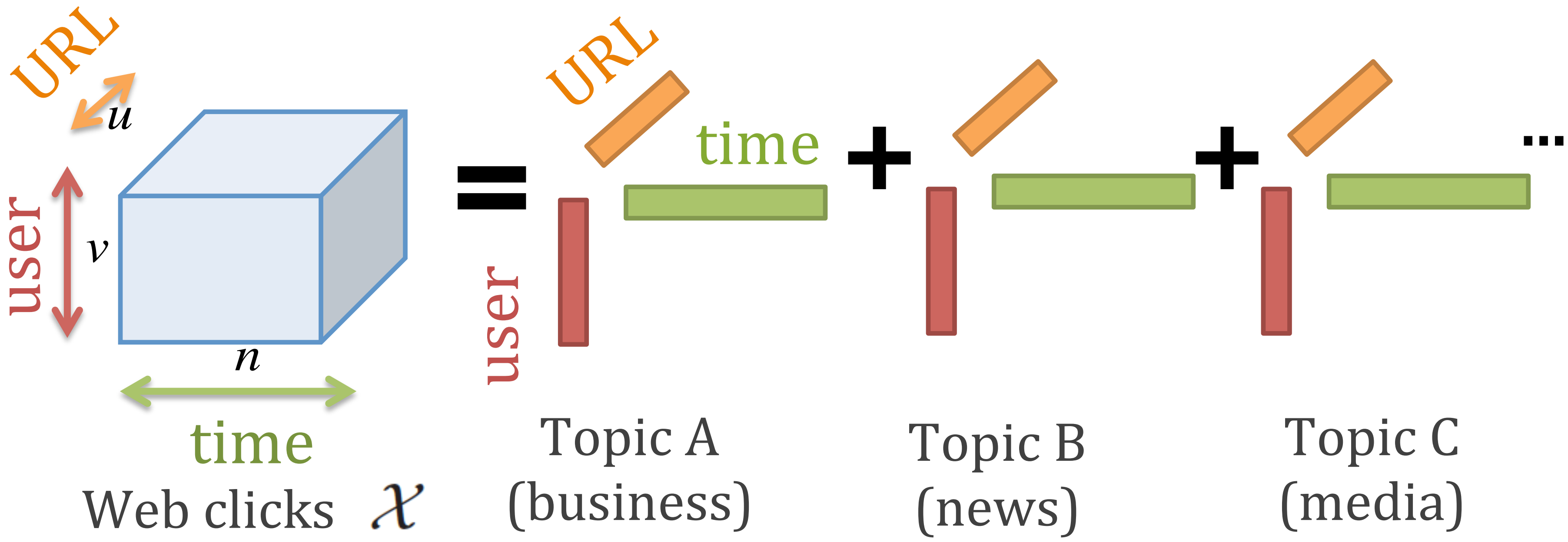
Tensors for time-series analysis

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



Tensors for time-series analysis

- Multi-aspect time-series analysis



Outline

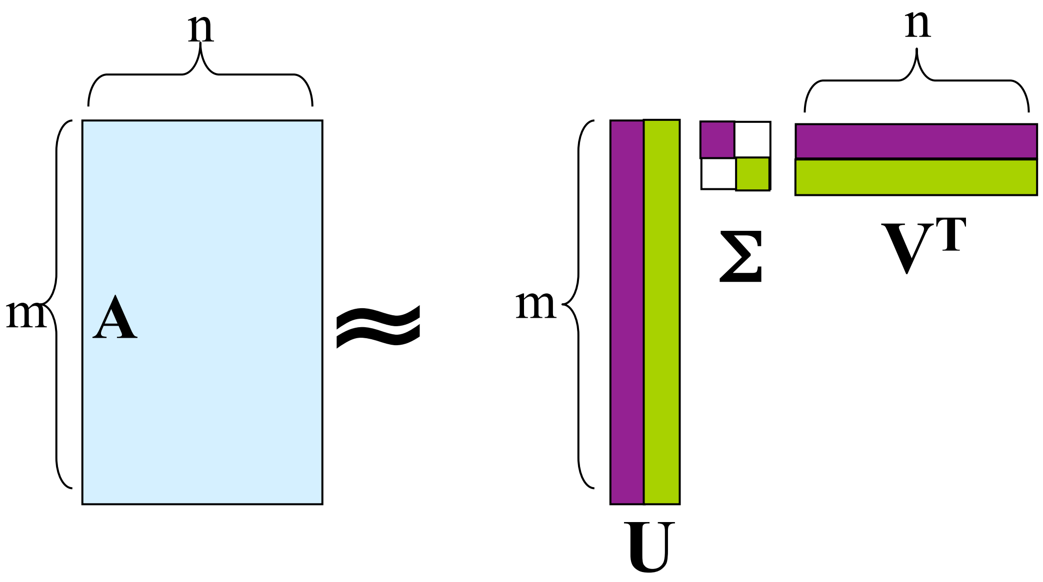
- Tensor decomposition
  - Motivation
  - Basic approaches
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

M A N T

Reminder: SVD

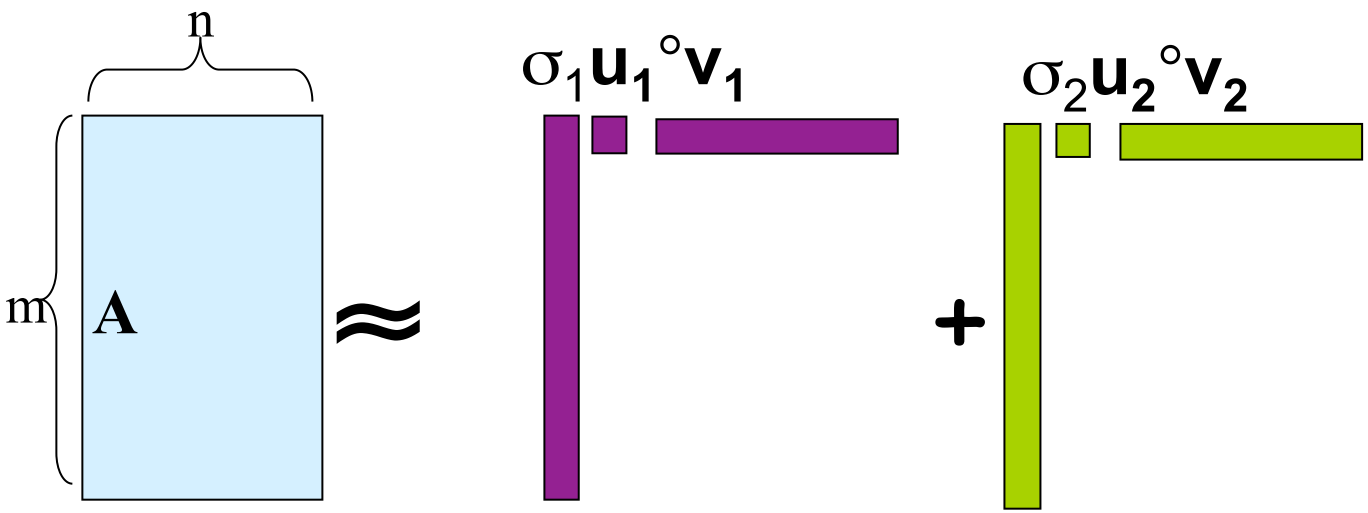
$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



– Best rank-k approximation in L2

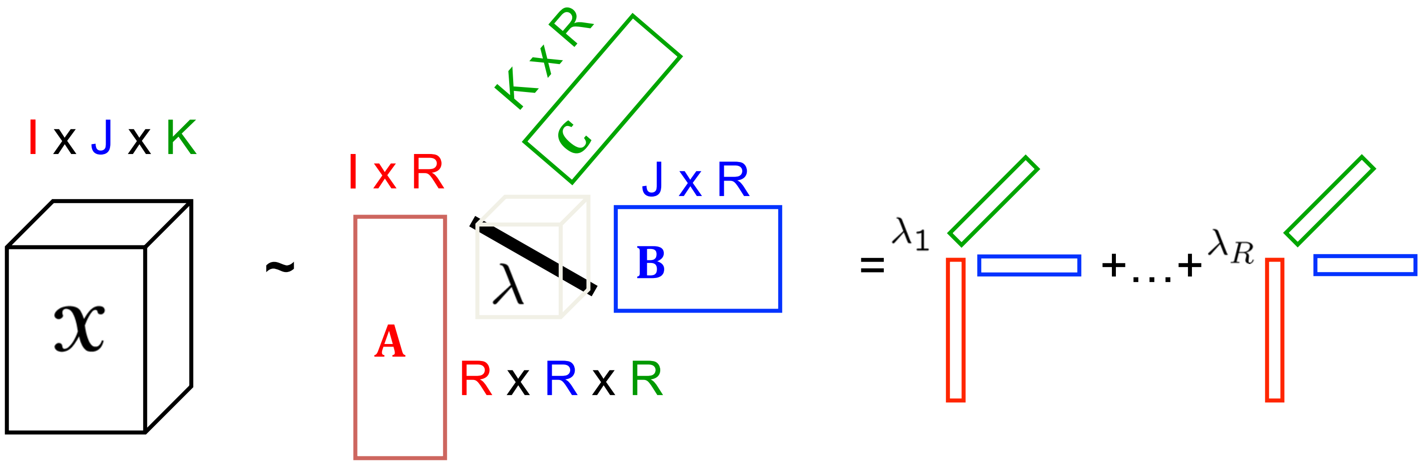
Reminder: SVD

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



– 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:

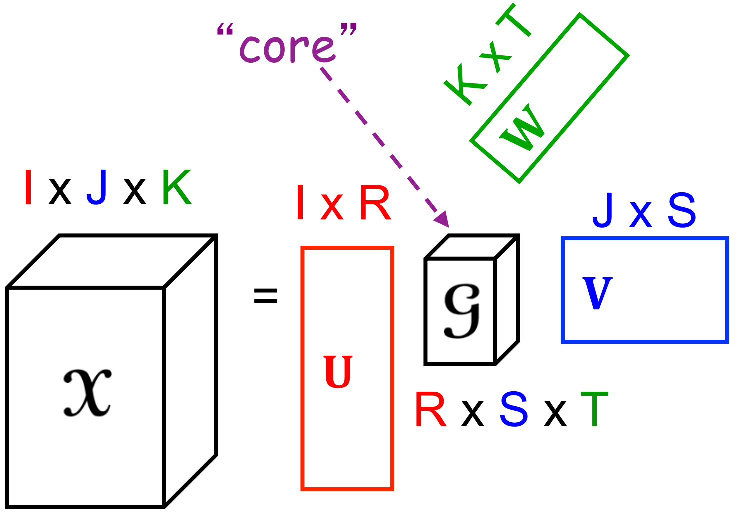
- 2 major types of tensor decompositions: PARAFAC and Tucker
- both can be solved with ``alternating least squares'' (ALS)
- Details follow

Specially Structured Tensors

- Tucker Tensor

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

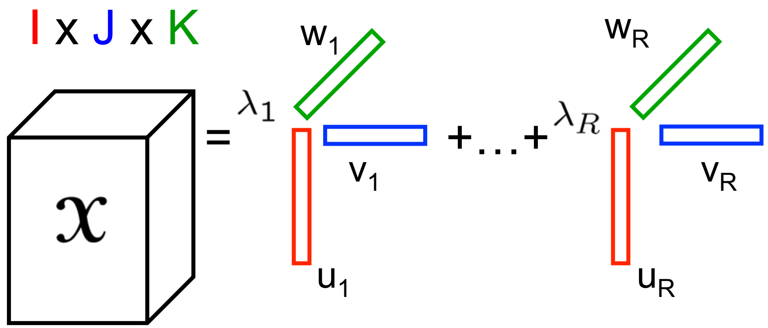
Our Notation



- 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

CMU CS

Specially Structured Tensors

- Tucker Tensor

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

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

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

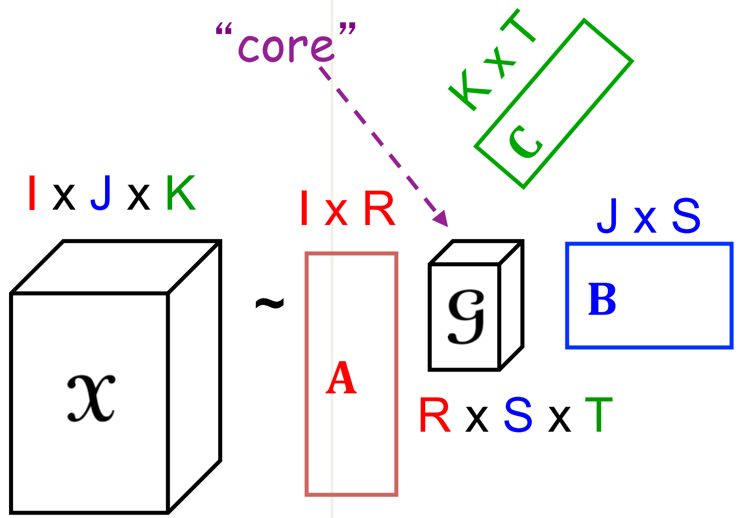
In matrix form:

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]

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

eg, terms x documents

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

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

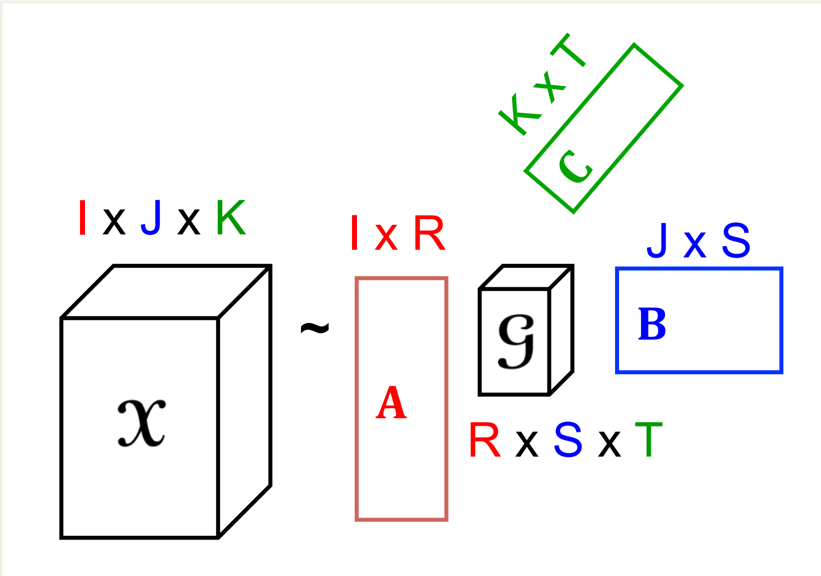
term group x
doc. group

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

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:

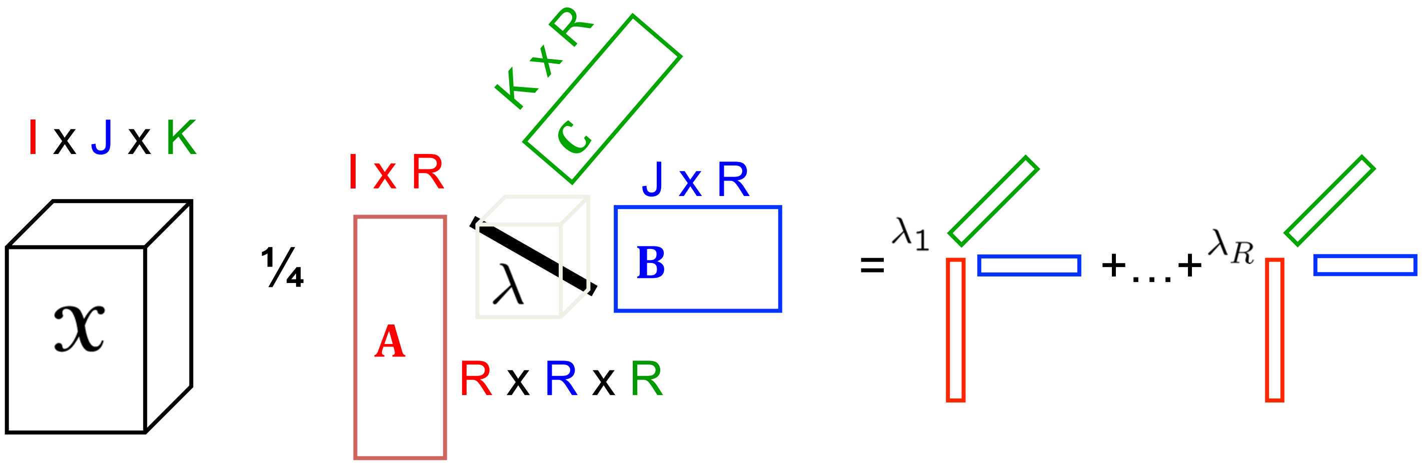
$$\mathcal{G} = [\![\mathcal{X}\,;\mathbf{A}^{\dagger},\mathbf{B}^{\dagger},\mathbf{C}^{\dagger}]\!]$$

- 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 <u>not</u> 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

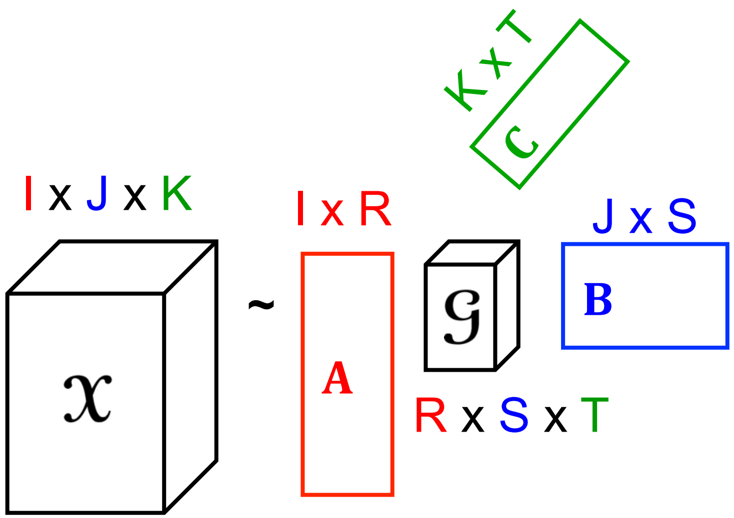


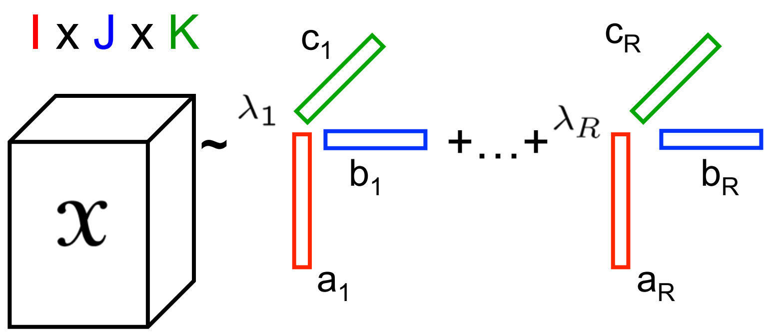
$$\mathcal{X} \approx [\![\lambda\,;\mathbf{A},\mathbf{B},\mathbf{C}]\!] = \sum_r \lambda_r\, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

- 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($\mathcal{X}$) > 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://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

[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 ?

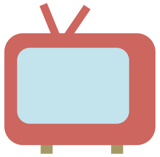
- news, tech, media, sports, etc...

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



Kumamoto U

CMU CS

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?

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

future clicks?

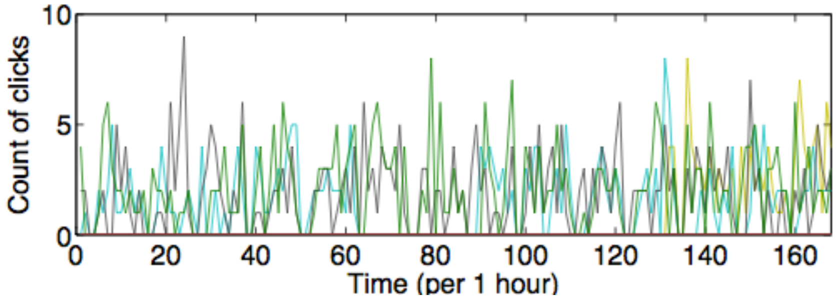
Motivation

Web click events – can we see any trends?

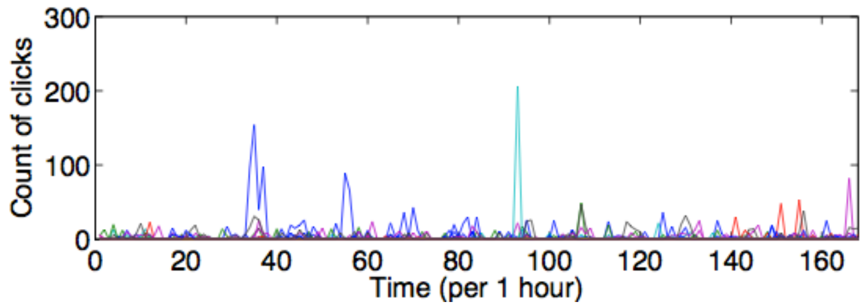
Original access counts of each URL

- 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

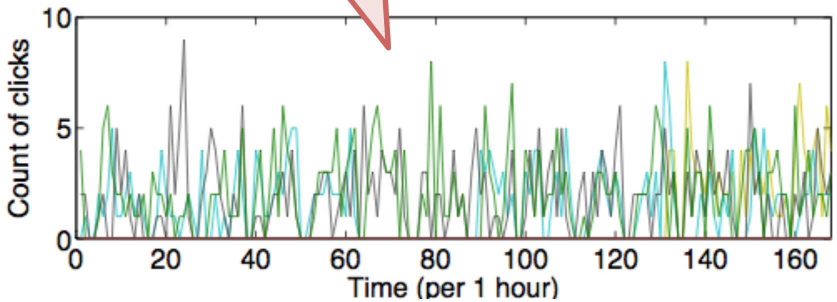
ndow size

Noisy ☹

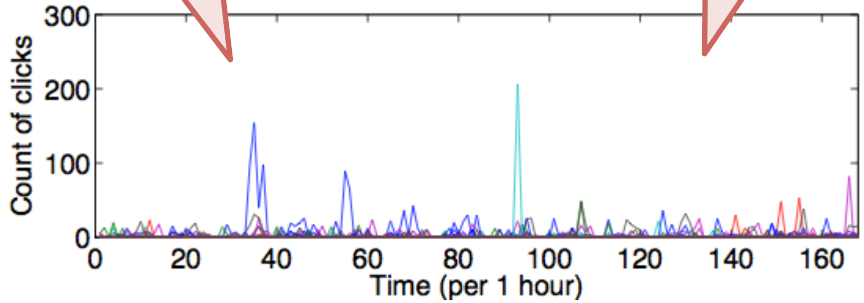
Sparse ☹

Bursty ☹

URL: mo y site



UR blog site



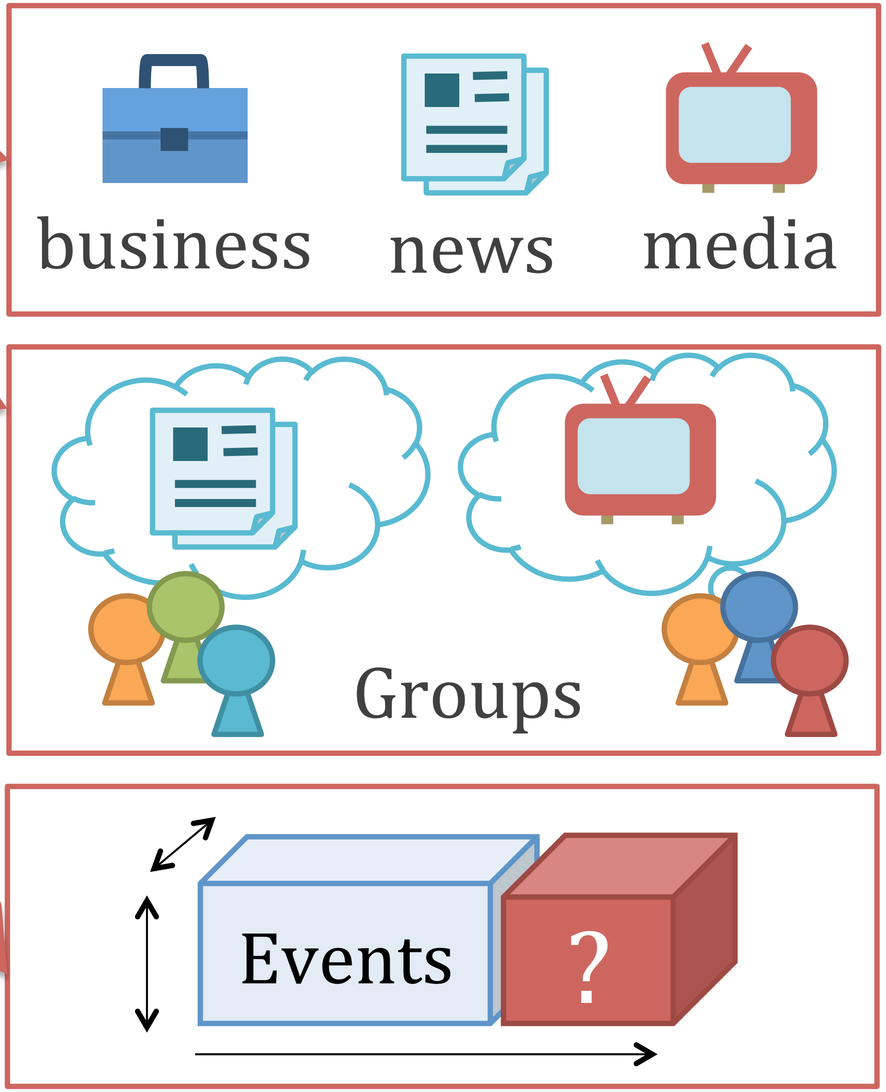
☹ We cannot see any trends !!

Our goals

Q1: Hidden topics

Q2: Groups

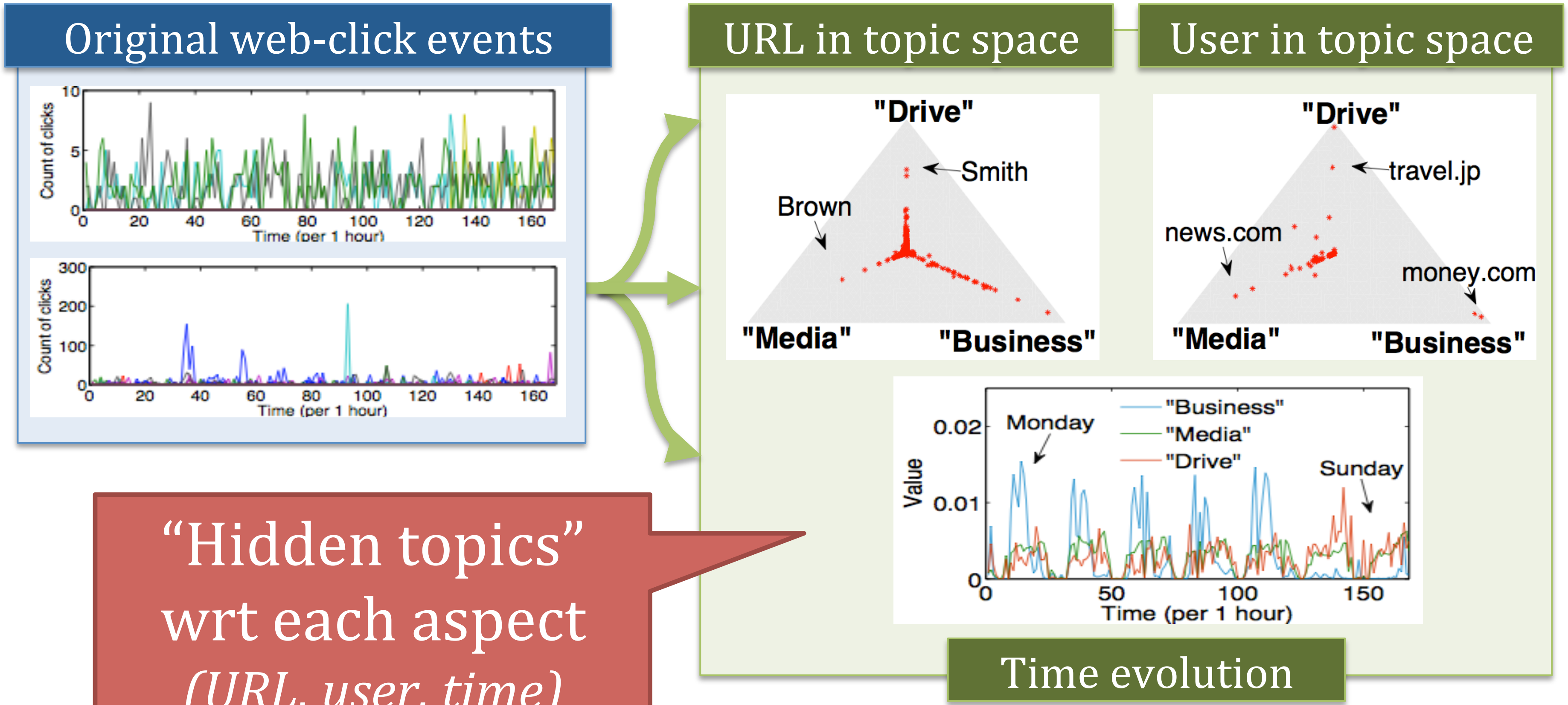
Q3: Forecasting



Problem definition

Given: a set of complex time-stamped events

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



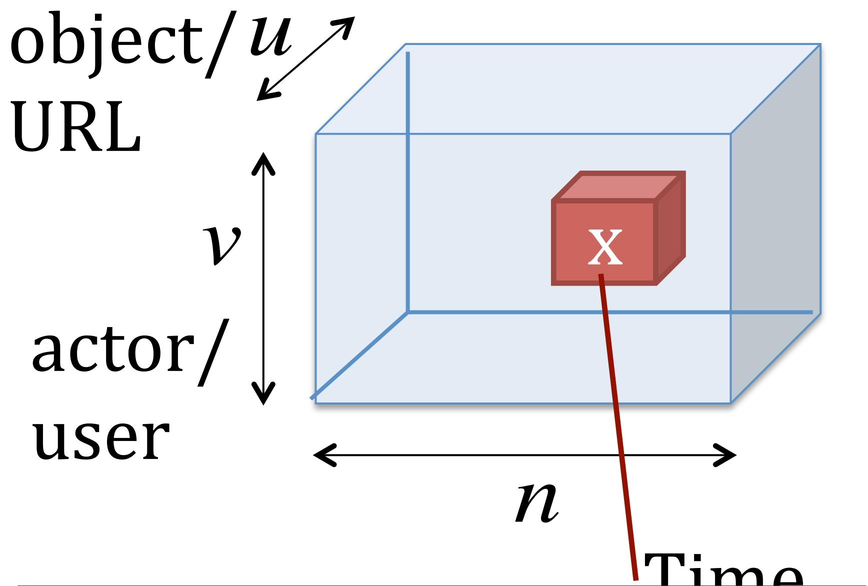
"Hidden topics" wrt each aspect
(URL, user, time)

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

M[th] order tensor (M=3)

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

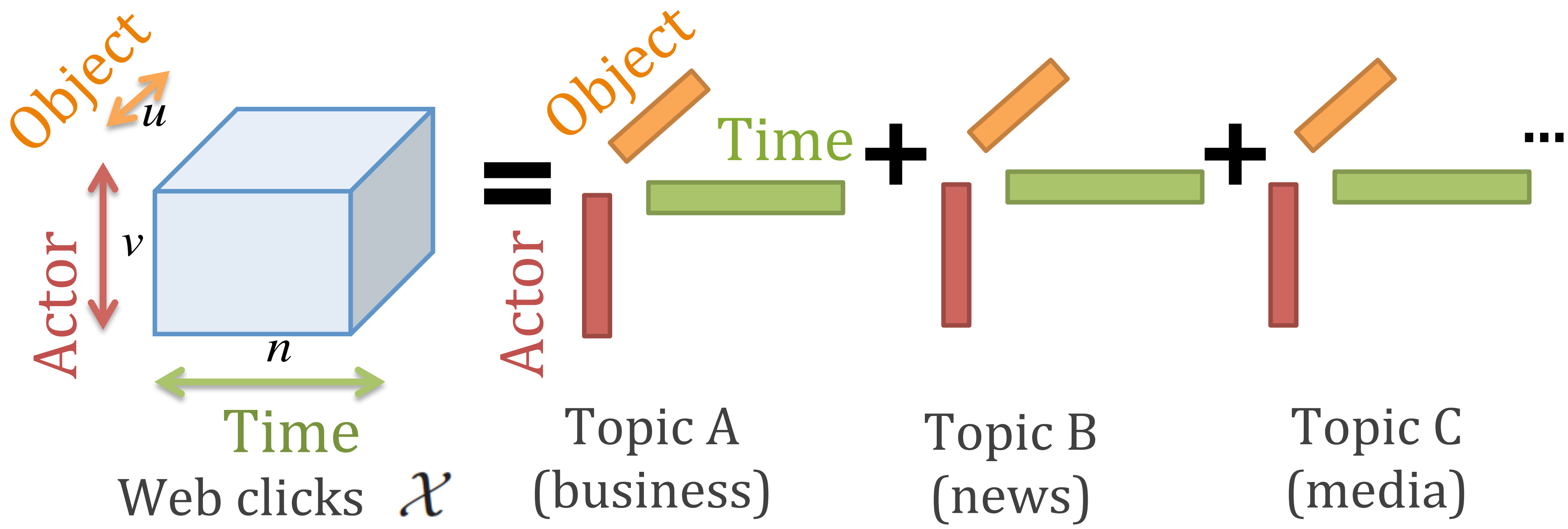
Element x: # of events

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

Main idea (1) : M-way analysis

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

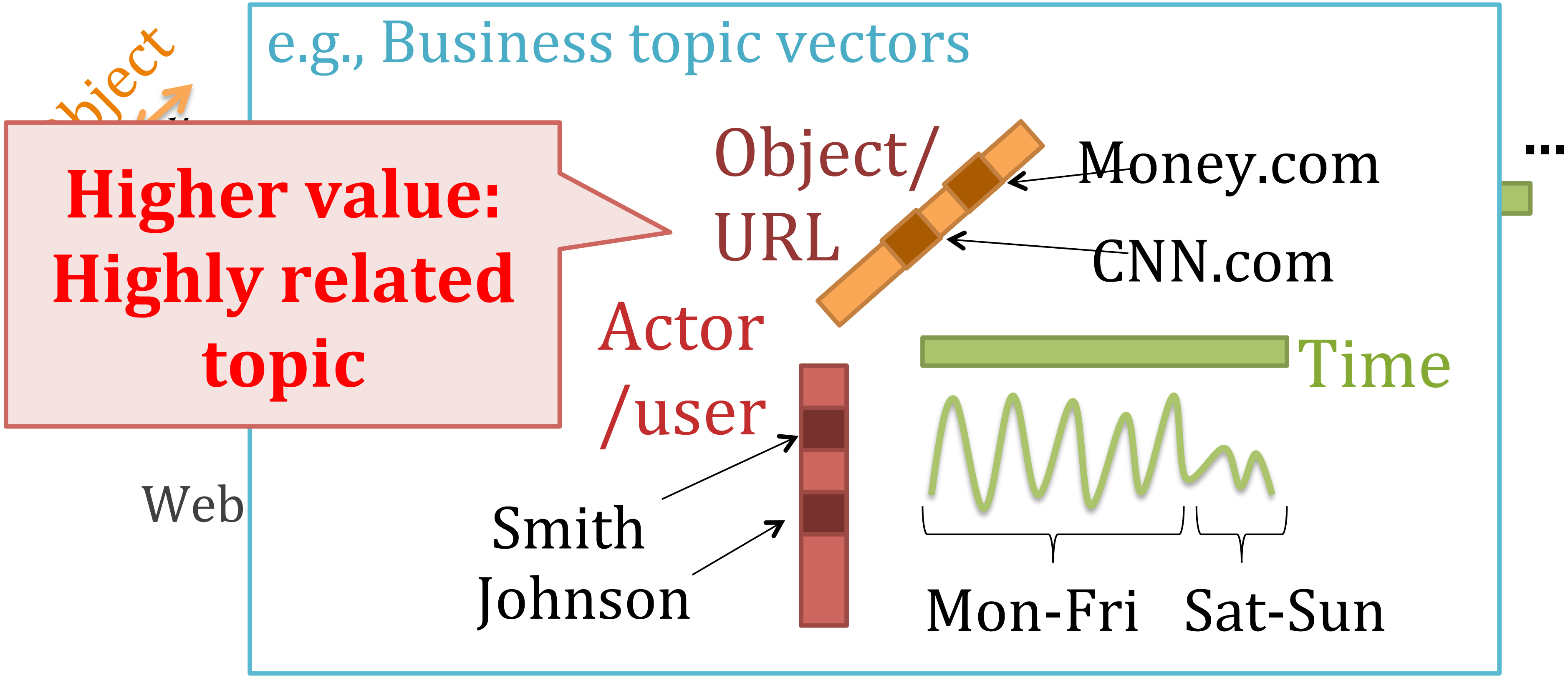
- **Object vector** **Actor vector** **Time vector**



Main idea (1) : M-way analysis

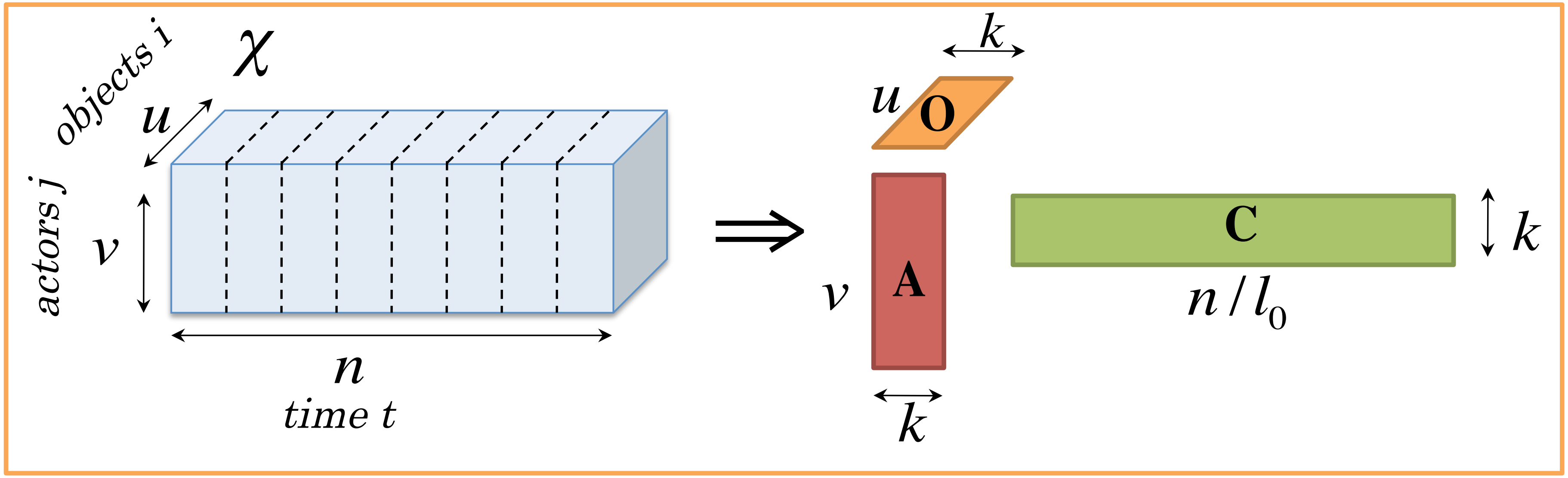
A. decompose to a set of **3 topic vectors:**

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



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

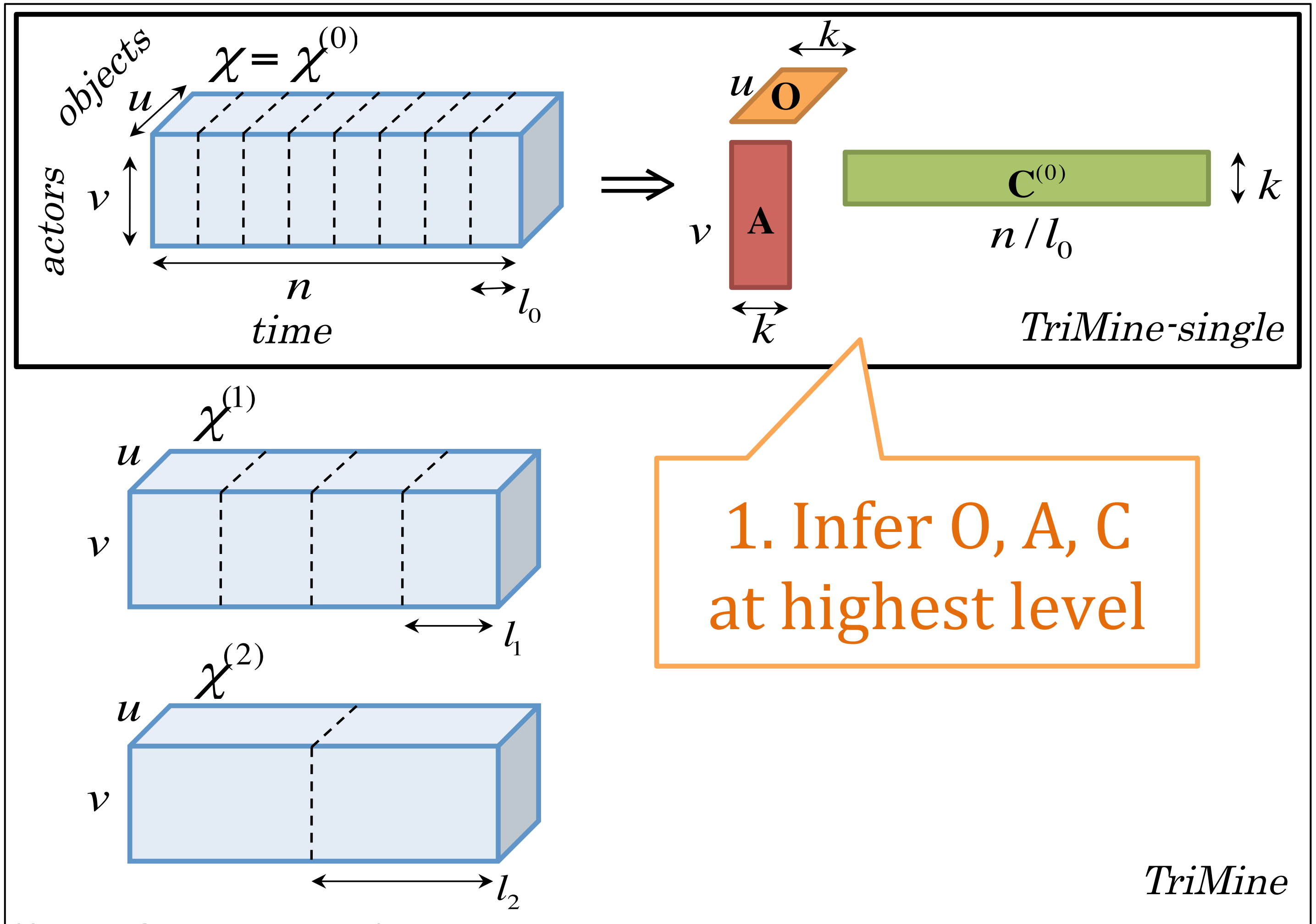
- M-way decomposition (M=3)
  - **[Gibbs sampling]** infer k hidden topics for each non-zero element of X, according to probability p



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

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

- Tensors with multiple window sizes



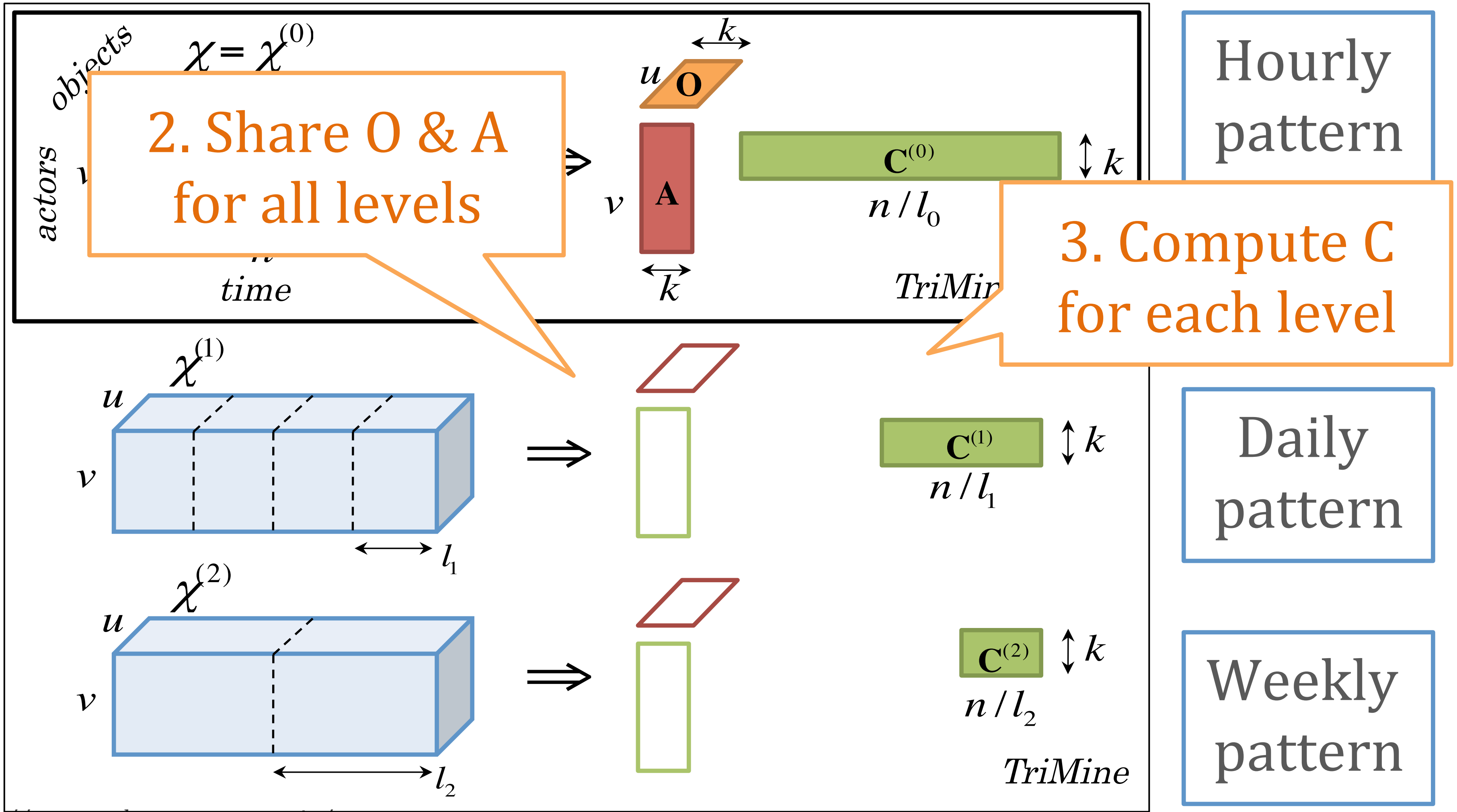
Hourly pattern

Daily pattern

Weekly pattern

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

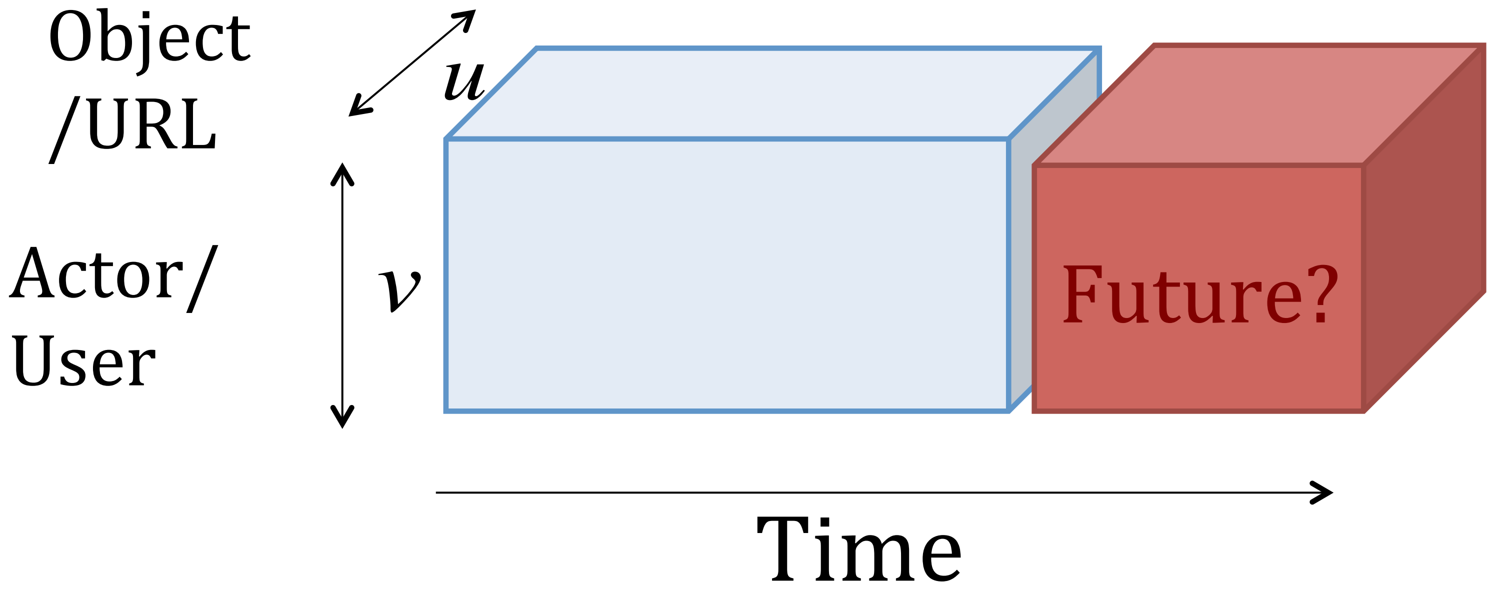
- Tensors with multiple window sizes



TriMine-Forecasts

Our final goal: “forecast future events”!

Q. How can we generate a realistic events?

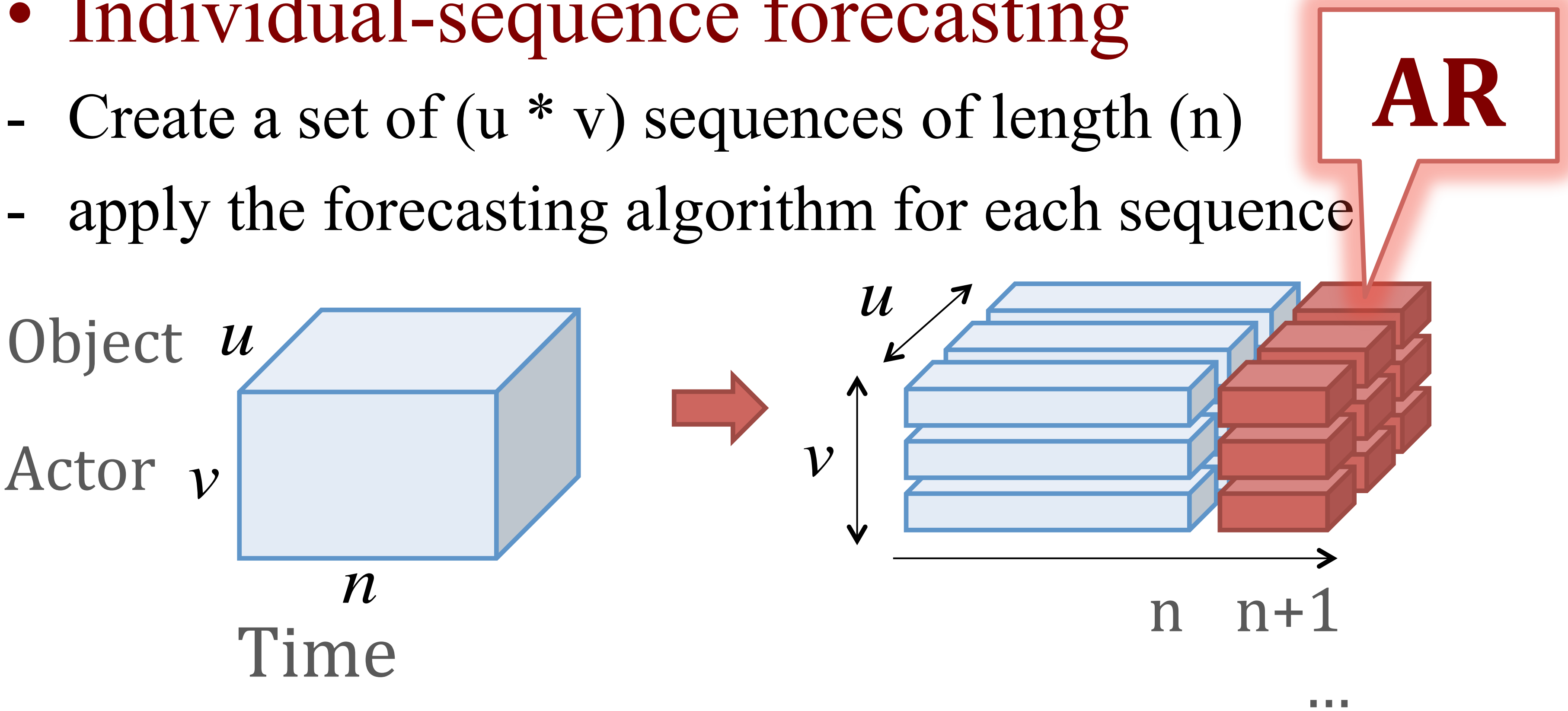


e.g., estimate the number of clicks for

user “smith”, to URL “CNN.com”, for next 10 days

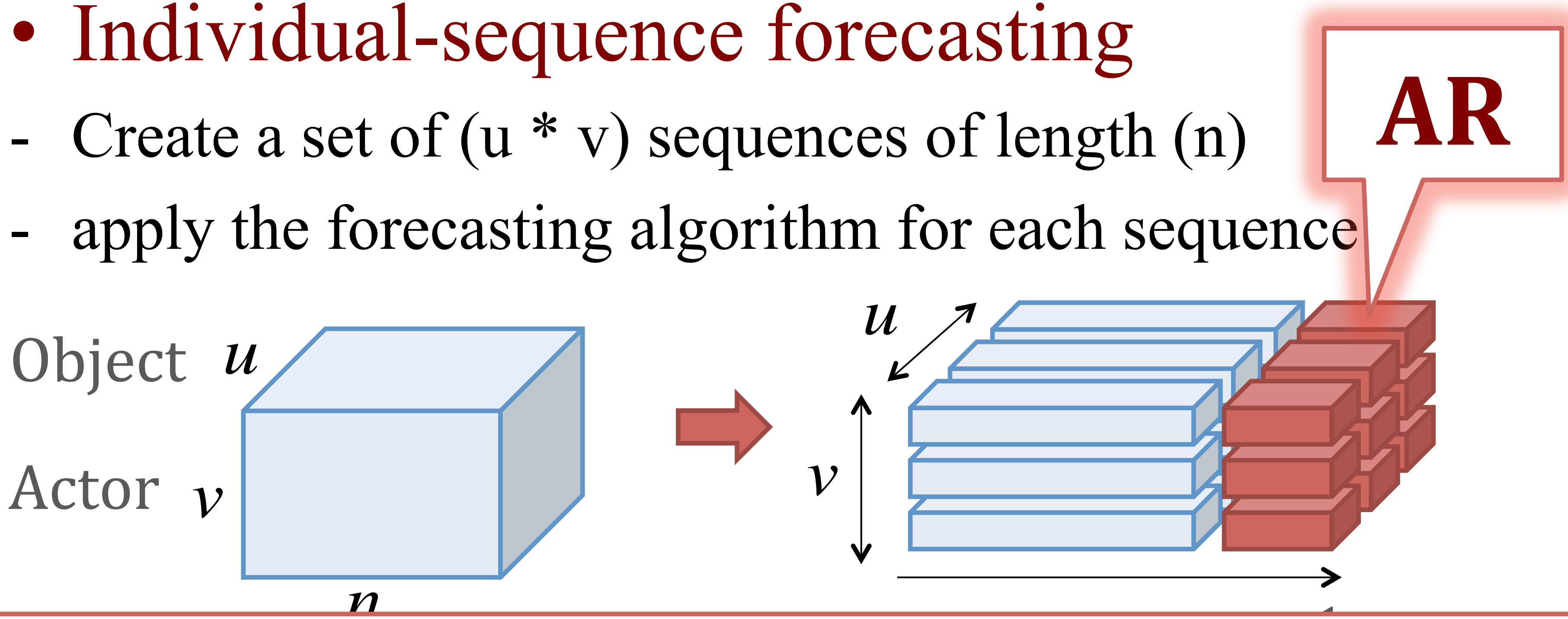
Why not naïve?

- 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



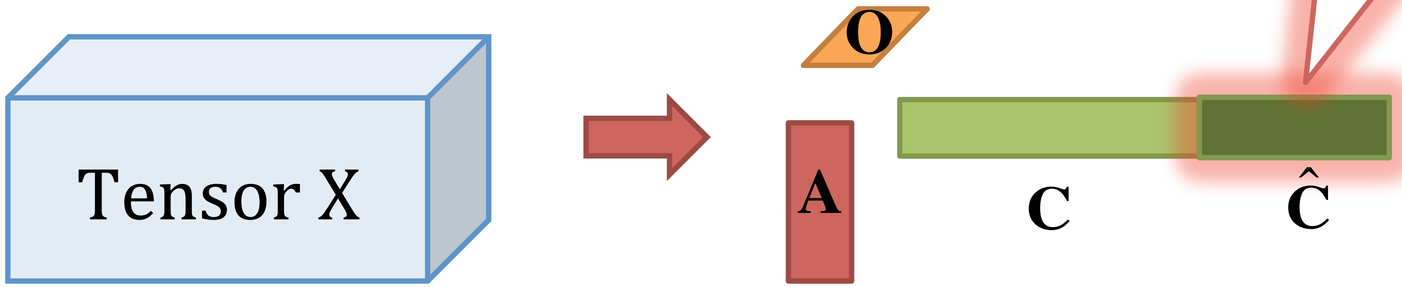
- ☹ **Scalability** : time complexity is at least $O(uvn)$
- ☹ **Accuracy** : each sequence "looks" like noise,
  (e.g., {0, 0, 0, 1, 0, 0, 2, 0, 0, ...}) -> hard to forecast

TriMine-F

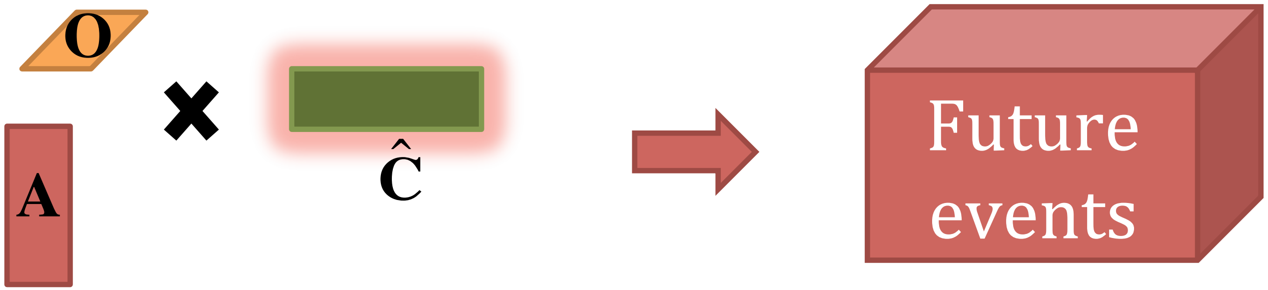
(Multi-scale)
AR

Our approach:

– Step 1: Forecast time-topic matrix:



– Step 2: Generate events using 3 matrices



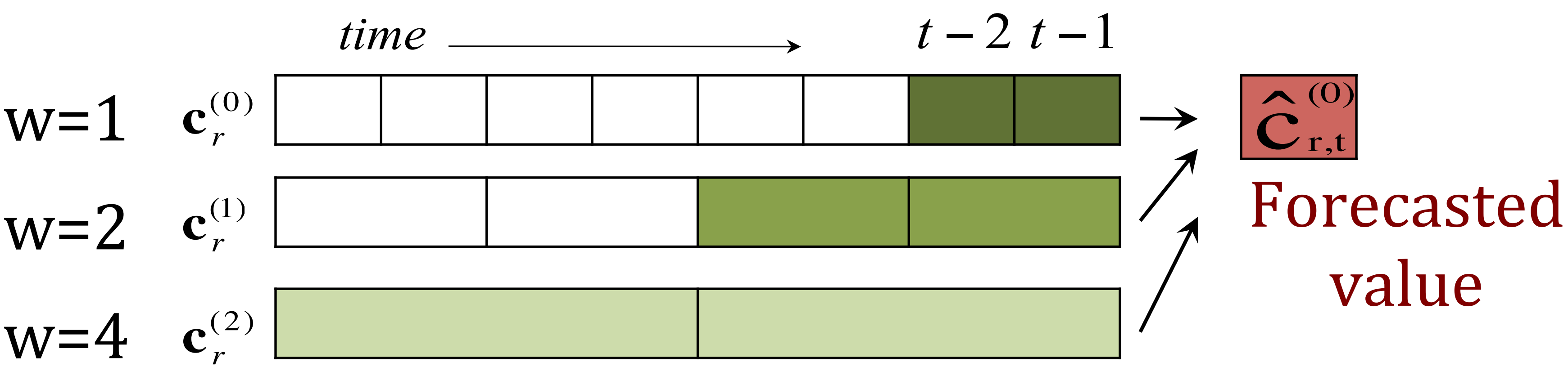
Forecast 'time-topic matrix' (details)

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

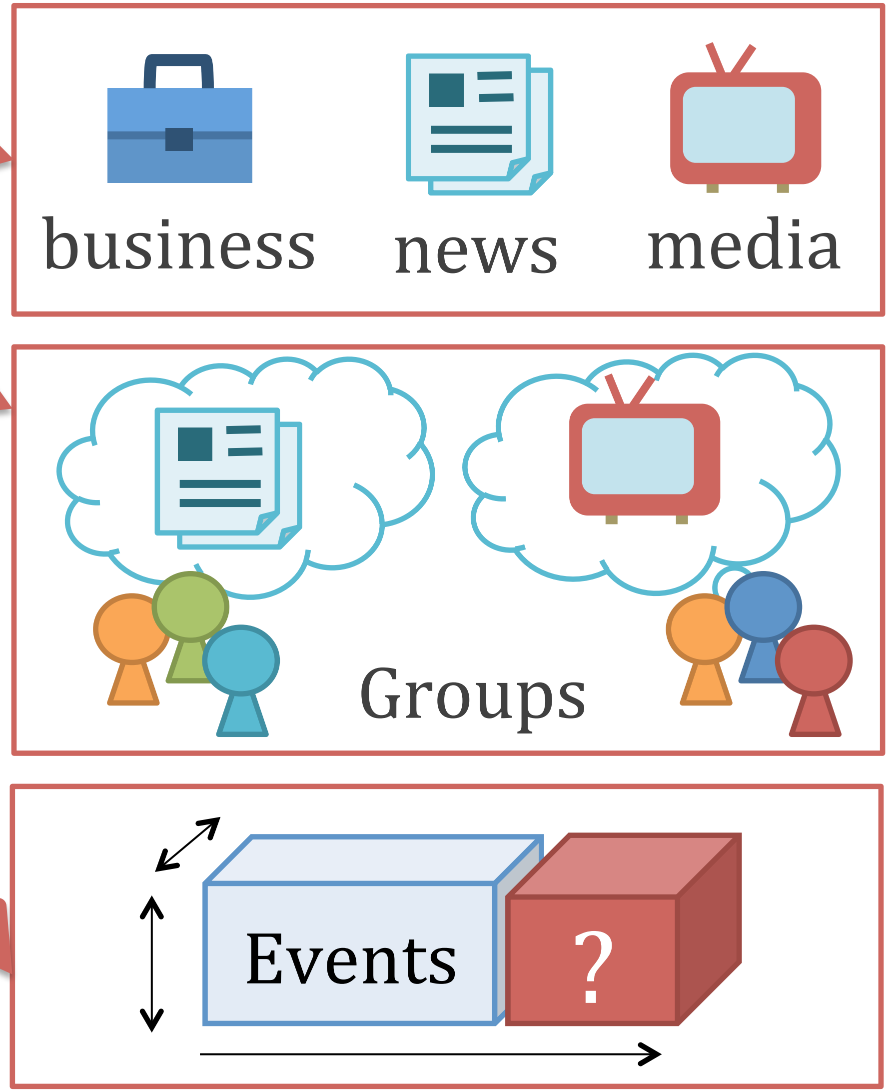


$$c^{(0)}_{r,t} = \sum_{h=0}^{\lceil \log n \rceil} \sum_{i=1}^{w} \lambda^{(h)}_{i,r} c^{(h)}_{r,t-i} + \epsilon_t .$$

(Details in paper)

Our goals

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

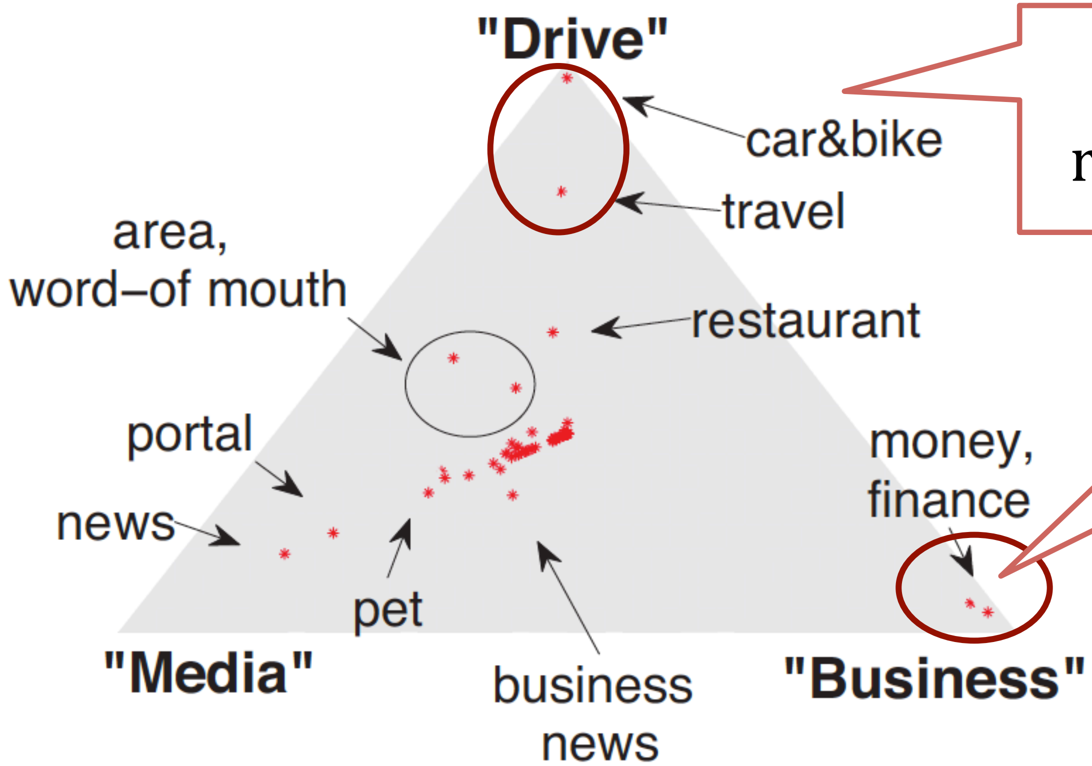


Q1&2. WebClick data

URL-topic matrix (O)

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

* Red point : each web site



Car & bike site is related to travel site

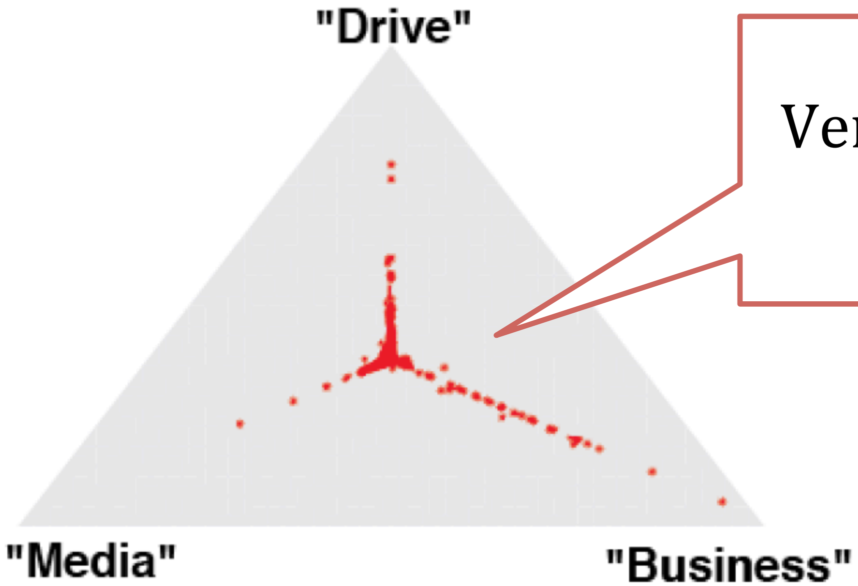
Money site& Finance site have similar trends

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

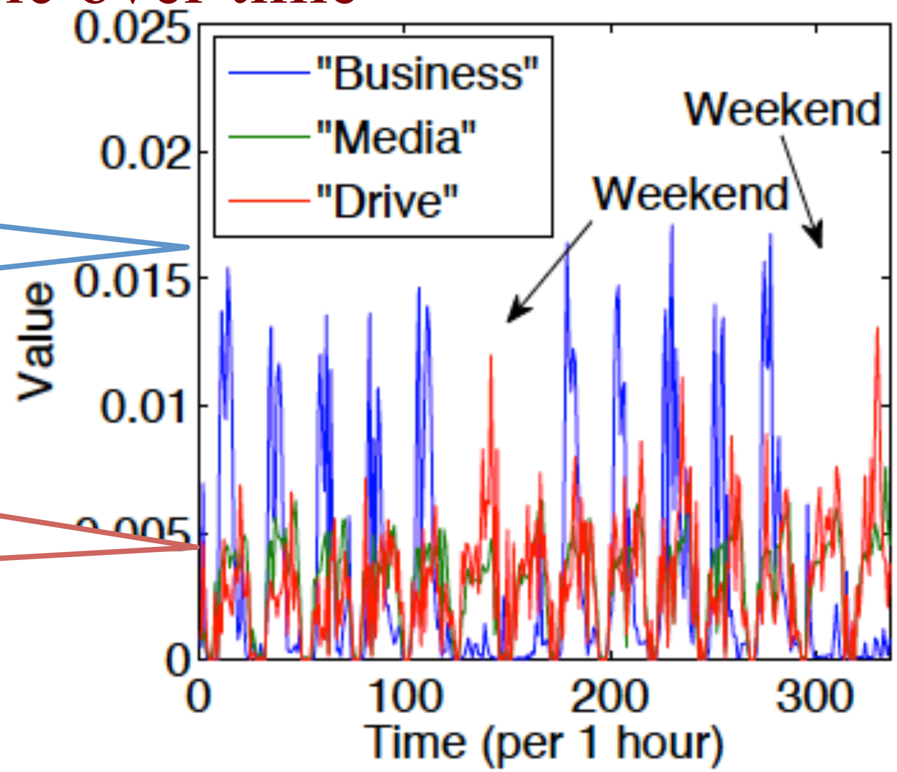
Time-topic matrix (C)

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

* Each sequence: each topic over time

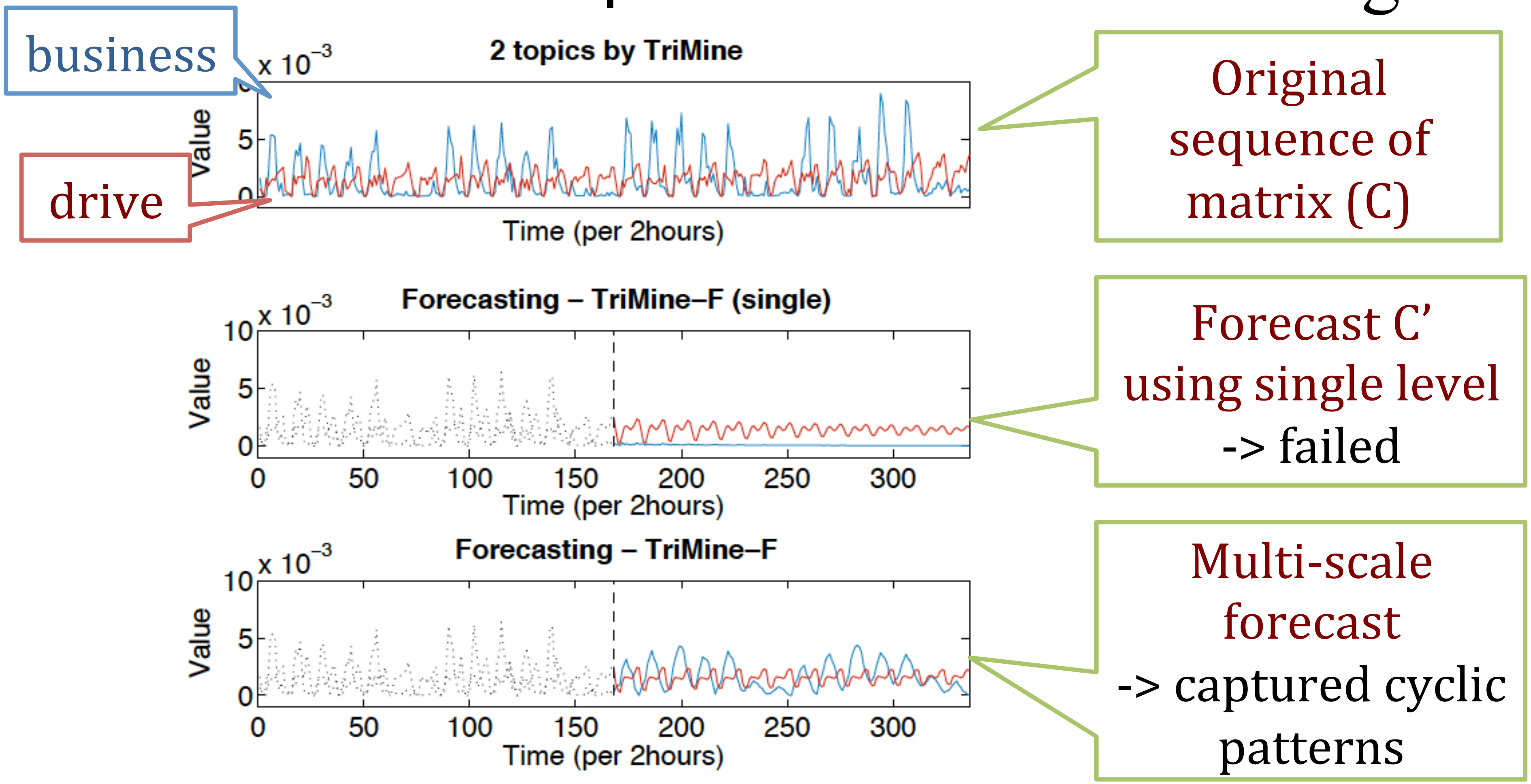
“**Business**” topic:
Less access during weekend

“**Drive**” topic:
Spikes during weekend



Q3. Forecasting accuracy

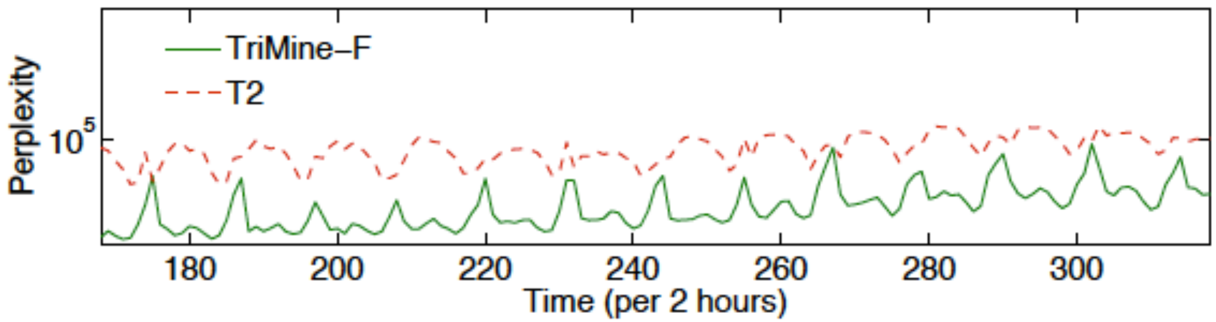
- Benefit of multiple time-scale forecasting



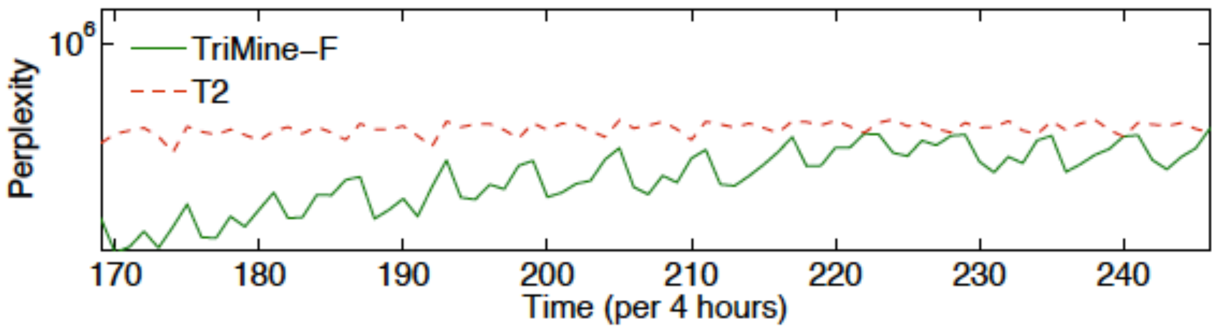
Q3. Forecasting accuracy

Temporal perplexity (entropy for each time-tick)

Lower perplexity: higher predictive accuracy



(a) *WebClick*



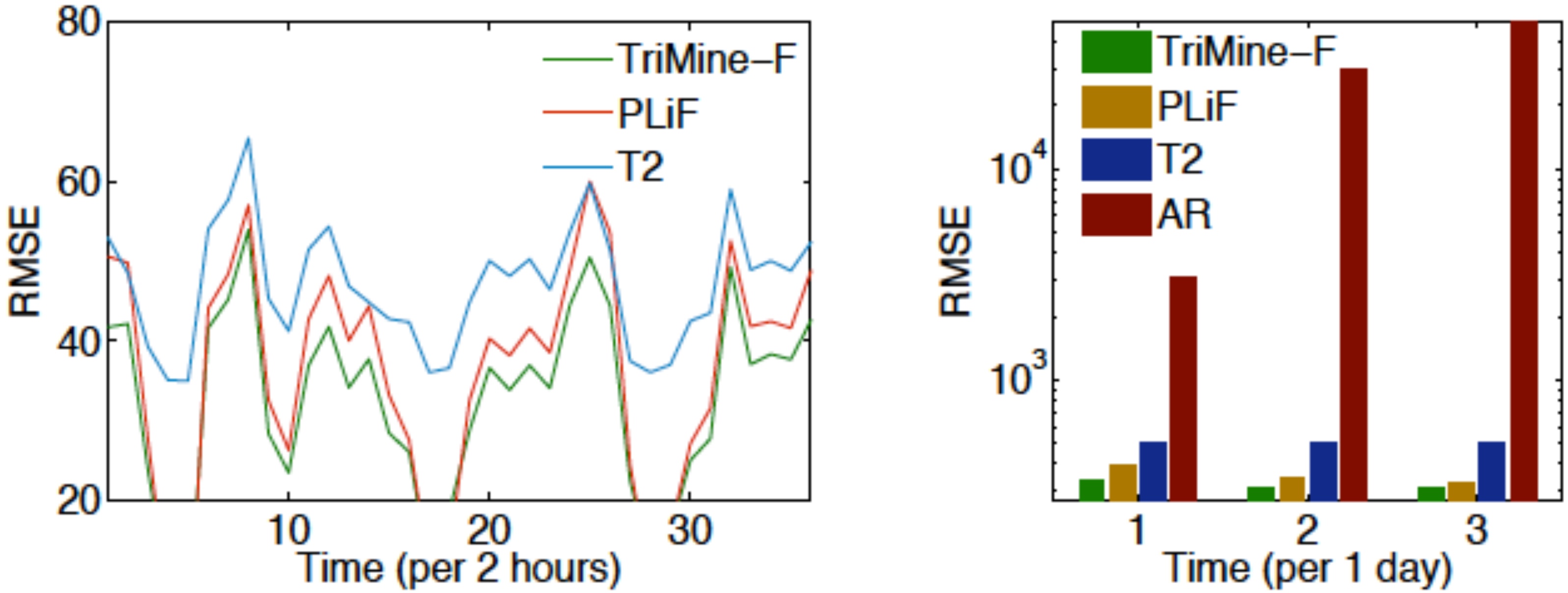
(b) *Ondemand TV*

T2: [Hong et al. KDD'11]

Q3. Forecasting accuracy

Accuracy of event forecasting

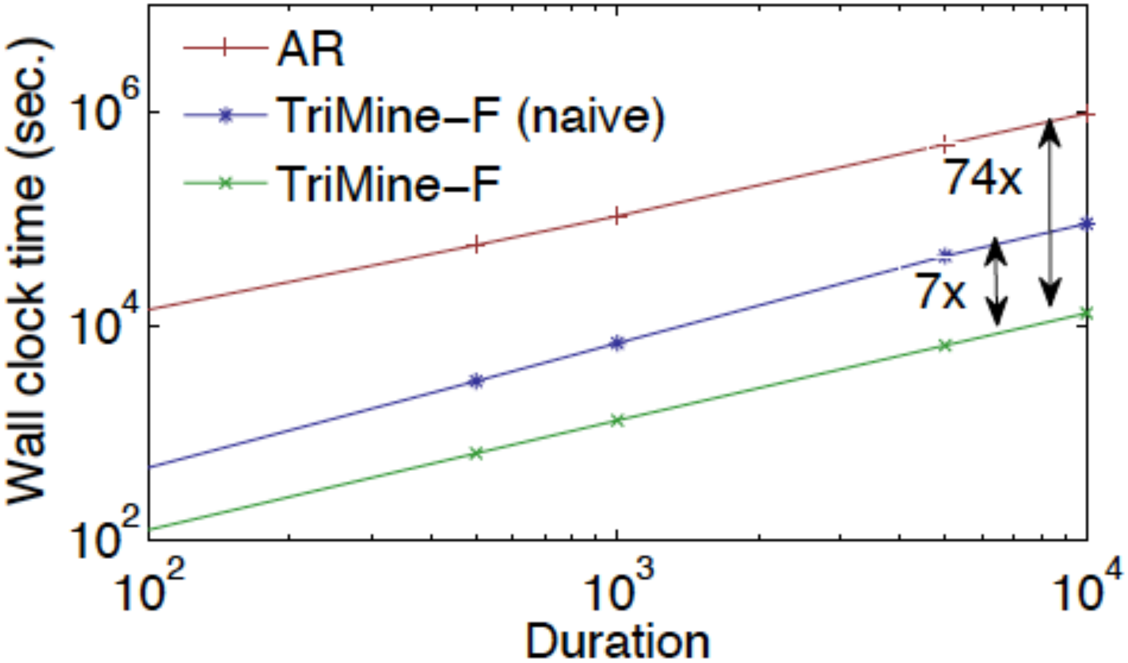
RMSE between original and forecasted events (lower is better)



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)

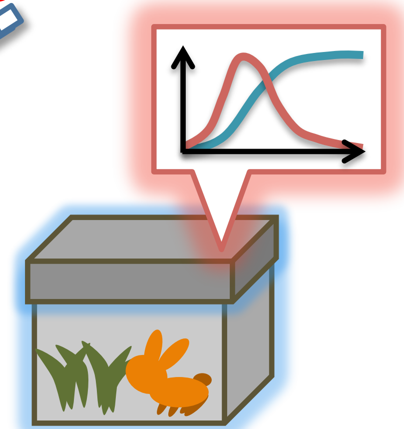
Outline

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

Multi-Aspect Non-linear Time-series

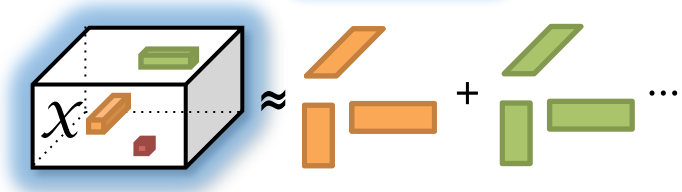
Non-linear tensor analysis

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]

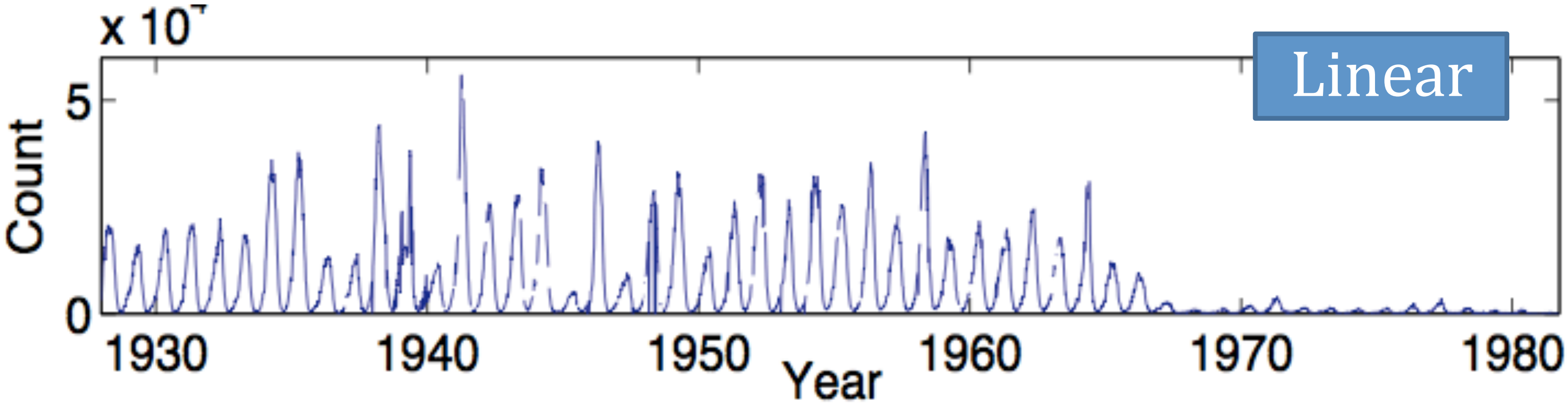
FUNNEL: Automatic Mining of Spatially Coevolving Epidemics

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.

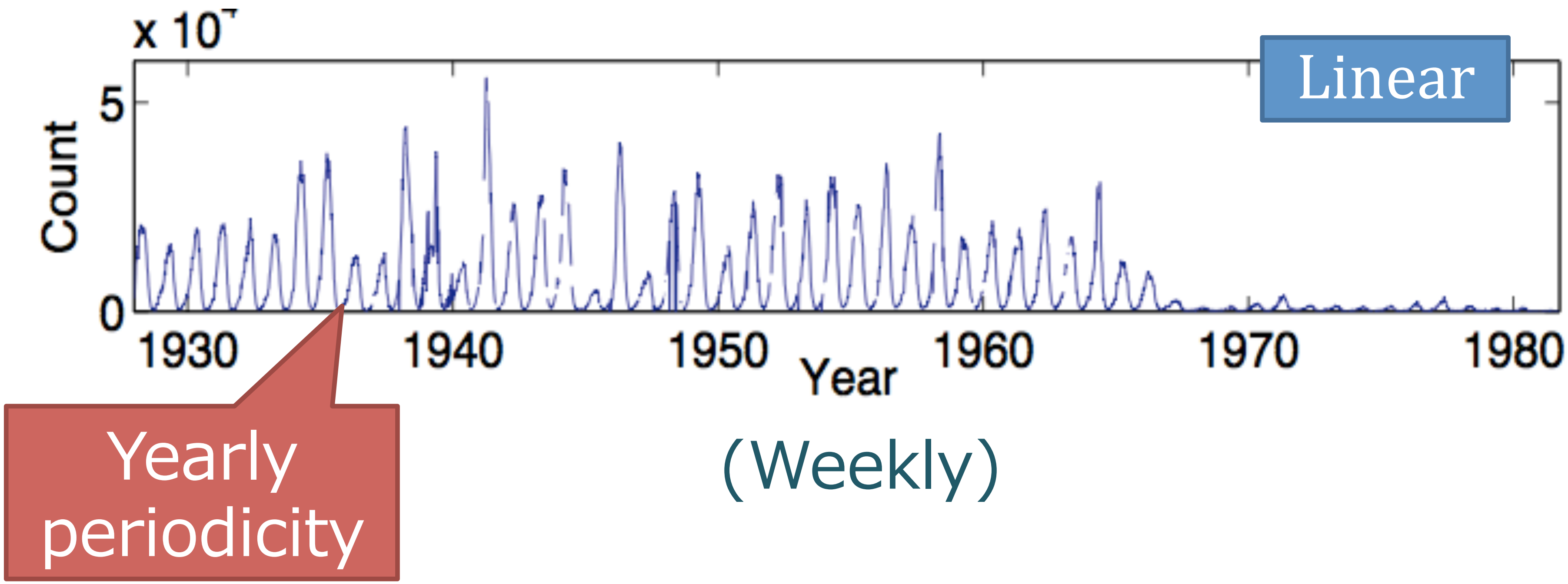


(Weekly)

Motivation

Given: Large set of epidemiological data

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

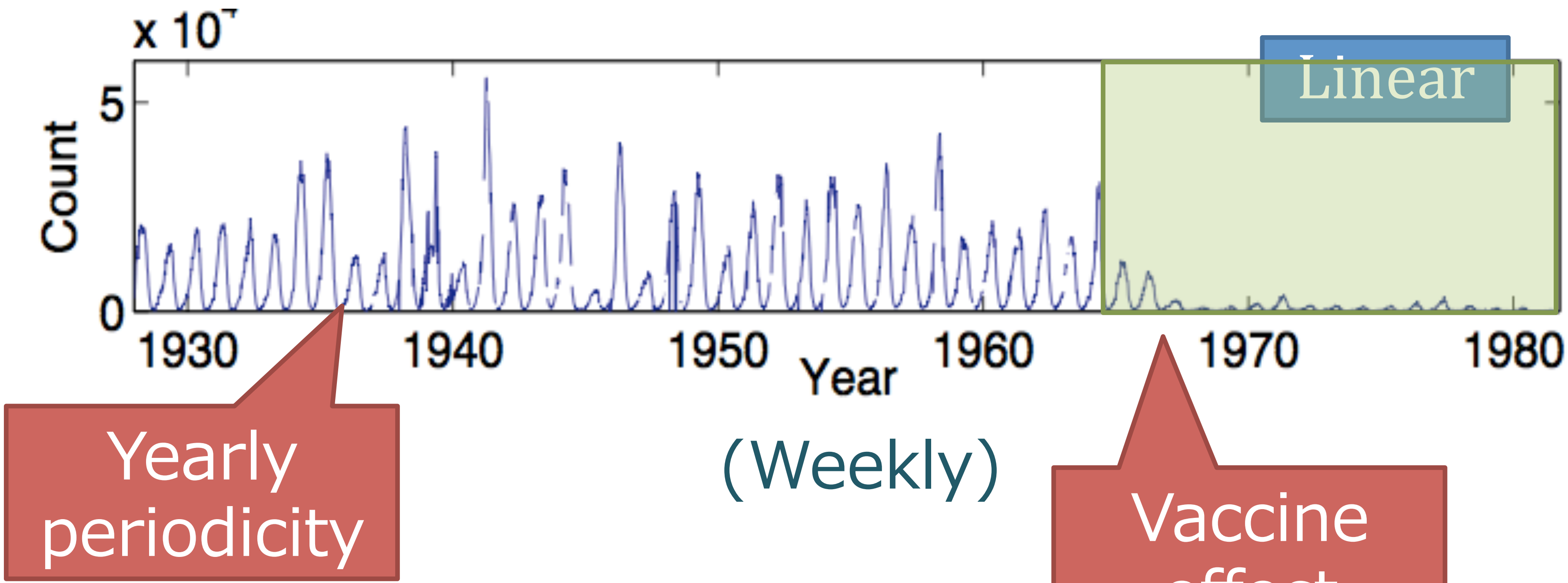


(Weekly)

Motivation

Given: Large set of epidemiological data

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

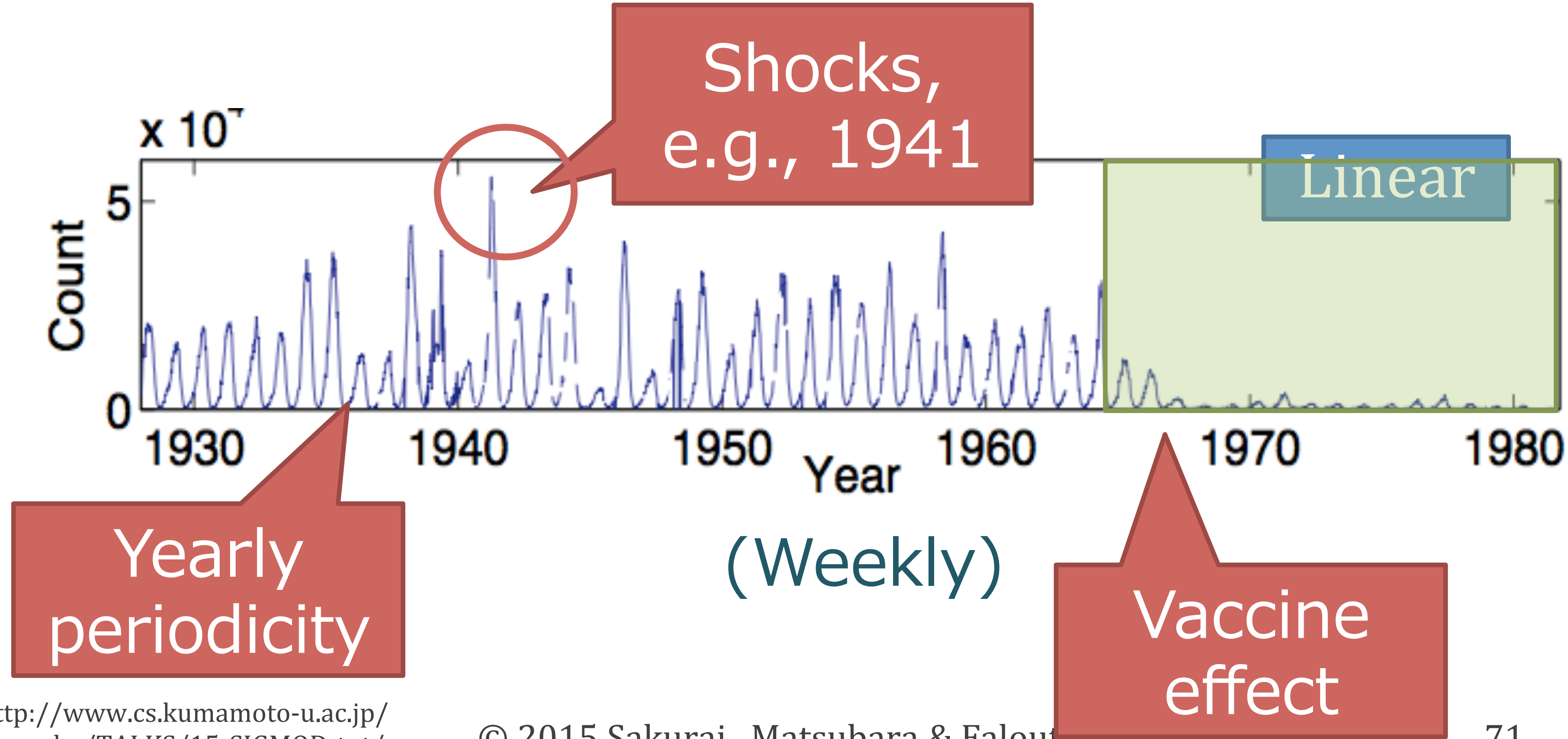


(Weekly)

Motivation

Given: Large set of epidemiological data

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

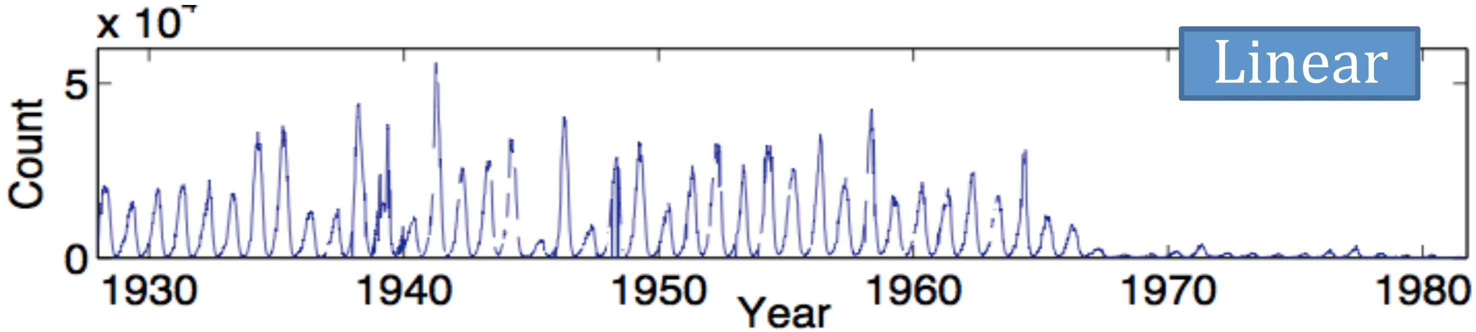


Motivation

Given: Large set of epidemiological data

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

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

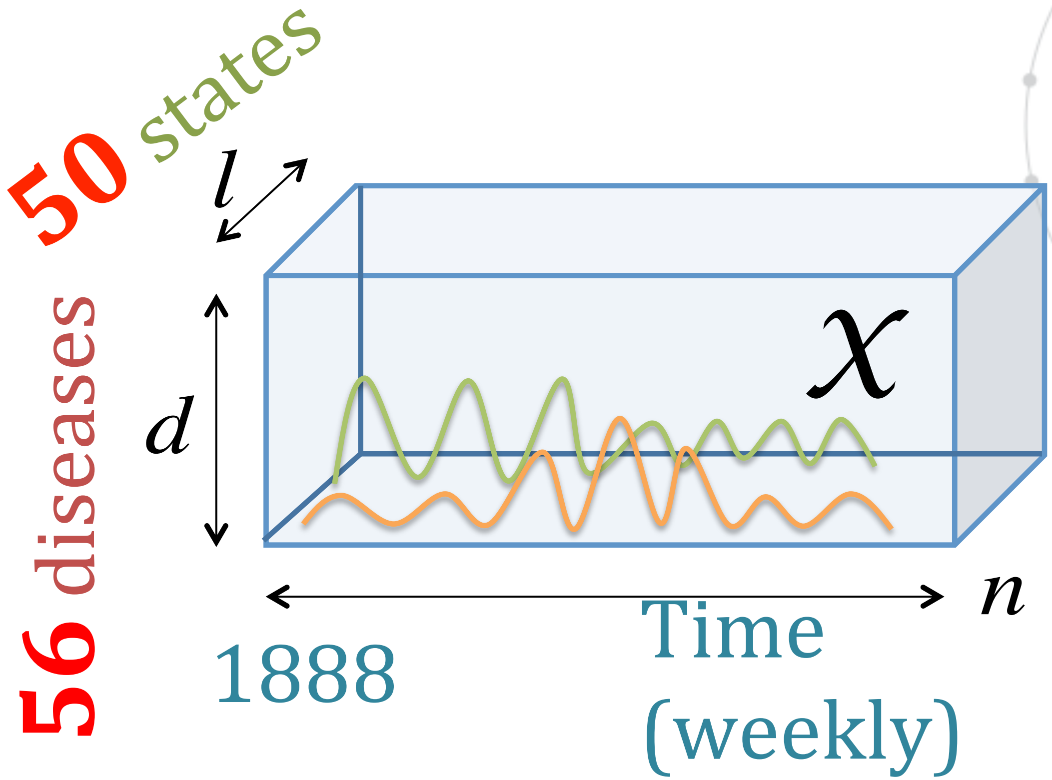


Kumamoto U

CMU CS

Data description

Project Tycho: infectious diseases in the U.S.



(> **125** years)

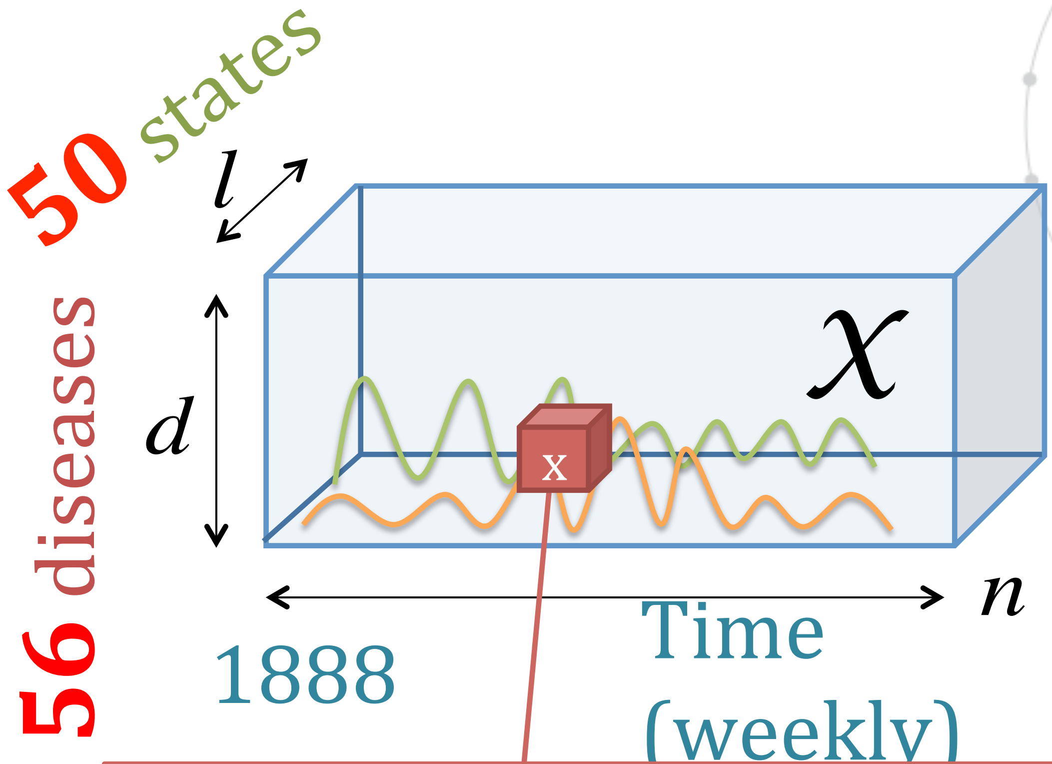
Kumamoto U

CMU CS

Data description

Project Tycho: infectious diseases in the U.S.

PROJECT
TYCHO
DATA FOR HEALTH



(> **125** years)

Element x : # of cases

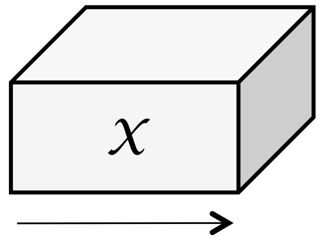
e.g., ‘measles’, ‘NY’, ‘April 1-7, 1931’, ‘4000’

http://www
~yasuko/T

Problem definition

Given:

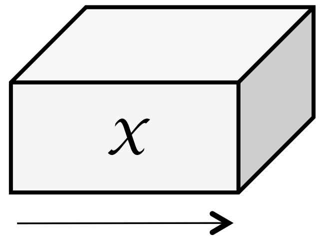
Tensor $\mathcal{X}$ (disease x state x time)



Problem definition

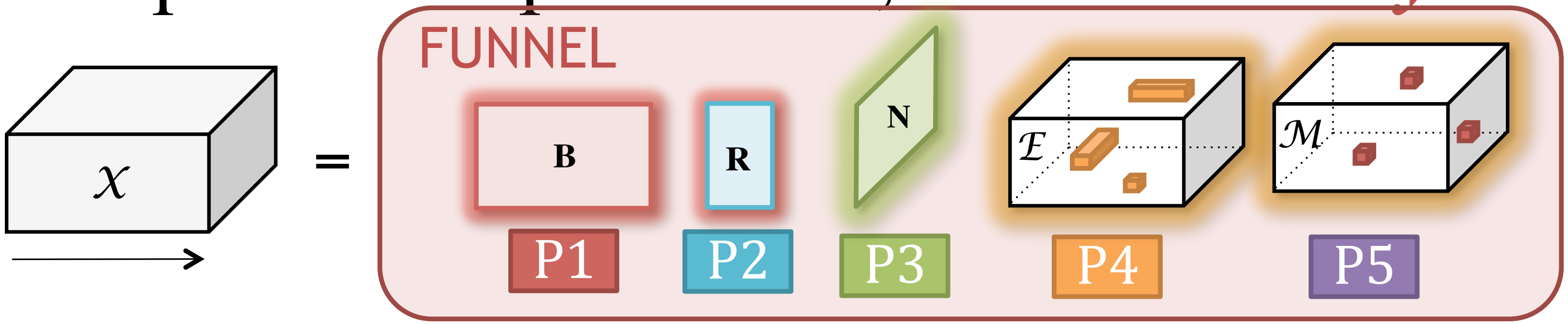
Given:

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



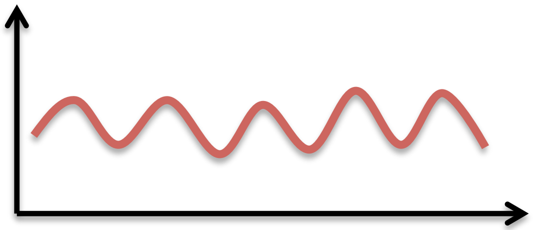
Find:

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

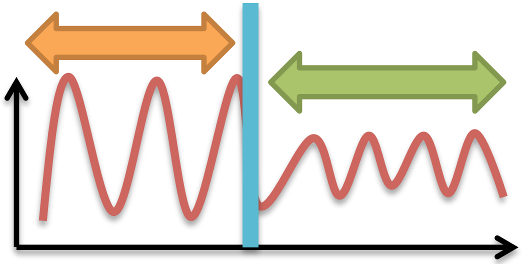


Problem definition

Seasonality



Discontinuities

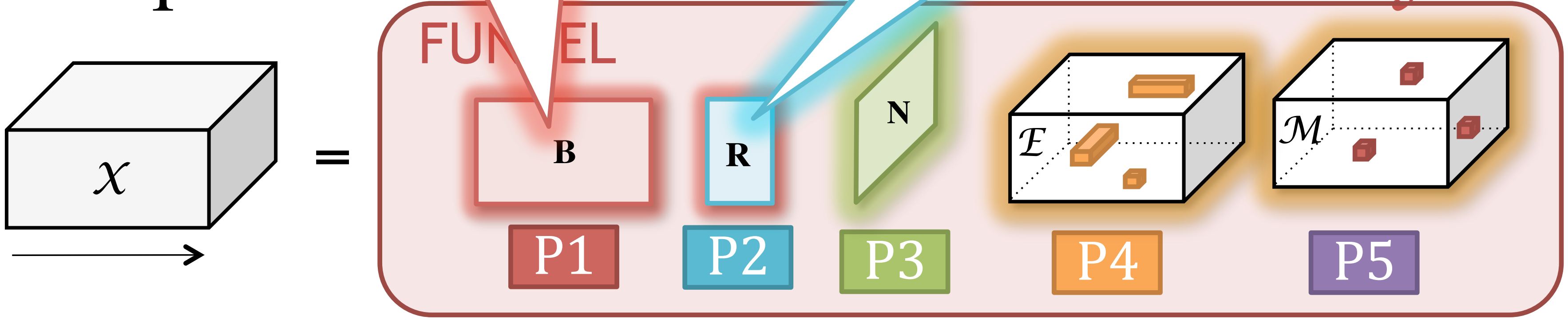


G

Te state

Find:

Compact desc tion of X automatically”



Kumamoto U

CMU CS

Problem definition

NO magic numbers !

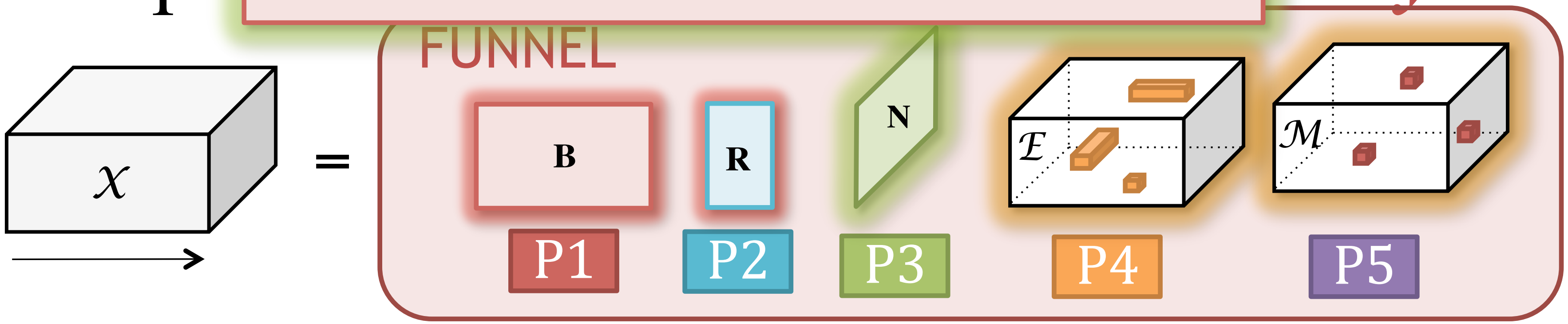
Parameter-free!

Given:

Tensor

Find:

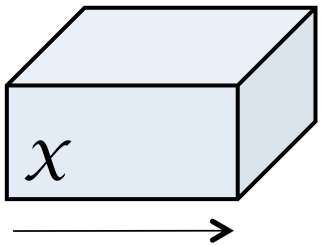
Compa… *…ically”*



Modeling power of FUNNEL

Questions about epidemics

Q1 Q2 Q3 Q4 Q5

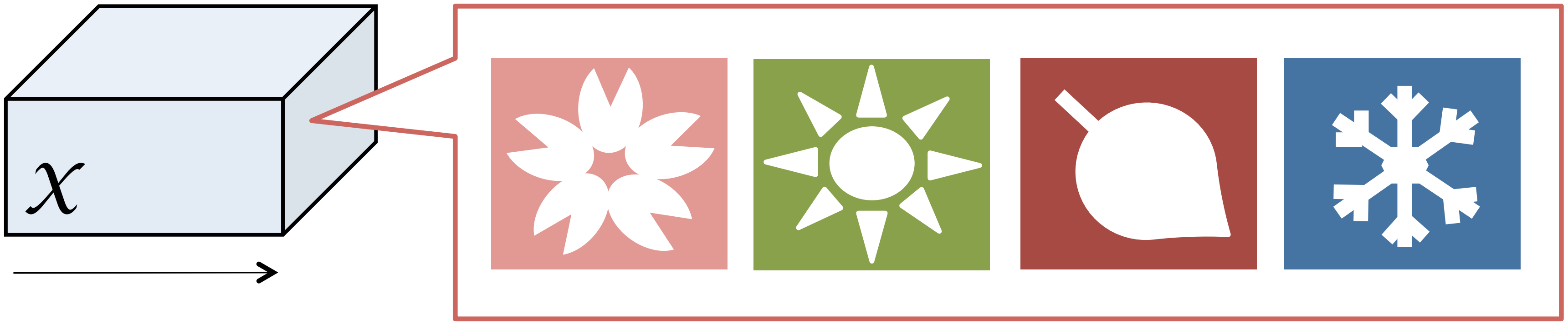


Questions

Q1

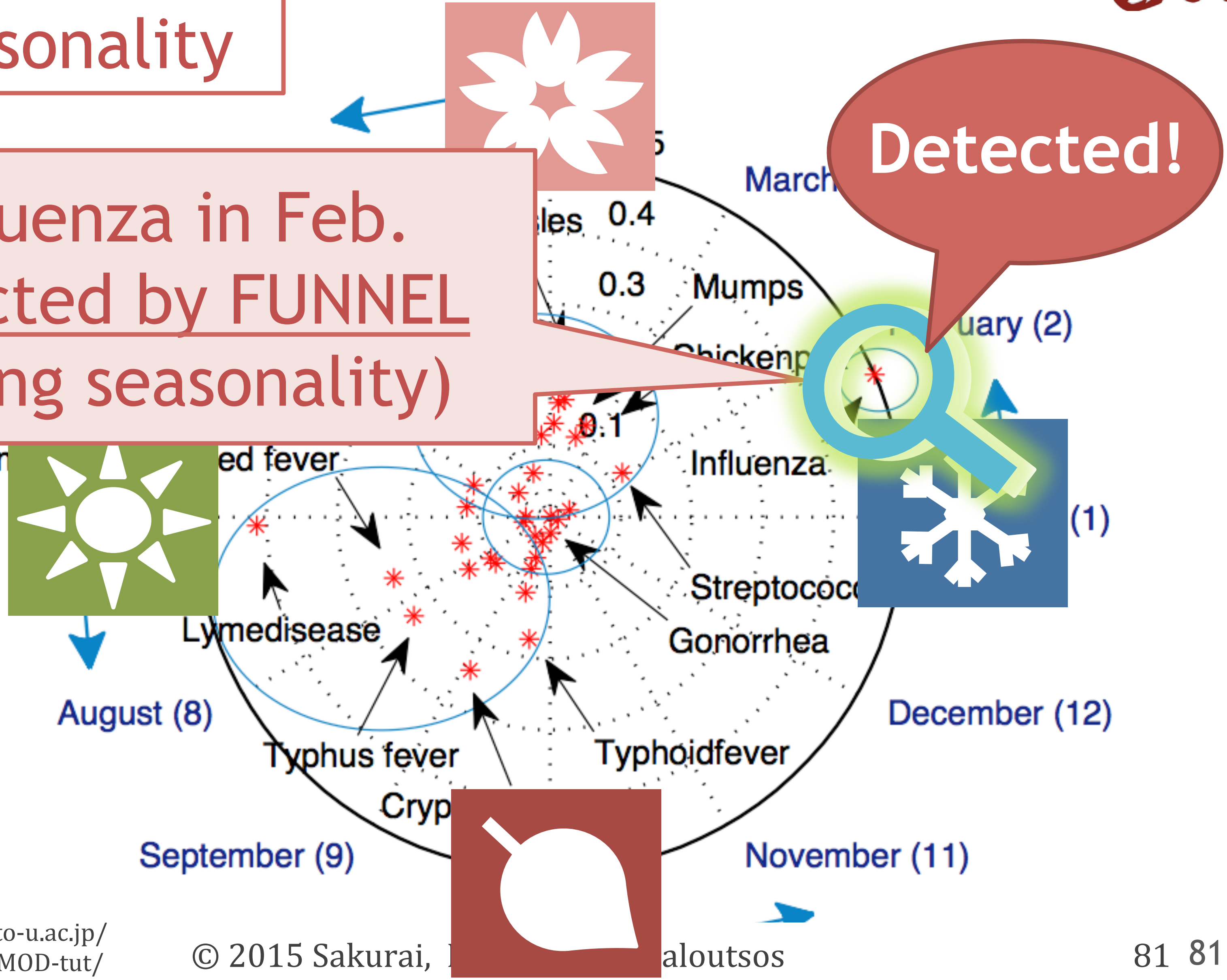
Are there any periodicities?

If yes, when is the peak season?



Answers

P1 Seasonality

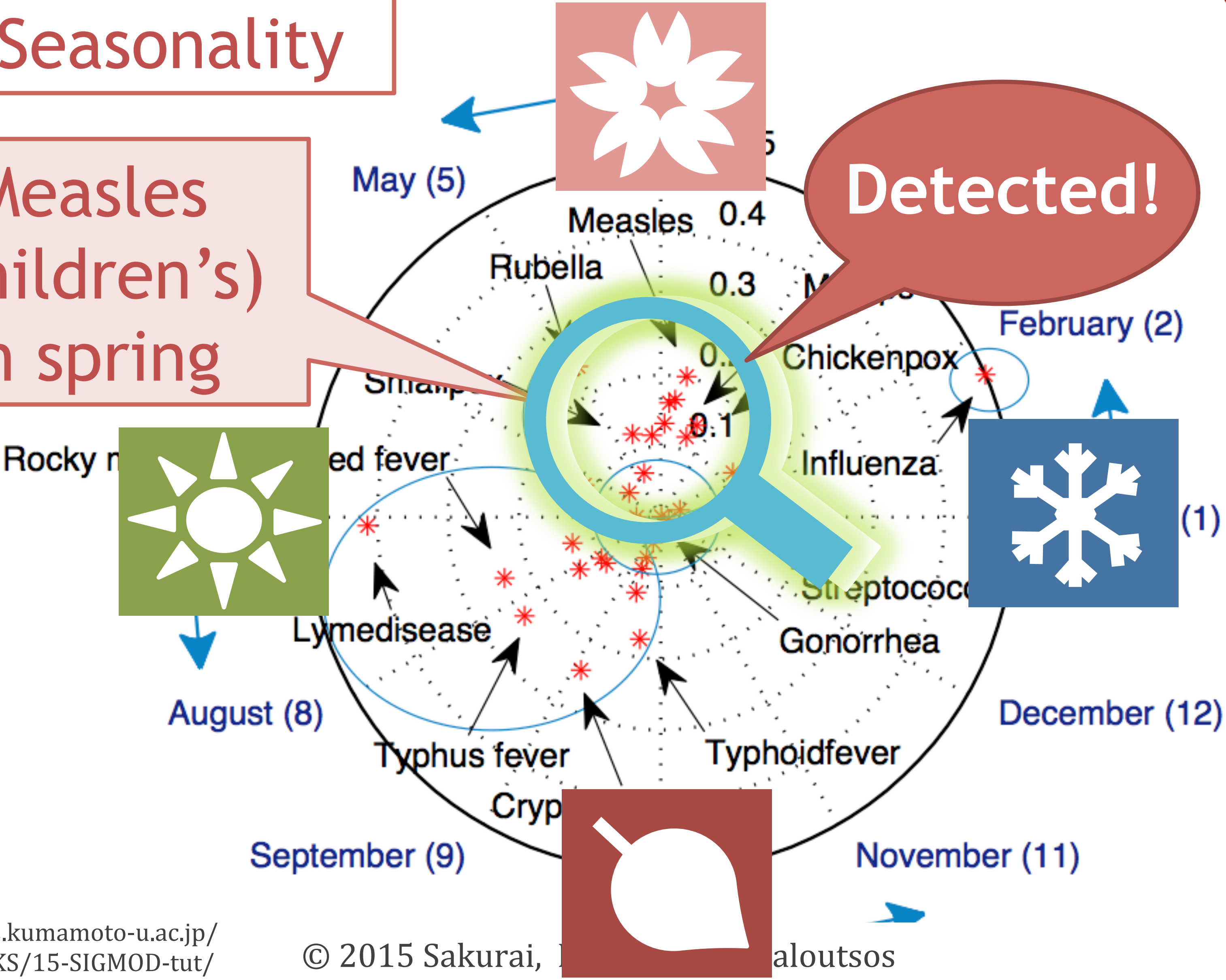


Influenza in Feb.
<u>Detected by FUNNEL</u>
(strong seasonality)

Answers

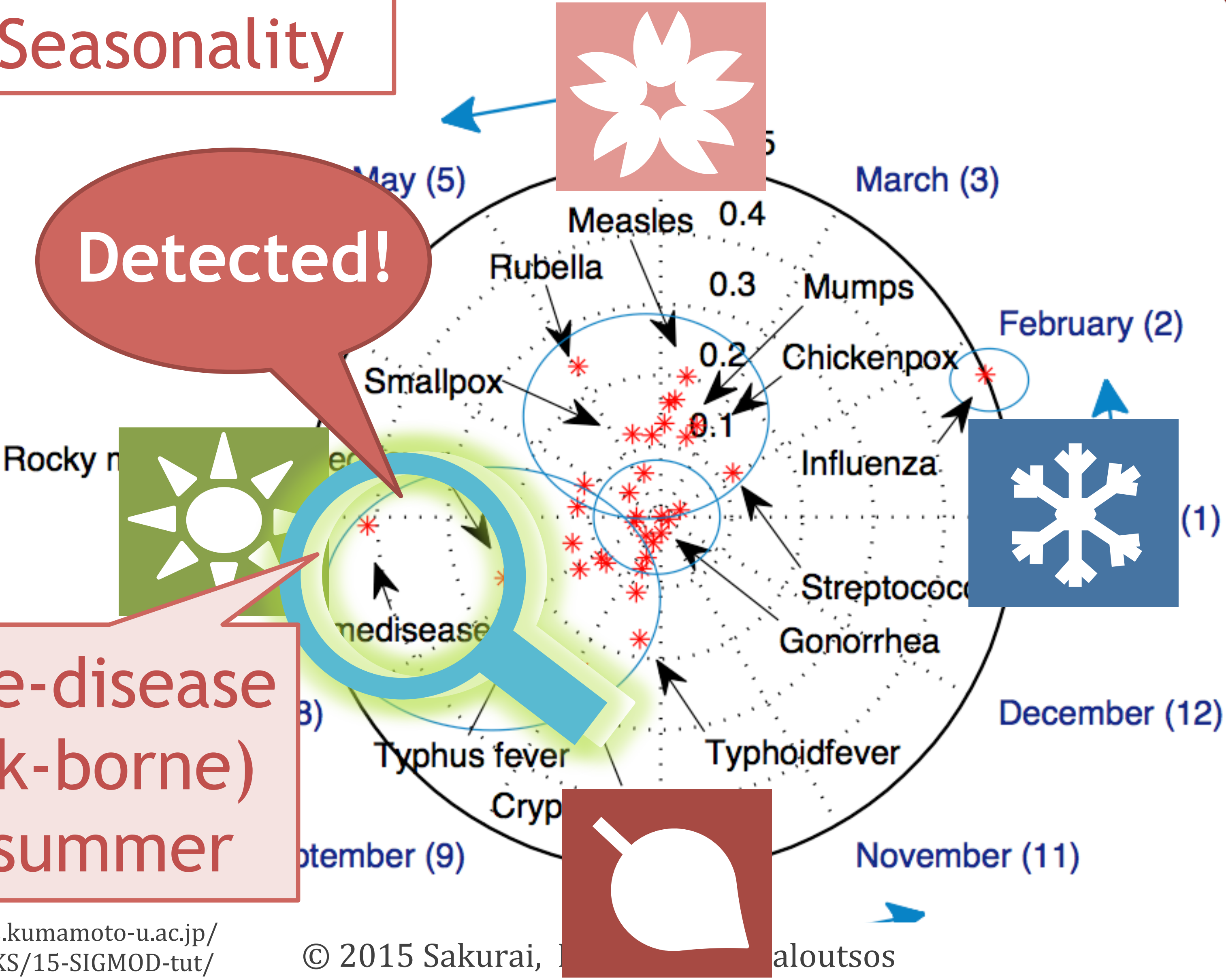
P1 Seasonality

Measles (children's) in spring



Answers

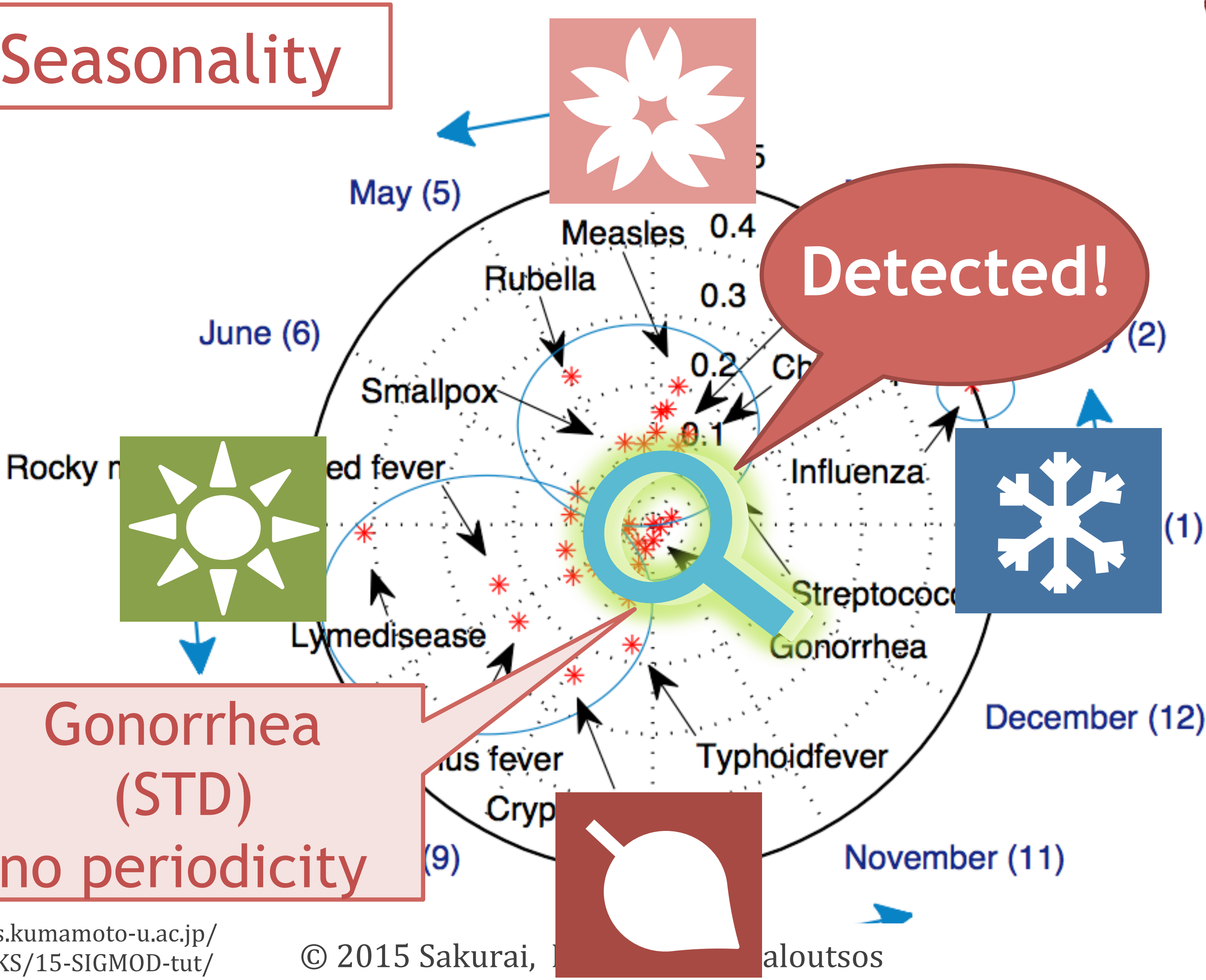
P1 Seasonality



Lyme-disease (tick-borne) in summer

Answers

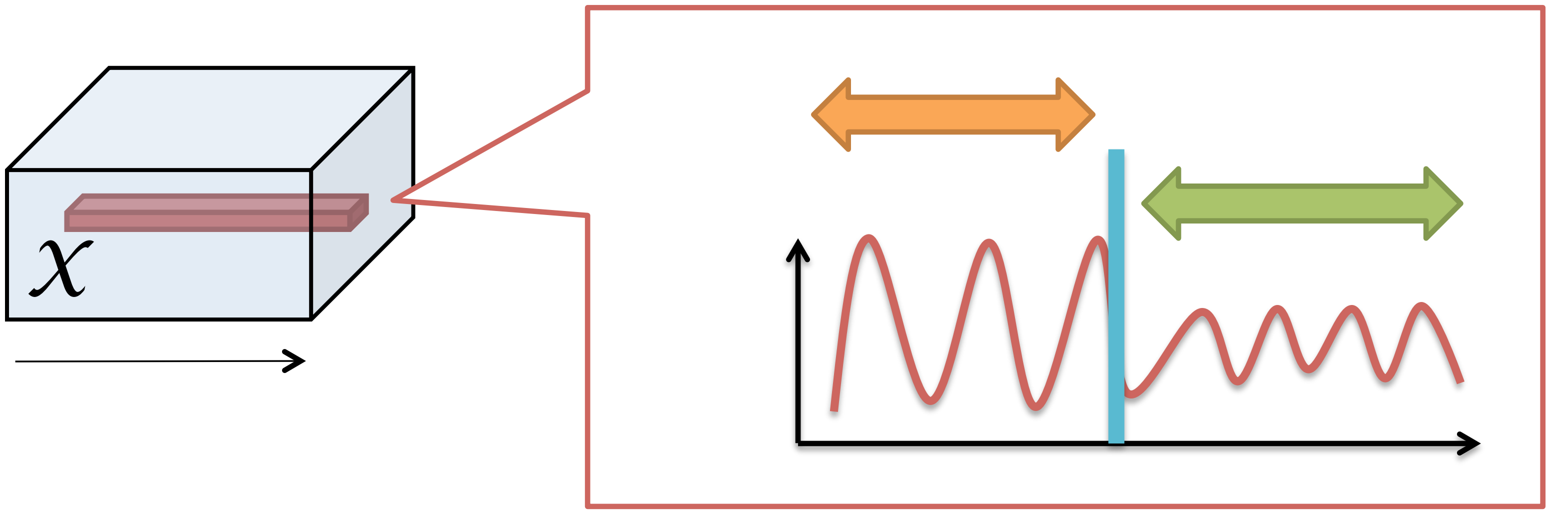
P1 Seasonality



Questions

Q2

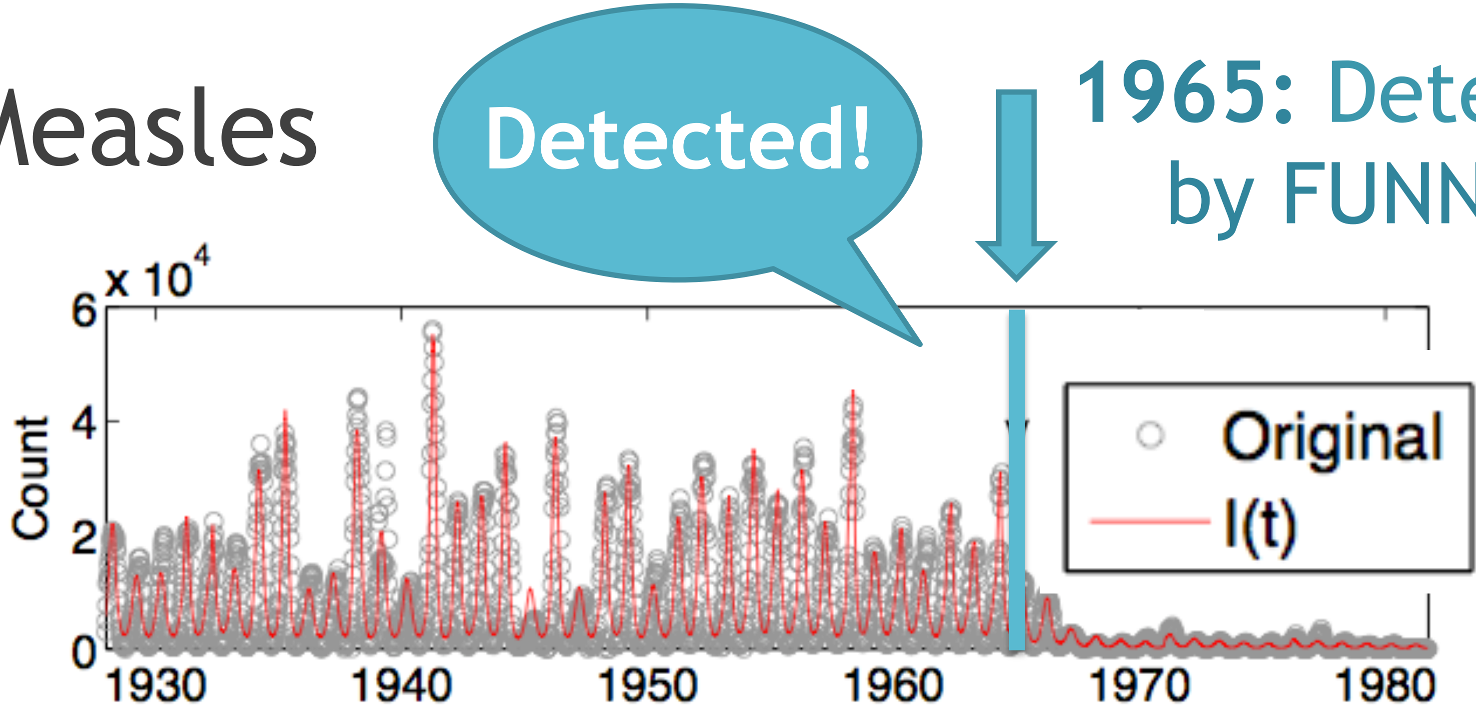
Can we see any discontinuities?



Answers

P2 Disease reduction effect

Measles



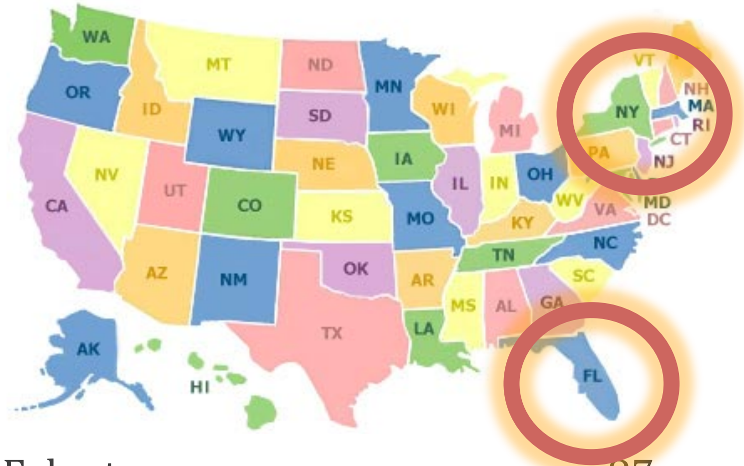
1965: Detected by FUNNEL

1963: Vaccine licensure

Questions

Q3

What's the difference between measles in NY and in FL?



Answers

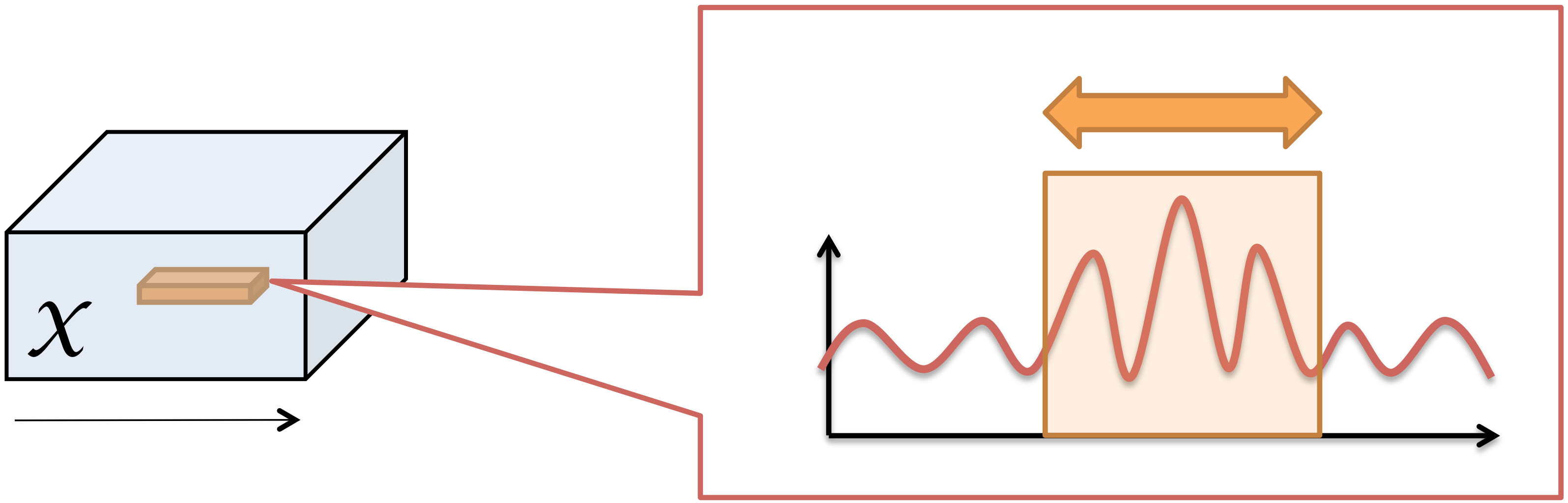
P3 Area sensitivity

FUNNEL's guess of susceptibles (measles)

Questions

Q4

Are there any external shock events, like wars?



Answers

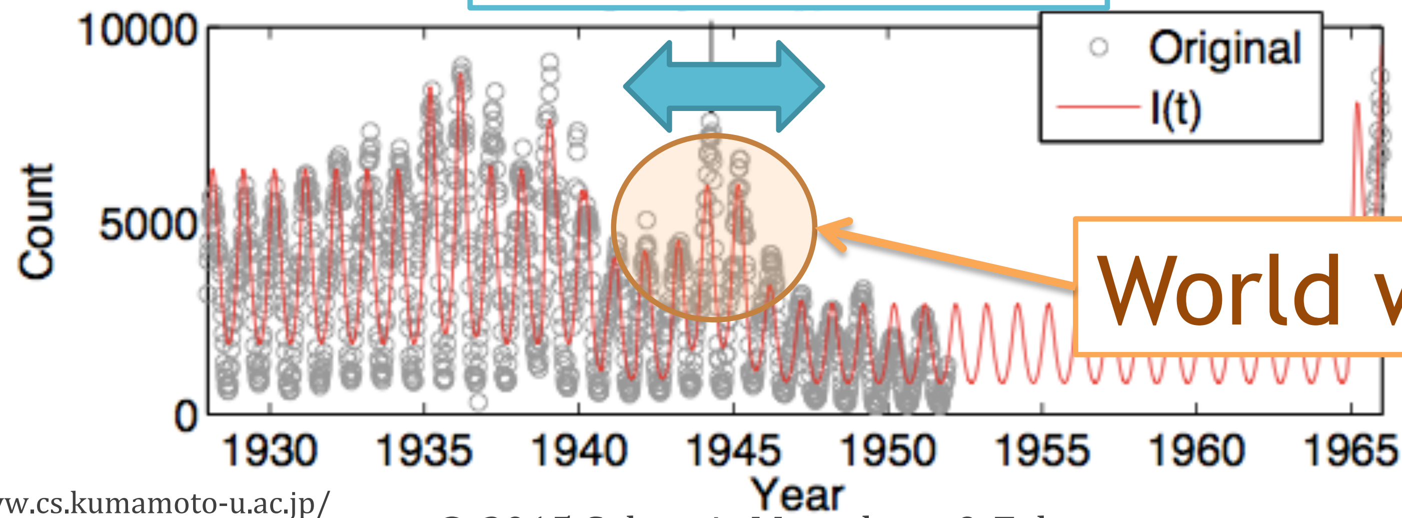
P4 External shock events

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

Scarlet fever

Detected by FUNNEL

Detected!

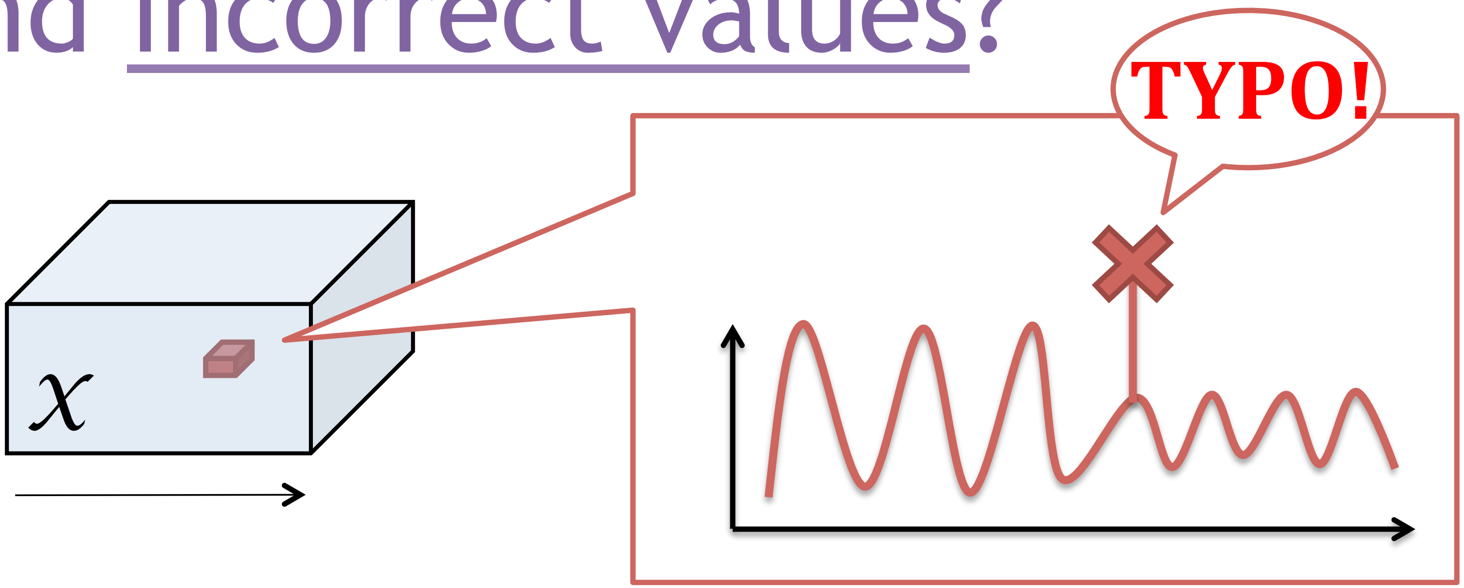


World war II

Questions

Q5

How can we remove mistakes and incorrect values?



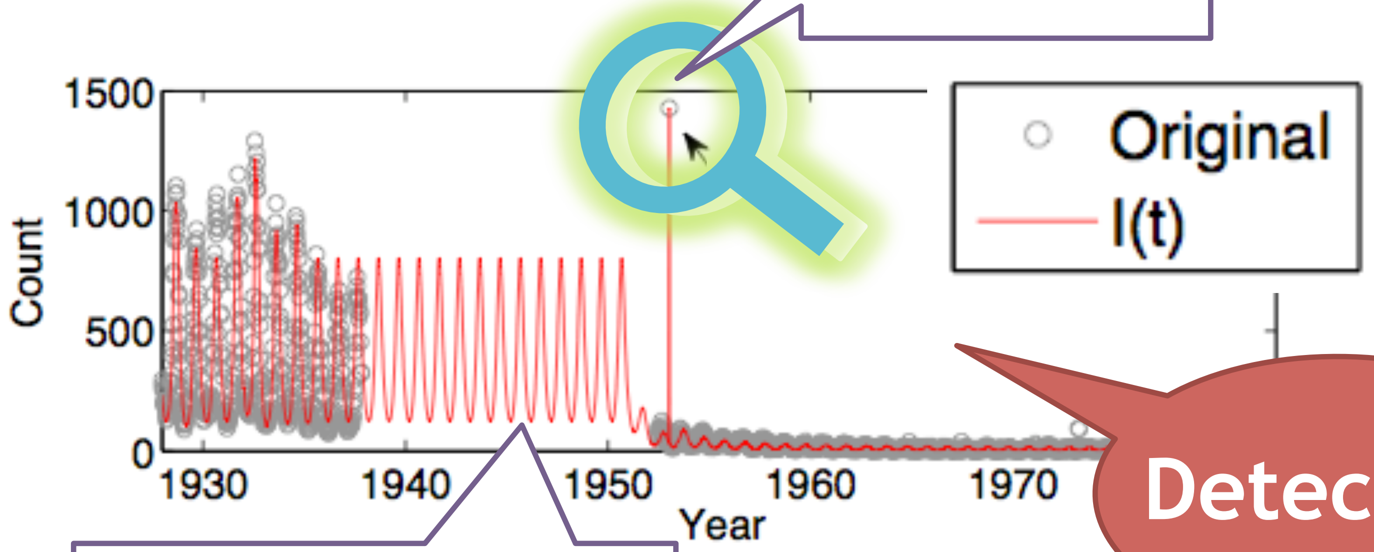
Answers

P5 Mistakes

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

Typhoid fever cases

Mistake



Detected!

Missing values

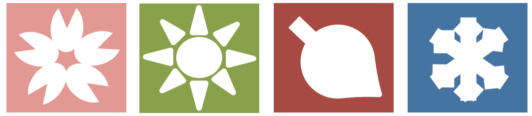
Kumamoto U

Modeling power of FUNNEL

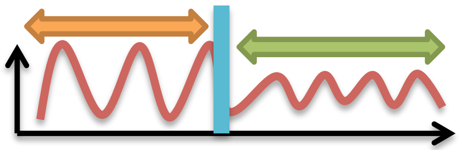
CMU CS

Our model can capture 5 properties

- **P1** Seasonality

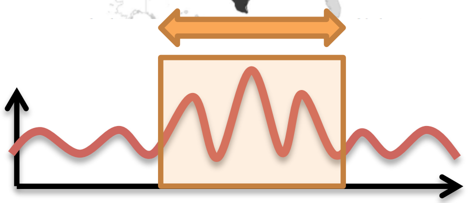


- **P2** Disease reductions



- **P3** Area sensitivity

- **P4** External events

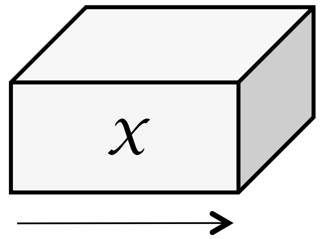


- **P5** Mistakes

Problem definition

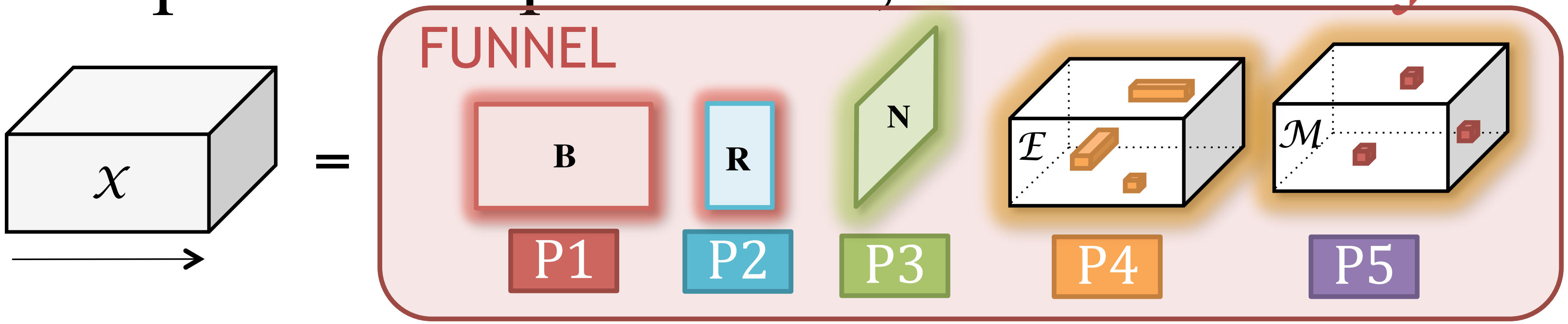
Given:

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



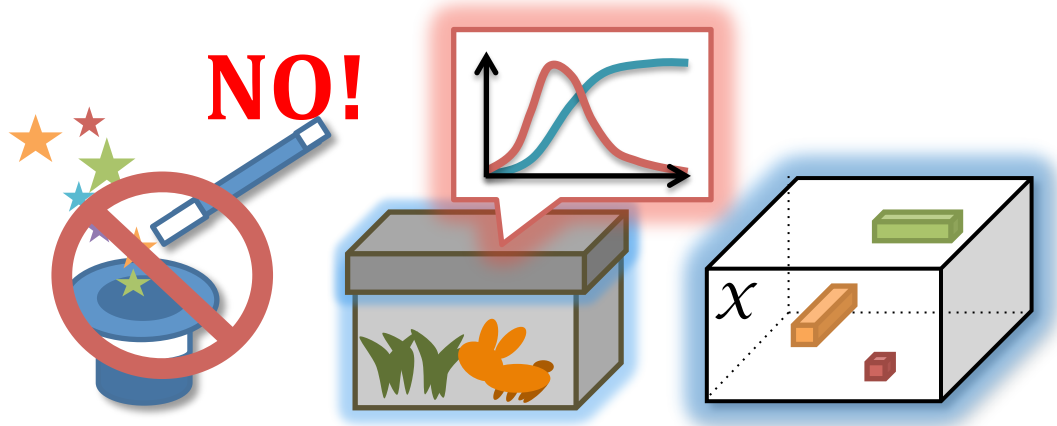
Find:

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



Main ideas

1. Automatic mining (no magic numbers!)
2. Non-linear (gray-box) modeling
3. Tensor analysis



M A N T

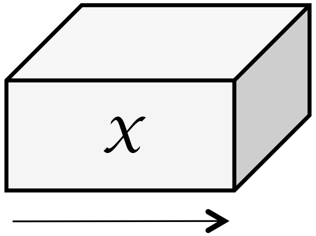
New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

Problem definition

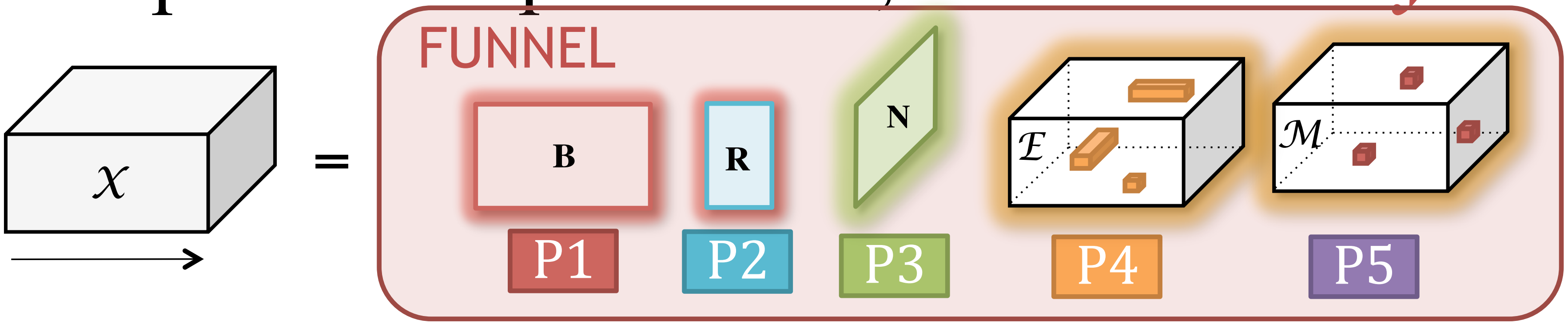
Given:

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

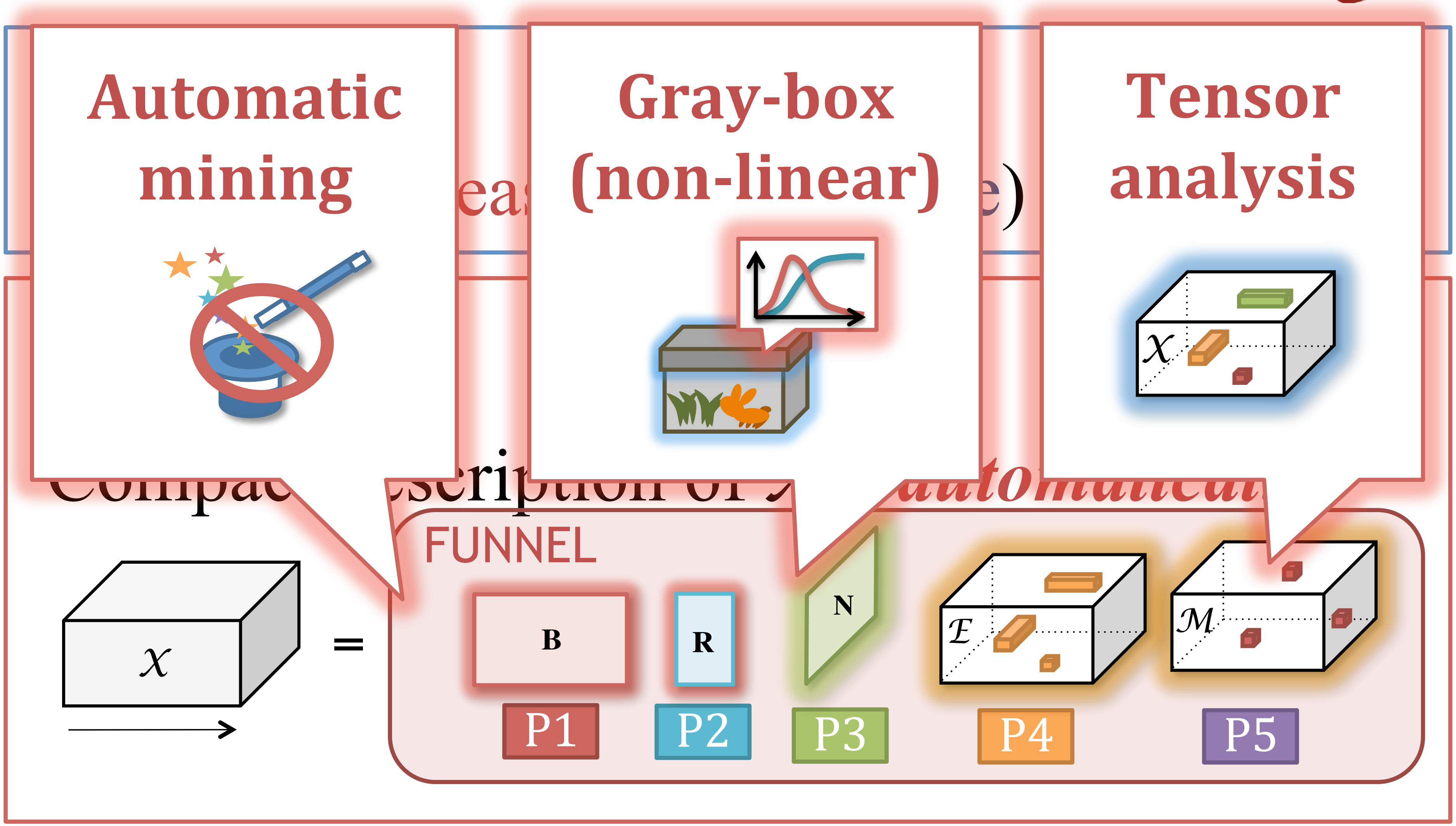


Find:

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

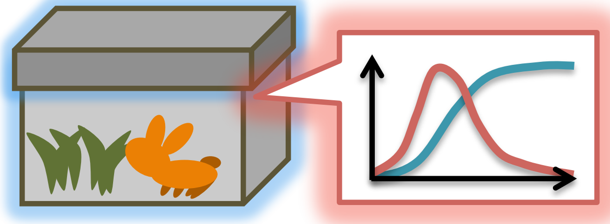


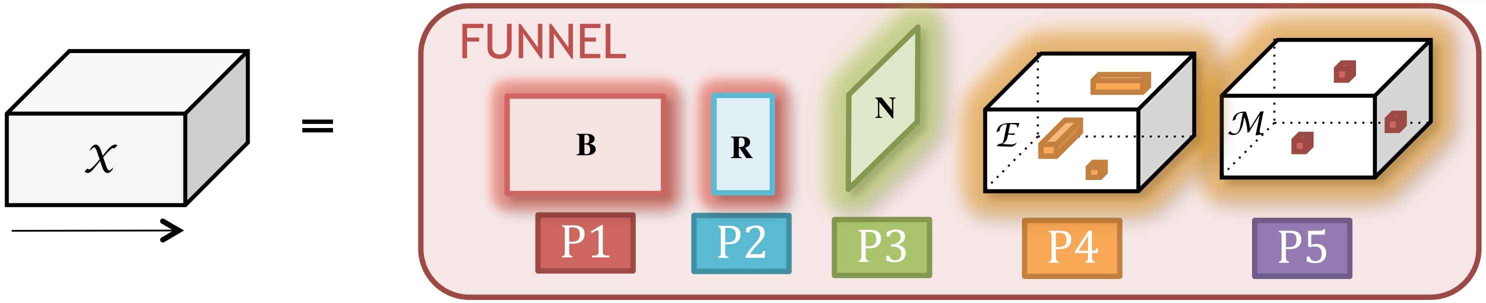
Problem definition



Two main ideas

Idea #1: Grey-box model



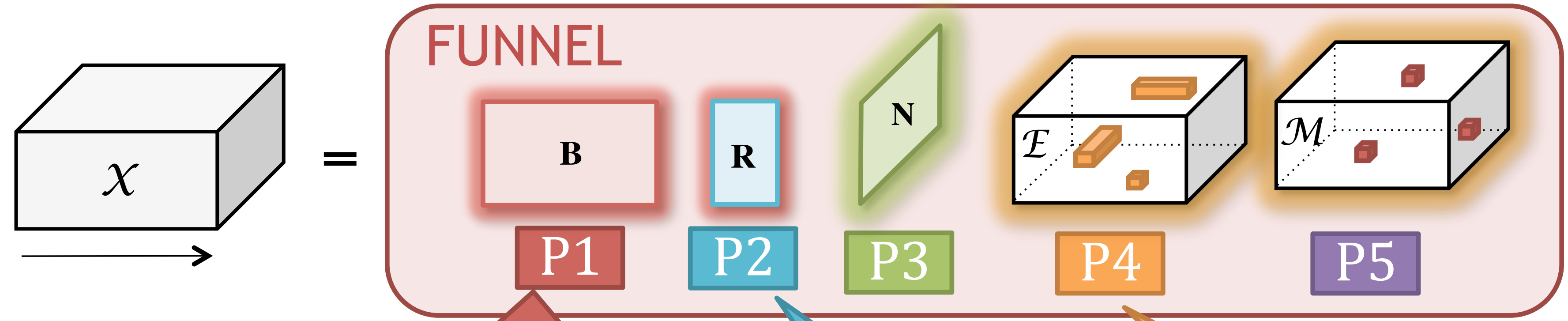


Idea #2: MDL for fitting

NO magic numbers ! (parameter-free)

Two main ideas

Idea #1: Grey-box model - domain knowledge

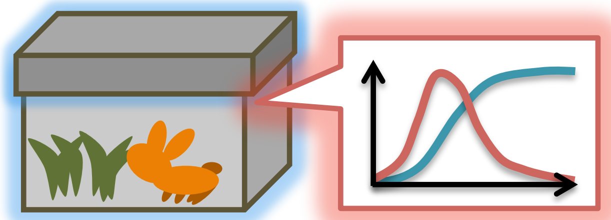


(SIRS+) : 6 parameters

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

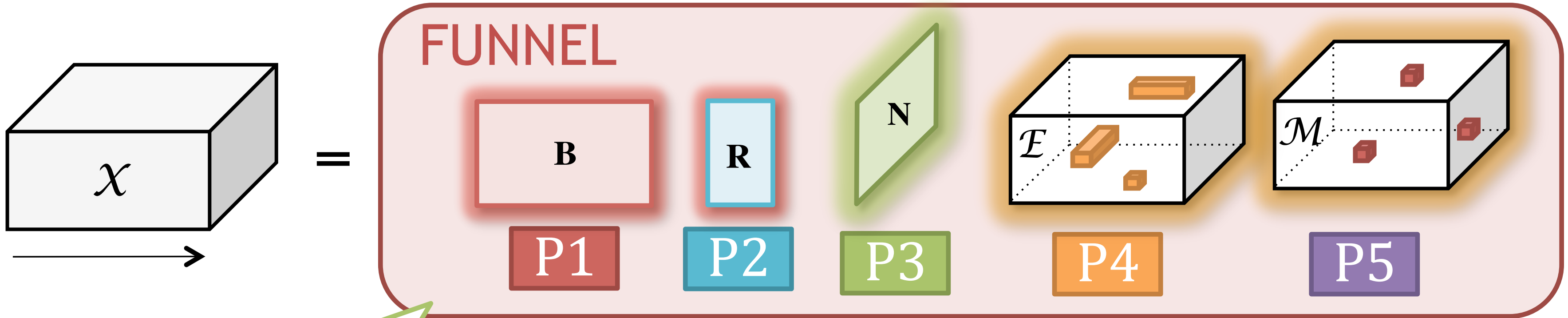
Vaccine

Shocks



Two main ideas

Idea #2: Fitting with MDL -> parameter free!



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

Cost function

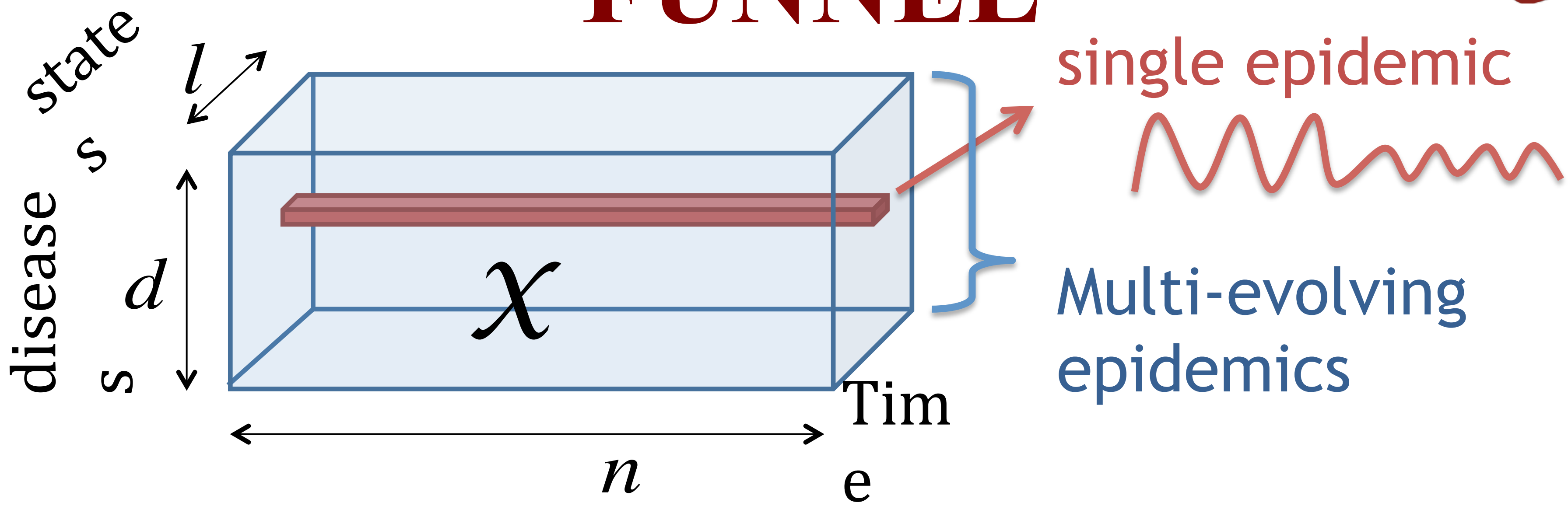
NO magic numbers

Parameter-free!

Kumamoto U

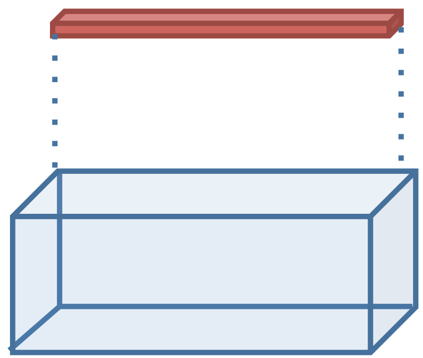
CMU CS

Proposed model: FUNNEL

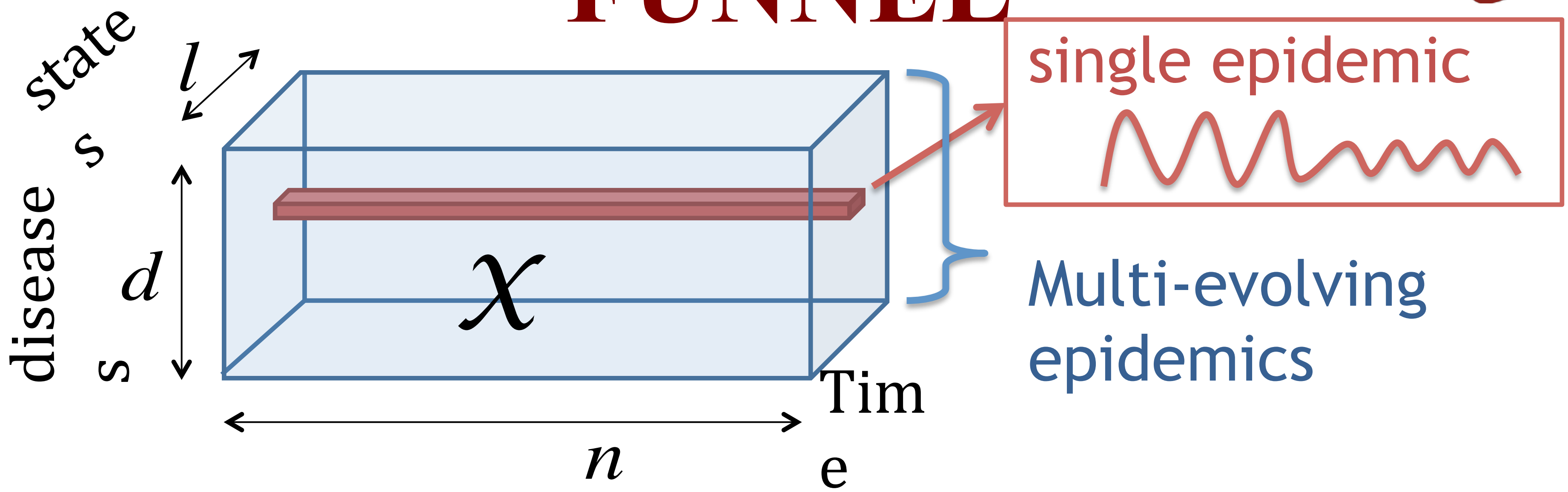


(a) FUNNEL-single

(b) FUNNEL-full (tensor)

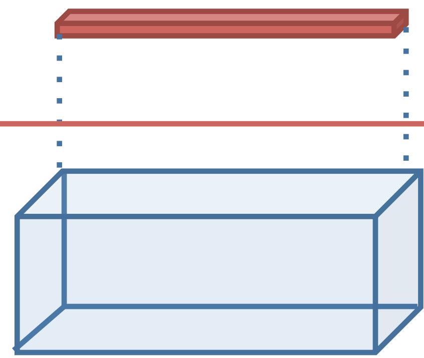


Proposed model: FUNNEL



(a) FUNNEL-single

(b) FUNNEL-full (tensor)

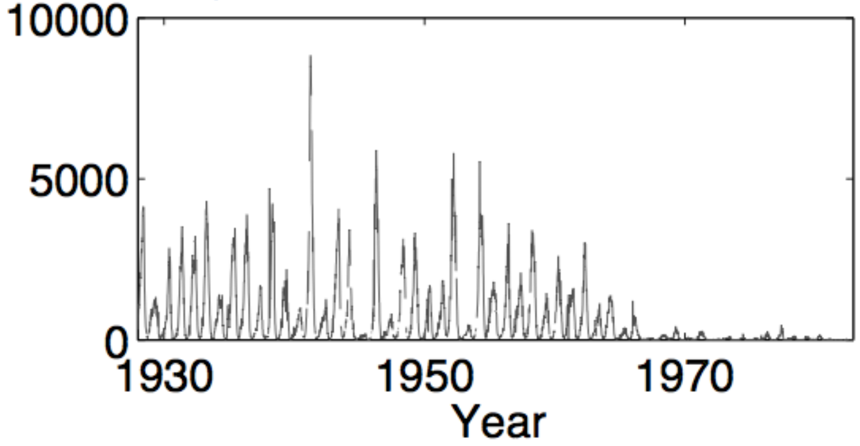


FUNNEL – with a single epidemic

Given:

"single" epidemic sequence

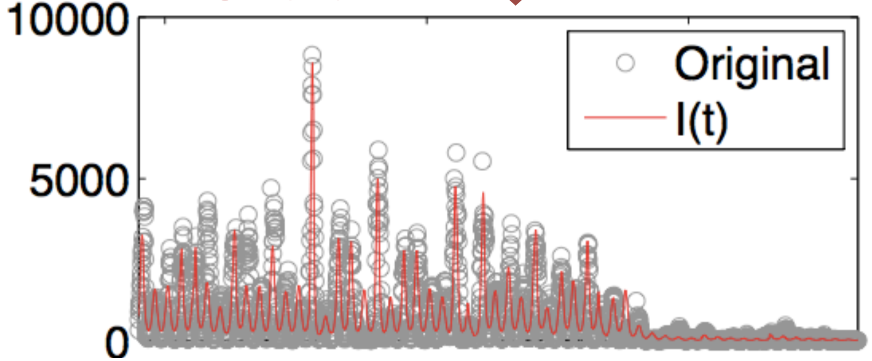
e.g., measles in NY



Find:

nonlinear equation, model parameters

FUNNEL

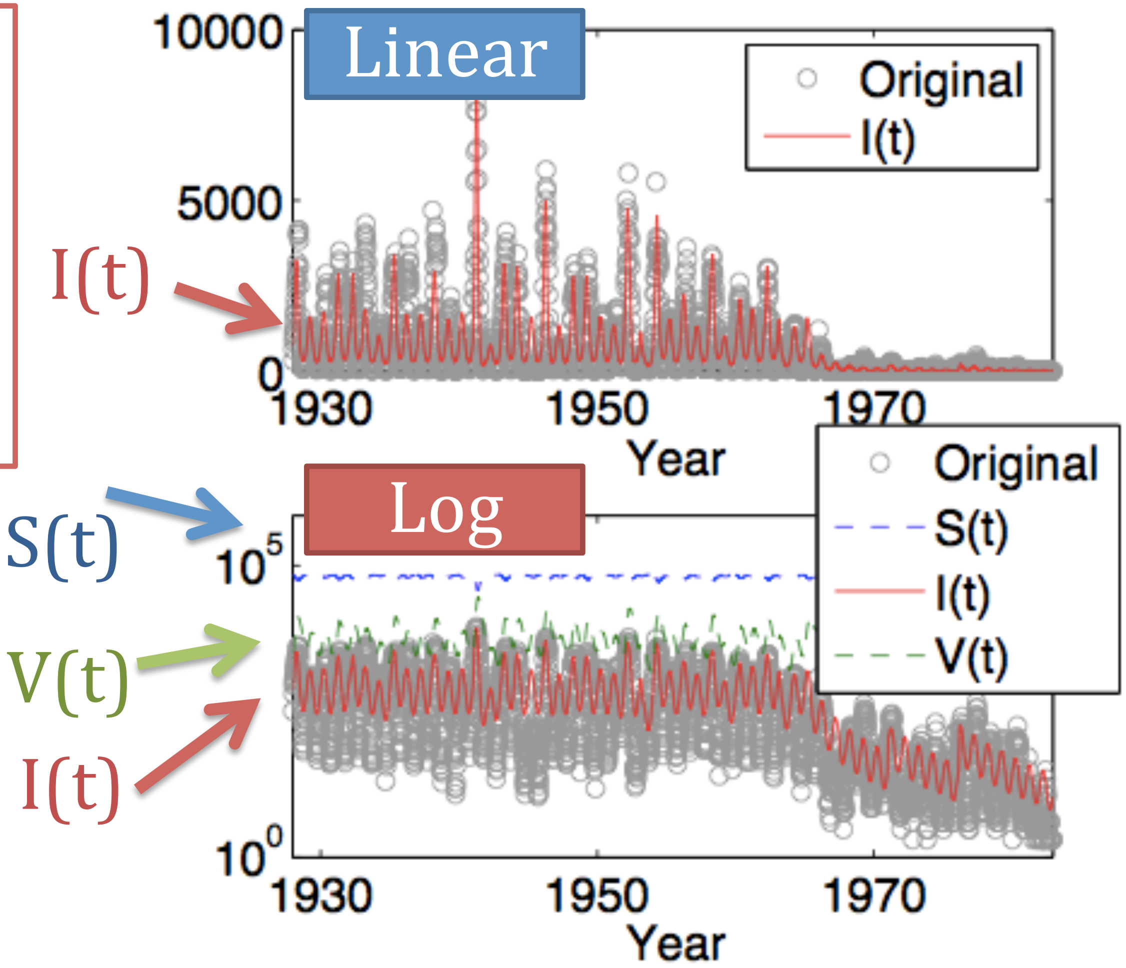


FUNNEL – with a single epidemic

Details

With a single epidemic: Funnel-RE

People of 3 classes
- S : Susceptible
- I : Infected
- V : Vigilant/ vaccinated



Kumamoto U

CMU CS

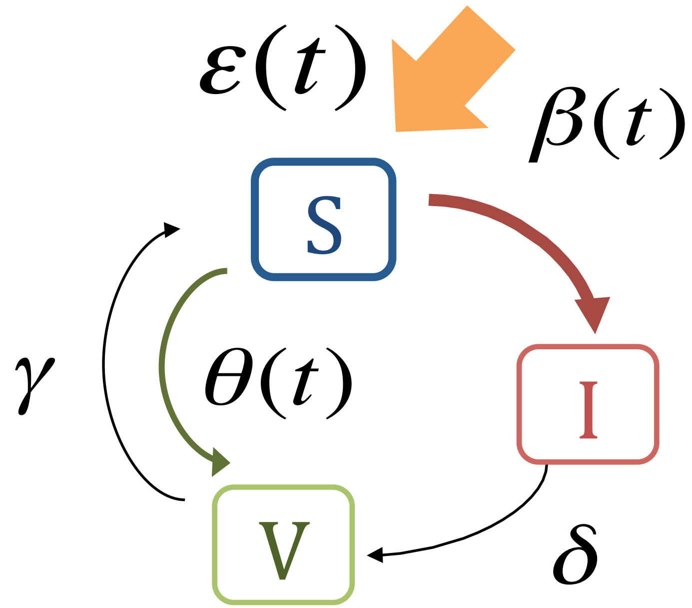
FUNNEL – with a single epidemic

Details

With a single epidemic: Funnel-RE

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

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



FUNNEL – with a single epidemic

Details

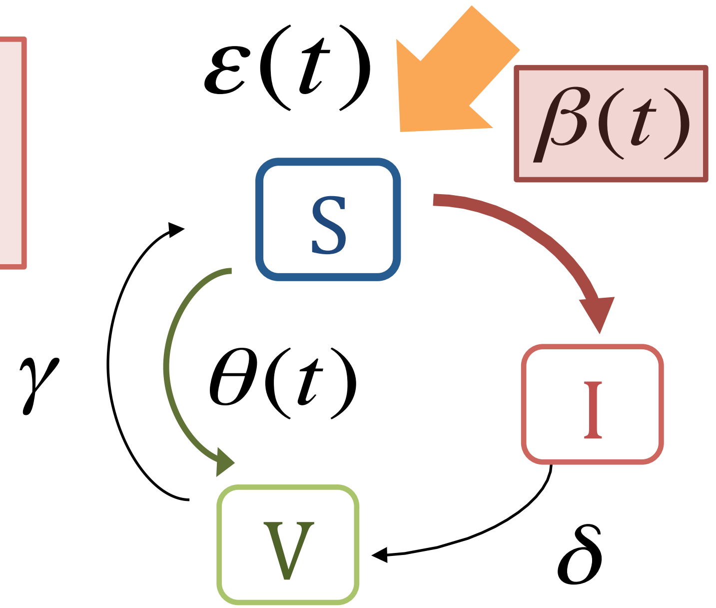
With a single epidemic: Funnel-RE

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

$\beta(t)$: 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

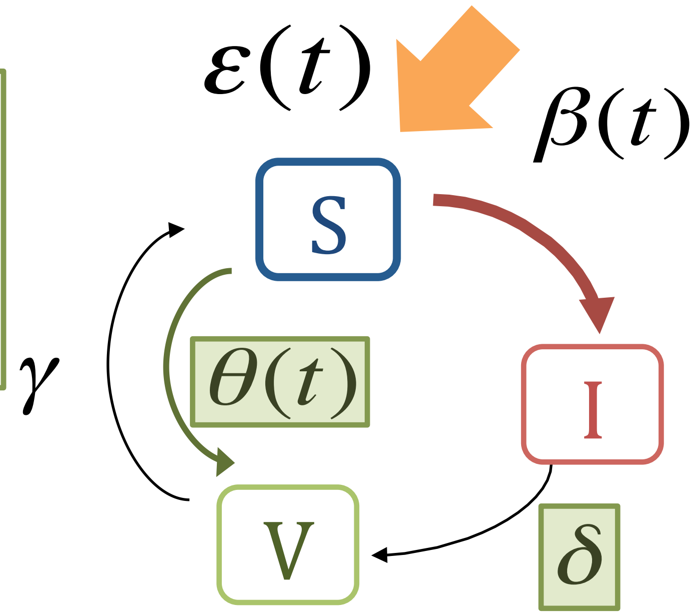
With a single epidemic: Funnel-RE

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

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

Kumamoto U

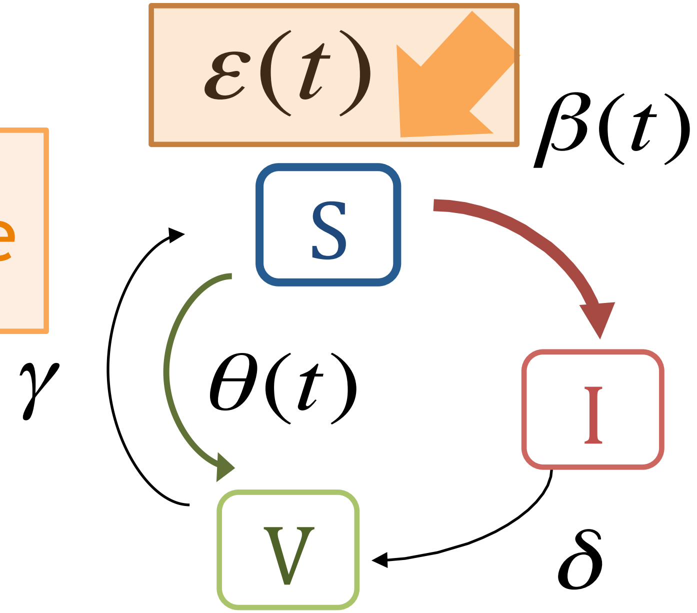
CMU CS

Details

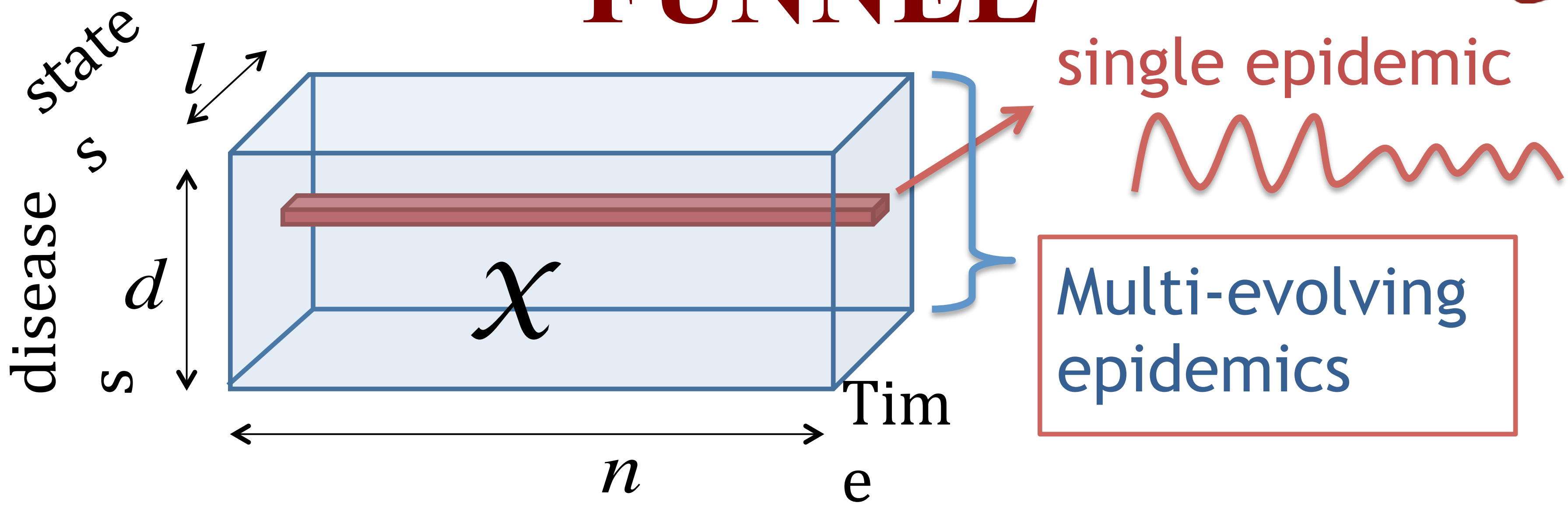
With a single epidemic: Funnel-RE

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

$\varepsilon(t)$: temporal susceptible rate

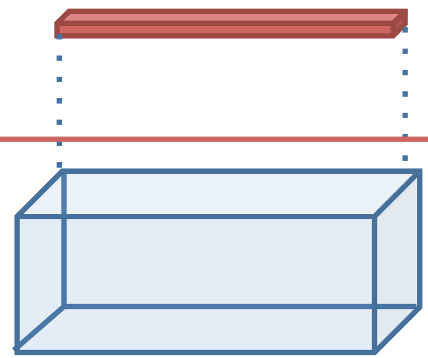


Proposed model: FUNNEL

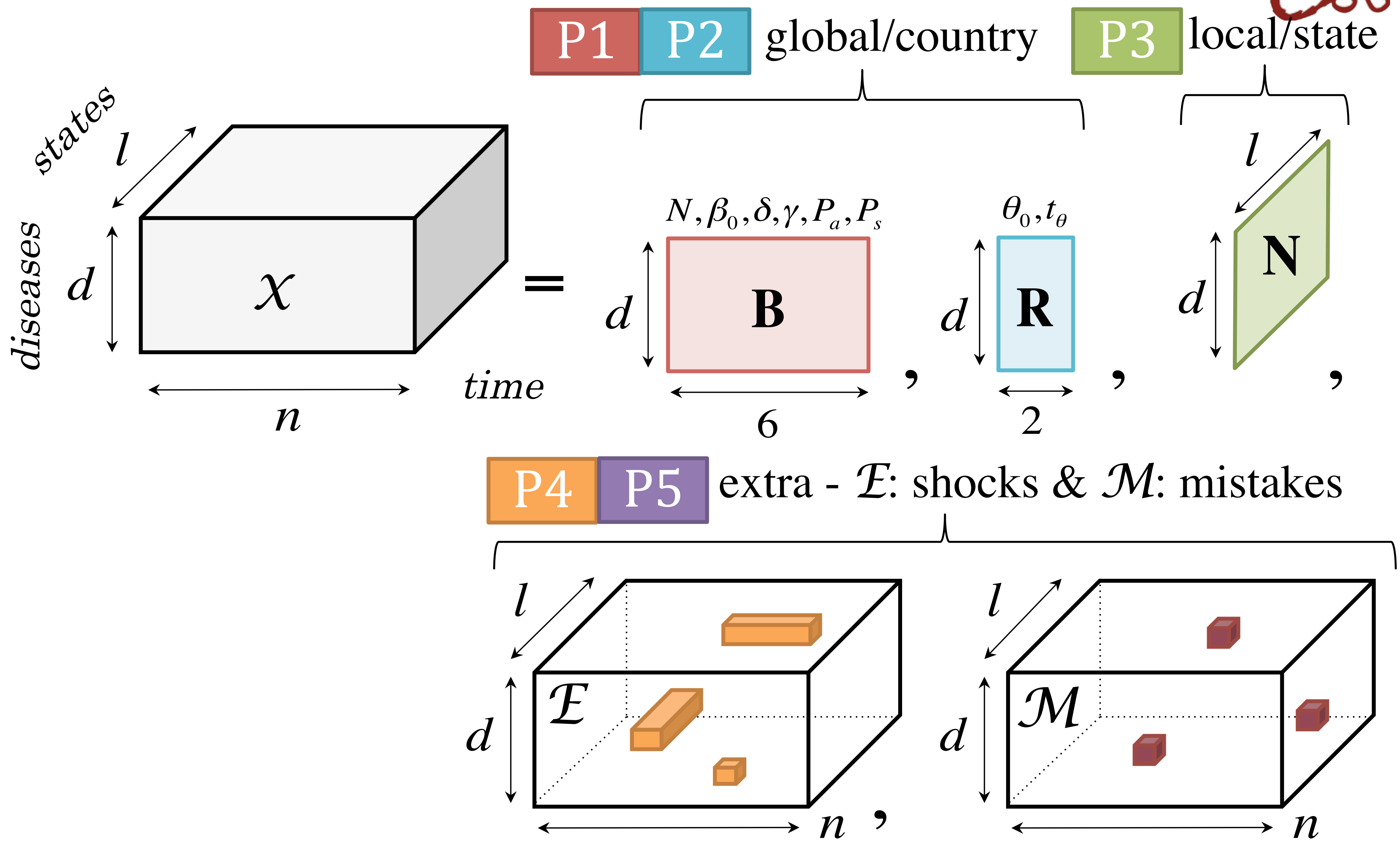


(a) FUNNEL-single

(b) FUNNEL-full (tensor)

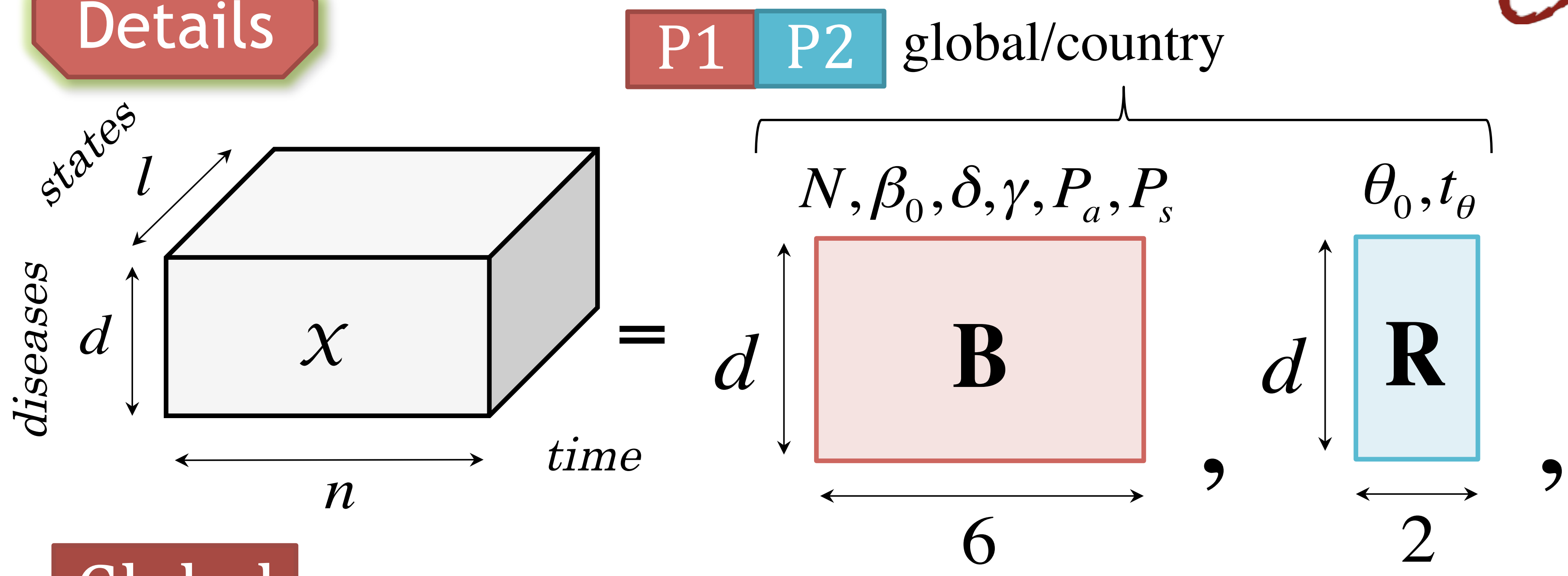


FUNNEL-full


P1 P2 global/country
P3 local/state
states
diseases
l
d
X
=
time
n
N, β0, δ, γ, Pa, Ps
d
B
6
θ0, tθ
d
R
2
l
d
N
P4 P5 extra - ℰ: shocks & ℳ: mistakes
l
d
ℰ
n
l
d
ℳ
n

FUNNEL-full

Details



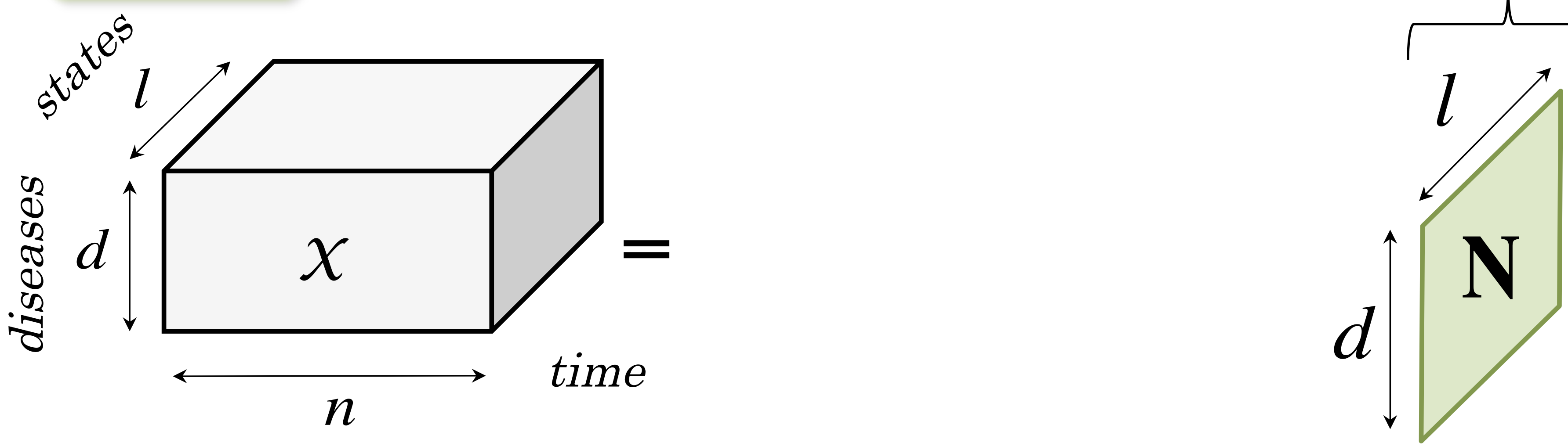
Global

- P1 Base matrix **B** (d x 6)
- P2 Disease reduction matrix **R** (d x 2)

FUNNEL-full

Details

P3 local/state



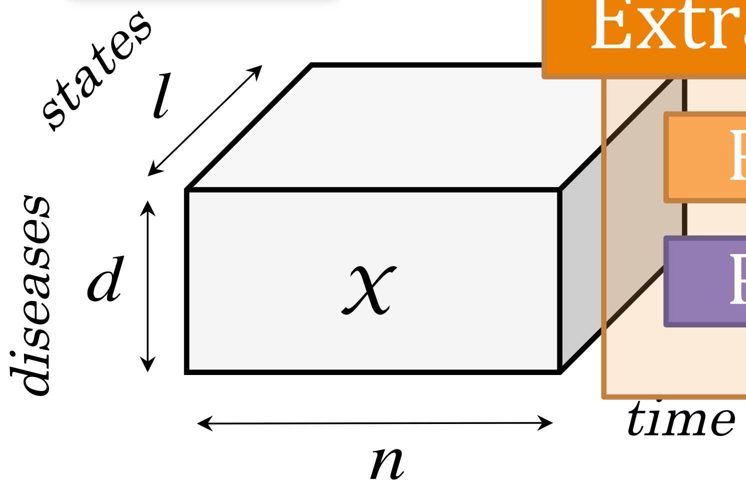
Local

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

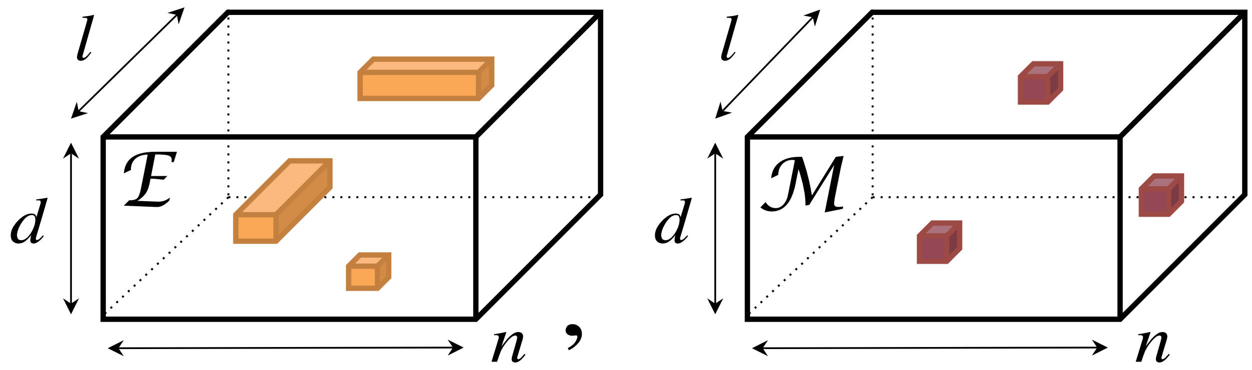


Extra

P4 External shock tensor $\mathcal{E}$

P5 Mistake tensor $\mathcal{M}$

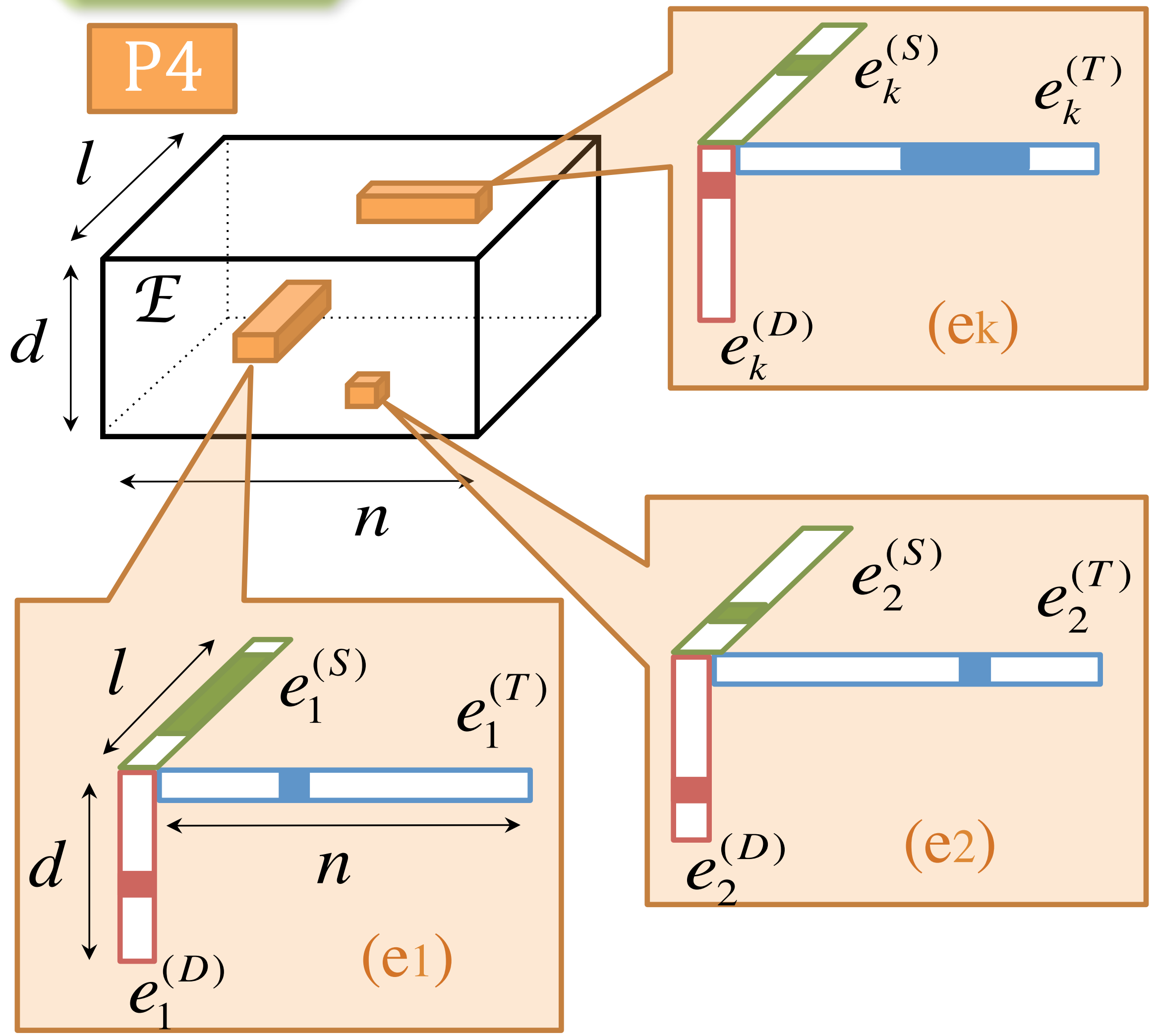
P4 P5 extra - $\mathcal{E}$: shocks & $\mathcal{M}$: mistakes

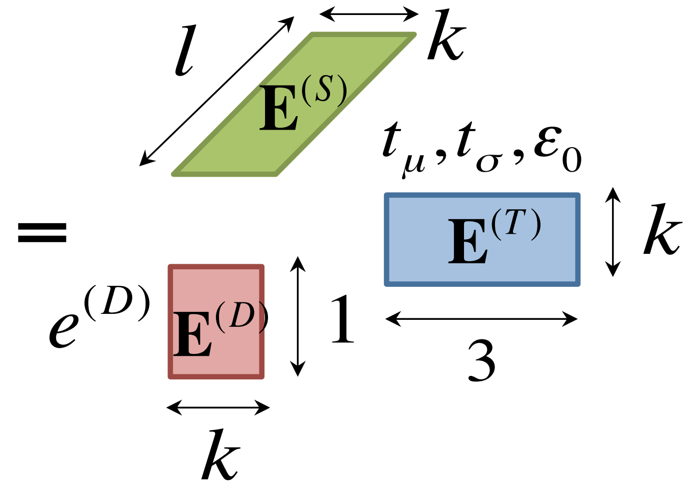


FUNNEL-full

Details

P4





$$\mathcal{E} = \{\mathbf{E}^{(D)}, \mathbf{E}^{(T)}, \mathbf{E}^{(S)}\}$$

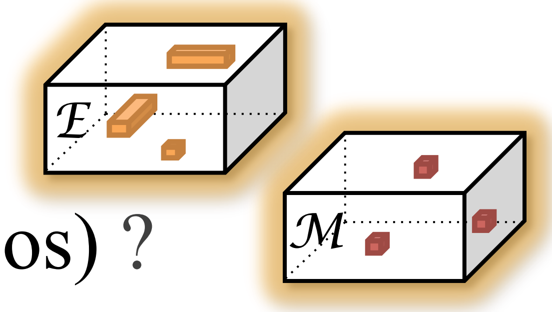
Disease matrix

Time matrix

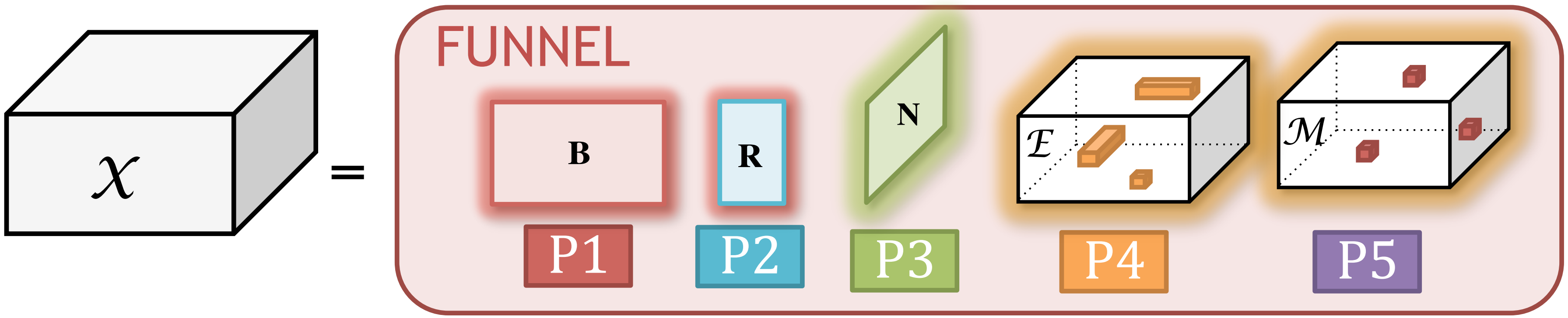
State matrix

Challenges

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

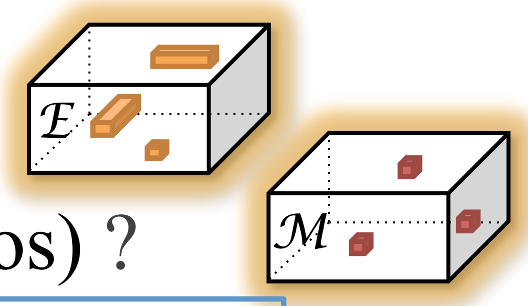


Q2. How to efficiently estimate model parameters ?



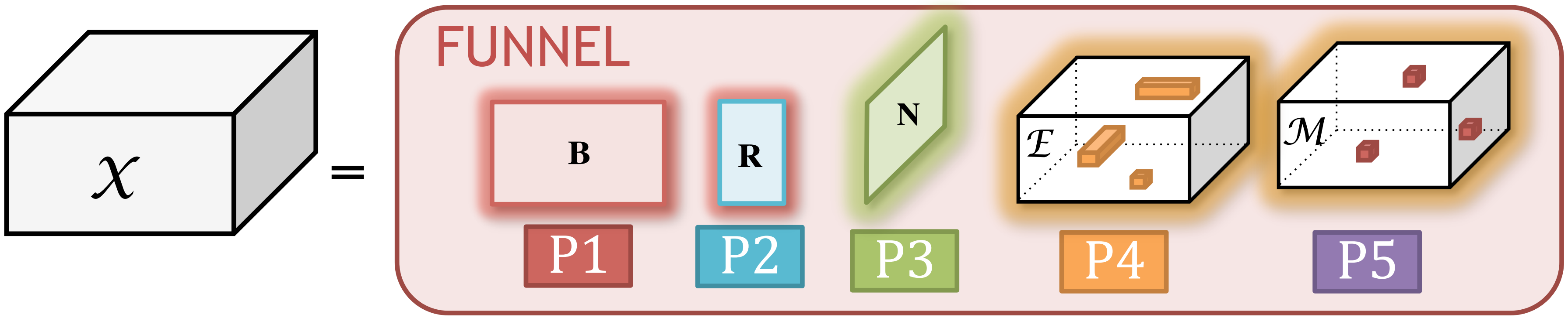
Challenges

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



Idea (1) : Model description cost

Q2. How to efficiently estimate model parameters ?



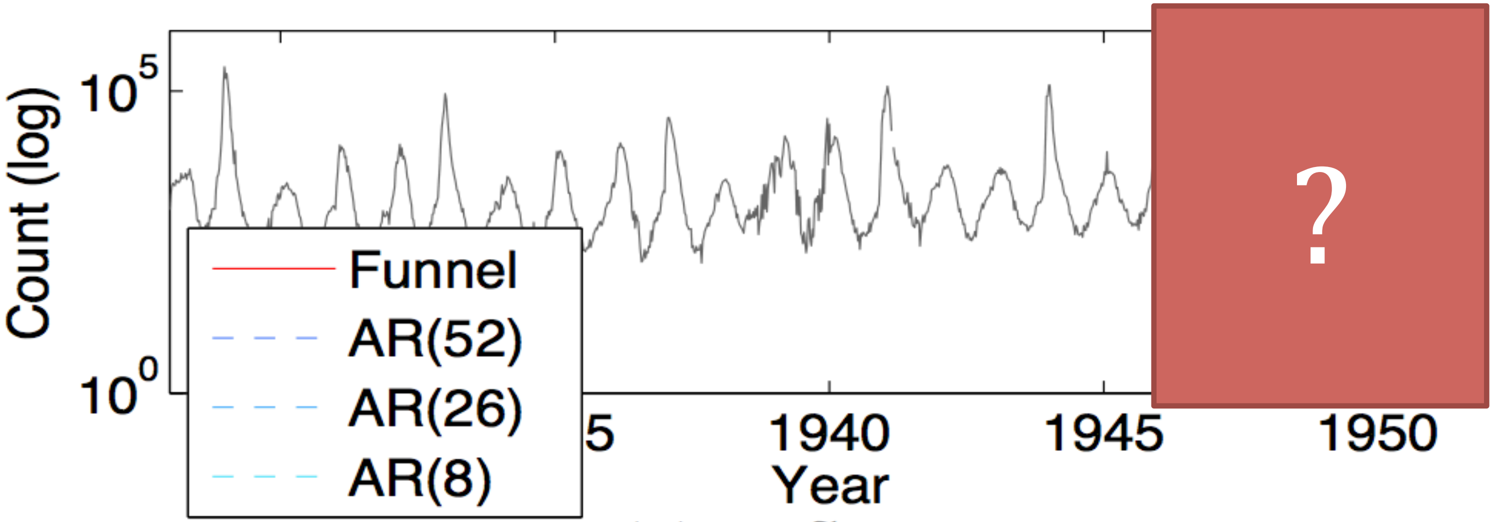
Idea (2): Multi-layer optimization - *O(dln)*

FUNNEL at work - forecasting

Forecasting future epidemics

Train:
2/3 sequences

Forecast:
1/3 following years



(a) Influenza

Kumamoto U

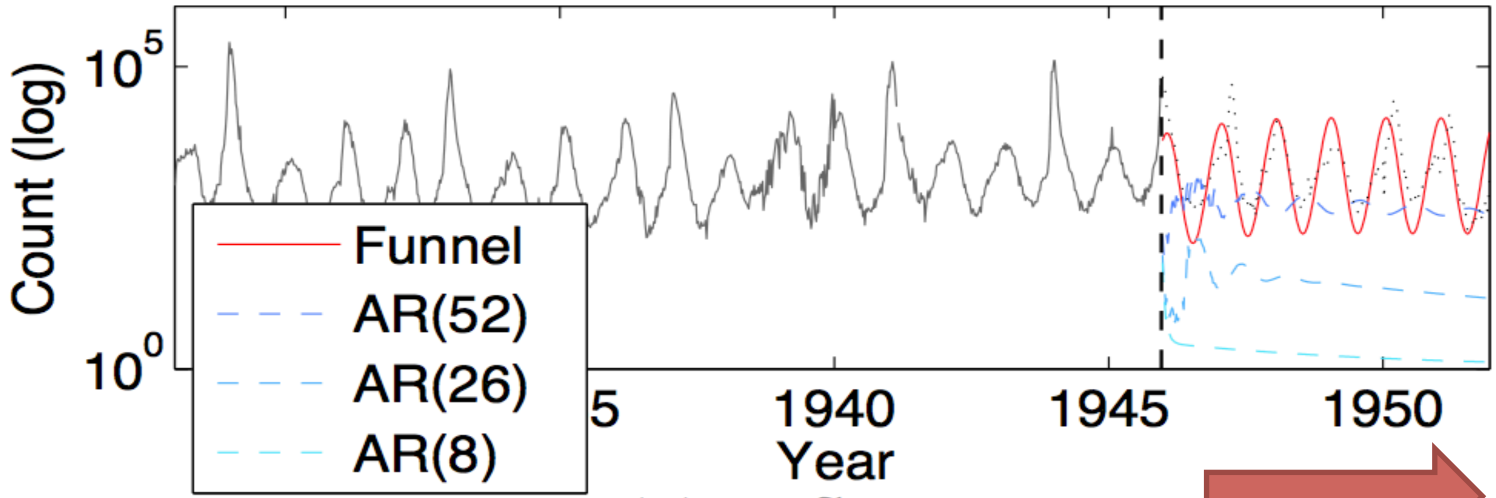
FUNNEL at work - forecasting

CMU CS

Forecasting future epidemics

Train:
2/3 sequences

Forecast:
1/3 following years



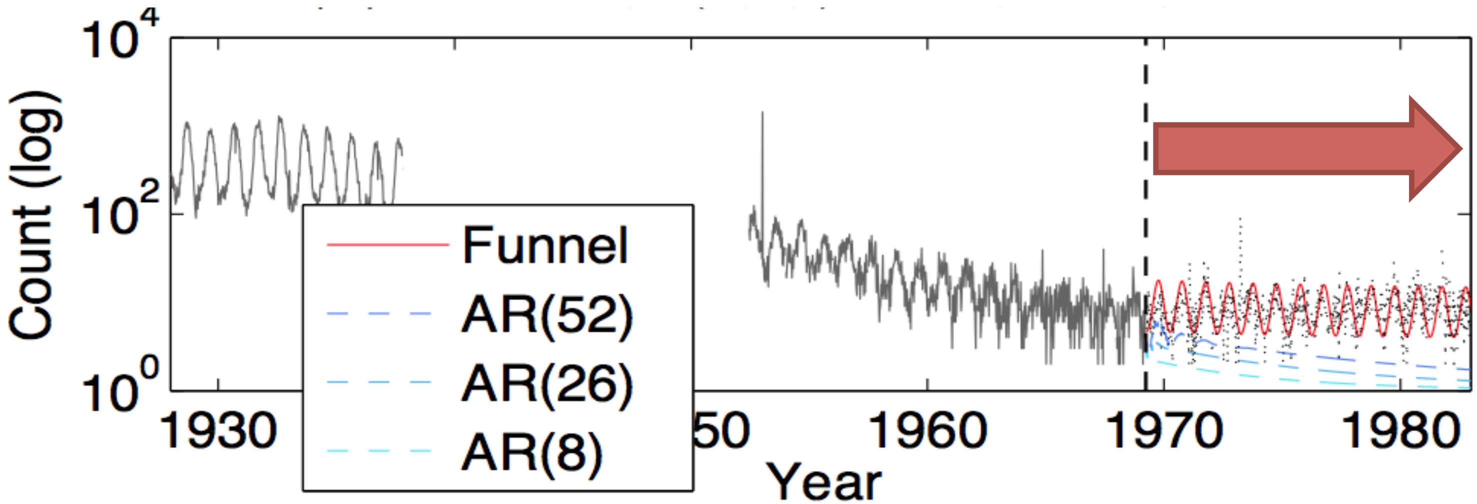
(a) Influenza

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

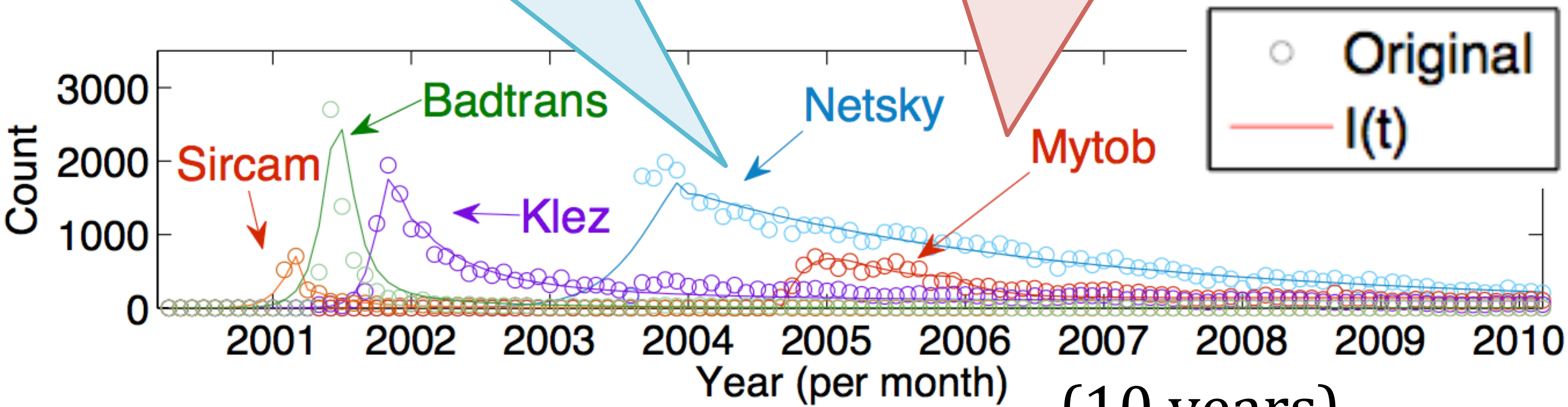
Funnel can capture future epidemics (AR: fail)

Generality of FUNNEL

Epidemics on computer networks

Spread via email attachment

Spread through corporate networks



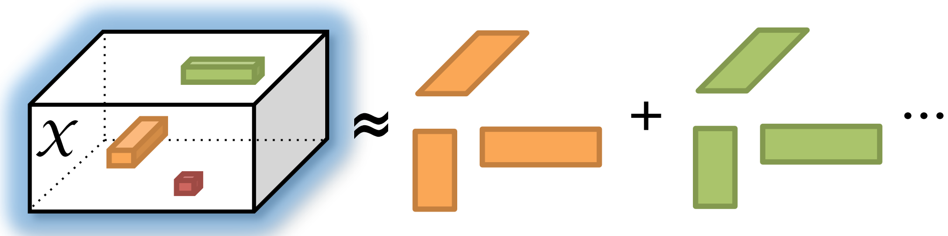
(10 years)

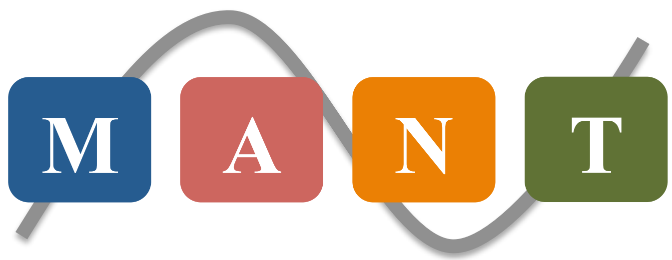
Funnel is general: it fits computer virus very well!

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

- 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

Part 3

Extension of time-series: tensor analysis

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