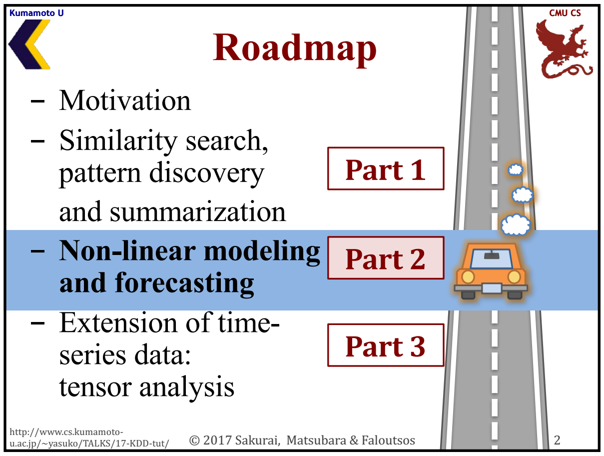
Smart Analytics for Big Time-series Data

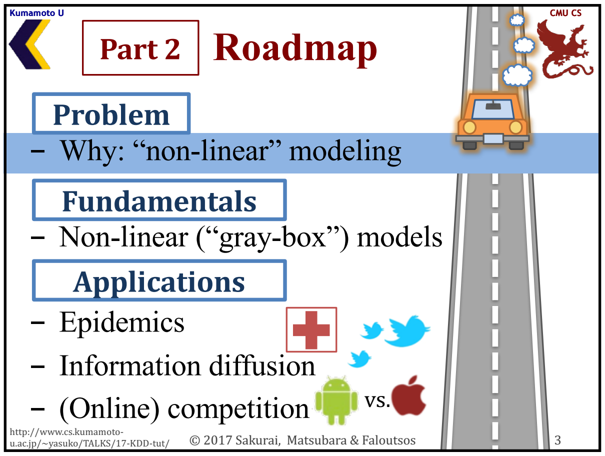
Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)



Kumamoto U CMU CS

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**



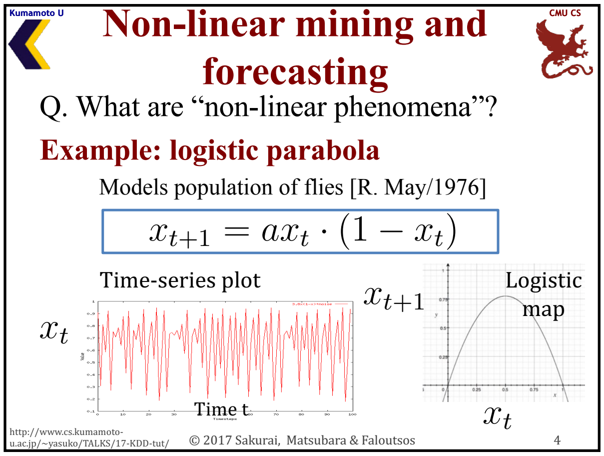
Kumamoto U CMU CS

Part 2 Roadmap

Problem
- Why: "non-linear" modeling

Fundamentals
- Non-linear ("gray-box") models

Applications
- Epidemics
- Information diffusion
- (Online) competition



Kumamoto U CMU CS

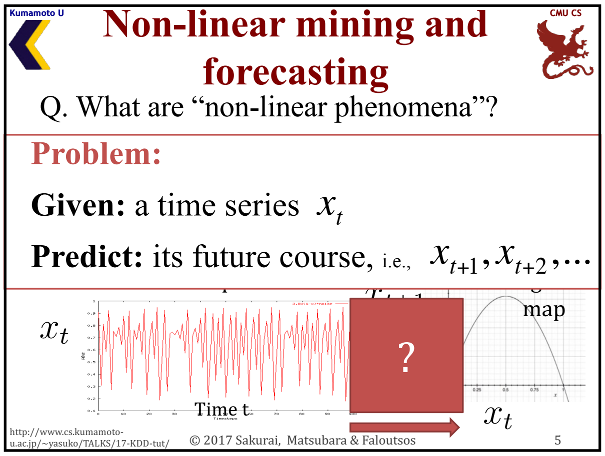
Non-linear mining and forecasting

Q. What are "non-linear phenomena"?

Example: logistic parabola

Models population of flies [R. May/1976]

$$x_{t+1} = a x_t \cdot (1 - x_t)$$



Kumamoto U CMU CS

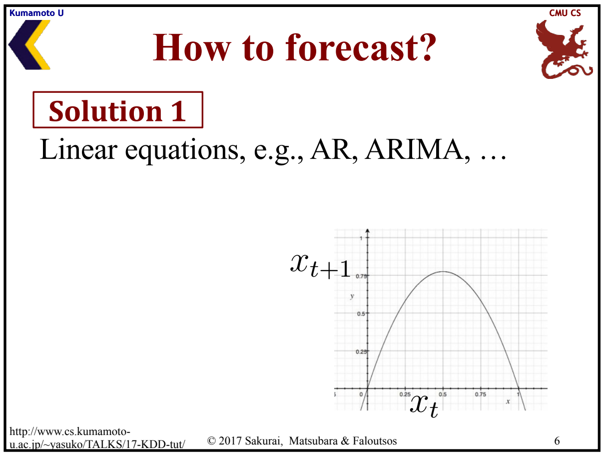
Non-linear mining and forecasting

Q. What are "non-linear phenomena"?

Problem:

Given: a time series x_t

Predict: its future course, i.e., $x_{t+1}, x_{t+2}, \ldots$

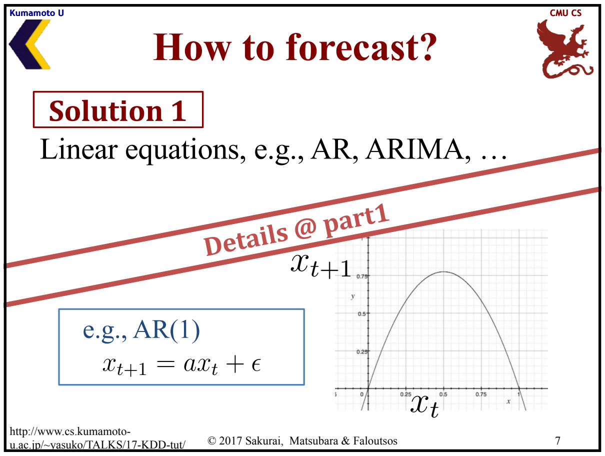


Kumamoto U CMU CS

How to forecast?

Solution 1

Linear equations, e.g., AR, ARIMA, …



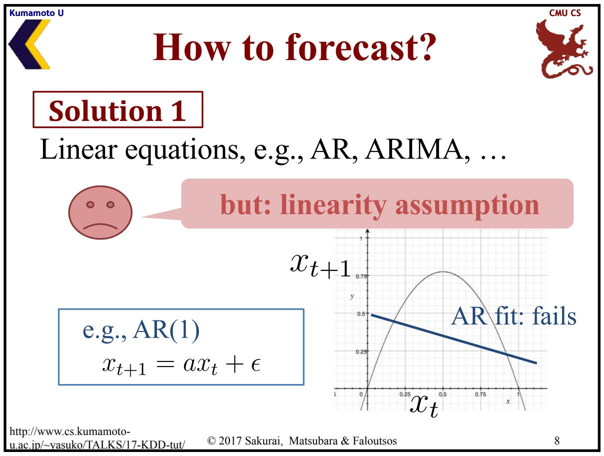
How to forecast?

Solution 1

Linear equations, e.g., AR, ARIMA, …

Details @ part1

e.g., AR(1)
$x_{t+1} = a x_t + \epsilon$



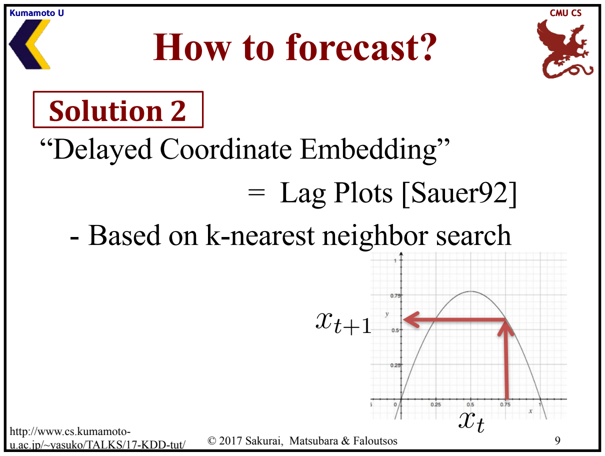
How to forecast?

Solution 1

Linear equations, e.g., AR, ARIMA, …

but: linearity assumption

e.g., AR(1)
$x_{t+1} = a x_t + \epsilon$



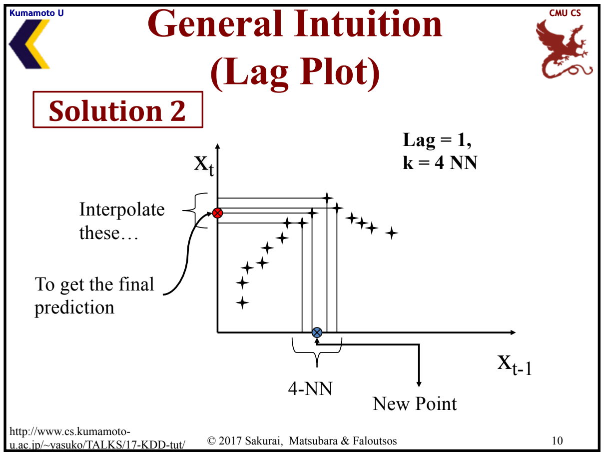
How to forecast?

Solution 2

"Delayed Coordinate Embedding"

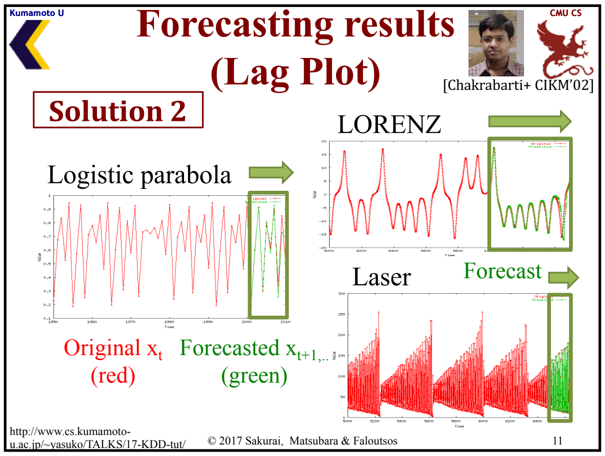
= Lag Plots [Sauer92]

- Based on k-nearest neighbor search



General Intuition (Lag Plot)

Solution 2



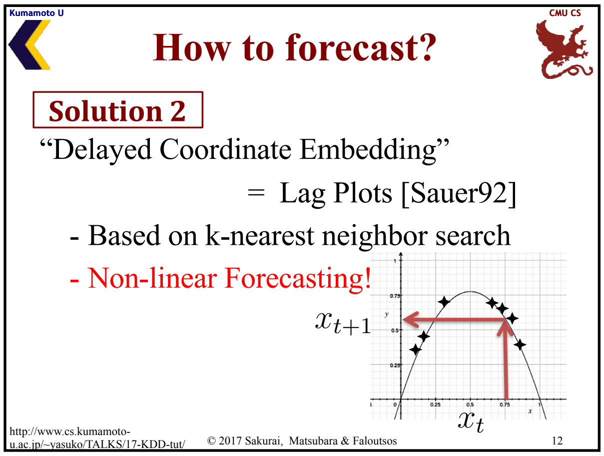
Forecasting results (Lag Plot)

[Chakrabarti+ CIKM'02]

Solution 2

Original x_t (red)

Forecasted x_{t+1},… (green)



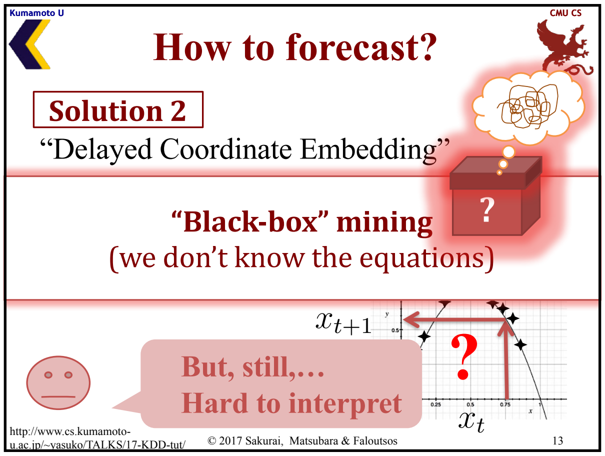
How to forecast?

Solution 2

"Delayed Coordinate Embedding"

= Lag Plots [Sauer92]

- Based on k-nearest neighbor search
- Non-linear Forecasting!



How to forecast?

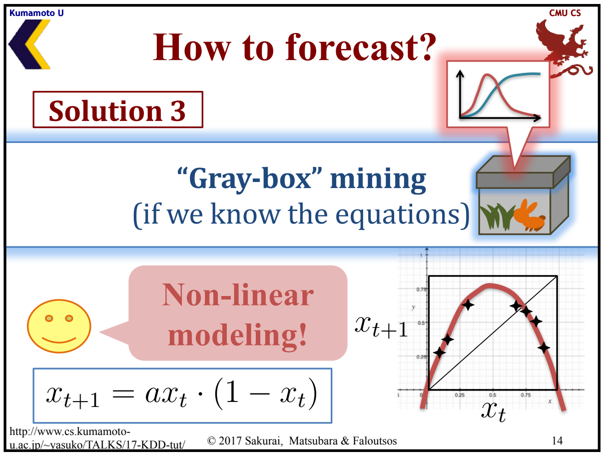
Solution 2

"Delayed Coordinate Embedding"

"Black-box" mining

(we don't know the equations)

But, still,…
Hard to interpret



How to forecast?

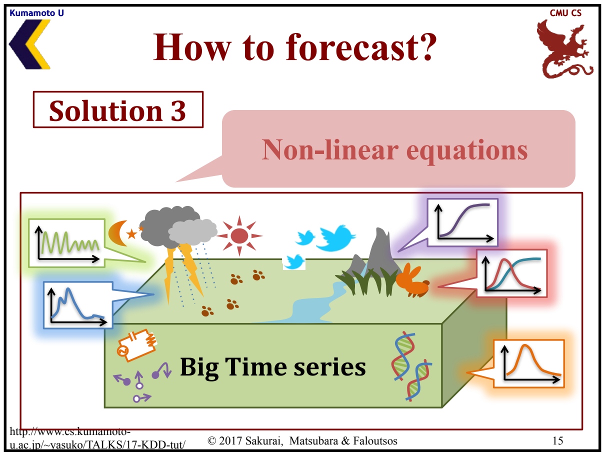
Solution 3

"Gray-box" mining

(if we know the equations)

Non-linear modeling!

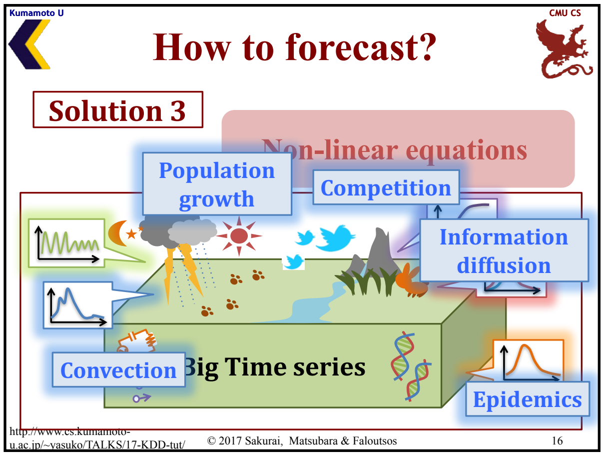
$$x_{t+1} = a x_t \cdot (1 - x_t)$$



How to forecast?

Solution 3

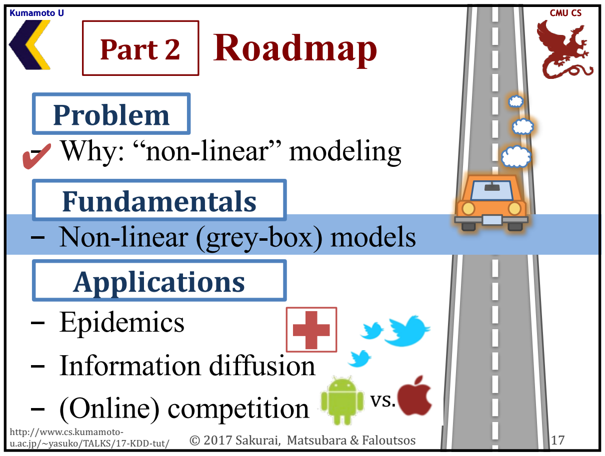
Non-linear equations



How to forecast?

Solution 3

Non-linear equations



Part 2 Roadmap

Problem

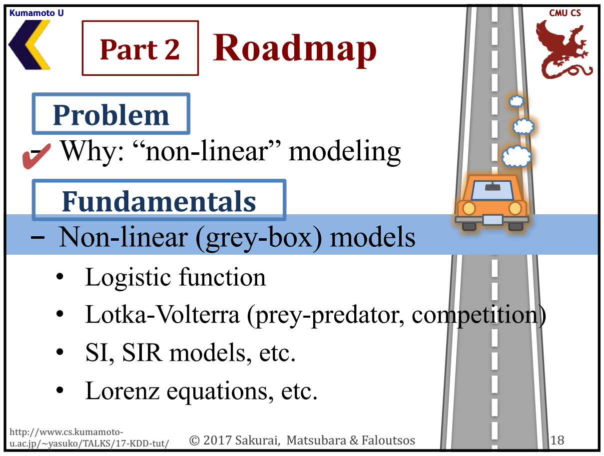
✓ Why: "non-linear" modeling

Fundamentals

- Non-linear (grey-box) models

Applications

- Epidemics
- Information diffusion
- (Online) competition



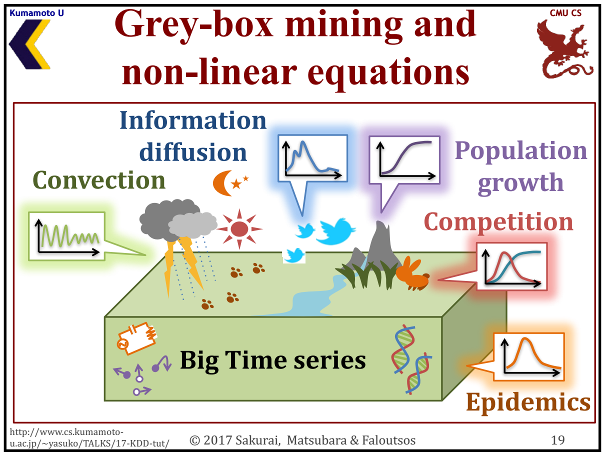
Part 2 Roadmap

Problem

✓ Why: "non-linear" modeling

Fundamentals

- Non-linear (grey-box) models
  - Logistic function
  - Lotka-Volterra (prey-predator, competition)
  - SI, SIR models, etc.
  - Lorenz equations, etc.

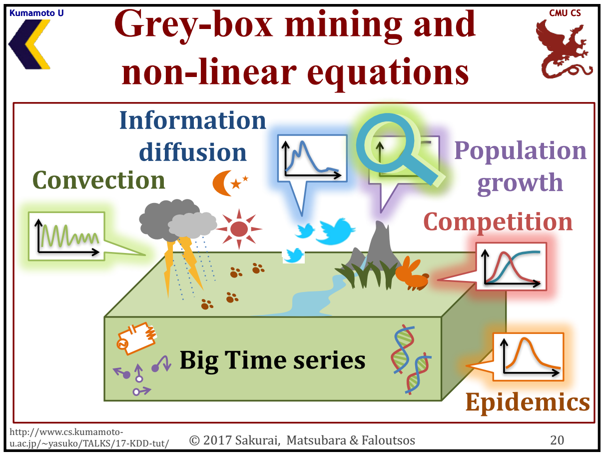


Kumamoto U

CMU CS

Grey-box mining and non-linear equations

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 19

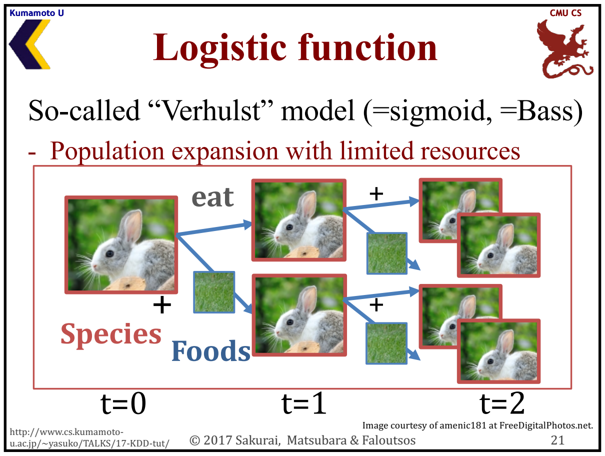


Kumamoto U

CMU CS

Grey-box mining and non-linear equations

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 20



Kumamoto U

CMU CS

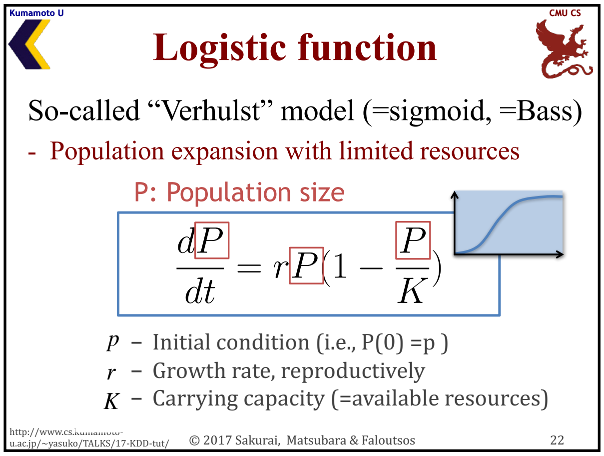
Logistic function

So-called "Verhulst" model (=sigmoid, =Bass)

- Population expansion with limited resources

Image courtesy of amenic181 at FreeDigitalPhotos.net.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 21



Kumamoto U

CMU CS

Logistic function

So-called "Verhulst" model (=sigmoid, =Bass)

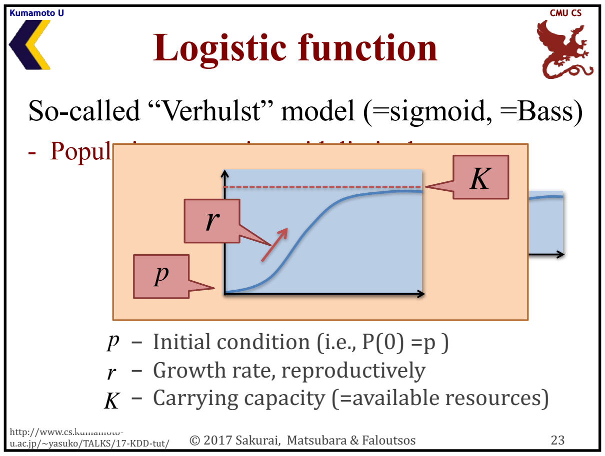
- Population expansion with limited resources

P: Population size

$$\frac{d\boxed{P}}{dt} = r\boxed{P}(1 - \frac{\boxed{P}}{K})$$

p – Initial condition (i.e., P(0) =p)
r – Growth rate, reproductively
K – Carrying capacity (=available resources)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 22



Kumamoto U

CMU CS

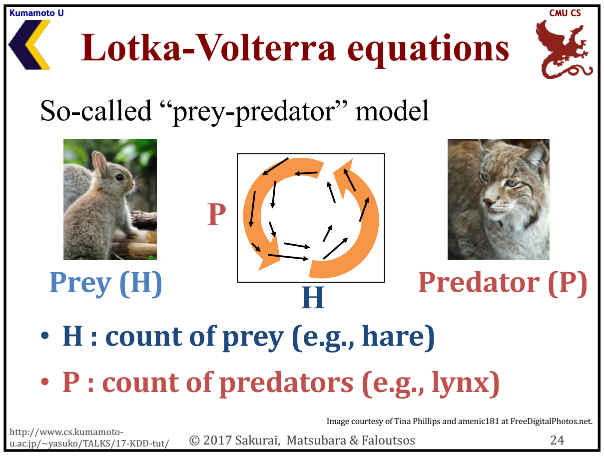
Logistic function

So-called "Verhulst" model (=sigmoid, =Bass)

- Popul

p – Initial condition (i.e., P(0) =p)
r – Growth rate, reproductively
K – Carrying capacity (=available resources)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 23



Kumamoto U

CMU CS

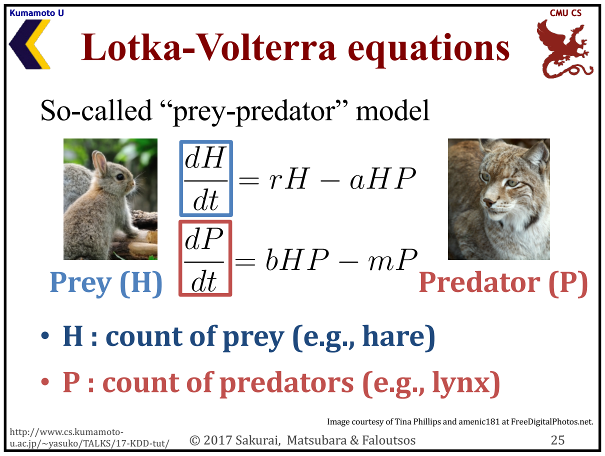
Lotka-Volterra equations

So-called "prey-predator" model

- **H : count of prey (e.g., hare)**
- **P : count of predators (e.g., lynx)**

Image courtesy of Tina Phillips and amenic181 at FreeDigitalPhotos.net.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 24



Lotka-Volterra equations

So-called “prey-predator” model

$$\boxed{\frac{dH}{dt}} = rH - aHP$$

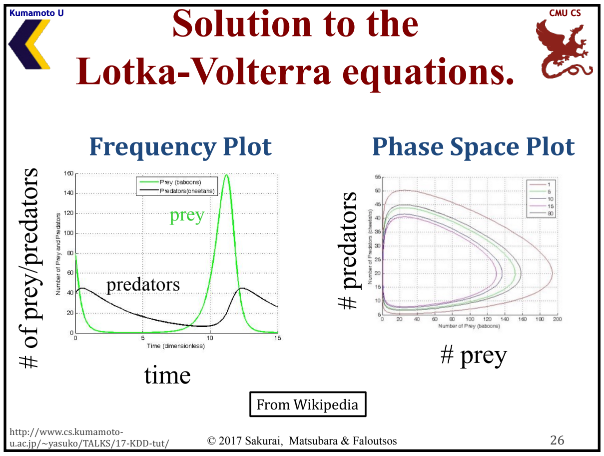
$$\boxed{\frac{dP}{dt}} = bHP - mP$$

Prey (H) **Predator (P)**

- **H : count of prey (e.g., hare)**
- **P : count of predators (e.g., lynx)**

Image courtesy of Tina Phillips and amenic181 at FreeDigitalPhotos.net.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 25

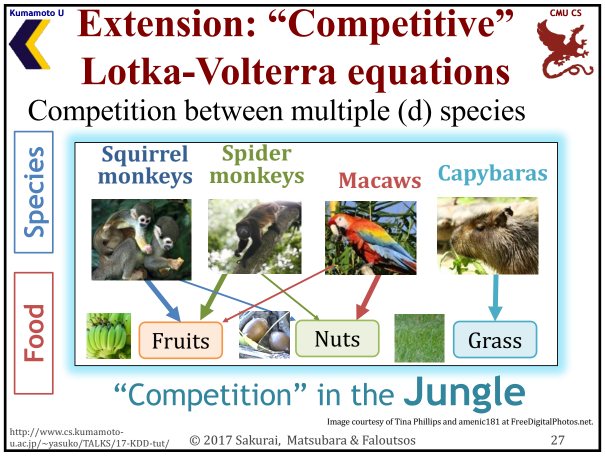


Solution to the Lotka-Volterra equations.

Frequency Plot **Phase Space Plot**

From Wikipedia

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 26



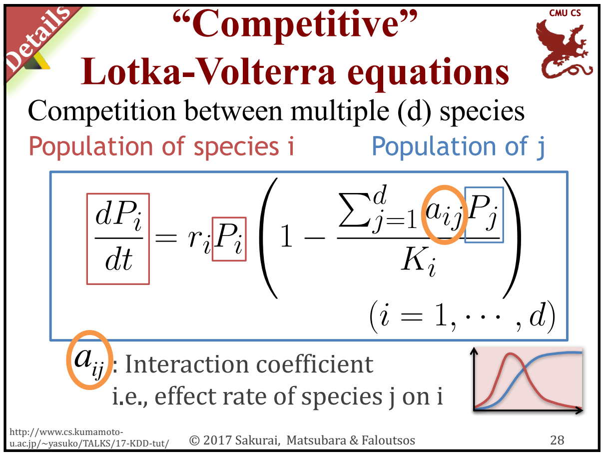
Extension: “Competitive” Lotka-Volterra equations

Competition between multiple (d) species

“Competition” in the **Jungle**

Image courtesy of Tina Phillips and amenic181 at FreeDigitalPhotos.net.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 27



“Competitive” Lotka-Volterra equations

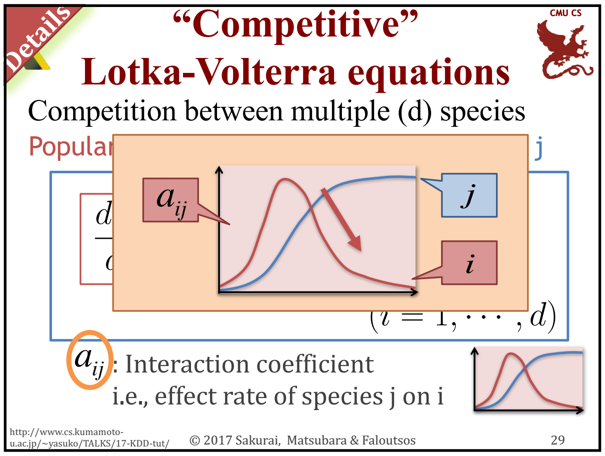
Competition between multiple (d) species

Population of species i   Population of j

$$\boxed{\frac{dP_i}{dt}} = r_i \boxed{P_i} \left(1 - \frac{\sum_{j=1}^{d} a_{ij} \boxed{P_j}}{K_i} \right) \quad (i = 1, \cdots, d)$$

a_{ij}: Interaction coefficient
i.e., effect rate of species j on i

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 28



“Competitive” Lotka-Volterra equations

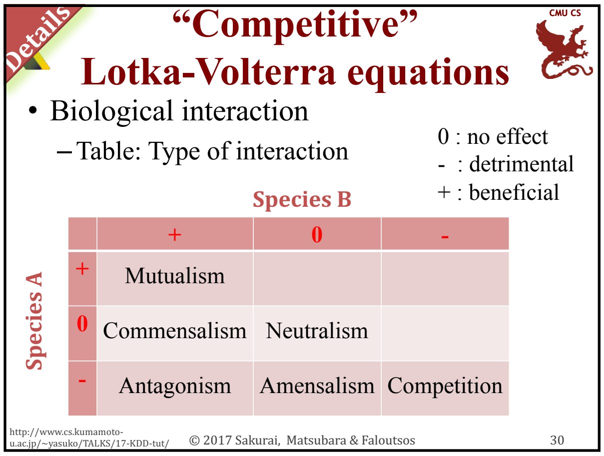
Competition between multiple (d) species

Popular ... j

$(i = 1, \cdots, d)$

a_{ij}: Interaction coefficient
i.e., effect rate of species j on i

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 29



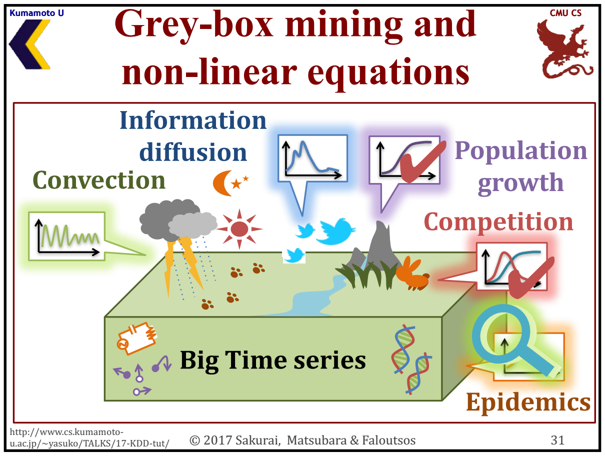
“Competitive” Lotka-Volterra equations

- Biological interaction
  - Table: Type of interaction

0 : no effect
- : detrimental
+ : beneficial

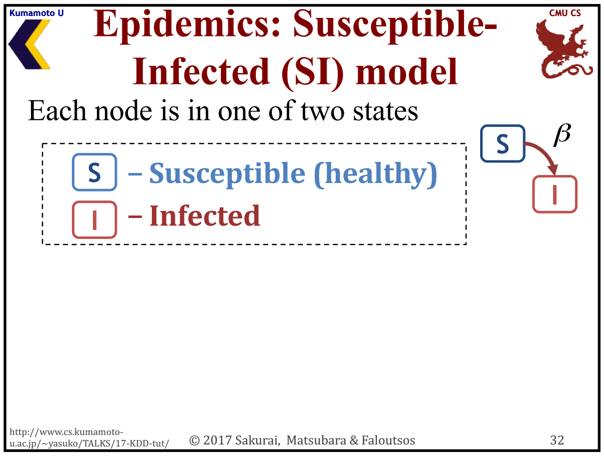
Species A \ Species B	+	0	-
+	Mutualism		
0	Commensalism	Neutralism	
-	Antagonism	Amensalism	Competition

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 30



Grey-box mining and non-linear equations

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 31

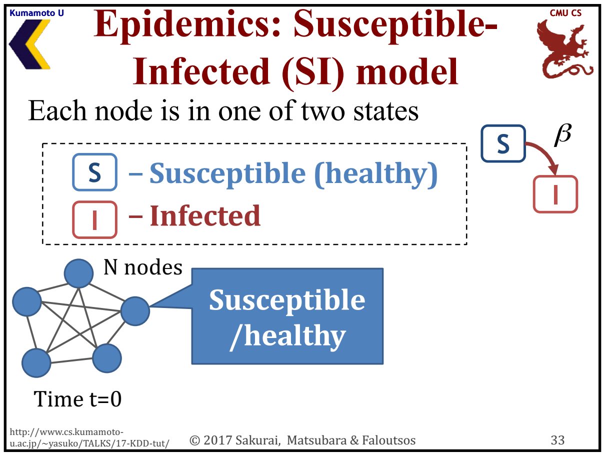


Epidemics: Susceptible-Infected (SI) model

Each node is in one of two states

S – Susceptible (healthy)
I – Infected

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 32

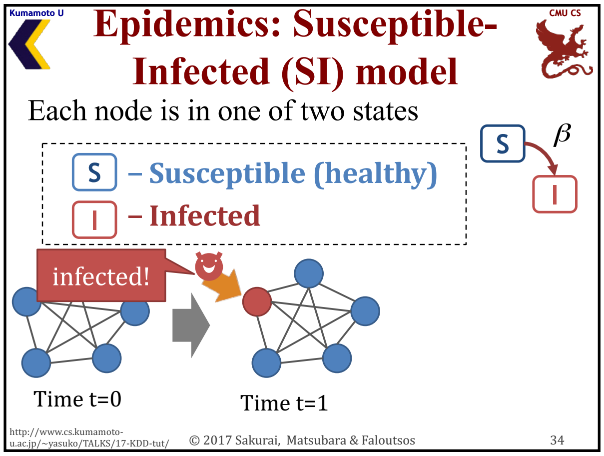


Epidemics: Susceptible-Infected (SI) model

Each node is in one of two states

S – Susceptible (healthy)
I – Infected

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 33

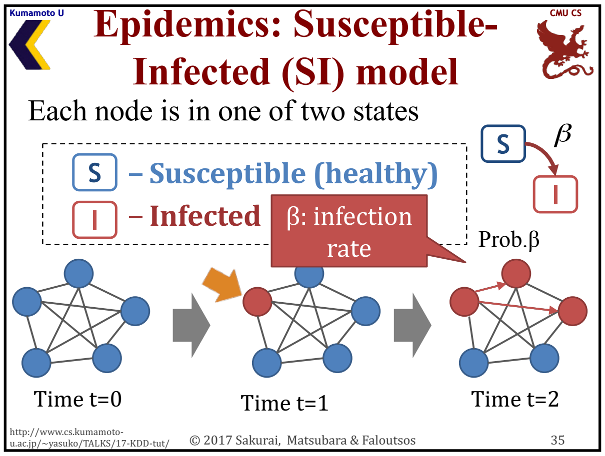


Epidemics: Susceptible-Infected (SI) model

Each node is in one of two states

S – Susceptible (healthy)
I – Infected

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 34



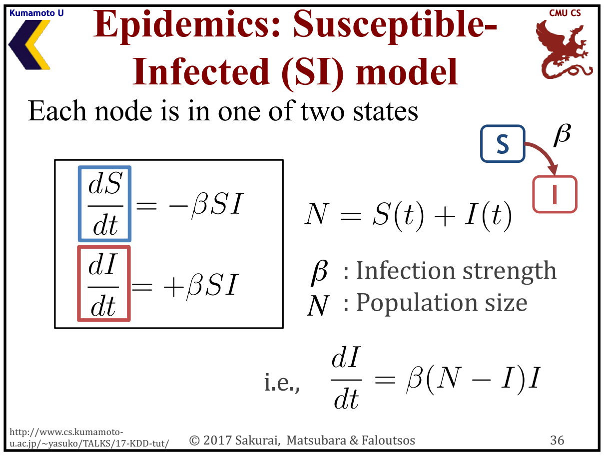
Epidemics: Susceptible-Infected (SI) model

Each node is in one of two states

S – Susceptible (healthy)
I – Infected

β: infection rate

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 35



Epidemics: Susceptible-Infected (SI) model

Each node is in one of two states

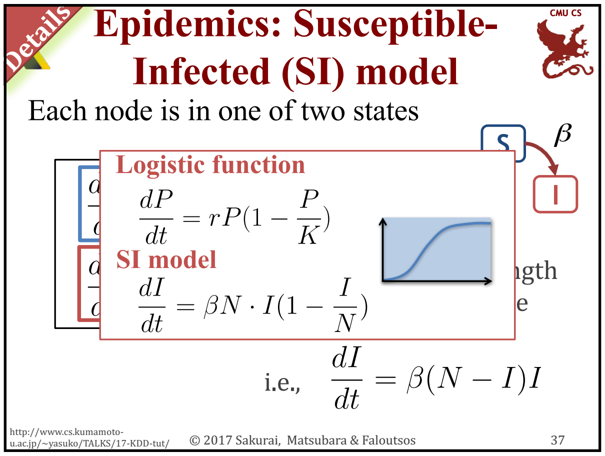
$$\boxed{\begin{aligned}\frac{dS}{dt} &= -\beta S I \\ \frac{dI}{dt} &= +\beta S I\end{aligned}}$$

$$N = S(t) + I(t)$$

β : Infection strength
N : Population size

i.e., $$\frac{dI}{dt} = \beta(N - I)I$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 36



Epidemics: Susceptible-Infected (SI) model

Each node is in one of two states

Logistic function

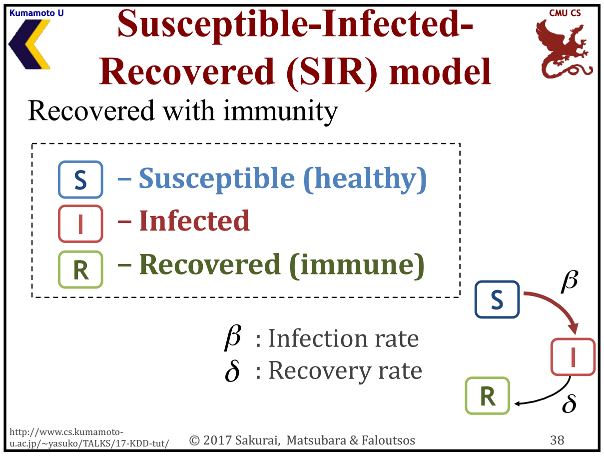
$$\frac{dP}{dt} = rP(1 - \frac{P}{K})$$

SI model

$$\frac{dI}{dt} = \beta N \cdot I(1 - \frac{I}{N})$$

i.e., $$\frac{dI}{dt} = \beta(N - I)I$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 37



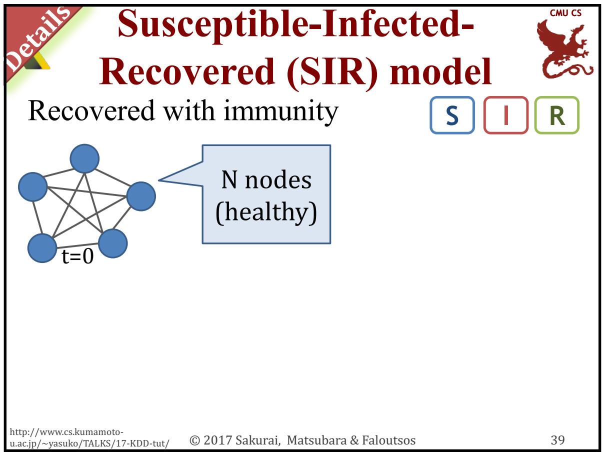
Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

- S – Susceptible (healthy)
- I – Infected
- R – Recovered (immune)

β : Infection rate
δ : Recovery rate

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 38

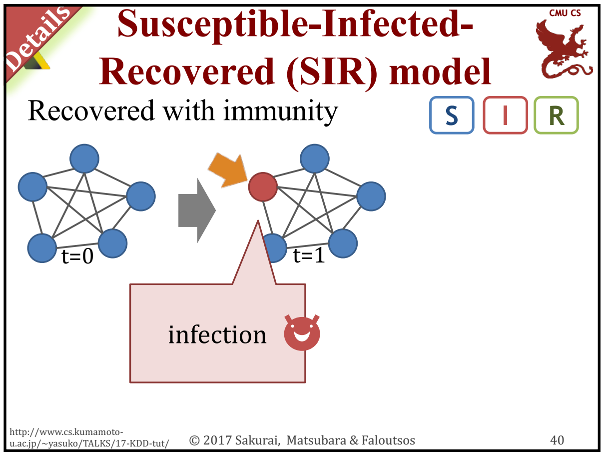


Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

N nodes
(healthy)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 39

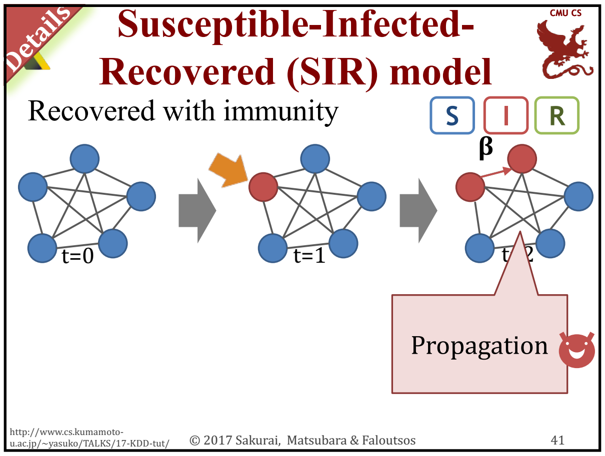


Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

infection

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 40

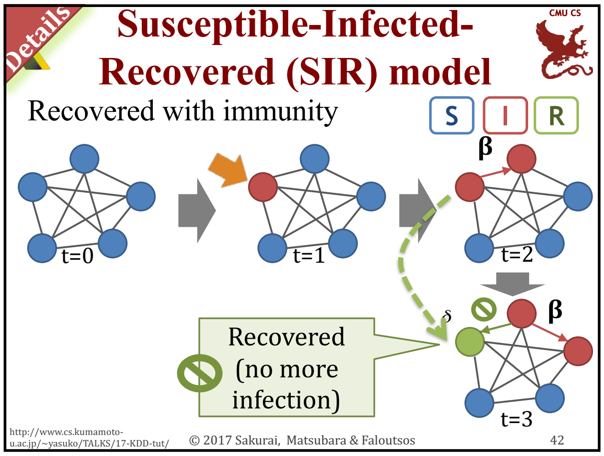


Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

Propagation

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 41

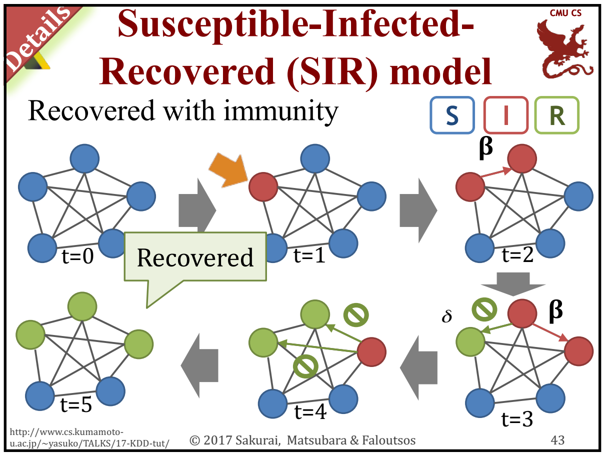


Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

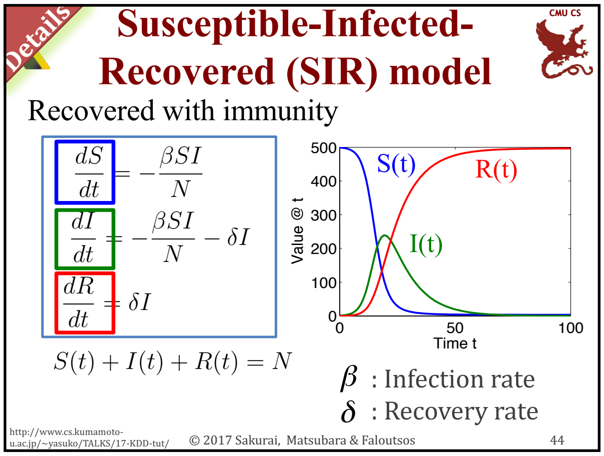
Recovered
(no more
infection)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 42



Susceptible-Infected-Recovered (SIR) model

Recovered with immunity



Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

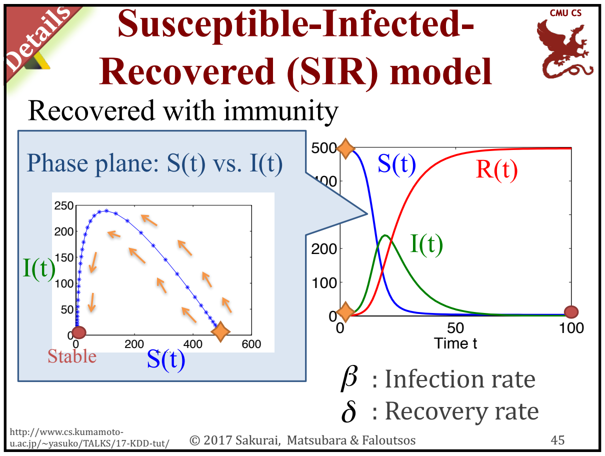
$$\frac{dI}{dt} = \frac{\beta SI}{N} - \delta I$$

$$\frac{dR}{dt} = \delta I$$

$$S(t) + I(t) + R(t) = N$$

β : Infection rate

δ : Recovery rate



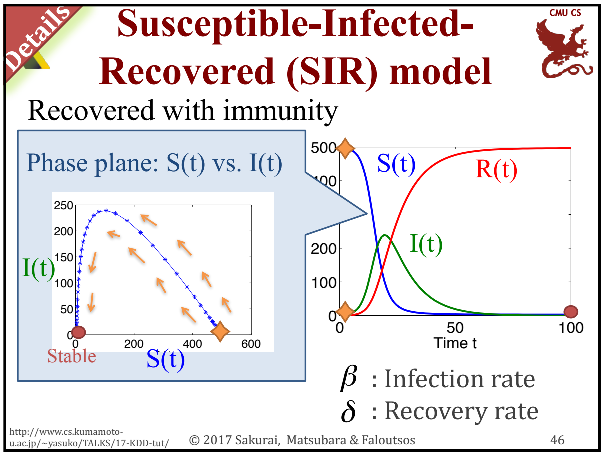
Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

Phase plane: S(t) vs. I(t)

β : Infection rate

δ : Recovery rate



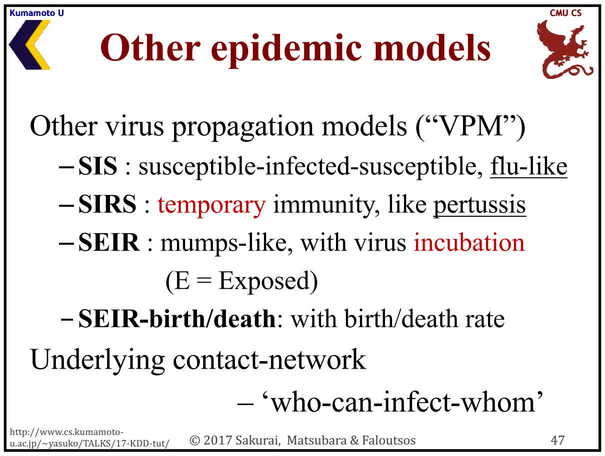
Susceptible-Infected-Recovered (SIR) model

Recovered with immunity

Phase plane: S(t) vs. I(t)

β : Infection rate

δ : Recovery rate



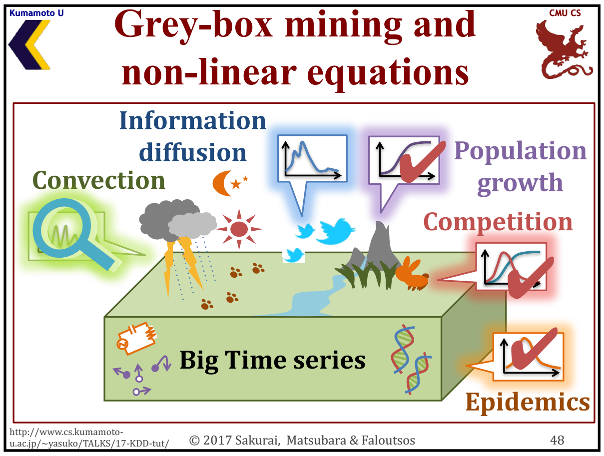
Other epidemic models

Other virus propagation models ("VPM")

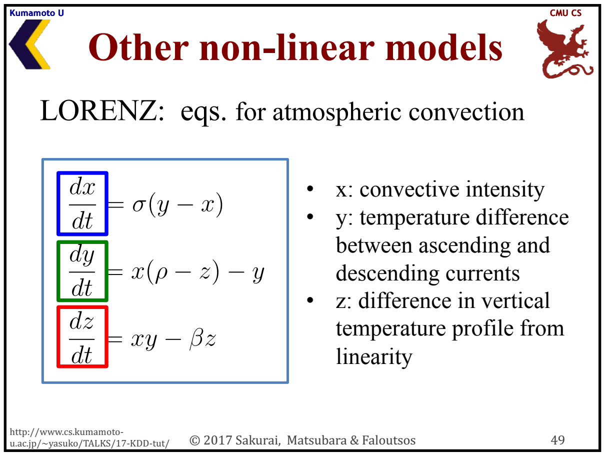
- **SIS** : susceptible-infected-susceptible, flu-like
- **SIRS** : temporary immunity, like pertussis
- **SEIR** : mumps-like, with virus incubation
  (E = Exposed)
- **SEIR-birth/death**: with birth/death rate

Underlying contact-network

– 'who-can-infect-whom'



Grey-box mining and non-linear equations

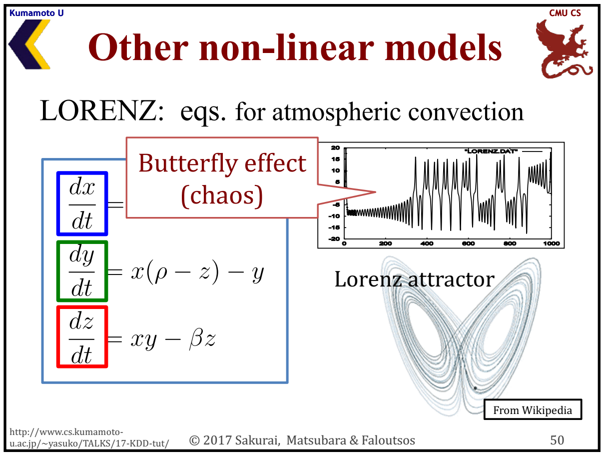


Other non-linear models

LORENZ: eqs. for atmospheric convection

$$\boxed{\frac{dx}{dt}} = \sigma(y-x)$$
$$\boxed{\frac{dy}{dt}} = x(\rho-z)-y$$
$$\boxed{\frac{dz}{dt}} = xy-\beta z$$

- x: convective intensity
- y: temperature difference between ascending and descending currents
- z: difference in vertical temperature profile from linearity



Other non-linear models

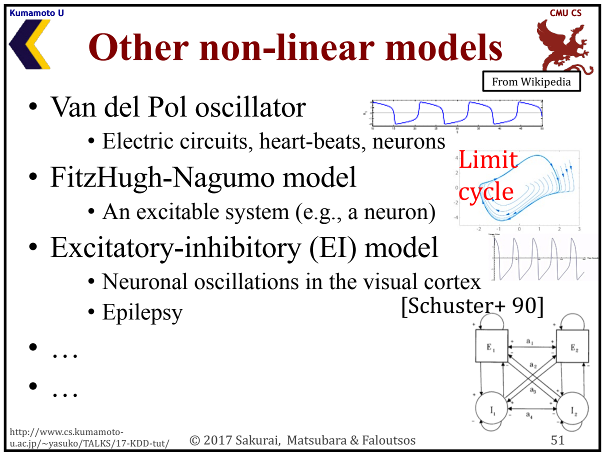
LORENZ: eqs. for atmospheric convection

$$\boxed{\frac{dx}{dt}} =$$
$$\boxed{\frac{dy}{dt}} = x(\rho-z)-y$$
$$\boxed{\frac{dz}{dt}} = xy-\beta z$$

Butterfly effect (chaos)

Lorenz attractor

From Wikipedia

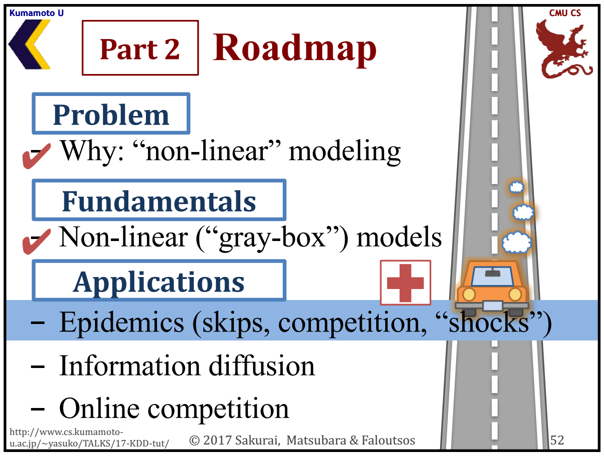


Other non-linear models

From Wikipedia

- Van del Pol oscillator
  - Electric circuits, heart-beats, neurons
- FitzHugh-Nagumo model
  - An excitable system (e.g., a neuron)
- Excitatory-inhibitory (EI) model
  - Neuronal oscillations in the visual cortex
  - Epilepsy [Schuster+ 90]
- …
- …

Limit cycle



Part 2 Roadmap

Problem

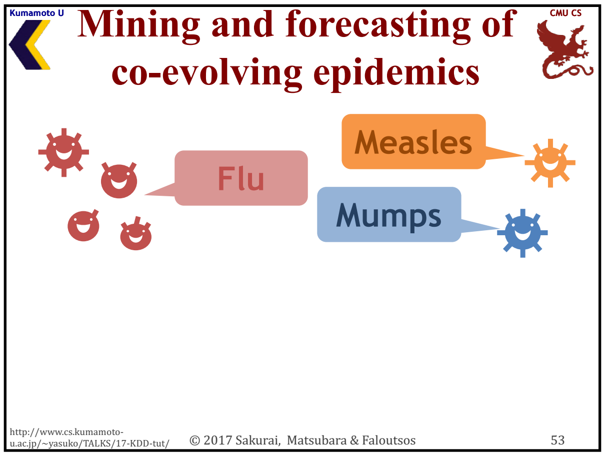
✓ Why: "non-linear" modeling

Fundamentals

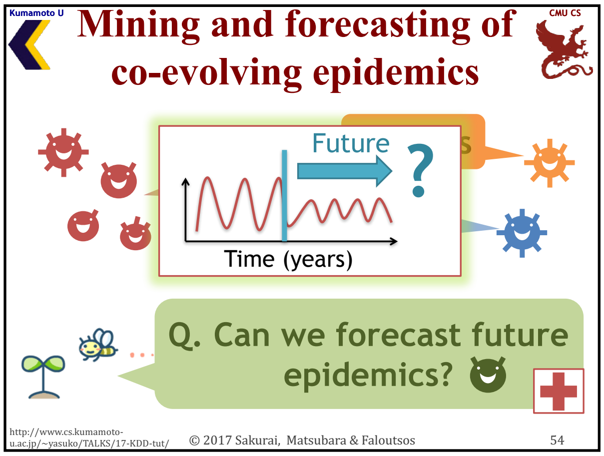
✓ Non-linear ("gray-box") models

Applications

- Epidemics (skips, competition, "shocks")
- Information diffusion
- Online competition

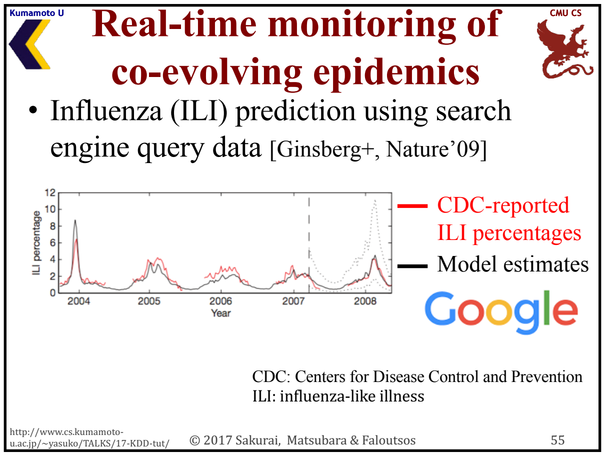


Mining and forecasting of co-evolving epidemics



Mining and forecasting of co-evolving epidemics

Q. Can we forecast future epidemics?

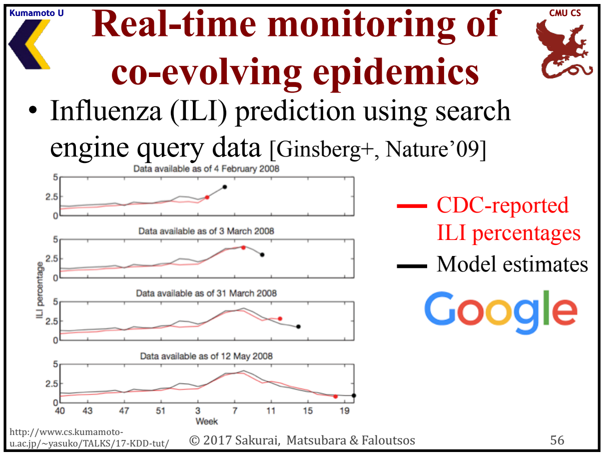


Real-time monitoring of co-evolving epidemics

- Influenza (ILI) prediction using search engine query data [Ginsberg+, Nature'09]

— CDC-reported ILI percentages
— Model estimates

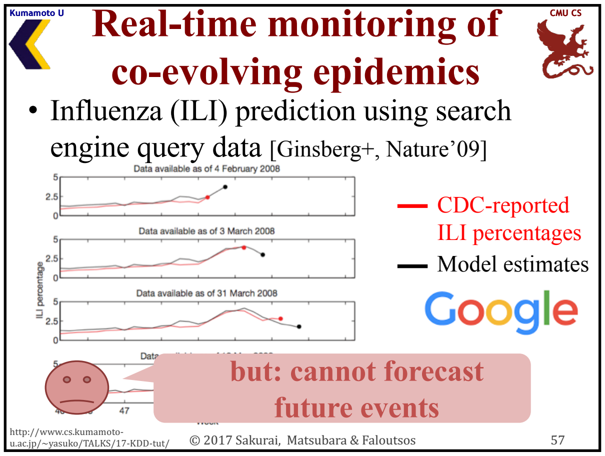
CDC: Centers for Disease Control and Prevention
ILI: influenza-like illness



Real-time monitoring of co-evolving epidemics

- Influenza (ILI) prediction using search engine query data [Ginsberg+, Nature'09]

— CDC-reported ILI percentages
— Model estimates

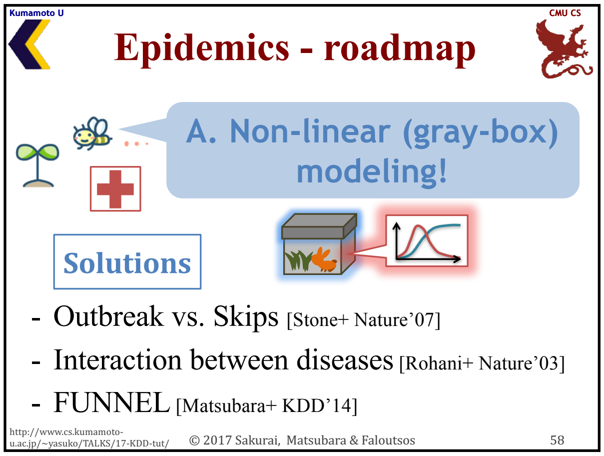


Real-time monitoring of co-evolving epidemics

- Influenza (ILI) prediction using search engine query data [Ginsberg+, Nature'09]

— CDC-reported ILI percentages
— Model estimates

but: cannot forecast future events

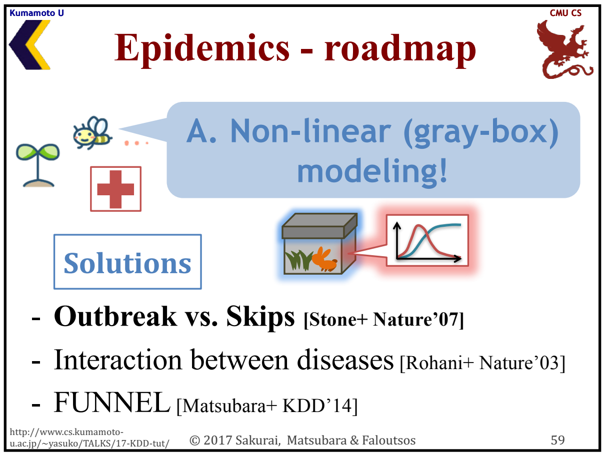


Epidemics - roadmap

A. Non-linear (gray-box) modeling!

Solutions

- Outbreak vs. Skips [Stone+ Nature'07]
- Interaction between diseases [Rohani+ Nature'03]
- FUNNEL [Matsubara+ KDD'14]

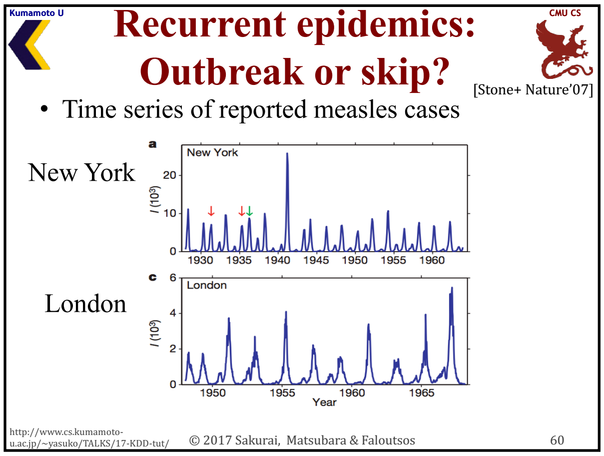


Epidemics - roadmap

A. Non-linear (gray-box) modeling!

Solutions

- **Outbreak vs. Skips [Stone+ Nature'07]**
- Interaction between diseases [Rohani+ Nature'03]
- FUNNEL [Matsubara+ KDD'14]



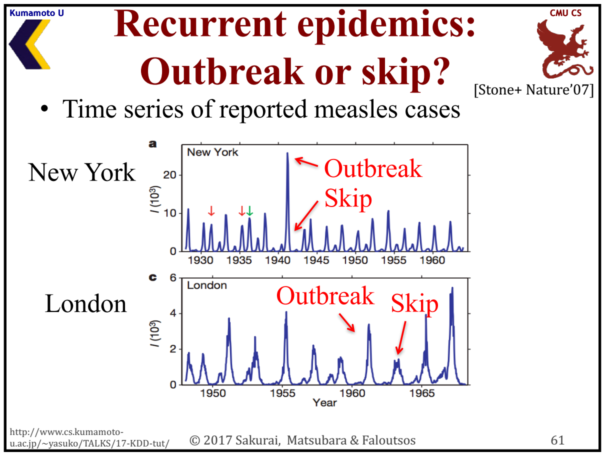
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Time series of reported measles cases

New York

London


Kumamoto U

CMU CS

Recurrent epidemics: Outbreak or skip?

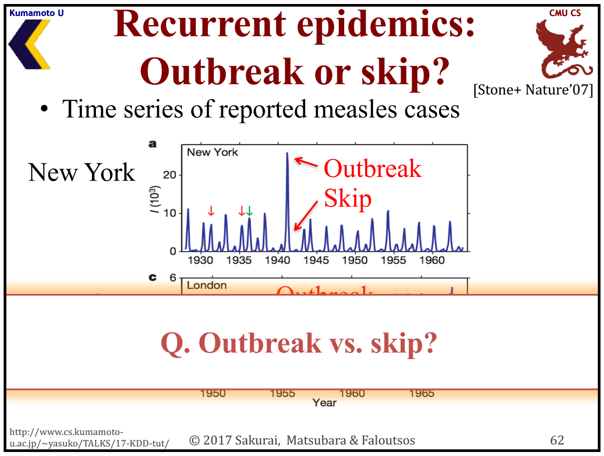
[Stone+ Nature'07]

- Time series of reported measles cases

New York

London

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 61


Kumamoto U

CMU CS

Recurrent epidemics: Outbreak or skip?

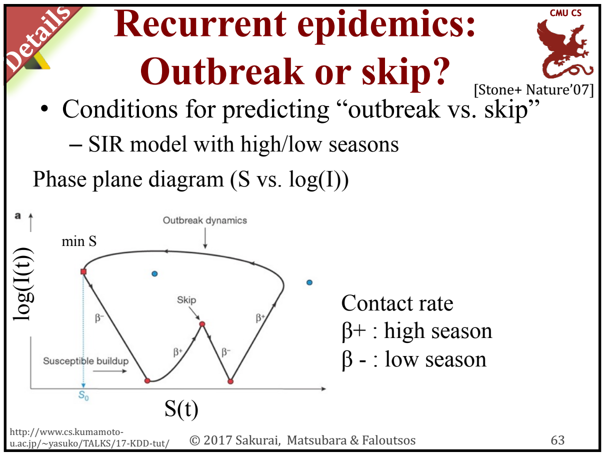
[Stone+ Nature'07]

- Time series of reported measles cases

New York

Q. Outbreak vs. skip?

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 62


Details

CMU CS

Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Conditions for predicting "outbreak vs. skip"
  - SIR model with high/low seasons

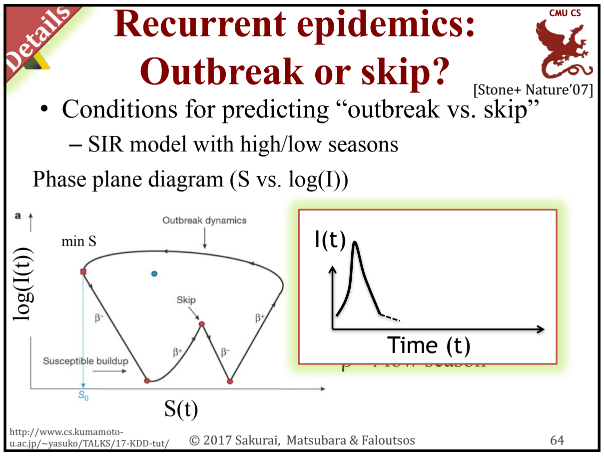
Phase plane diagram (S vs. log(I))

Contact rate

β+ : high season

β - : low season

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 63


Details

CMU CS

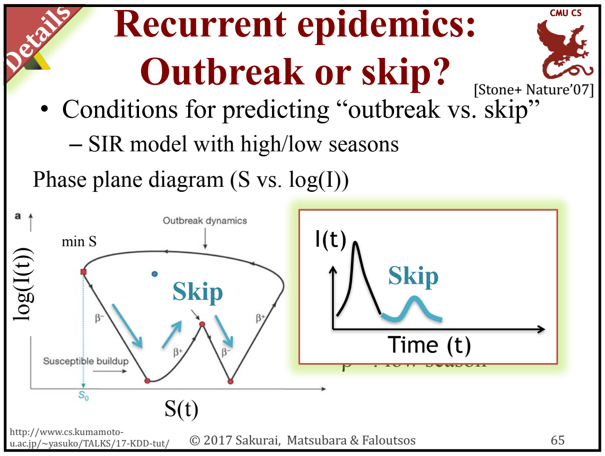
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Conditions for predicting "outbreak vs. skip"
  - SIR model with high/low seasons

Phase plane diagram (S vs. log(I))

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 64


Details

CMU CS

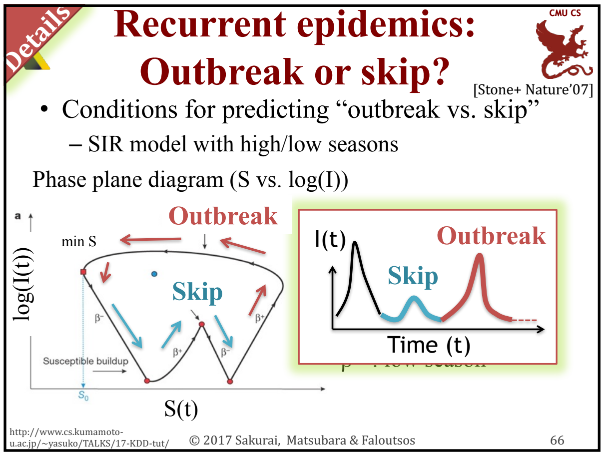
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Conditions for predicting "outbreak vs. skip"
  - SIR model with high/low seasons

Phase plane diagram (S vs. log(I))

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 65


Details

CMU CS

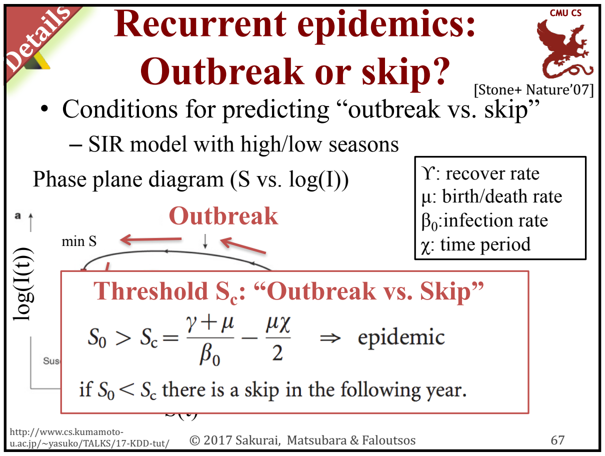
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Conditions for predicting "outbreak vs. skip"
  - SIR model with high/low seasons

Phase plane diagram (S vs. log(I))

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 66



Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Conditions for predicting "outbreak vs. skip"
  - SIR model with high/low seasons

Phase plane diagram (S vs. log(I))

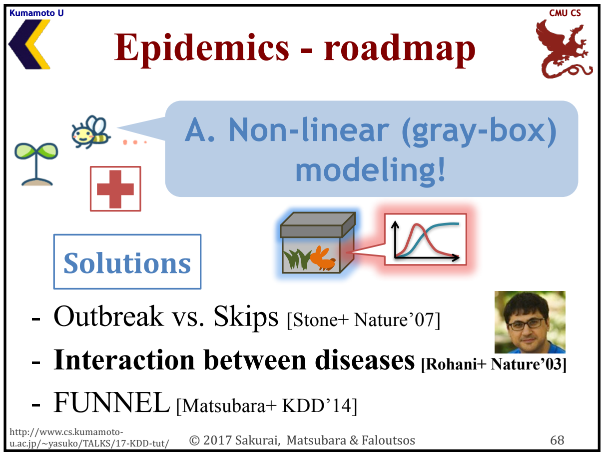
ϒ: recover rate
μ: birth/death rate
β_0:infection rate
χ: time period

Outbreak

Threshold S_c: "Outbreak vs. Skip"

$$S_0 > S_c = \frac{\gamma+\mu}{\beta_0} - \frac{\mu\chi}{2} \Rightarrow \text{epidemic}$$

if $S_0 < S_c$ there is a skip in the following year.



Epidemics - roadmap

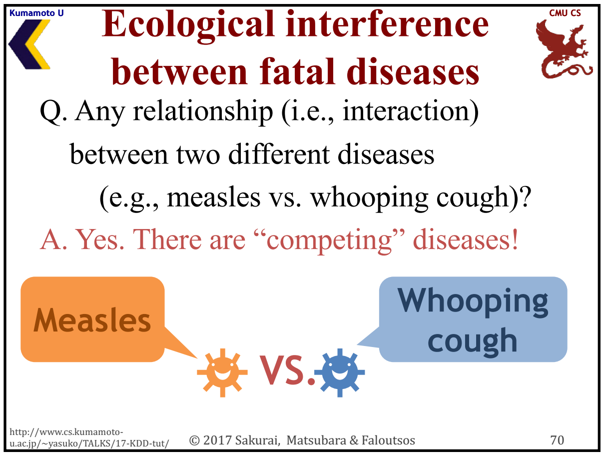
A. Non-linear (gray-box) modeling!

Solutions

- Outbreak vs. Skips [Stone+ Nature'07]
- **Interaction between diseases [Rohani+ Nature'03]**
- FUNNEL [Matsubara+ KDD'14]

Ecological interference between fatal diseases

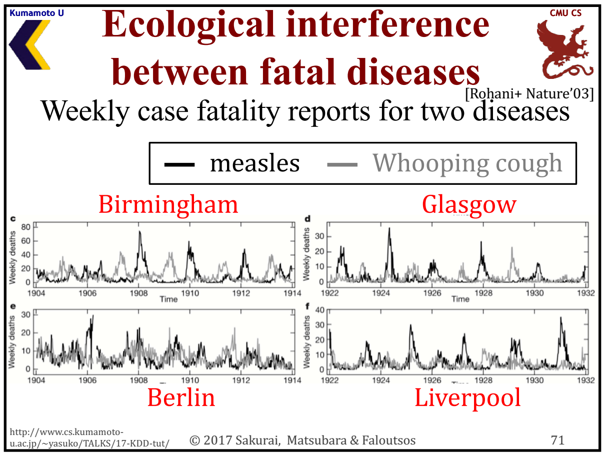
Q. Any relationship (i.e., interaction) between two different diseases (e.g., measles vs. whooping cough)?



Ecological interference between fatal diseases

Q. Any relationship (i.e., interaction) between two different diseases (e.g., measles vs. whooping cough)?

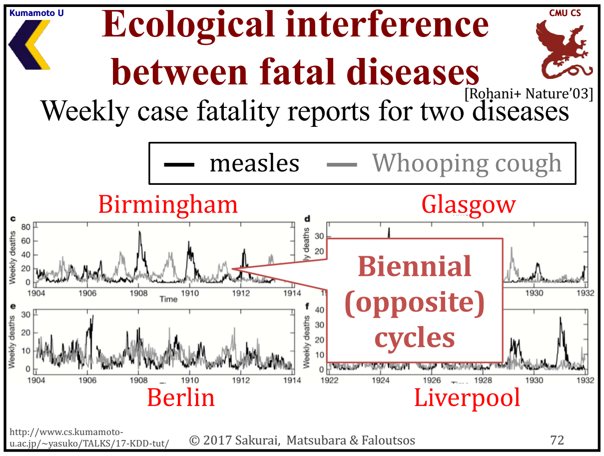
A. Yes. There are "competing" diseases!



Ecological interference between fatal diseases

[Rohani+ Nature'03]

Weekly case fatality reports for two diseases

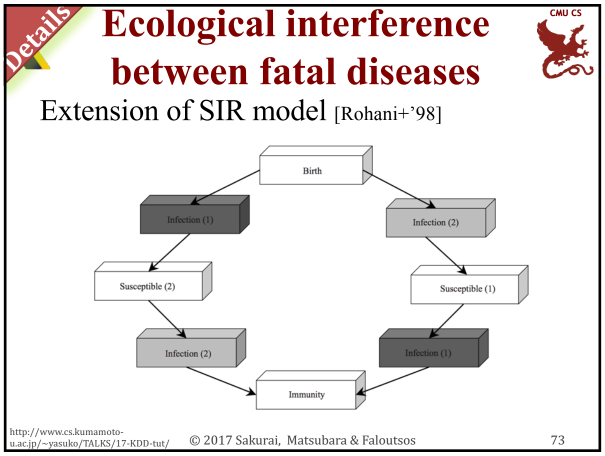


Ecological interference between fatal diseases

[Rohani+ Nature'03]

Weekly case fatality reports for two diseases

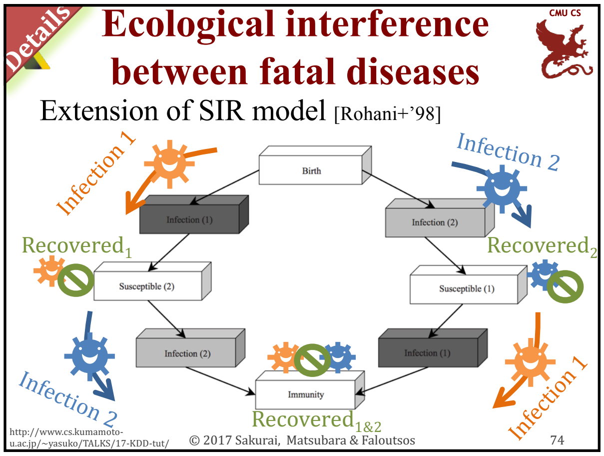
Biennial (opposite) cycles



Ecological interference between fatal diseases

Extension of SIR model [Rohani+'98]

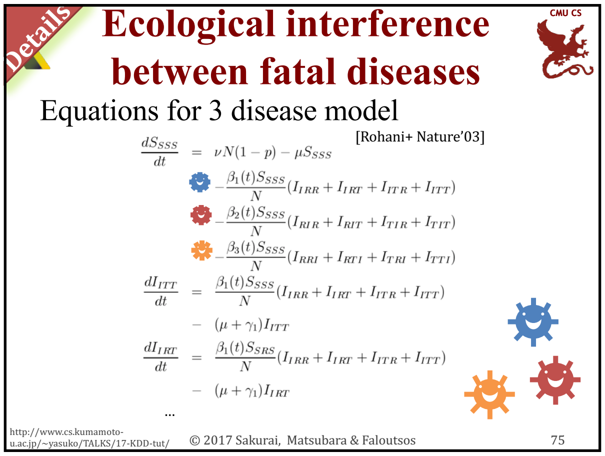
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 73



Ecological interference between fatal diseases

Extension of SIR model [Rohani+'98]

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 74



Ecological interference between fatal diseases

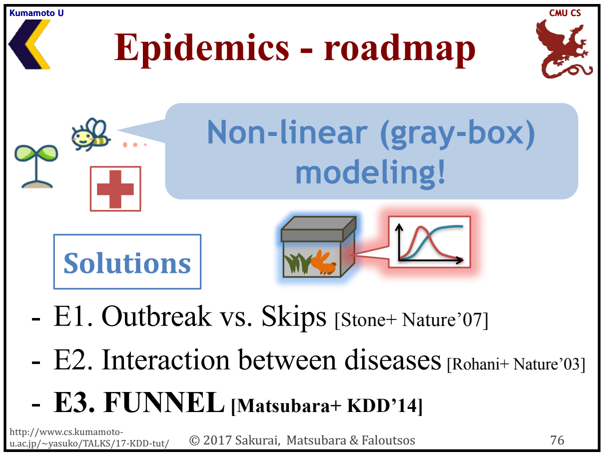
Equations for 3 disease model

[Rohani+ Nature'03]

$$\begin{aligned}\frac{dS_{SSS}}{dt} &= \nu N(1-p) - \mu S_{SSS}\\ &\quad -\frac{\beta_1(t)S_{SSS}}{N}(I_{IRR}+I_{IRT}+I_{ITR}+I_{ITT})\\ &\quad -\frac{\beta_2(t)S_{SSS}}{N}(I_{RIR}+I_{RIT}+I_{TIR}+I_{TIT})\\ &\quad -\frac{\beta_3(t)S_{SSS}}{N}(I_{RRI}+I_{RTI}+I_{TRI}+I_{TTI})\\ \frac{dI_{ITT}}{dt} &= \frac{\beta_1(t)S_{SSS}}{N}(I_{IRR}+I_{IRT}+I_{ITR}+I_{ITT})\\ &- (\mu+\gamma_1)I_{ITT}\\ \frac{dI_{IRT}}{dt} &= \frac{\beta_1(t)S_{SRS}}{N}(I_{IRR}+I_{IRT}+I_{ITR}+I_{ITT})\\ &- (\mu+\gamma_1)I_{IRT}\end{aligned}$$

...

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 75



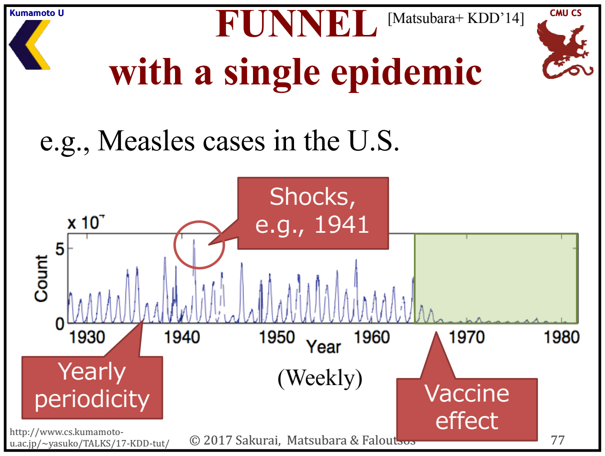
Epidemics - roadmap

Non-linear (gray-box) modeling!

Solutions

- E1. Outbreak vs. Skips [Stone+ Nature'07]
- E2. Interaction between diseases [Rohani+ Nature'03]
- **E3. FUNNEL [Matsubara+ KDD'14]**

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 76

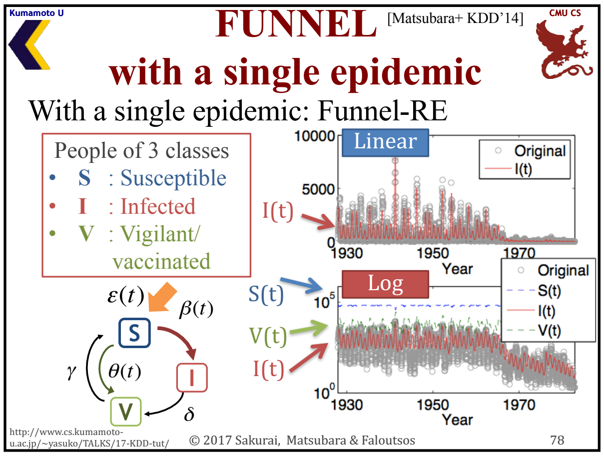


FUNNEL [Matsubara+ KDD'14] with a single epidemic

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

(Weekly)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 77

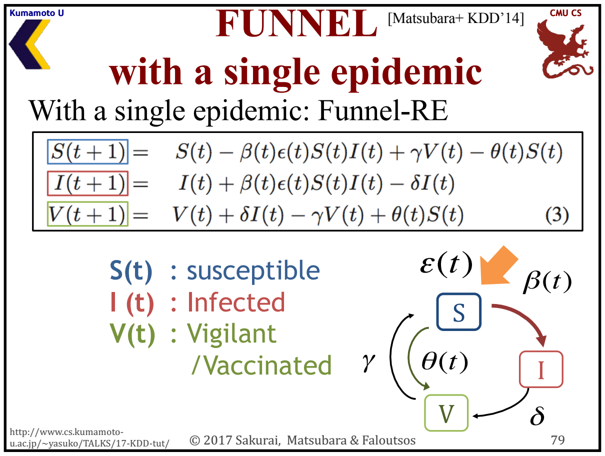


FUNNEL [Matsubara+ KDD'14] with a single epidemic

With a single epidemic: Funnel-RE

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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 78



Kumamoto U CMU CS

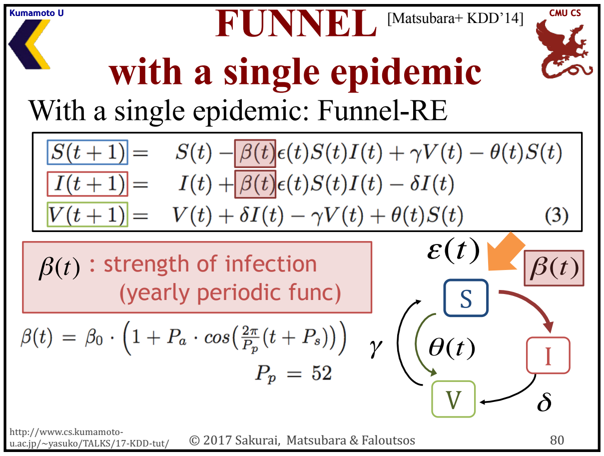
FUNNEL [Matsubara+ KDD'14]
with a single epidemic

With a single epidemic: Funnel-RE

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

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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 79



Kumamoto U CMU CS

FUNNEL [Matsubara+ KDD'14]
with a single epidemic

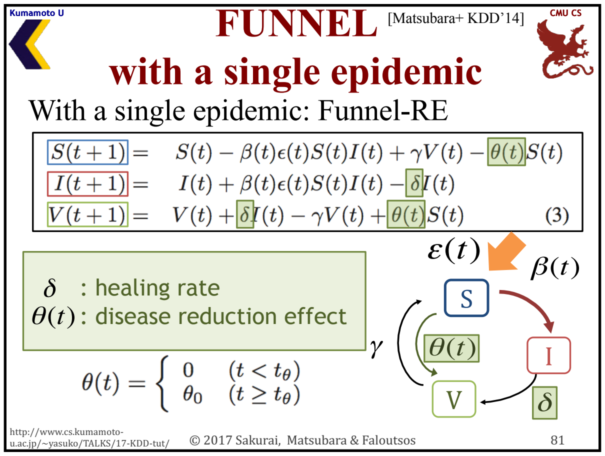
With a single epidemic: Funnel-RE

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

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

$$\beta(t) = \beta_0 \cdot \left(1 + P_a \cdot cos\left(\frac{2\pi}{P_p}(t + P_s)\right)\right)$$
$$P_p = 52$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 80



Kumamoto U CMU CS

FUNNEL [Matsubara+ KDD'14]
with a single epidemic

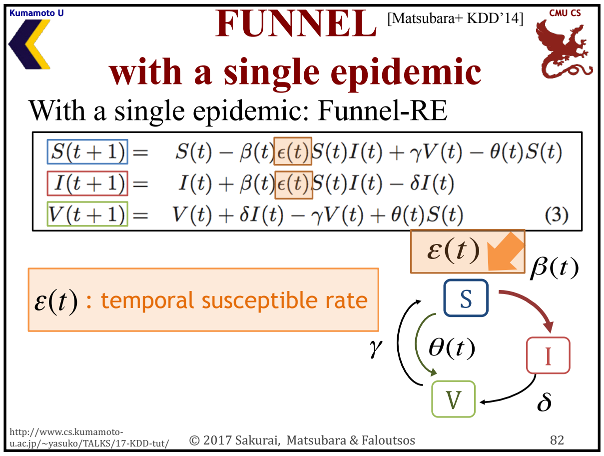
With a single epidemic: Funnel-RE

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

δ : healing rate
$\theta(t)$: disease reduction effect

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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 81



Kumamoto U CMU CS

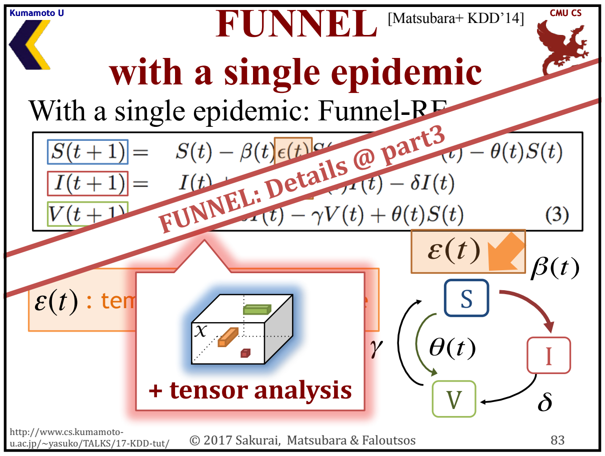
FUNNEL [Matsubara+ KDD'14]
with a single epidemic

With a single epidemic: Funnel-RE

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

$\varepsilon(t)$: temporal susceptible rate

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 82



Kumamoto U CMU CS

FUNNEL [Matsubara+ KDD'14]
with a single epidemic

With a single epidemic: Funnel-RE

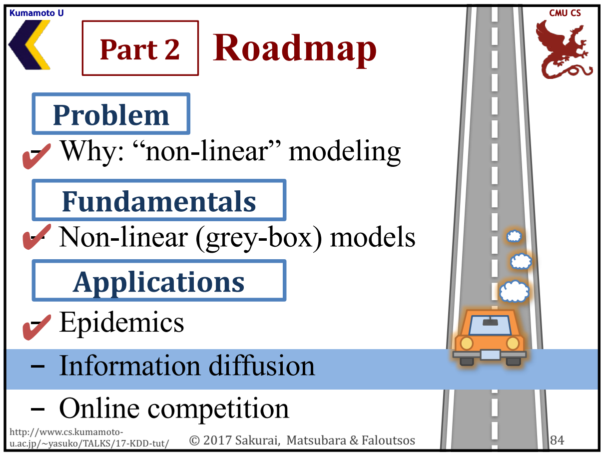
$$\boxed{S(t+1)} = S(t) - \beta(t)\boxed{\epsilon(t)}S \quad \theta(t)S(t)$$
$$\boxed{I(t+1)} = I(t \quad I(t) - \delta I(t)$$
$$\boxed{V(t+1} \quad I(t) - \gamma V(t) + \theta(t)S(t) \quad (3)$$

FUNNEL: Details @ part3

$\varepsilon(t)$: tem

+ tensor analysis

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 83



Kumamoto U CMU CS

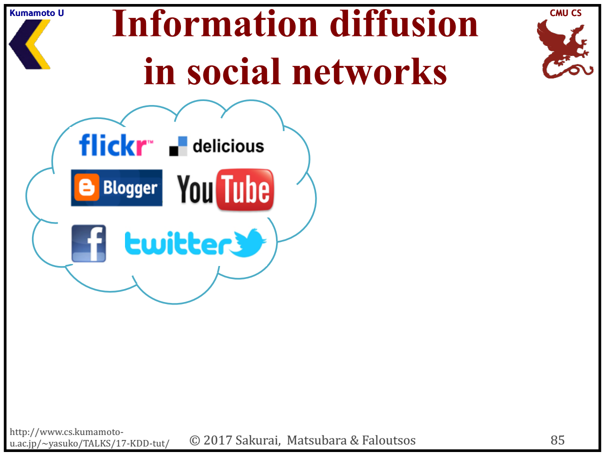
Part 2 Roadmap

Problem
✓ Why: "non-linear" modeling

Fundamentals
✓ Non-linear (grey-box) models

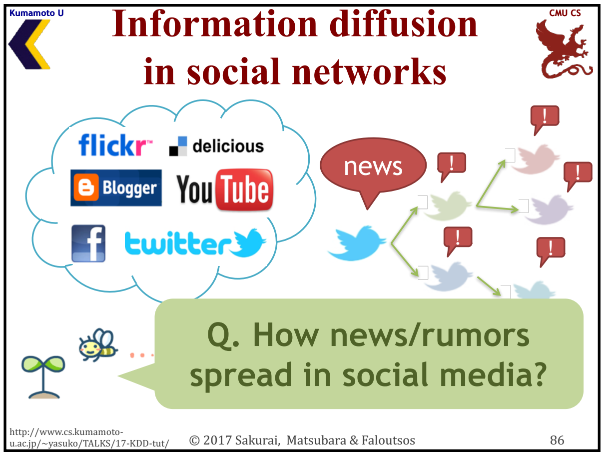
Applications
✓ Epidemics
– Information diffusion
– Online competition

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 84



Information diffusion in social networks

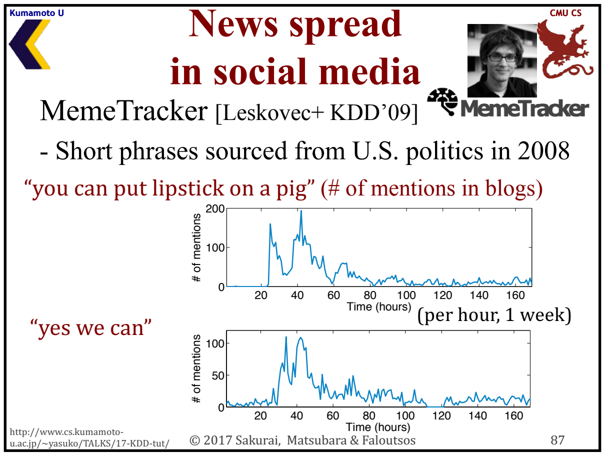
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 85



Information diffusion in social networks

Q. How news/rumors spread in social media?

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 86



News spread in social media

MemeTracker [Leskovec+ KDD'09]

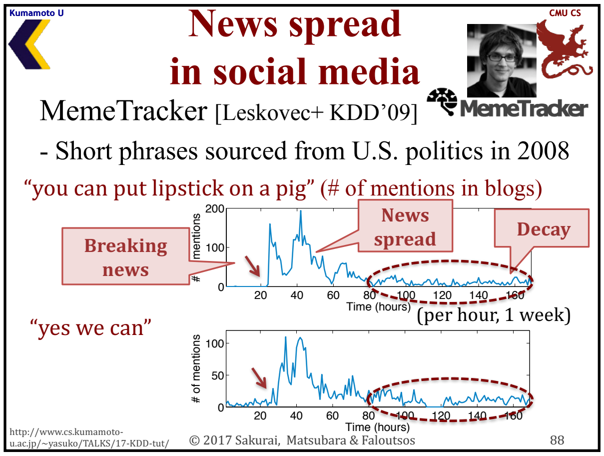
- Short phrases sourced from U.S. politics in 2008

"you can put lipstick on a pig" (# of mentions in blogs)

(per hour, 1 week)

"yes we can"

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 87



News spread in social media

MemeTracker [Leskovec+ KDD'09]

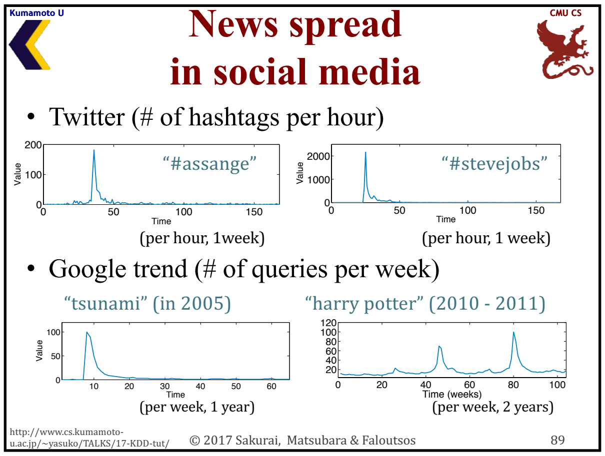
- Short phrases sourced from U.S. politics in 2008

"you can put lipstick on a pig" (# of mentions in blogs)

(per hour, 1 week)

"yes we can"

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 88



News spread in social media

- Twitter (# of hashtags per hour)

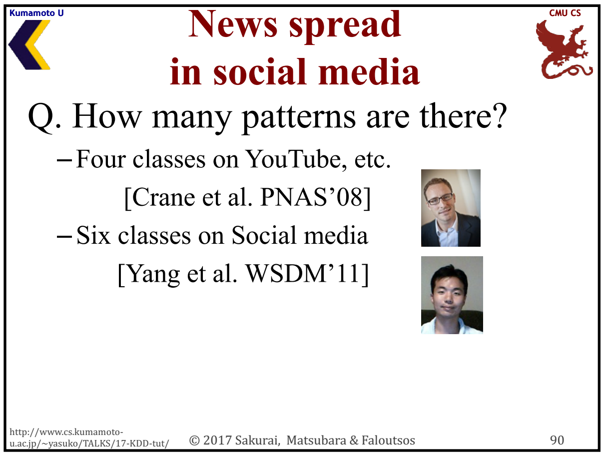
(per hour, 1week) (per hour, 1 week)

- Google trend (# of queries per week)

"tsunami" (in 2005) "harry potter" (2010 - 2011)

(per week, 1 year) (per week, 2 years)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 89

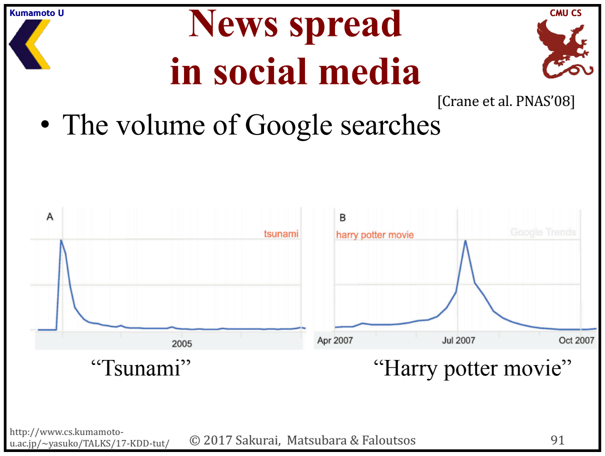


News spread in social media

Q. How many patterns are there?

- Four classes on YouTube, etc. [Crane et al. PNAS'08]
- Six classes on Social media [Yang et al. WSDM'11]

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 90

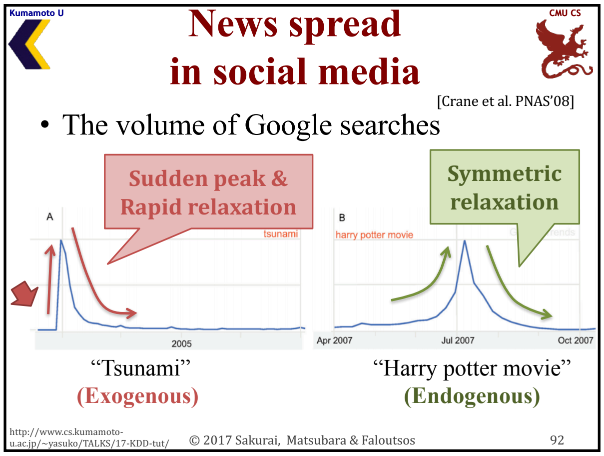


News spread in social media

[Crane et al. PNAS'08]

- The volume of Google searches

"Tsunami" "Harry potter movie"



News spread in social media

[Crane et al. PNAS'08]

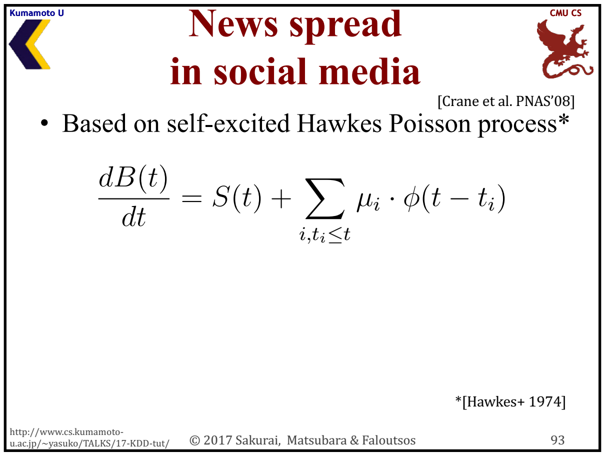
- The volume of Google searches

Sudden peak & Rapid relaxation

Symmetric relaxation

"Tsunami" (Exogenous)

"Harry potter movie" (Endogenous)



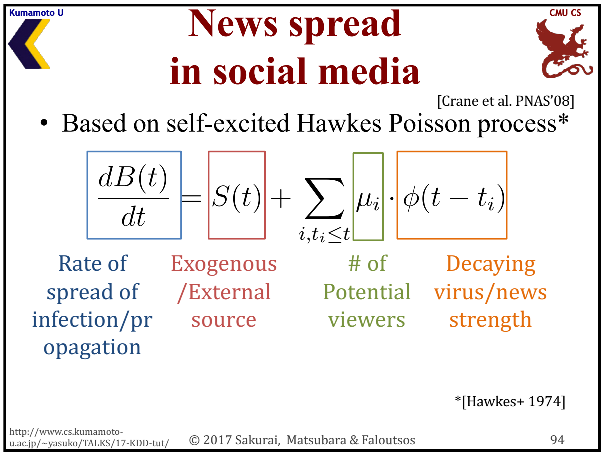
News spread in social media

[Crane et al. PNAS'08]

- Based on self-excited Hawkes Poisson process*

$$\frac{dB(t)}{dt} = S(t) + \sum_{i,t_i \le t} \mu_i \cdot \phi(t - t_i)$$

*[Hawkes+ 1974]



News spread in social media

[Crane et al. PNAS'08]

- Based on self-excited Hawkes Poisson process*

$$\boxed{\frac{dB(t)}{dt}} = \boxed{S(t)} + \sum_{i,t_i \le t} \boxed{\mu_i} \cdot \boxed{\phi(t - t_i)}$$

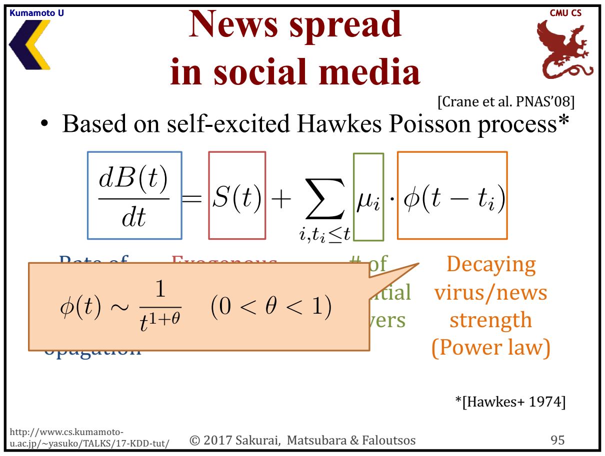
Rate of spread of infection/propagation

Exogenous /External source

# of Potential viewers

Decaying virus/news strength

*[Hawkes+ 1974]



News spread in social media

[Crane et al. PNAS'08]

- Based on self-excited Hawkes Poisson process*

$$\boxed{\frac{dB(t)}{dt}} = \boxed{S(t)} + \sum_{i,t_i \le t} \boxed{\mu_i} \cdot \boxed{\phi(t - t_i)}$$

Rate of Exogenous # of

tial

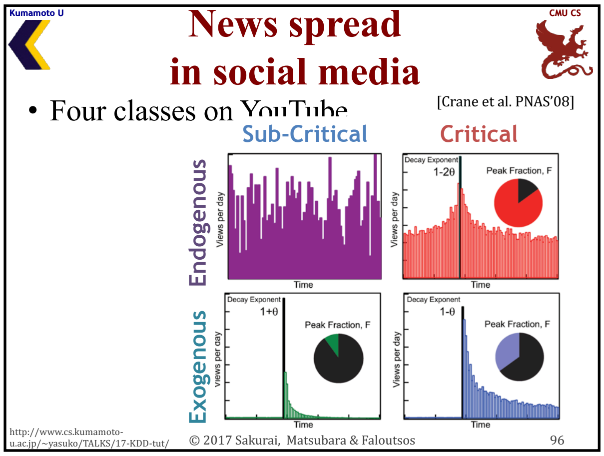
vers

opagation

$$\phi(t) \sim \frac{1}{t^{1+\theta}} \quad (0 < \theta < 1)$$

Decaying virus/news strength (Power law)

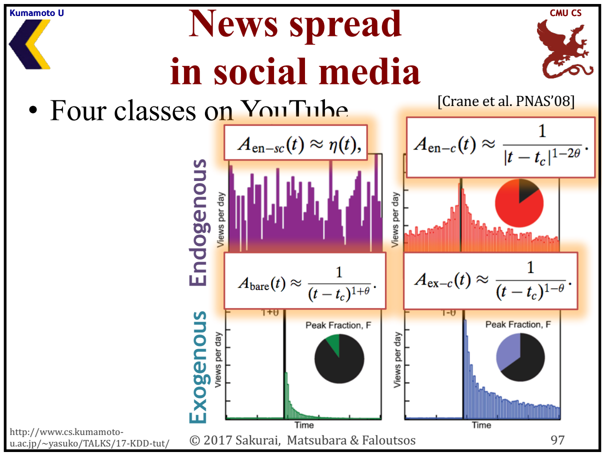
*[Hawkes+ 1974]



News spread in social media

[Crane et al. PNAS'08]

- Four classes on YouTube



News spread in social media

- Four classes on YouTube

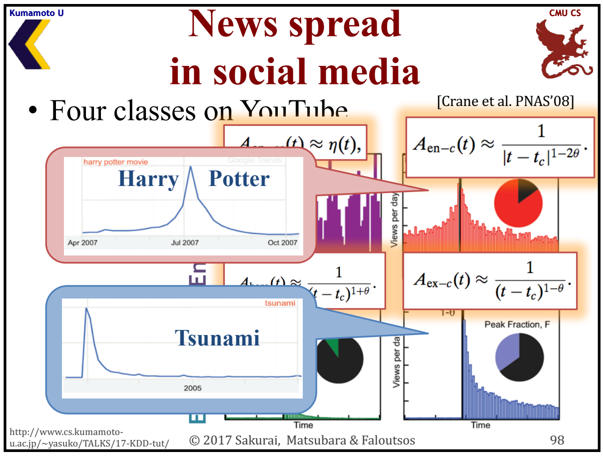
[Crane et al. PNAS'08]

$A_{en-sc}(t) \approx \eta(t),$

$A_{en-c}(t) \approx \frac{1}{|t-t_c|^{1-2\theta}}.$

$A_{bare}(t) \approx \frac{1}{(t-t_c)^{1+\theta}}.$

$A_{ex-c}(t) \approx \frac{1}{(t-t_c)^{1-\theta}}.$



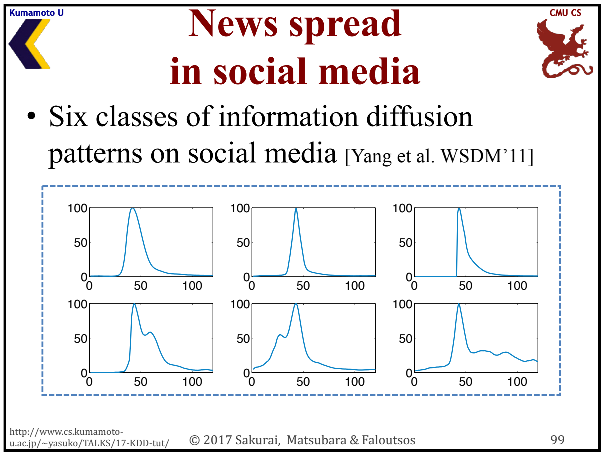
News spread in social media

- Four classes on YouTube

[Crane et al. PNAS'08]

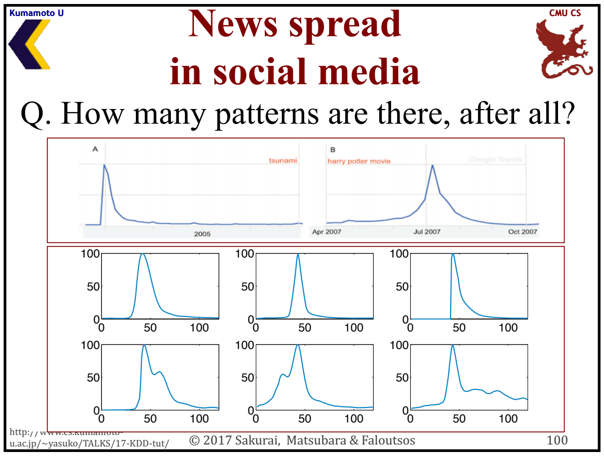
$A_{en-c}(t) \approx \frac{1}{|t-t_c|^{1-2\theta}}.$

$A_{ex-c}(t) \approx \frac{1}{(t-t_c)^{1-\theta}}.$



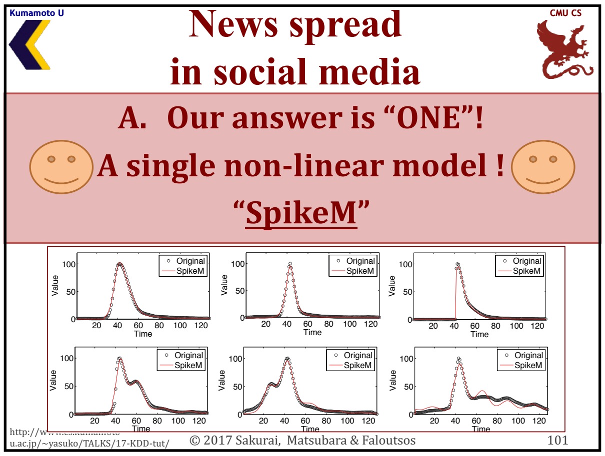
News spread in social media

- Six classes of information diffusion patterns on social media [Yang et al. WSDM'11]



News spread in social media

Q. How many patterns are there, after all?

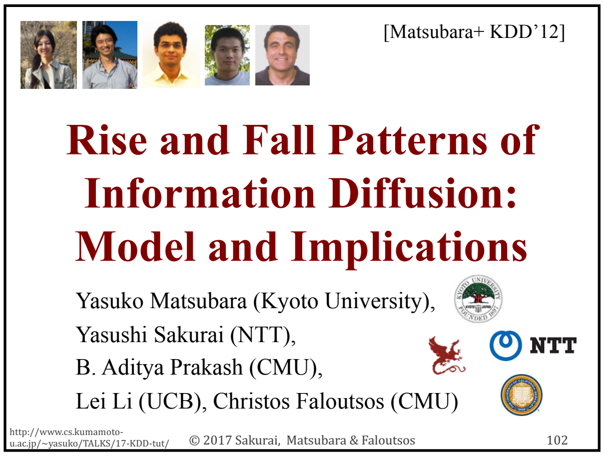


News spread in social media

A. Our answer is "ONE"!

A single non-linear model !

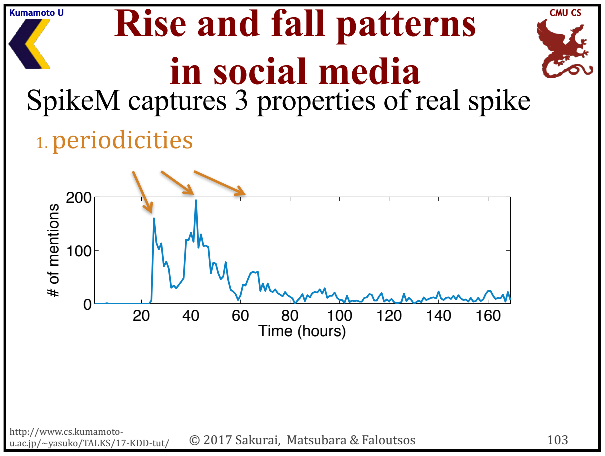
"SpikeM"



[Matsubara+ KDD'12]

Rise and Fall Patterns of Information Diffusion: Model and Implications

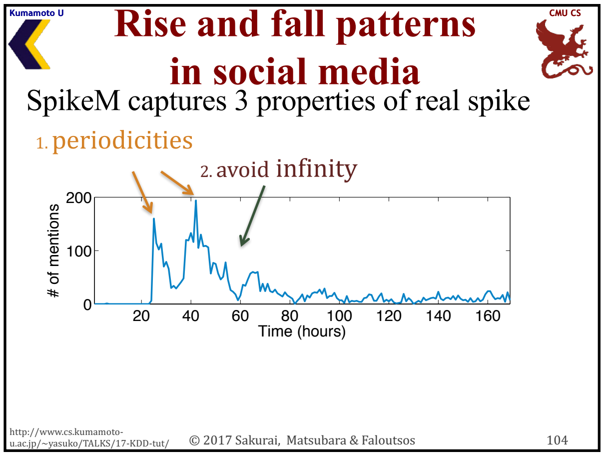
Yasuko Matsubara (Kyoto University),
Yasushi Sakurai (NTT),
B. Aditya Prakash (CMU),
Lei Li (UCB), Christos Faloutsos (CMU)



Rise and fall patterns in social media

SpikeM captures 3 properties of real spike

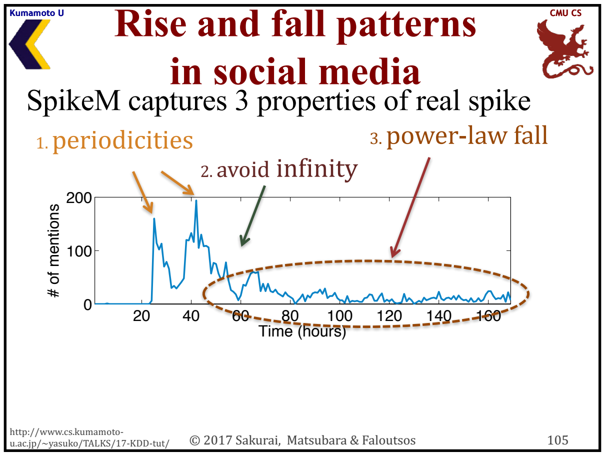
1. periodicities



Rise and fall patterns in social media

SpikeM captures 3 properties of real spike

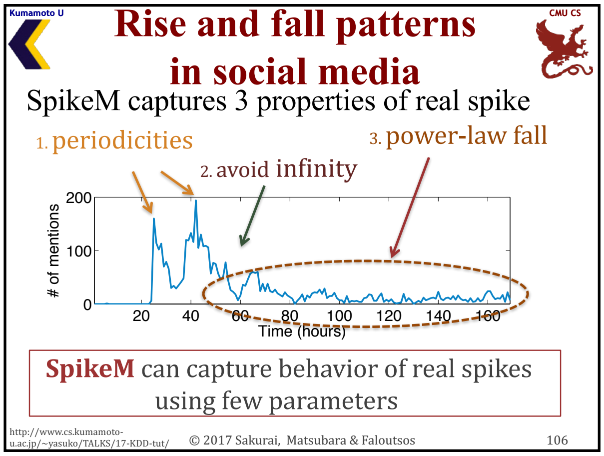
1. periodicities
2. avoid infinity



Rise and fall patterns in social media

SpikeM captures 3 properties of real spike

1. periodicities
2. avoid infinity
3. power-law fall

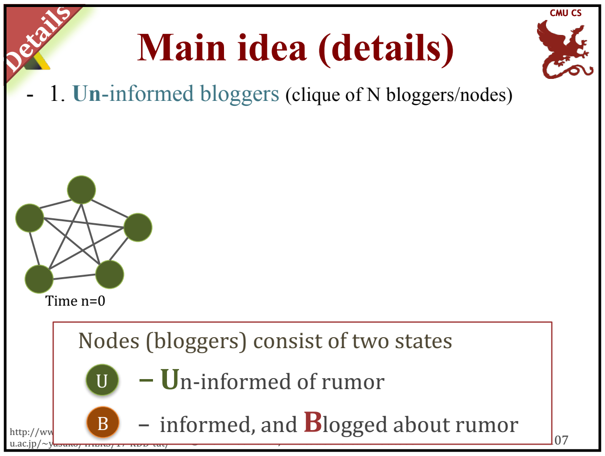


Rise and fall patterns in social media

SpikeM captures 3 properties of real spike

1. periodicities
2. avoid infinity
3. power-law fall

SpikeM can capture behavior of real spikes using few parameters



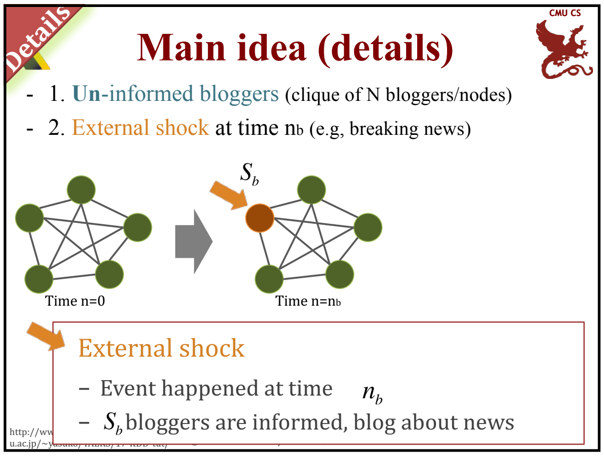
Main idea (details)

- 1. **Un**-informed bloggers (clique of N bloggers/nodes)

Nodes (bloggers) consist of two states

U – **U**n-informed of rumor

B – informed, and **B**logged about rumor

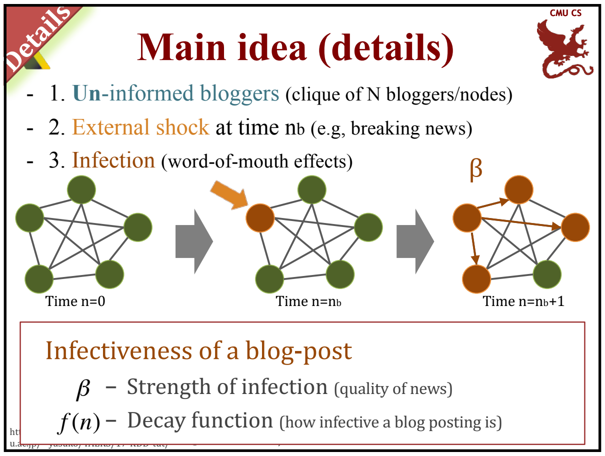


Main idea (details)

- 1. **Un**-informed bloggers (clique of N bloggers/nodes)
- 2. External shock at time n_b (e.g, breaking news)

External shock

- Event happened at time n_b
- S_b bloggers are informed, blog about news



Details

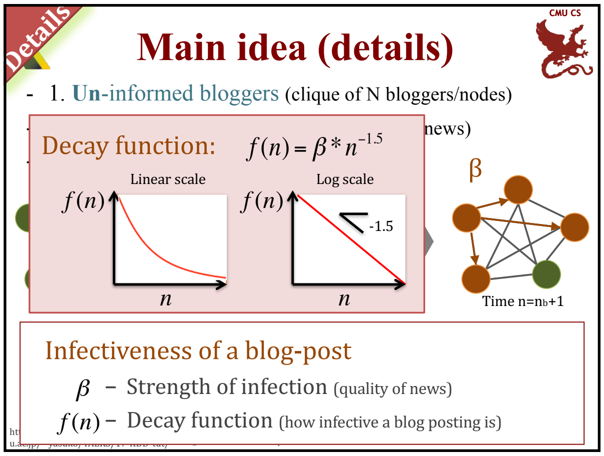
Main idea (details)

- 1. Un-informed bloggers (clique of N bloggers/nodes)
- 2. External shock at time nb (e.g, breaking news)
- 3. Infection (word-of-mouth effects)

Infectiveness of a blog-post

β – Strength of infection (quality of news)

$f(n)$ – Decay function (how infective a blog posting is)



Details

Main idea (details)

- 1. Un-informed bloggers (clique of N bloggers/nodes)

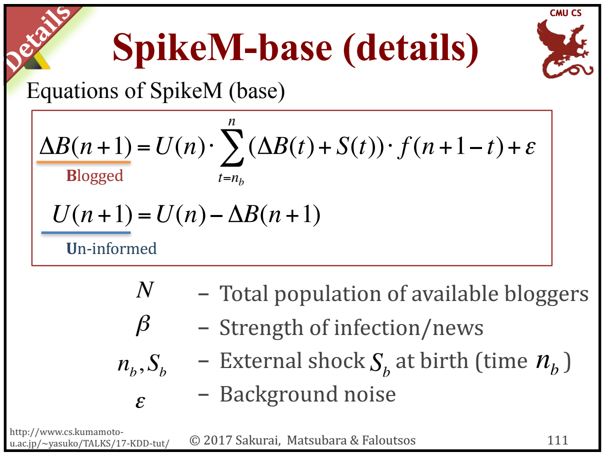
news)

Decay function: $f(n) = \beta * n^{-1.5}$

Infectiveness of a blog-post

β – Strength of infection (quality of news)

$f(n)$ – Decay function (how infective a blog posting is)



Details

SpikeM-base (details)

Equations of SpikeM (base)

$$\underbrace{\Delta B(n+1)}_{\text{Blogged}} = U(n)\cdot\sum_{t=n_b}^{n}(\Delta B(t)+S(t))\cdot f(n+1-t)+\varepsilon$$

$$\underbrace{U(n+1)}_{\text{Un-informed}} = U(n)-\Delta B(n+1)$$

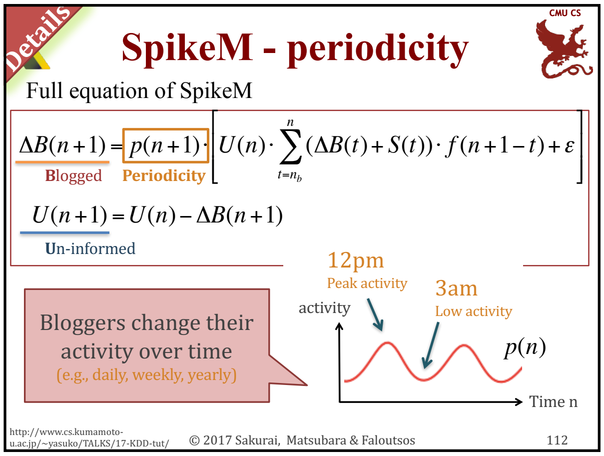
N – Total population of available bloggers

β – Strength of infection/news

n_b, S_b – External shock S_b at birth (time n_b)

ε – Background noise

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 111



Details

SpikeM - periodicity

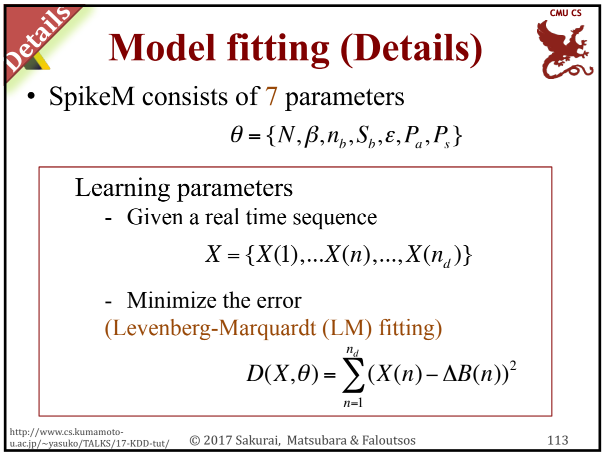
Full equation of SpikeM

$$\underbrace{\Delta B(n+1)}_{\text{Blogged}} = \underbrace{\boxed{p(n+1)}}_{\text{Periodicity}}\cdot\left[U(n)\cdot\sum_{t=n_b}^{n}(\Delta B(t)+S(t))\cdot f(n+1-t)+\varepsilon\right]$$

$$\underbrace{U(n+1)}_{\text{Un-informed}} = U(n)-\Delta B(n+1)$$

Bloggers change their activity over time (e.g., daily, weekly, yearly)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 112



Details

Model fitting (Details)

- SpikeM consists of 7 parameters

$$\theta = \{N, \beta, n_b, S_b, \varepsilon, P_a, P_s\}$$

Learning parameters

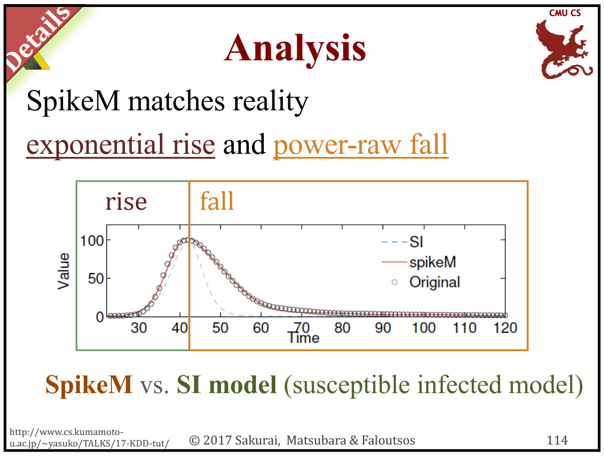
- Given a real time sequence

$$X = \{X(1), \ldots X(n), \ldots, X(n_d)\}$$

- Minimize the error (Levenberg-Marquardt (LM) fitting)

$$D(X,\theta) = \sum_{n=1}^{n_d}(X(n)-\Delta B(n))^2$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 113



Details

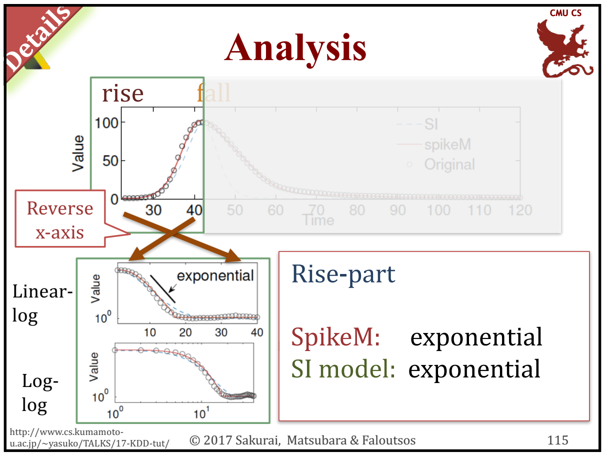
Analysis

SpikeM matches reality

exponential rise and power-raw fall

SpikeM vs. SI model (susceptible infected model)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 114



Analysis

Reverse x-axis

Linear-log

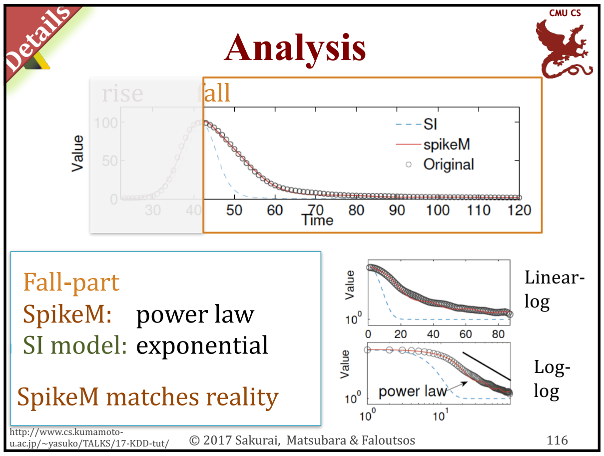
Log-log

exponential

Rise-part

SpikeM: exponential

SI model: exponential



Analysis

Fall-part

SpikeM: power law

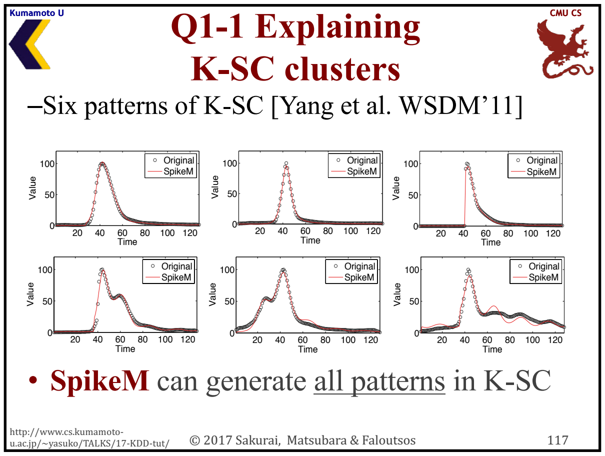
SI model: exponential

SpikeM matches reality

Linear-log

Log-log

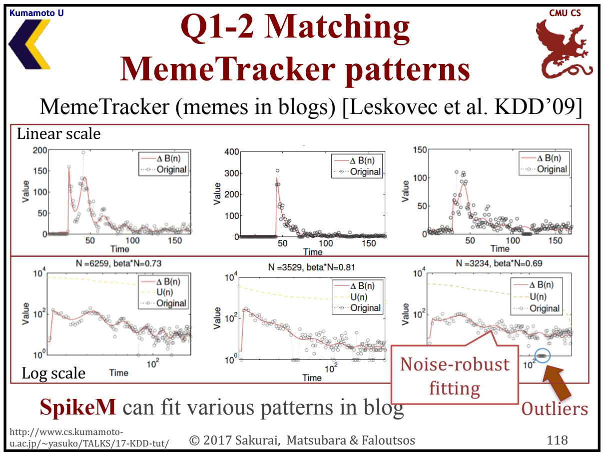
power law



Q1-1 Explaining K-SC clusters

–Six patterns of K-SC [Yang et al. WSDM'11]

- **SpikeM** can generate all patterns in K-SC



Q1-2 Matching MemeTracker patterns

MemeTracker (memes in blogs) [Leskovec et al. KDD'09]

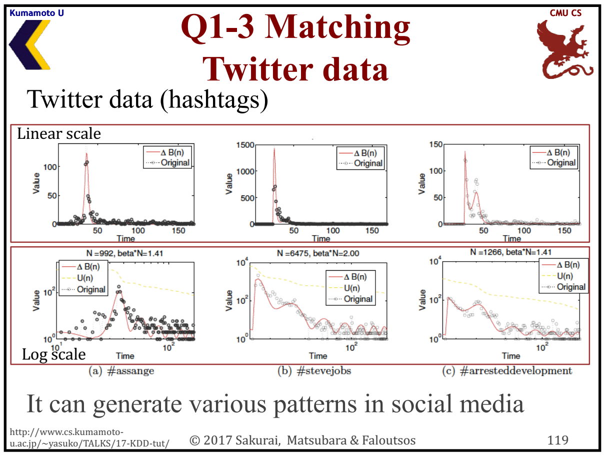
Linear scale

Log scale

Noise-robust fitting

Outliers

SpikeM can fit various patterns in blog



Q1-3 Matching Twitter data

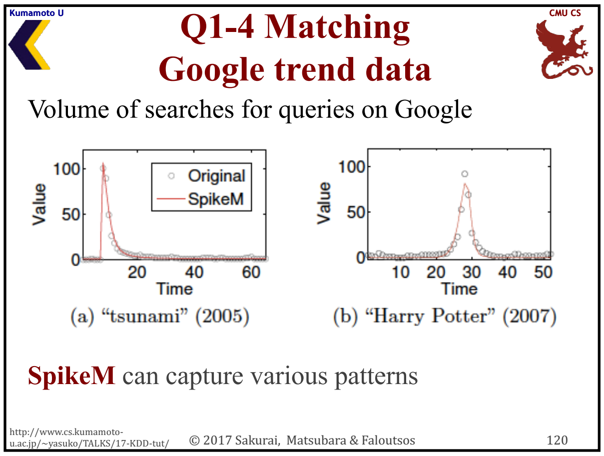
Twitter data (hashtags)

Linear scale

Log scale

(a) #assange (b) #stevejobs (c) #arresteddevelopment

It can generate various patterns in social media

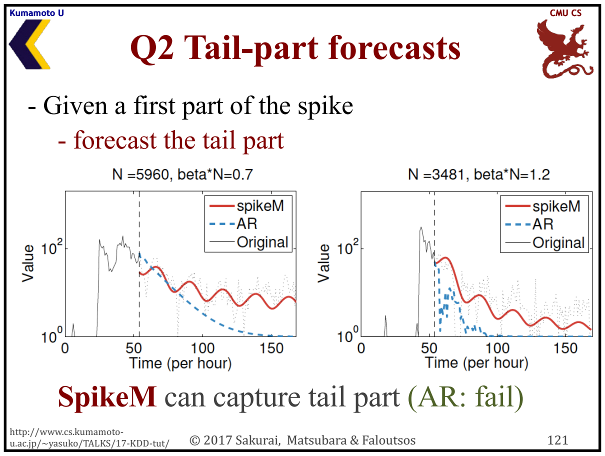


Q1-4 Matching Google trend data

Volume of searches for queries on Google

(a) "tsunami" (2005) (b) "Harry Potter" (2007)

SpikeM can capture various patterns

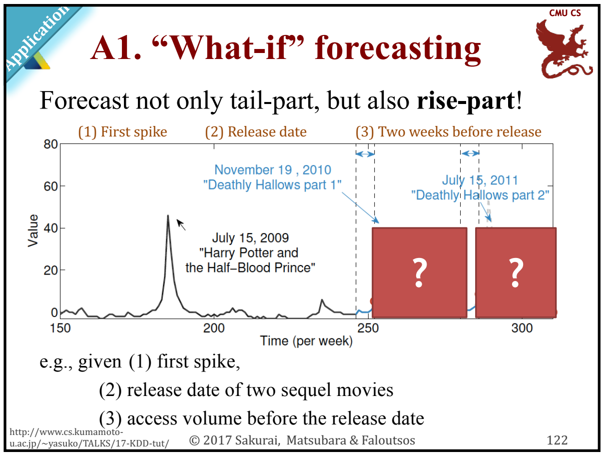


Q2 Tail-part forecasts

- Given a first part of the spike
  - forecast the tail part

SpikeM can capture tail part (AR: fail)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 121



A1. “What-if” forecasting

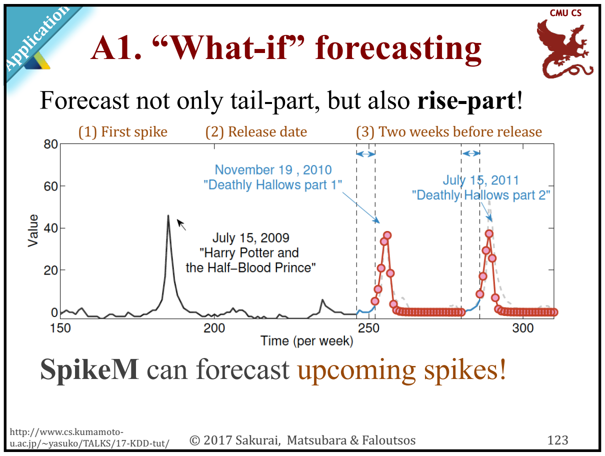
Forecast not only tail-part, but also **rise-part**!

e.g., given (1) first spike,

(2) release date of two sequel movies

(3) access volume before the release date

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 122

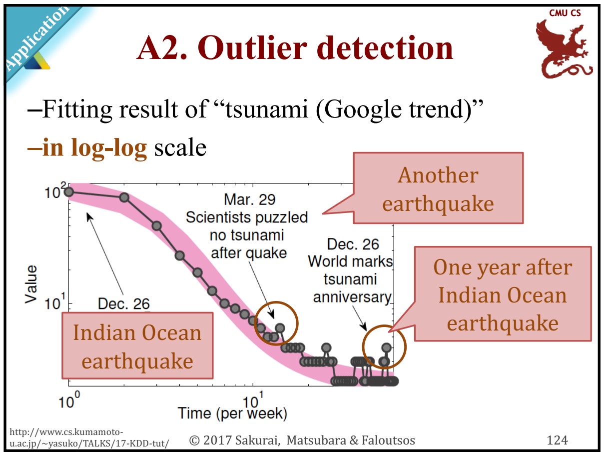


A1. “What-if” forecasting

Forecast not only tail-part, but also **rise-part**!

SpikeM can forecast upcoming spikes!

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 123

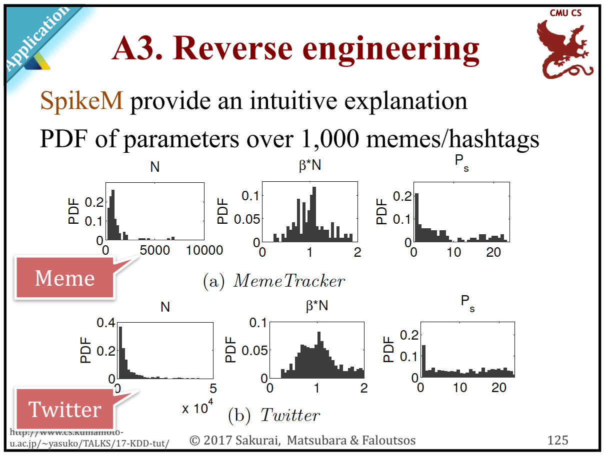


A2. Outlier detection

–Fitting result of “tsunami (Google trend)”

–**in log-log** scale

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 124

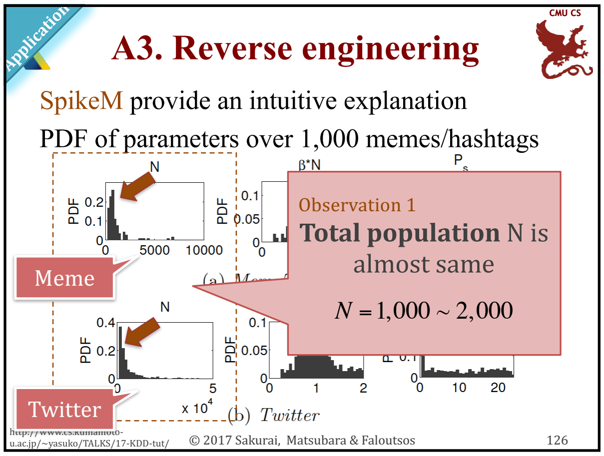


A3. Reverse engineering

SpikeM provide an intuitive explanation

PDF of parameters over 1,000 memes/hashtags

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 125



A3. Reverse engineering

SpikeM provide an intuitive explanation

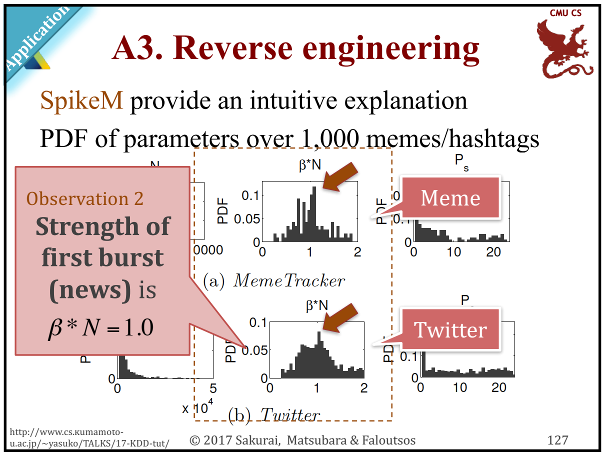
PDF of parameters over 1,000 memes/hashtags

Observation 1

Total population N is almost same

$N = 1{,}000 \sim 2{,}000$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 126



A3. Reverse engineering

SpikeM provide an intuitive explanation

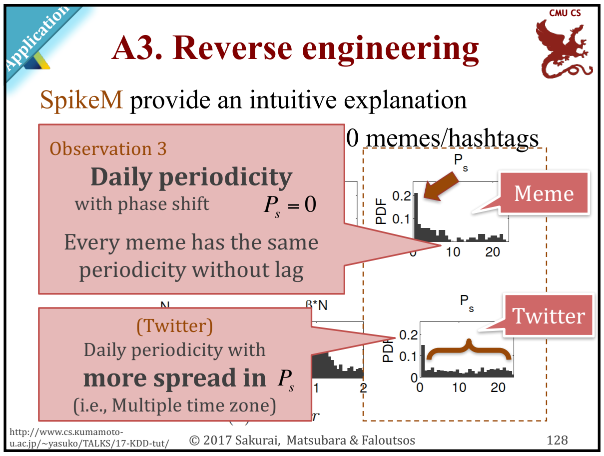
PDF of parameters over 1,000 memes/hashtags

Observation 2

Strength of first burst (news) is $\beta * N = 1.0$

(a) *MemeTracker*

(b) *Twitter*



A3. Reverse engineering

SpikeM provide an intuitive explanation

0 memes/hashtags

Observation 3

Daily periodicity

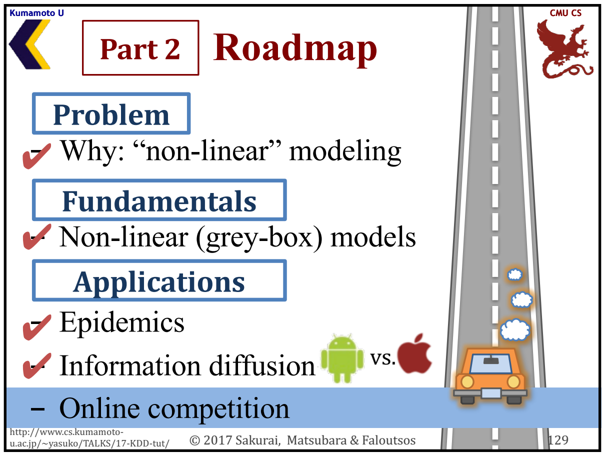
with phase shift $P_s = 0$

Every meme has the same periodicity without lag

(Twitter)

Daily periodicity with **more spread in** P_s

(i.e., Multiple time zone)



Part 2 Roadmap

Problem

✓ Why: “non-linear” modeling

Fundamentals

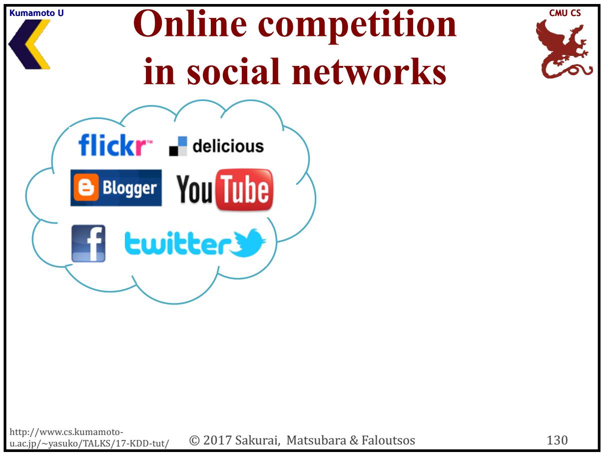
✓ Non-linear (grey-box) models

Applications

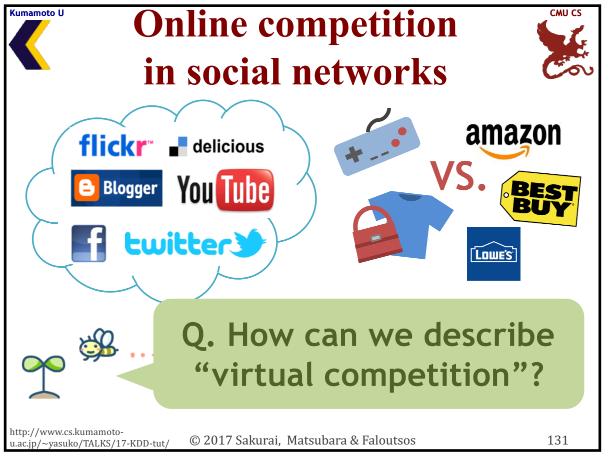
✓ Epidemics

✓ Information diffusion vs.

– Online competition

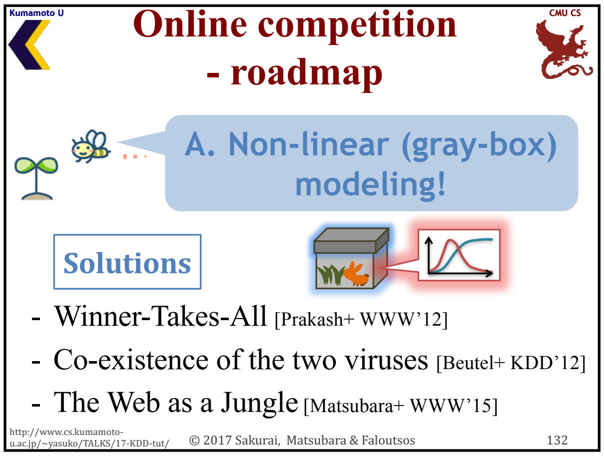


Online competition in social networks



Online competition in social networks

Q. How can we describe “virtual competition”?

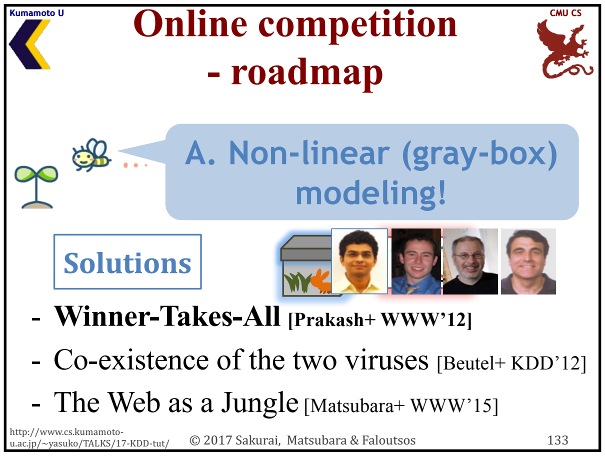


Online competition - roadmap

A. Non-linear (gray-box) modeling!

Solutions

- Winner-Takes-All [Prakash+ WWW’12]
- Co-existence of the two viruses [Beutel+ KDD’12]
- The Web as a Jungle [Matsubara+ WWW’15]



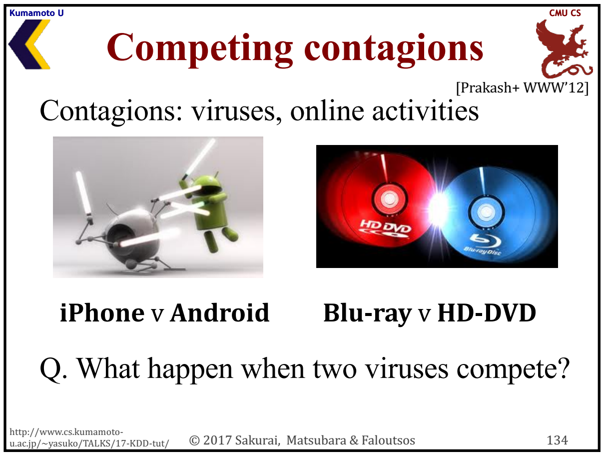
Online competition - roadmap

A. Non-linear (gray-box) modeling!

Solutions

- **Winner-Takes-All [Prakash+ WWW'12]**
- Co-existence of the two viruses [Beutel+ KDD'12]
- The Web as a Jungle [Matsubara+ WWW'15]

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 133



Competing contagions

[Prakash+ WWW'12]

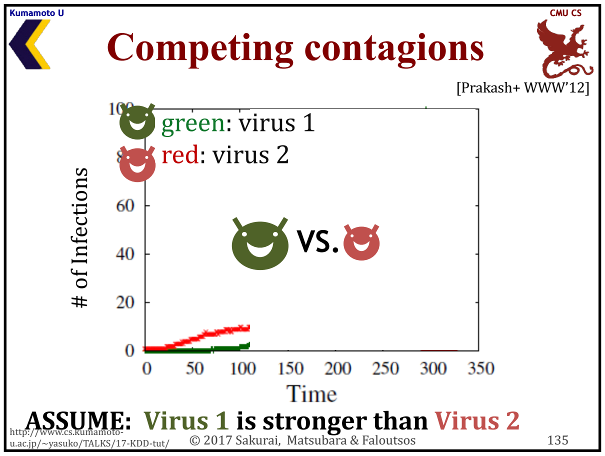
Contagions: viruses, online activities

iPhone v **Android**

Blu-ray v **HD-DVD**

Q. What happen when two viruses compete?

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 134

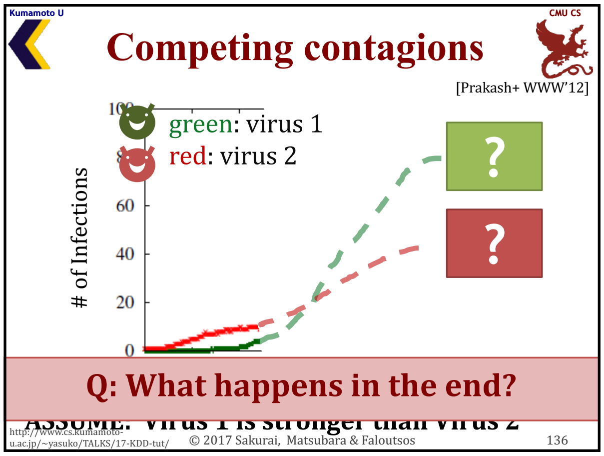


Competing contagions

[Prakash+ WWW'12]

ASSUME: Virus 1 is stronger than Virus 2

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 135



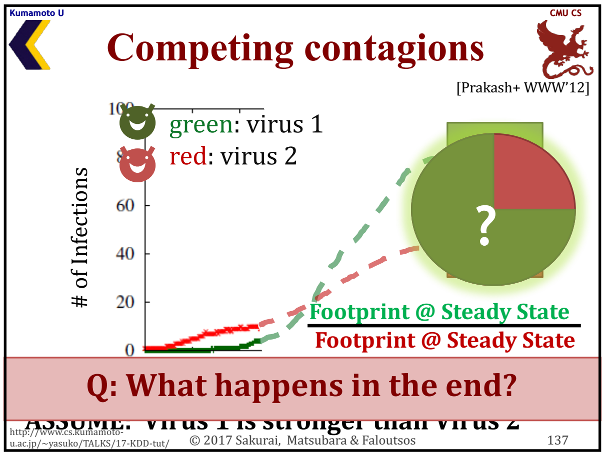
Competing contagions

[Prakash+ WWW'12]

Q: What happens in the end?

ASSUME: Virus 1 is stronger than Virus 2

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 136



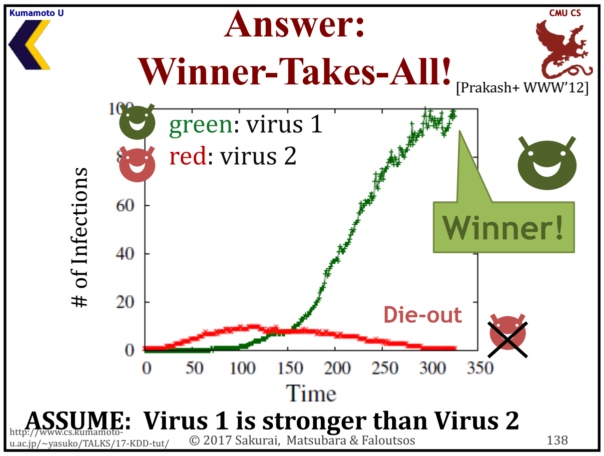
Competing contagions

[Prakash+ WWW'12]

Q: What happens in the end?

ASSUME: Virus 1 is stronger than Virus 2

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 137



Answer: Winner-Takes-All!

[Prakash+ WWW'12]

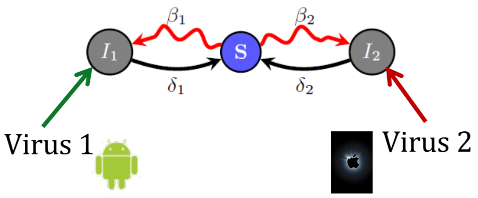
ASSUME: Virus 1 is stronger than Virus 2

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 138

A simple model

[Prakash+ WWW'12]

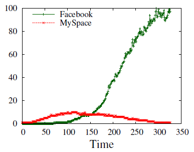
- Modified flu-like (SIS) model
- Mutual Immunity ("pick one of the two")
- Susceptible-Infected1-Infected2-Susceptible



Result: Winner-Takes-All

[Prakash+ WWW'12]

Given this model,
and ***any graph***,
the weaker virus always
dies-out, *completely*

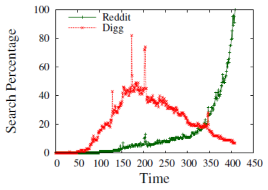


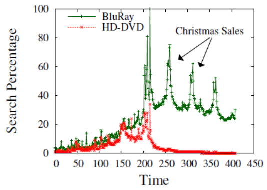
1. The stronger survives only if it is above threshold
2. Virus 1 is stronger than Virus 2, if:
   strength(Virus 1) > strength(Virus 2)
3. Strength(Virus) = λ β / δ → same as before!

Real Examples of "WTA"

[Prakash+ WWW'12]

[Google Search Trends data]



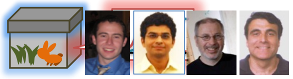


Reddit v **Digg**

Blu-Ray v **HD-DVD**

Online competition in social networks

A. Non-linear (gray-box) modeling!



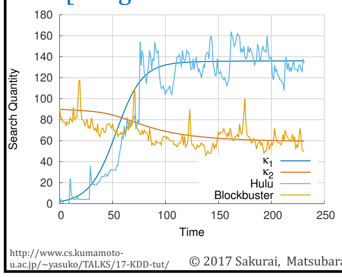
Solutions

- Winner-Takes-All [Prakash+ WWW'12]
- **Co-existence of the two viruses [Beutel+ KDD'12]**
- The Web as a Jungle [Matsubara+ WWW'15]

Interacting Viruses: Can Both Survive?

Real example of "co-existence"

[Google Search Trends data]

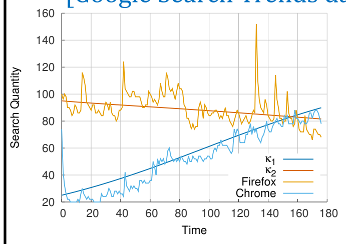


Hulu v **Blockbuster**

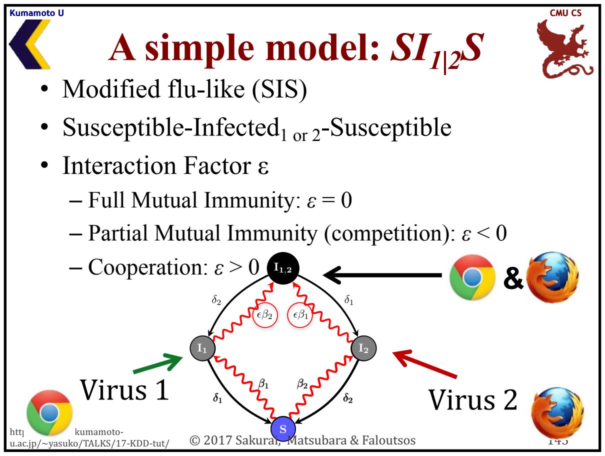
Interacting Viruses: Can Both Survive?

Real example of "co-existence"

[Google Search Trends data]



Chrome v **Firefox**



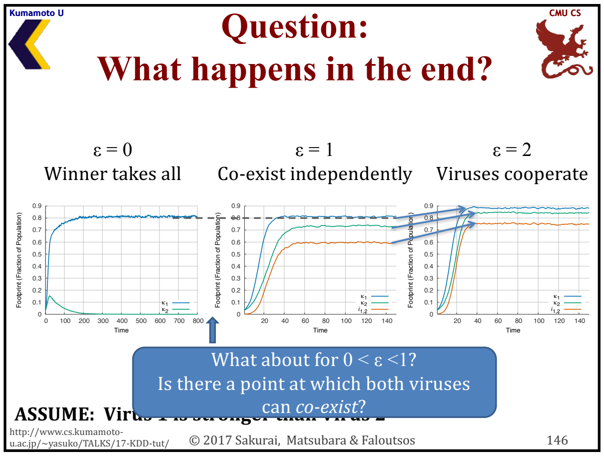
A simple model: $SI_{1|2}S$

- Modified flu-like (SIS)
- Susceptible-Infected$_{1\text{ or }2}$-Susceptible
- Interaction Factor ε
  - Full Mutual Immunity: $\varepsilon = 0$
  - Partial Mutual Immunity (competition): $\varepsilon < 0$
  - Cooperation: $\varepsilon > 0$

Virus 1

Virus 2

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 145



Question: What happens in the end?

ε = 0
Winner takes all

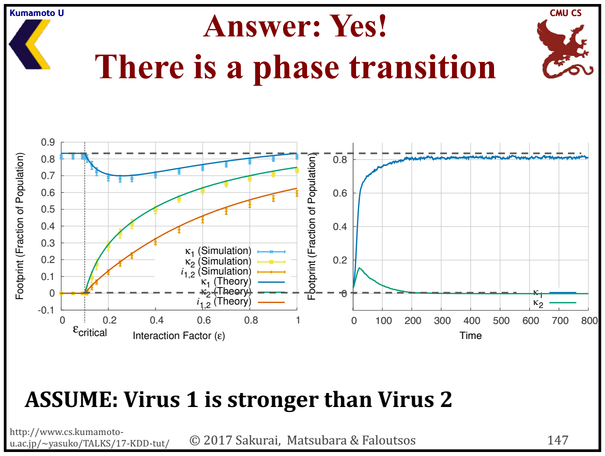
ε = 1
Co-exist independently

ε = 2
Viruses cooperate

What about for 0 < ε <1?
Is there a point at which both viruses can *co-exist*?

ASSUME: Virus 1 is stronger than Virus 2

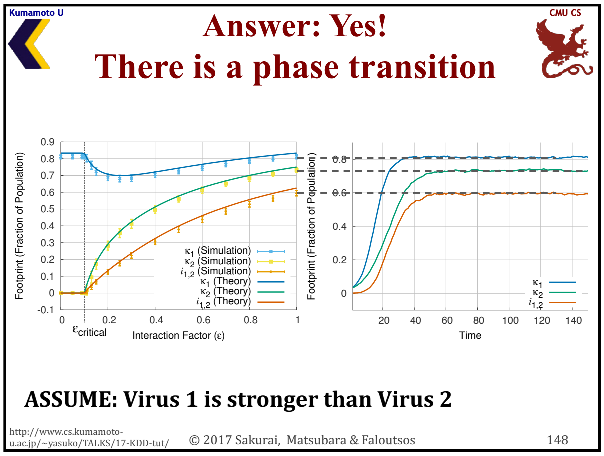
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 146



Answer: Yes! There is a phase transition

ASSUME: Virus 1 is stronger than Virus 2

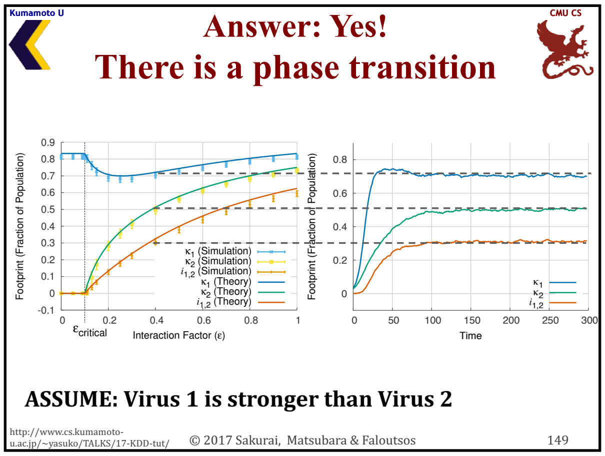
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 147



Answer: Yes! There is a phase transition

ASSUME: Virus 1 is stronger than Virus 2

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 148



Answer: Yes! There is a phase transition

ASSUME: Virus 1 is stronger than Virus 2

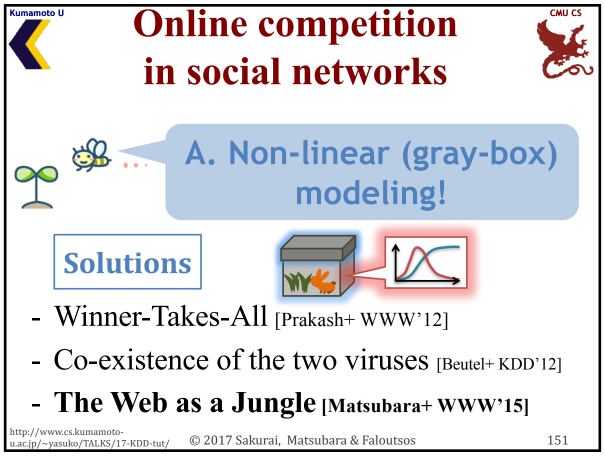
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 149

Result: Viruses can Co-exist

> Given this model and a fully connected graph, there exists an $\varepsilon_{critical}$ such that for $\varepsilon \geq \varepsilon_{critical}$, there is a fixed point where both viruses survive.

1. The stronger survives only if it is above threshold
2. Virus 1 is stronger than Virus 2, if:
   strength(Virus 1) > strength(Virus 2)
3. Strength(Virus) $\sigma = N \beta / \delta$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 150



Kumamoto U

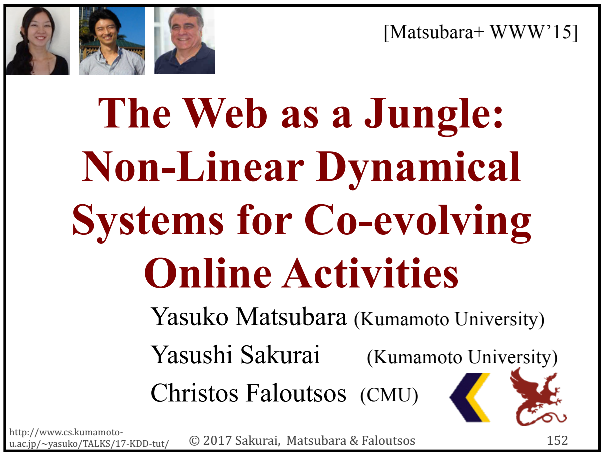
CMU CS

Online competition in social networks

A. Non-linear (gray-box) modeling!

Solutions

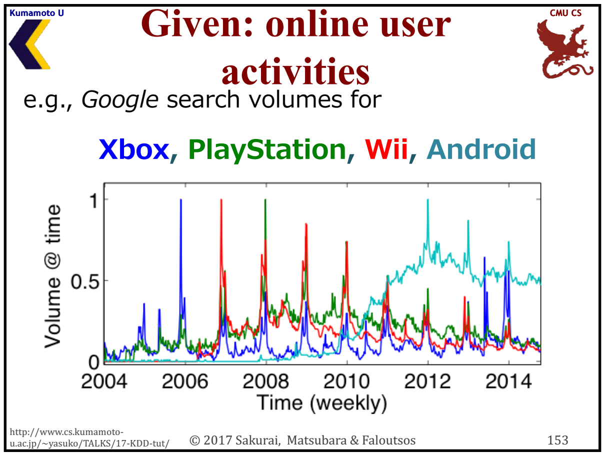
- Winner-Takes-All [Prakash+ WWW'12]
- Co-existence of the two viruses [Beutel+ KDD'12]
- **The Web as a Jungle [Matsubara+ WWW'15]**



[Matsubara+ WWW'15]

The Web as a Jungle: Non-Linear Dynamical Systems for Co-evolving Online Activities

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



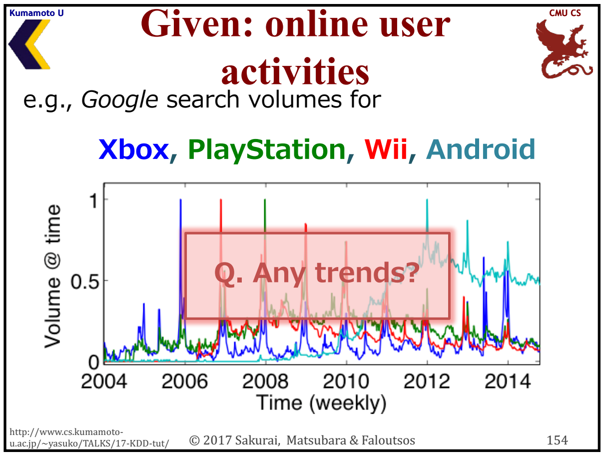
Kumamoto U

CMU CS

Given: online user activities

e.g., *Google* search volumes for

Xbox, PlayStation, Wii, Android



Kumamoto U

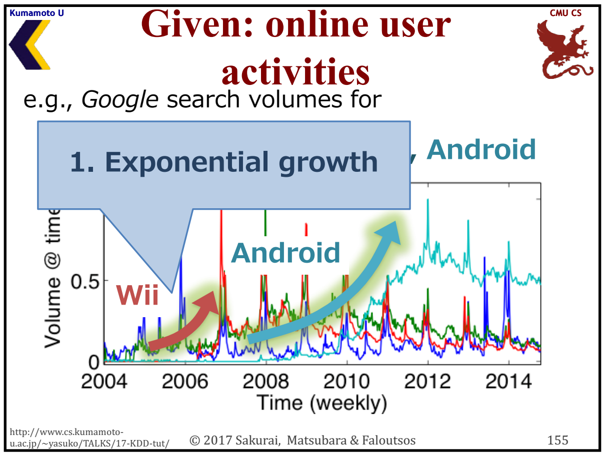
CMU CS

Given: online user activities

e.g., *Google* search volumes for

Xbox, PlayStation, Wii, Android

Q. Any trends?



Kumamoto U

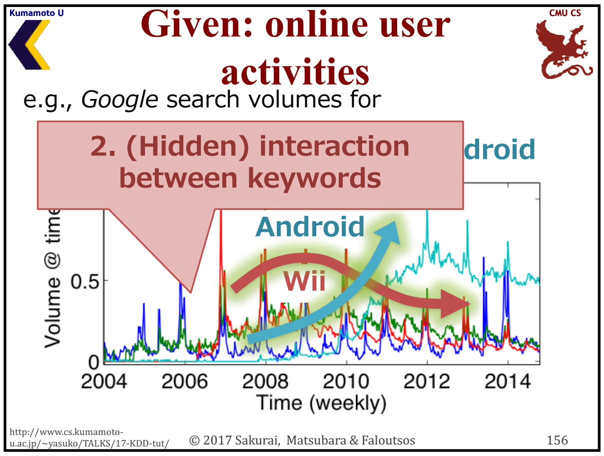
CMU CS

Given: online user activities

e.g., *Google* search volumes for

, Android

1. Exponential growth



Kumamoto U

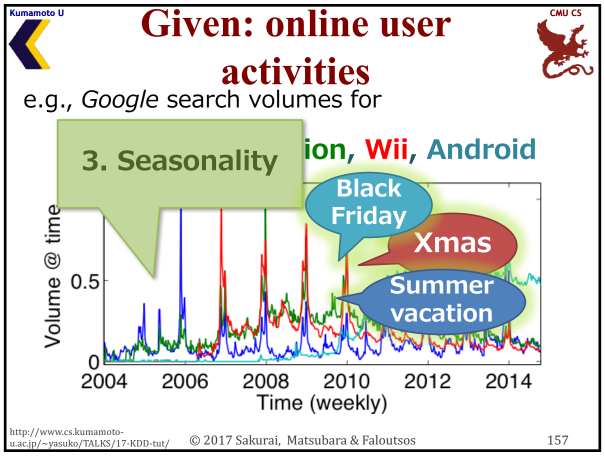
CMU CS

Given: online user activities

e.g., *Google* search volumes for

droid

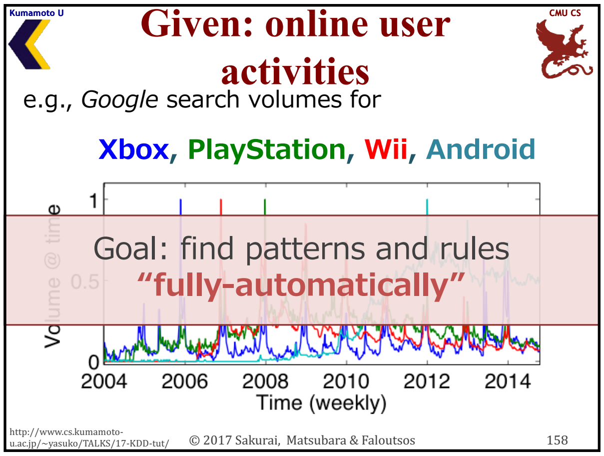
2. (Hidden) interaction between keywords



Given: online user activities

e.g., *Google* search volumes for

ion, Wii, Android

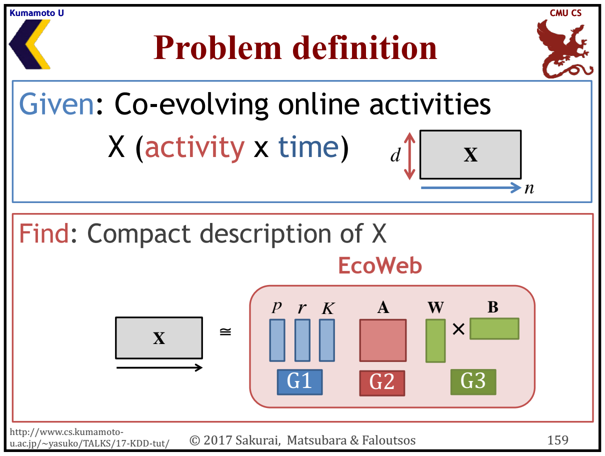


Given: online user activities

e.g., *Google* search volumes for

Xbox, PlayStation, Wii, Android

Goal: find patterns and rules
"fully-automatically"

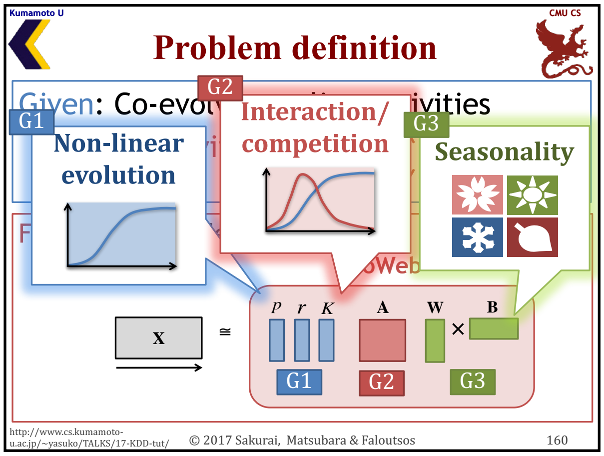


Problem definition

Given: Co-evolving online activities
X (activity x time)

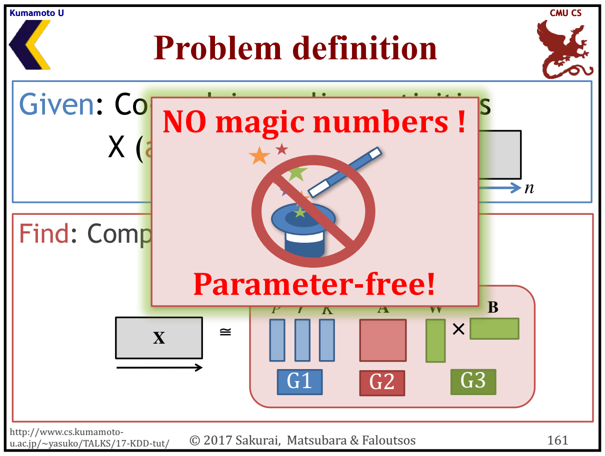
Find: Compact description of X

EcoWeb



Problem definition

Given: Co-evol ivities



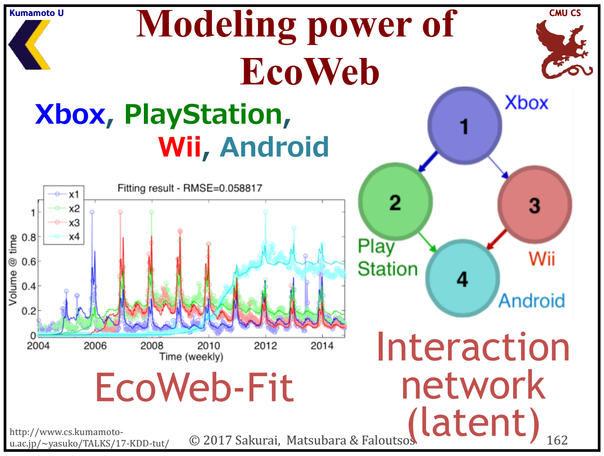
Problem definition

Given: Co s
X (

Find: Comp

NO magic numbers !

Parameter-free!

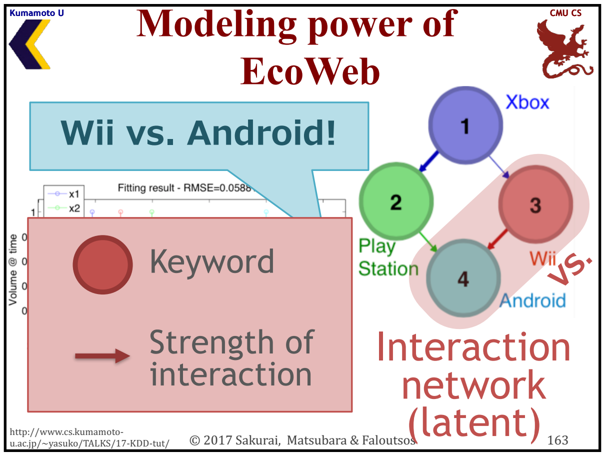


Modeling power of EcoWeb

Xbox, PlayStation, Wii, Android

EcoWeb-Fit

Interaction network (latent)



Kumamoto U

Modeling power of EcoWeb

CMU CS

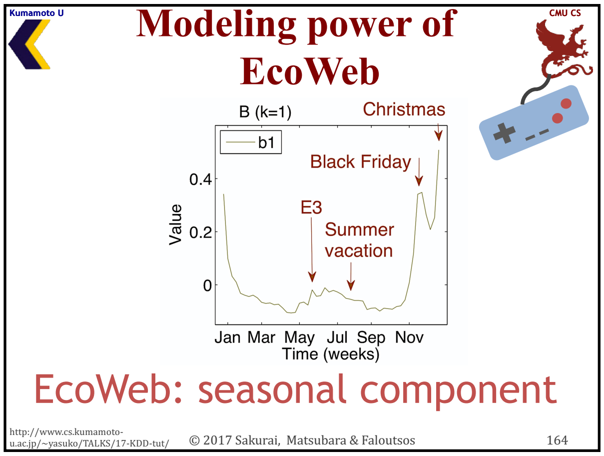
Wii vs. Android!

Keyword

Strength of interaction

Interaction network (latent)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 163



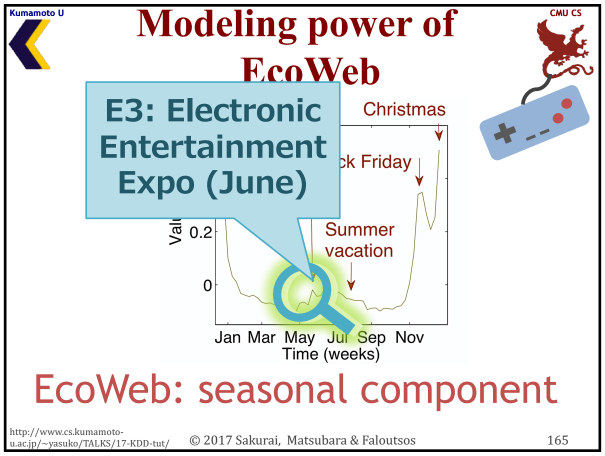
Kumamoto U

Modeling power of EcoWeb

CMU CS

EcoWeb: seasonal component

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 164



Kumamoto U

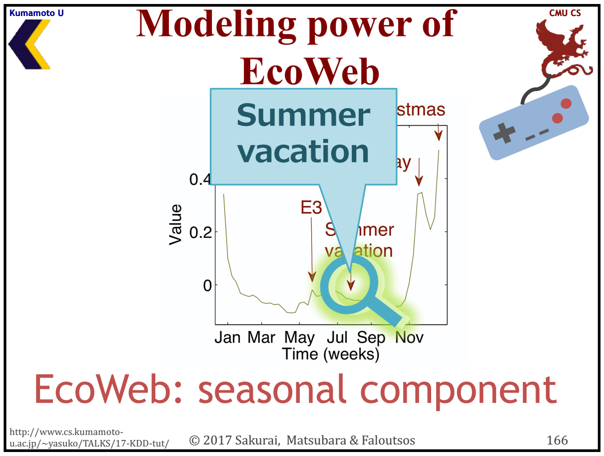
Modeling power of EcoWeb

CMU CS

E3: Electronic Entertainment Expo (June)

EcoWeb: seasonal component

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 165



Kumamoto U

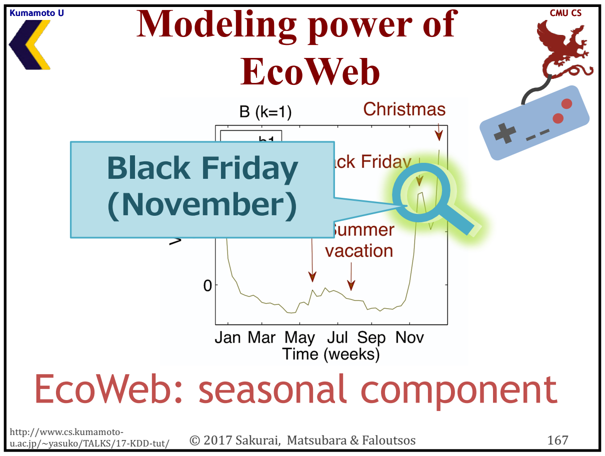
Modeling power of EcoWeb

CMU CS

Summer vacation

EcoWeb: seasonal component

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 166



Kumamoto U

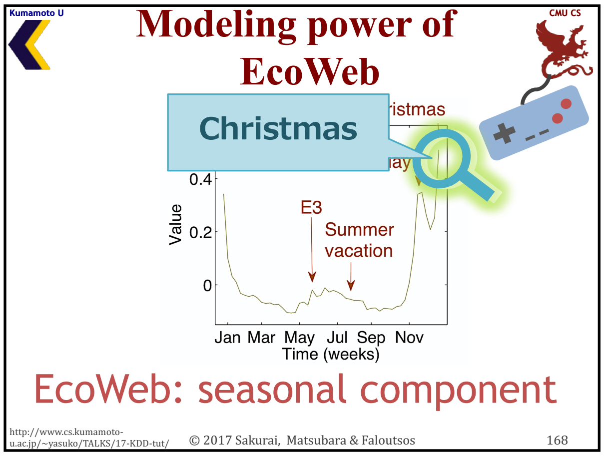
Modeling power of EcoWeb

CMU CS

Black Friday (November)

EcoWeb: seasonal component

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 167



Kumamoto U

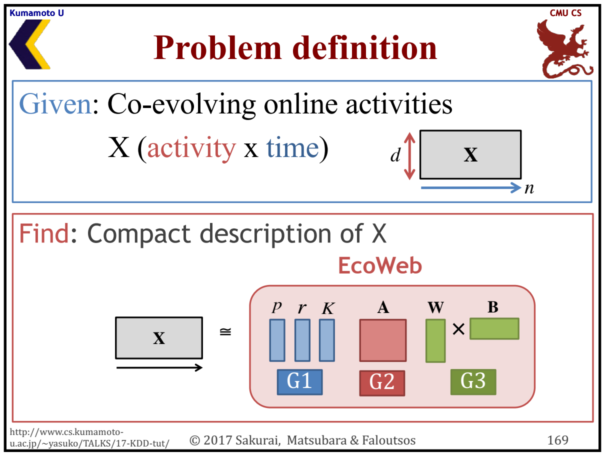
Modeling power of EcoWeb

CMU CS

Christmas

EcoWeb: seasonal component

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 168

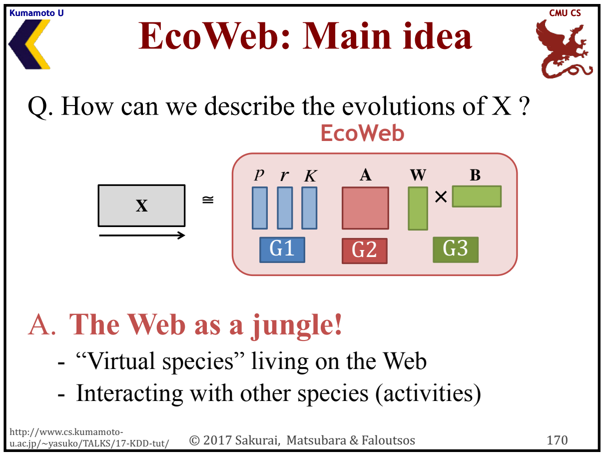


Problem definition

Given: Co-evolving online activities

X (activity x time)

Find: Compact description of X

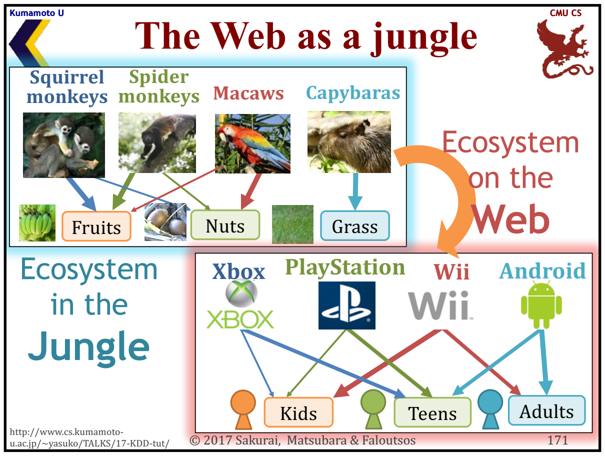


EcoWeb: Main idea

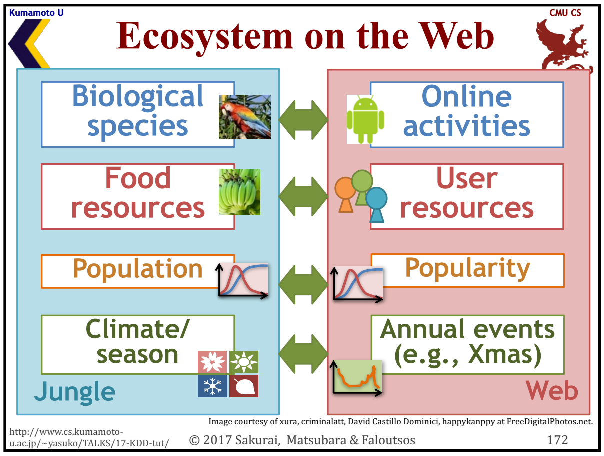
Q. How can we describe the evolutions of X ?

A. **The Web as a jungle!**

- "Virtual species" living on the Web
- Interacting with other species (activities)

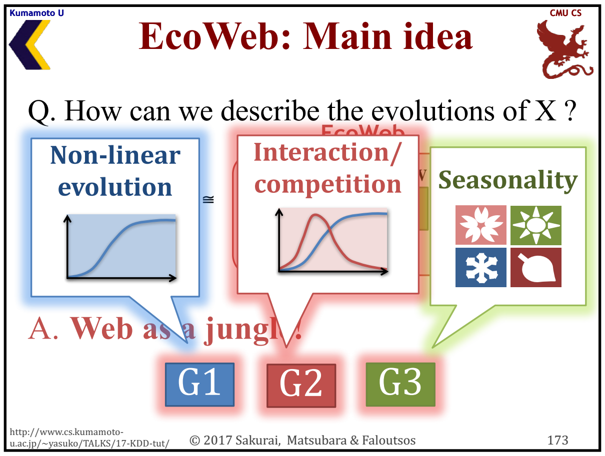


The Web as a jungle



Ecosystem on the Web

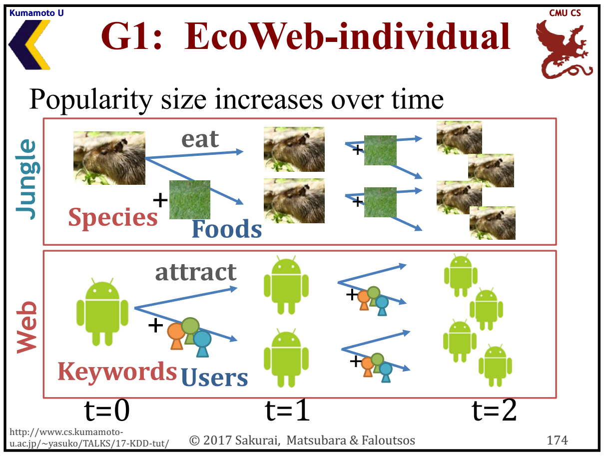
Image courtesy of xura, criminalatt, David Castillo Dominici, happykanppy at FreeDigitalPhotos.net.



EcoWeb: Main idea

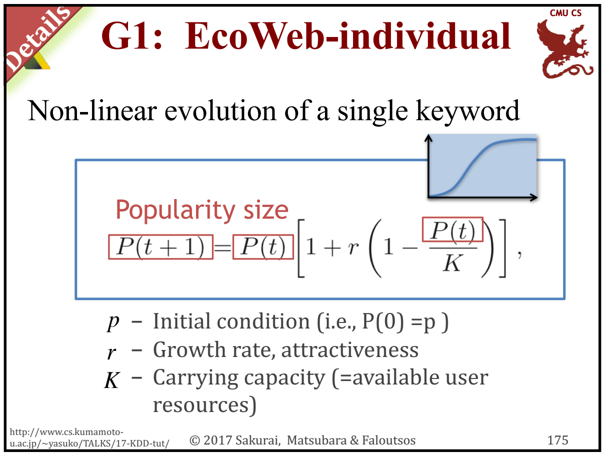
Q. How can we describe the evolutions of X ?

A. Web as a jungle!



G1: EcoWeb-individual

Popularity size increases over time



Details

G1: EcoWeb-individual

CMU CS

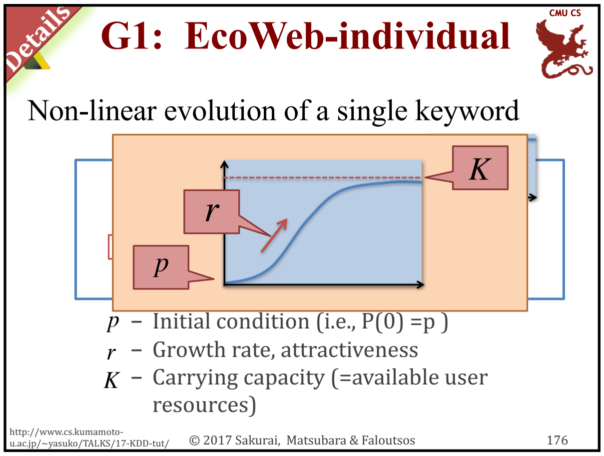
Non-linear evolution of a single keyword

Popularity size

$$P(t+1) = P(t)\left[1 + r\left(1 - \frac{P(t)}{K}\right)\right],$$

- p – Initial condition (i.e., P(0) =p)
- r – Growth rate, attractiveness
- K – Carrying capacity (=available user resources)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 175



Details

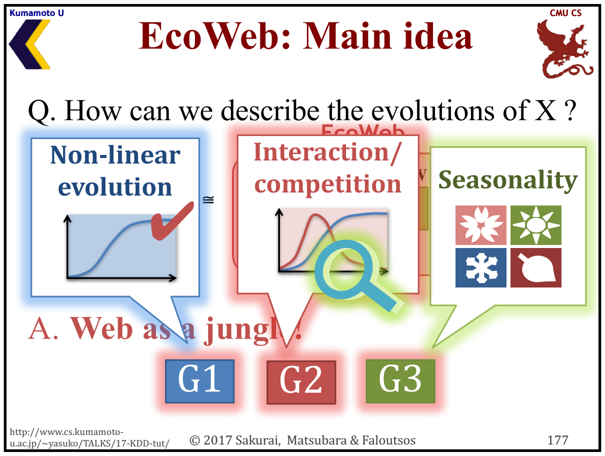
G1: EcoWeb-individual

CMU CS

Non-linear evolution of a single keyword

- p – Initial condition (i.e., P(0) =p)
- r – Growth rate, attractiveness
- K – Carrying capacity (=available user resources)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 176



Kumamoto U

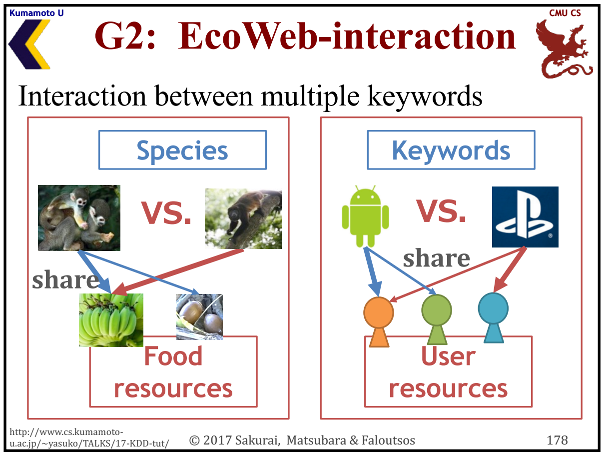
EcoWeb: Main idea

CMU CS

Q. How can we describe the evolutions of X ?

A. Web as a jungl !

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 177



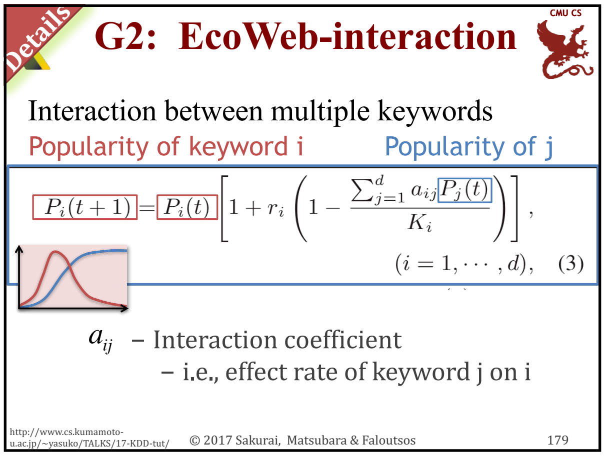
Kumamoto U

G2: EcoWeb-interaction

CMU CS

Interaction between multiple keywords

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 178



Details

G2: EcoWeb-interaction

CMU CS

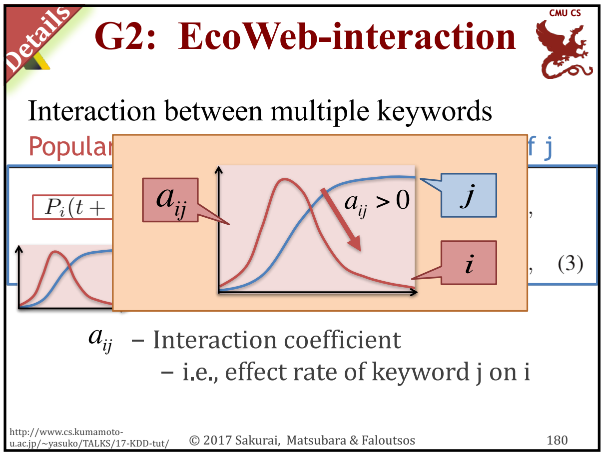
Interaction between multiple keywords

Popularity of keyword i Popularity of j

$$P_i(t+1) = P_i(t)\left[1 + r_i\left(1 - \frac{\sum_{j=1}^{d} a_{ij} P_j(t)}{K_i}\right)\right], \quad (i = 1, \cdots, d), \quad (3)$$

a_{ij} – Interaction coefficient
– i.e., effect rate of keyword j on i

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 179



Details

G2: EcoWeb-interaction

CMU CS

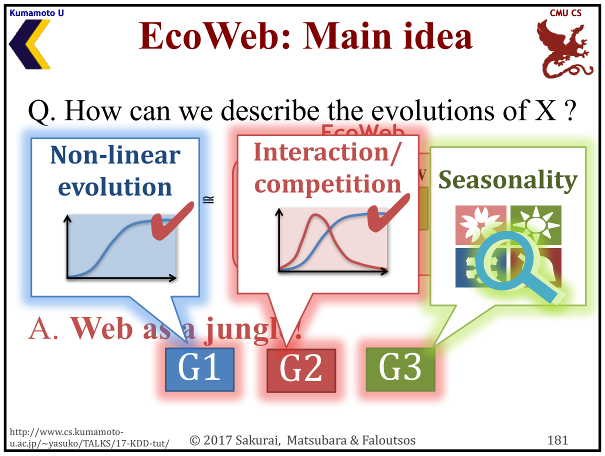
Interaction between multiple keywords

Popular f j

$P_i(t+$, (3)

a_{ij} – Interaction coefficient
– i.e., effect rate of keyword j on i

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 180

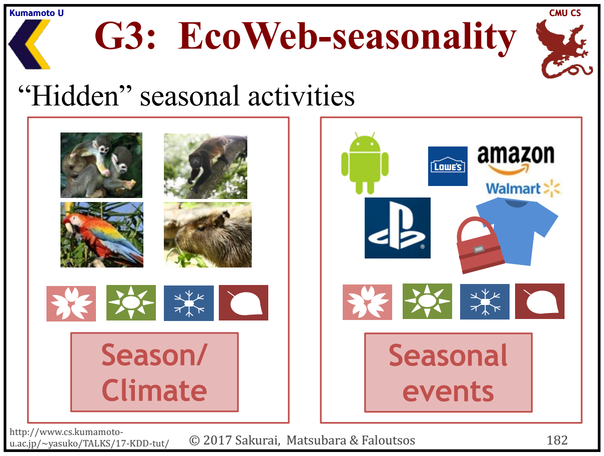


EcoWeb: Main idea

Q. How can we describe the evolutions of X ?

A. Web as a jungl !

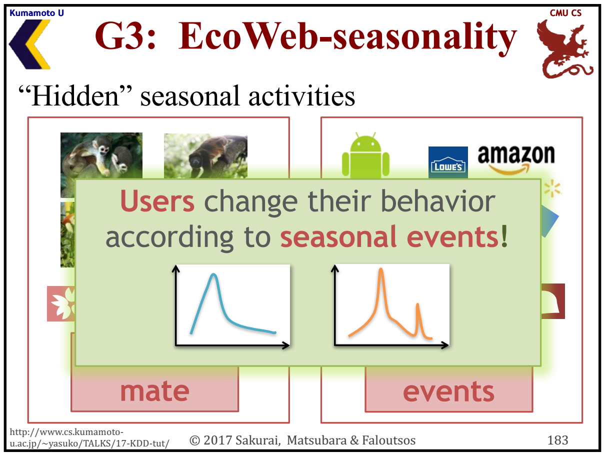
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 181



G3: EcoWeb-seasonality

"Hidden" seasonal activities

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 182

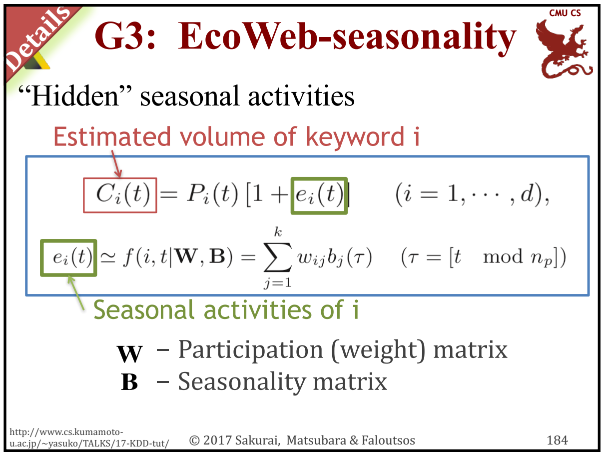


G3: EcoWeb-seasonality

"Hidden" seasonal activities

Users change their behavior according to **seasonal events!**

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 183



G3: EcoWeb-seasonality

Details

"Hidden" seasonal activities

Estimated volume of keyword i

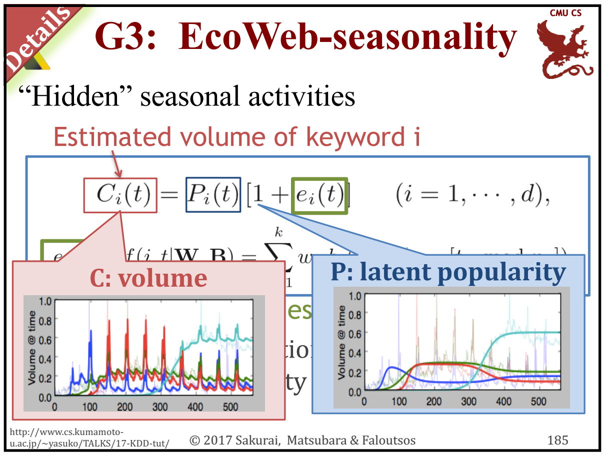
$$\boxed{C_i(t)} = P_i(t)\,[1 + \boxed{e_i(t)}] \qquad (i = 1, \cdots, d),$$

$$\boxed{e_i(t)} \simeq f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{j=1}^{k} w_{ij} b_j(\tau) \qquad (\tau = [t \mod n_p])$$

Seasonal activities of i

W – Participation (weight) matrix
B – Seasonality matrix

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 184



G3: EcoWeb-seasonality

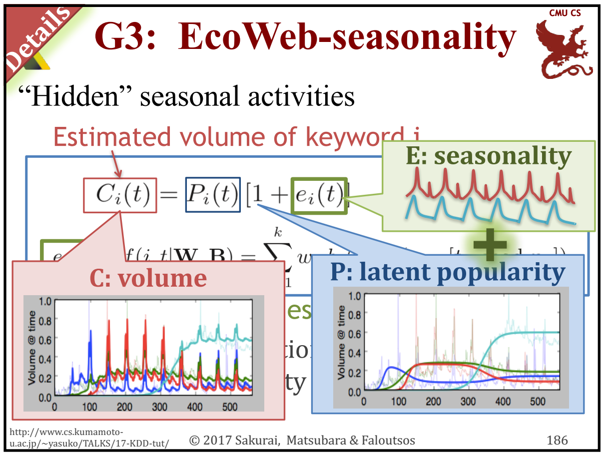
Details

"Hidden" seasonal activities

Estimated volume of keyword i

$$\boxed{C_i(t)} = \boxed{P_i(t)}\,[1 + \boxed{e_i(t)}] \qquad (i = 1, \cdots, d),$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 185



G3: EcoWeb-seasonality

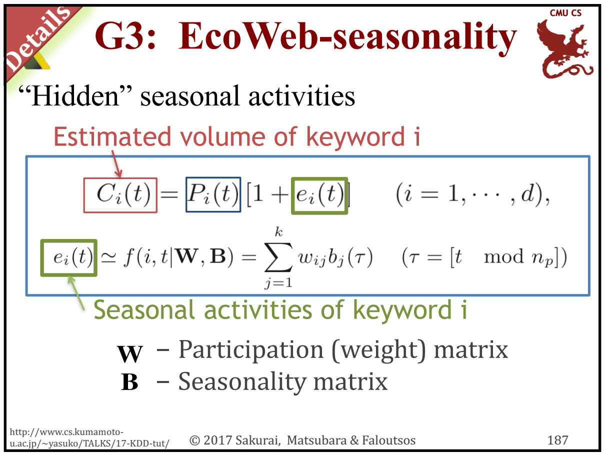
Details

"Hidden" seasonal activities

Estimated volume of keyword i

$$\boxed{C_i(t)} = \boxed{P_i(t)}\,[1 + \boxed{e_i(t)}]$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 186



G3: EcoWeb-seasonality

"Hidden" seasonal activities

Estimated volume of keyword i

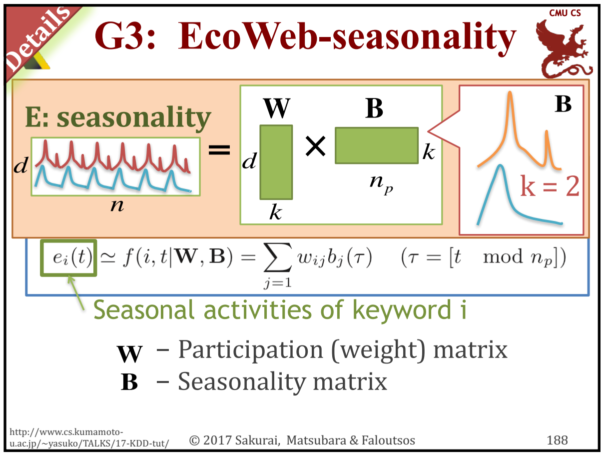
$$\boxed{C_i(t)} = \boxed{P_i(t)}[1 + \boxed{e_i(t)}] \quad (i = 1, \cdots, d),$$

$$\boxed{e_i(t)} \simeq f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{j=1}^{k} w_{ij} b_j(\tau) \quad (\tau = [t \mod n_p])$$

Seasonal activities of keyword i

W – Participation (weight) matrix
B – Seasonality matrix

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 187



G3: EcoWeb-seasonality

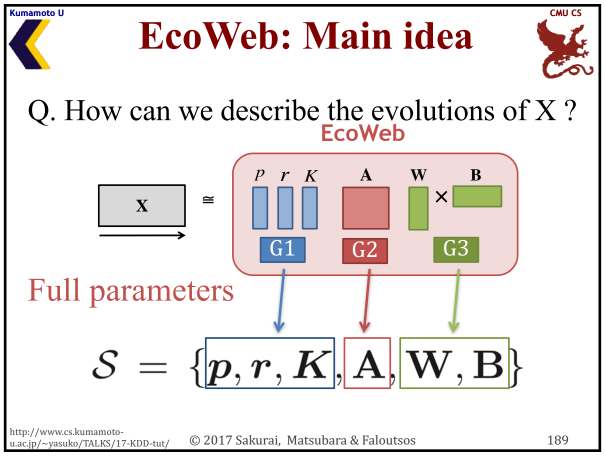
E: seasonality =

$$\boxed{e_i(t)} \simeq f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{j=1}^{k} w_{ij} b_j(\tau) \quad (\tau = [t \mod n_p])$$

Seasonal activities of keyword i

W – Participation (weight) matrix
B – Seasonality matrix

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 188



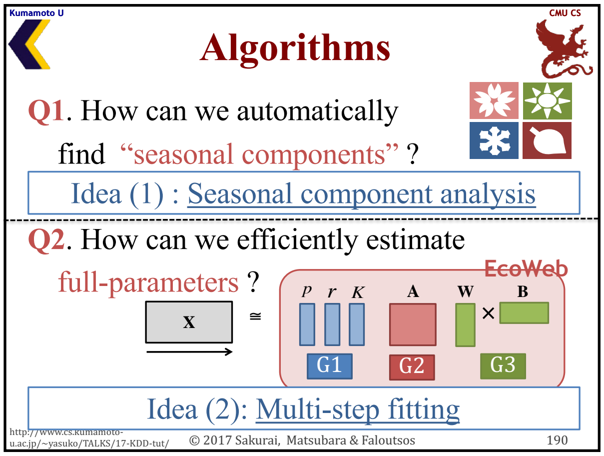
EcoWeb: Main idea

Q. How can we describe the evolutions of X ?

Full parameters

$$\mathcal{S} = \{\boxed{\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{K}}, \boxed{\mathbf{A}}, \boxed{\mathbf{W}, \mathbf{B}}\}$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 189



Algorithms

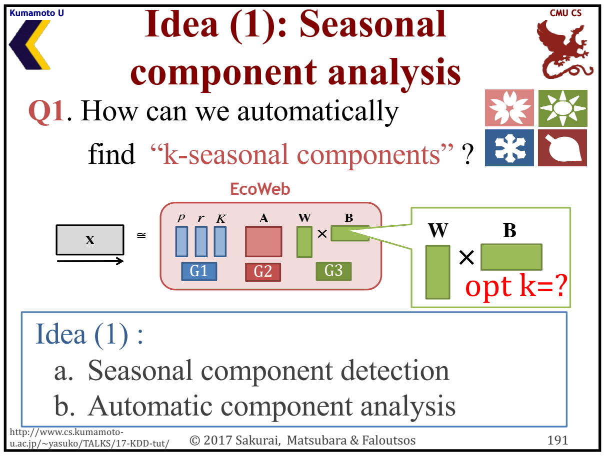
Q1. How can we automatically find "seasonal components" ?

Idea (1) : Seasonal component analysis

Q2. How can we efficiently estimate full-parameters ?

Idea (2): Multi-step fitting

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 190

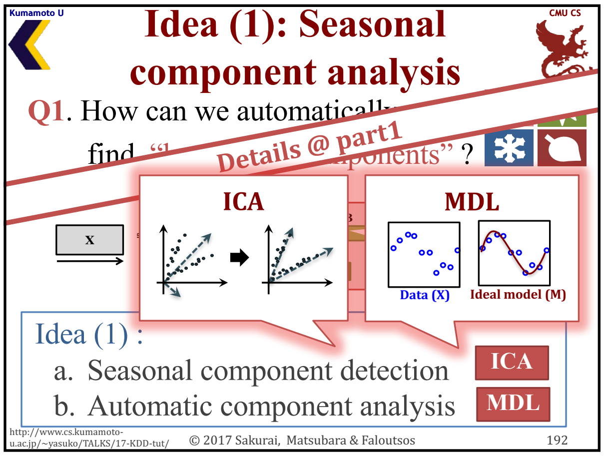


Idea (1): Seasonal component analysis

Q1. How can we automatically find "k-seasonal components" ?

Idea (1) :
a. Seasonal component detection
b. Automatic component analysis

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 191

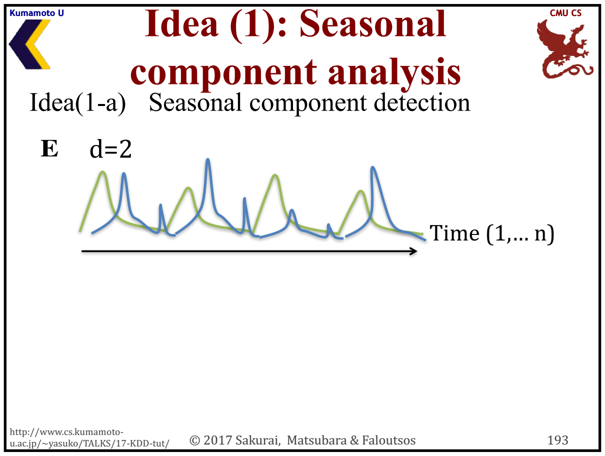


Idea (1): Seasonal component analysis

Q1. How can we automatically find "k-seasonal components" ?

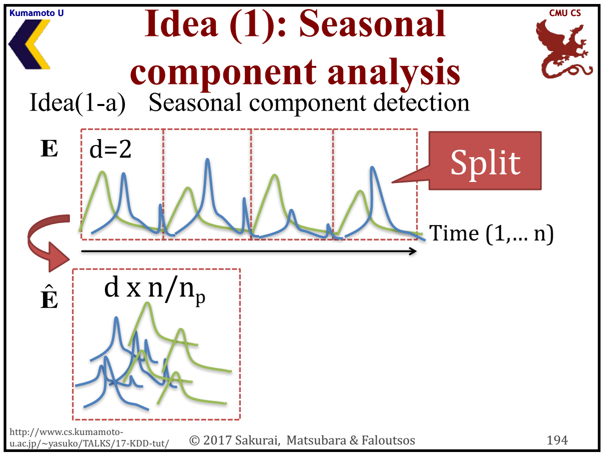
Idea (1) :
a. Seasonal component detection
b. Automatic component analysis

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 192



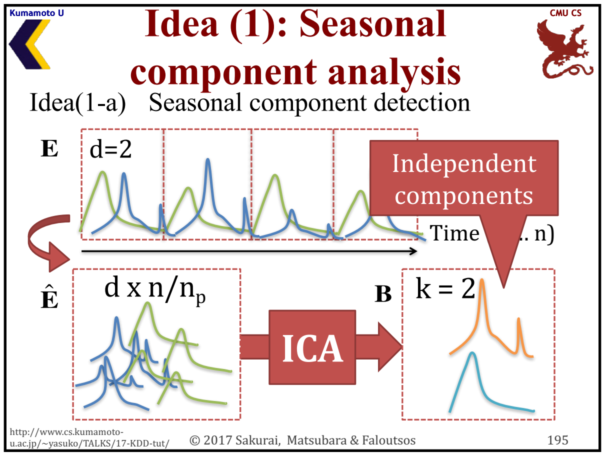
Idea (1): Seasonal component analysis

Idea(1-a) Seasonal component detection



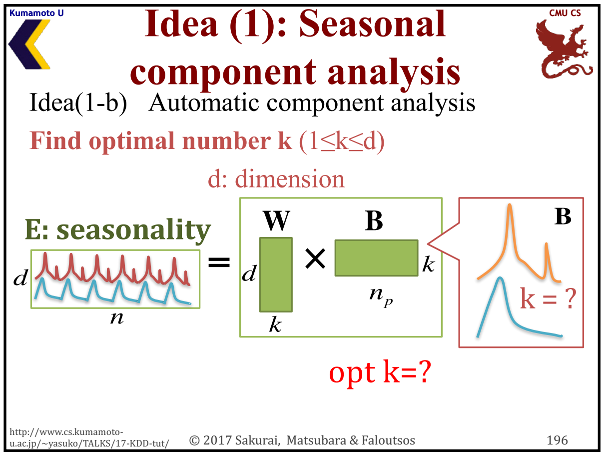
Idea (1): Seasonal component analysis

Idea(1-a) Seasonal component detection



Idea (1): Seasonal component analysis

Idea(1-a) Seasonal component detection

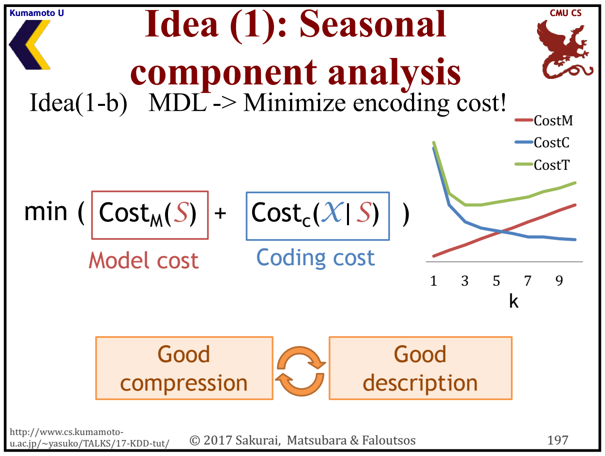


Idea (1): Seasonal component analysis

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

d: dimension



Idea (1): Seasonal component analysis

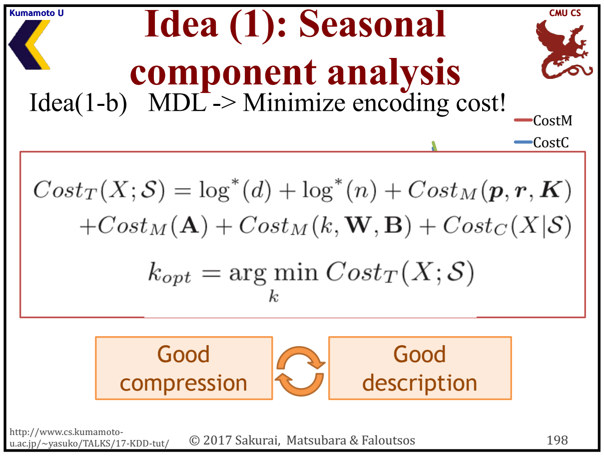
Idea(1-b) MDL -> Minimize encoding cost!

min ($\boxed{\mathrm{Cost}_M(\mathcal{S})}$ + $\boxed{\mathrm{Cost}_c(\mathcal{X} | \mathcal{S})}$)

Model cost

Coding cost

Good compression ⟳ Good description



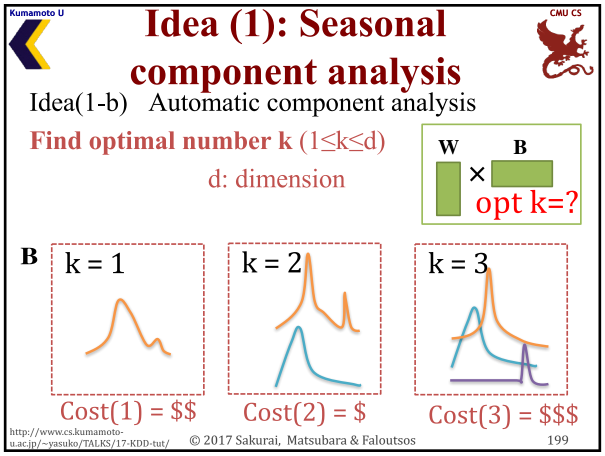
Idea (1): Seasonal component analysis

Idea(1-b) MDL -> Minimize encoding cost!

$$Cost_T(X;\mathcal{S}) = \log^*(d) + \log^*(n) + Cost_M(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{K})$$
$$+Cost_M(\mathbf{A}) + Cost_M(k, \mathbf{W}, \mathbf{B}) + Cost_C(X|\mathcal{S})$$

$$k_{opt} = \arg\min_k Cost_T(X;\mathcal{S})$$

Good compression ⟳ Good description



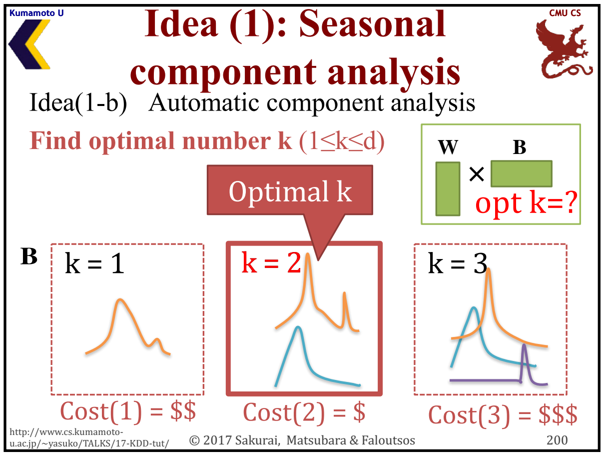
Idea (1): Seasonal component analysis

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

d: dimension

Cost(1) = $$ Cost(2) = $ Cost(3) = $$$

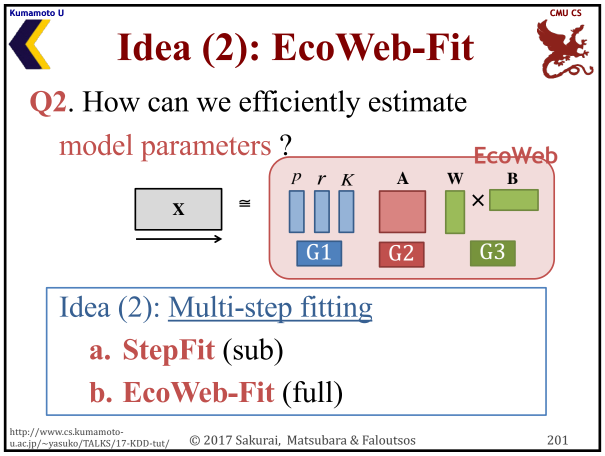


Idea (1): Seasonal component analysis

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

Cost(1) = $$ Cost(2) = $ Cost(3) = $$$

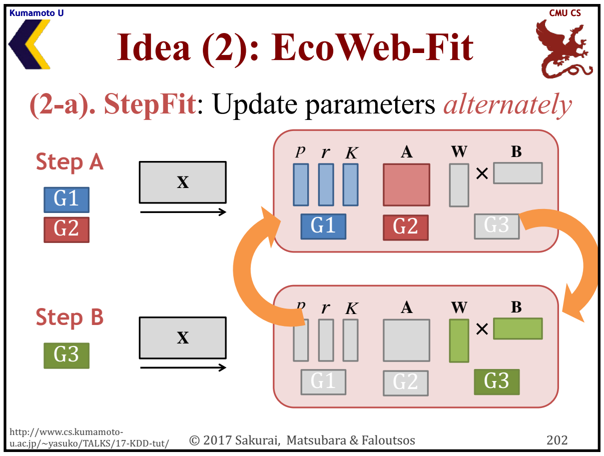


Idea (2): EcoWeb-Fit

Q2. How can we efficiently estimate model parameters ?

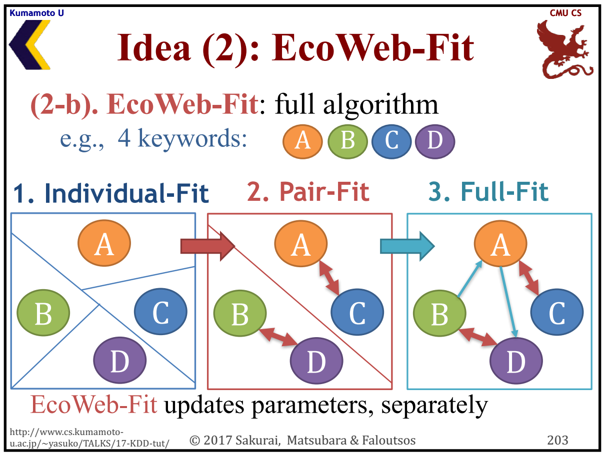
Idea (2): Multi-step fitting

a. **StepFit** (sub)
b. **EcoWeb-Fit** (full)



Idea (2): EcoWeb-Fit

(2-a). StepFit: Update parameters *alternately*



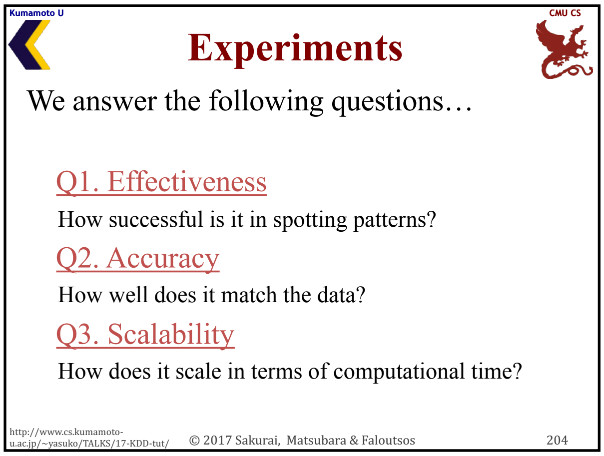
Idea (2): EcoWeb-Fit

(2-b). EcoWeb-Fit: full algorithm

e.g., 4 keywords: A B C D

1. Individual-Fit **2. Pair-Fit** **3. Full-Fit**

EcoWeb-Fit updates parameters, separately



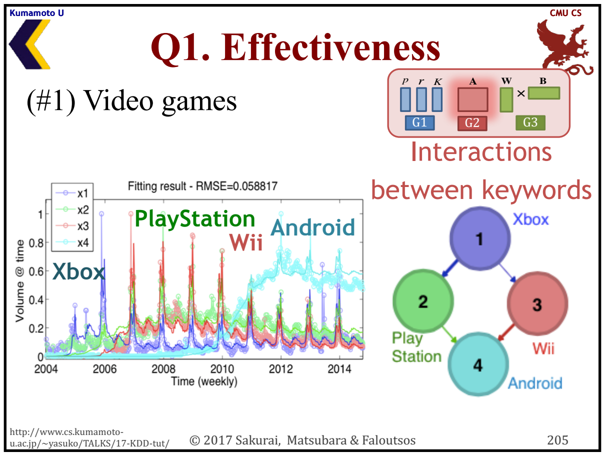
Experiments

We answer the following questions…

Q1. Effectiveness
How successful is it in spotting patterns?

Q2. Accuracy
How well does it match the data?

Q3. Scalability
How does it scale in terms of computational time?



Kumamoto U

CMU CS

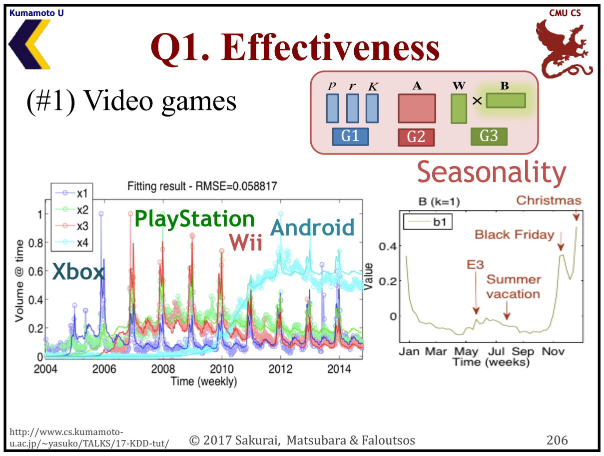
Q1. Effectiveness

(#1) Video games

Interactions between keywords

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/

205



Kumamoto U

CMU CS

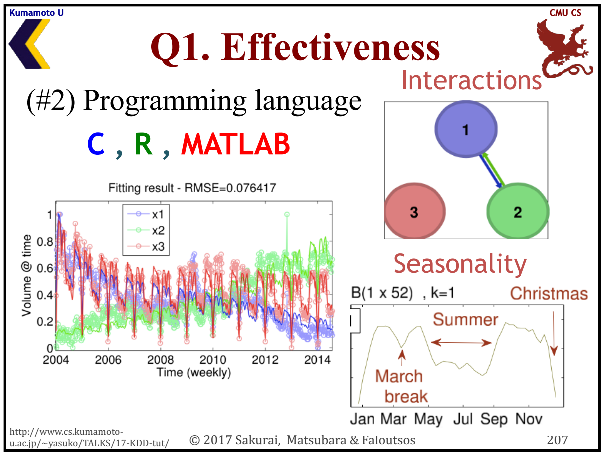
Q1. Effectiveness

(#1) Video games

Seasonality

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/

206



Kumamoto U

CMU CS

Q1. Effectiveness

(#2) Programming language

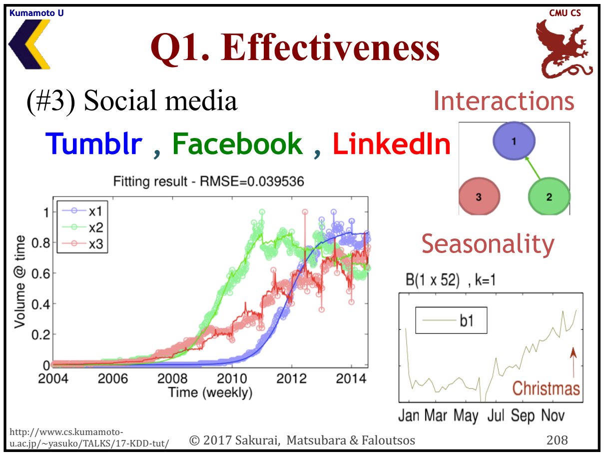
C , R , MATLAB

Interactions

Seasonality

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/

207



Kumamoto U

CMU CS

Q1. Effectiveness

(#3) Social media

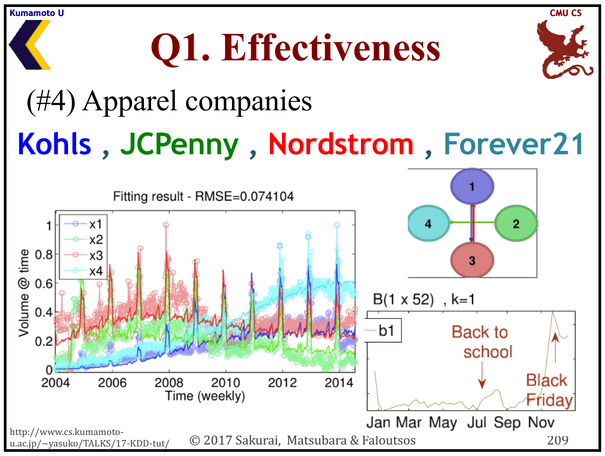
Tumblr , Facebook , LinkedIn

Interactions

Seasonality

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/

208



Kumamoto U

CMU CS

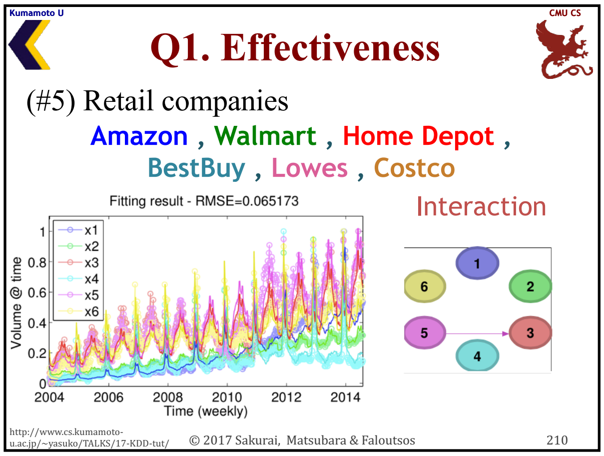
Q1. Effectiveness

(#4) Apparel companies

Kohls , JCPenny , Nordstrom , Forever21

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/

209



Kumamoto U

CMU CS

Q1. Effectiveness

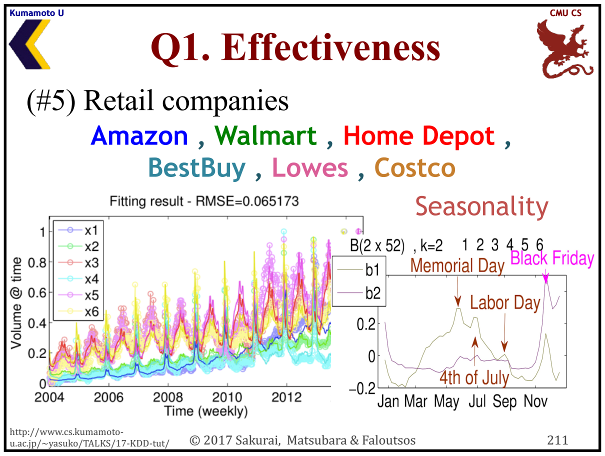
(#5) Retail companies

Amazon , Walmart , Home Depot , BestBuy , Lowes , Costco

Interaction

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/

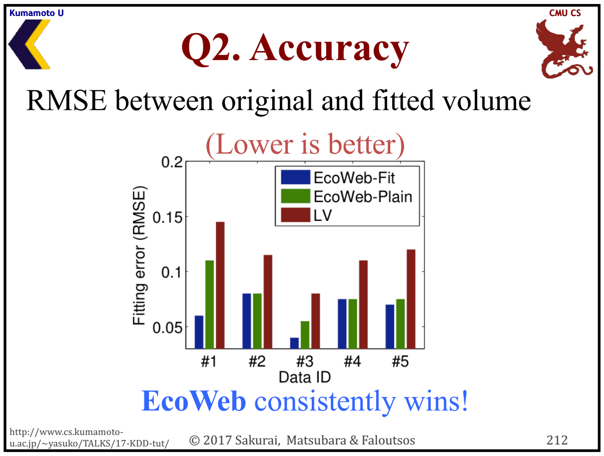
210



Q1. Effectiveness

(#5) Retail companies

Amazon , Walmart , Home Depot , BestBuy , Lowes , Costco

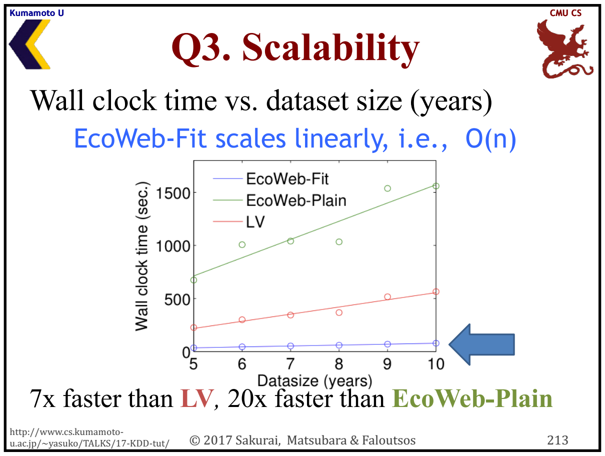


Q2. Accuracy

RMSE between original and fitted volume

(Lower is better)

EcoWeb consistently wins!

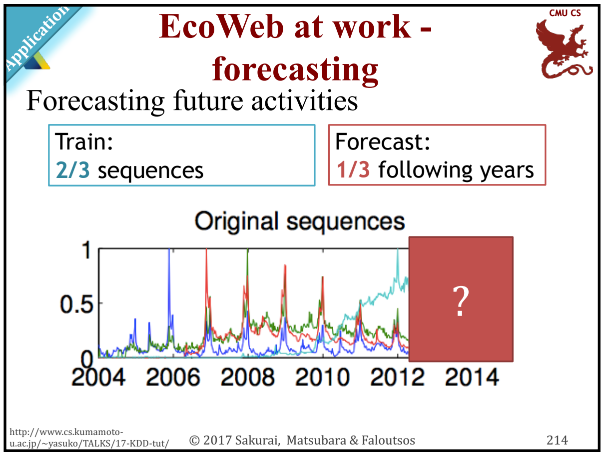


Q3. Scalability

Wall clock time vs. dataset size (years)

EcoWeb-Fit scales linearly, i.e., O(n)

7x faster than **LV**, 20x faster than **EcoWeb-Plain**

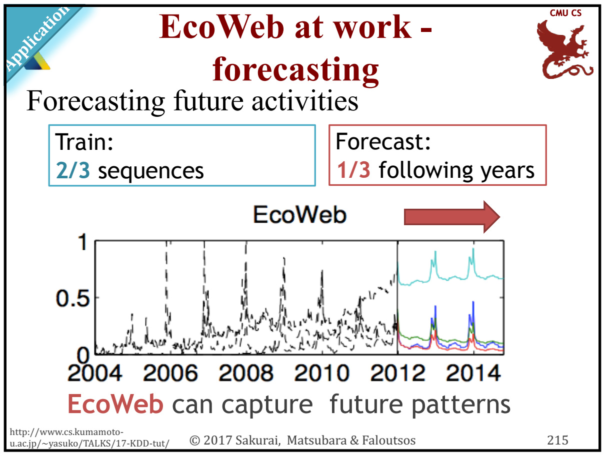


EcoWeb at work - forecasting

Forecasting future activities

Train:
2/3 sequences

Forecast:
1/3 following years



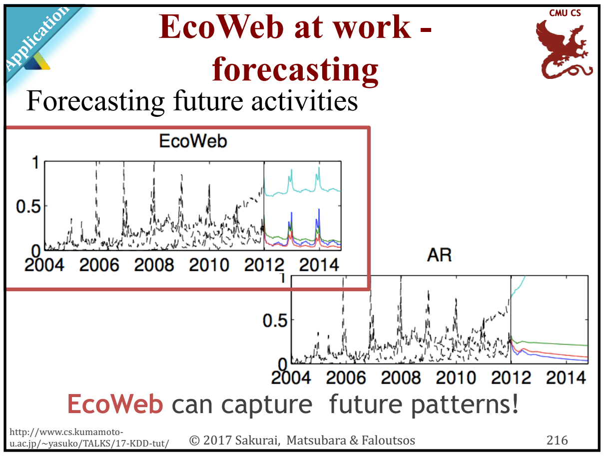
EcoWeb at work - forecasting

Forecasting future activities

Train:
2/3 sequences

Forecast:
1/3 following years

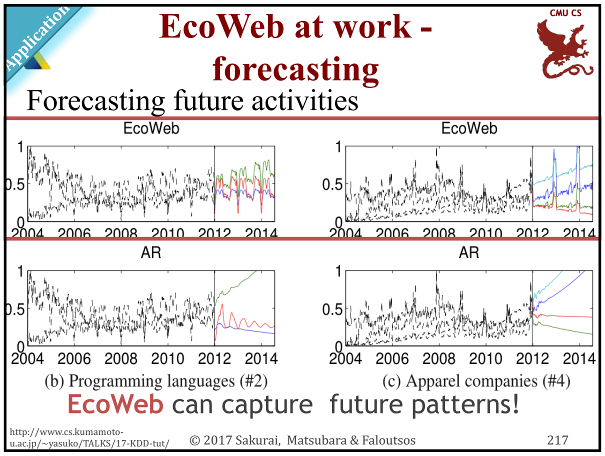
EcoWeb can capture future patterns



EcoWeb at work - forecasting

Forecasting future activities

EcoWeb can capture future patterns!



EcoWeb at work - forecasting

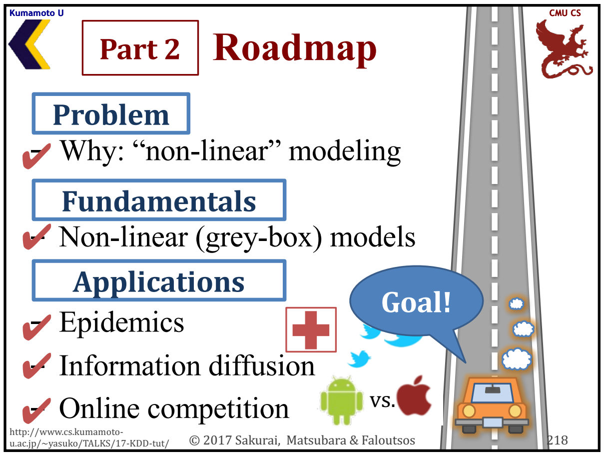
Forecasting future activities

(b) Programming languages (#2)

(c) Apparel companies (#4)

EcoWeb can capture future patterns!

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 217



Part 2 Roadmap

Problem

- ✓ Why: "non-linear" modeling

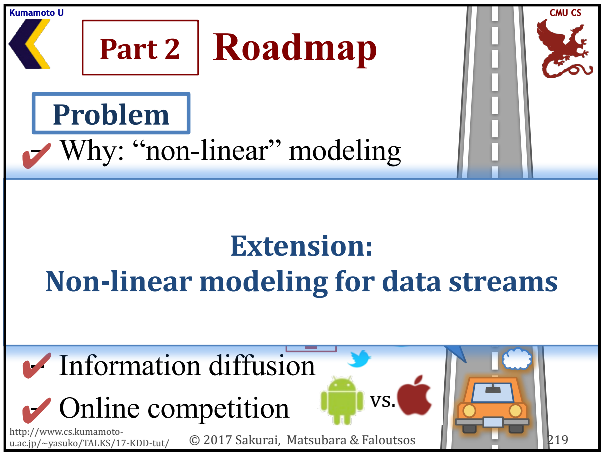
Fundamentals

- ✓ Non-linear (grey-box) models

Applications

- ✓ Epidemics
- ✓ Information diffusion
- ✓ Online competition

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 218



Part 2 Roadmap

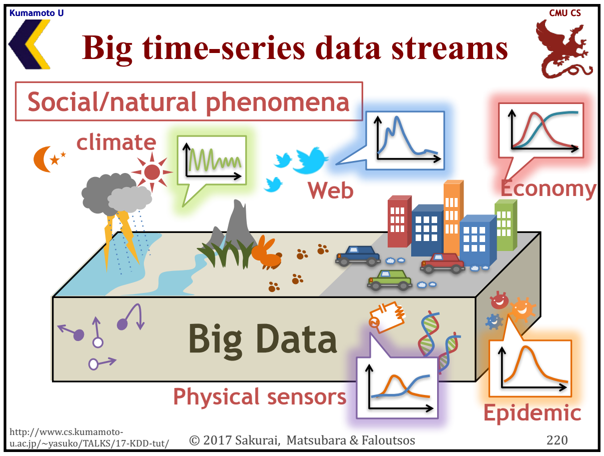
Problem

- ✓ Why: "non-linear" modeling

Extension:
Non-linear modeling for data streams

- ✓ Information diffusion
- ✓ Online competition

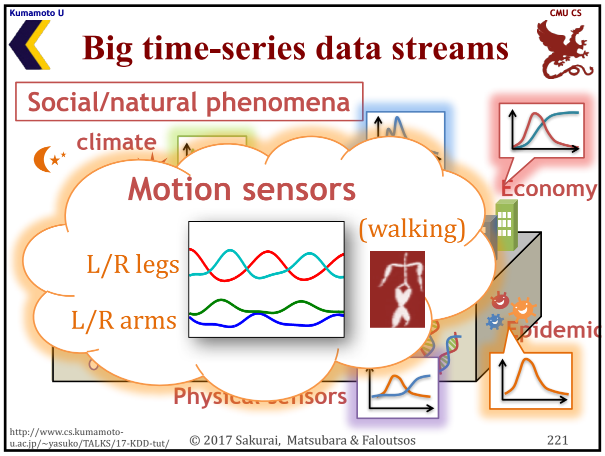
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 219



Big time-series data streams

Social/natural phenomena

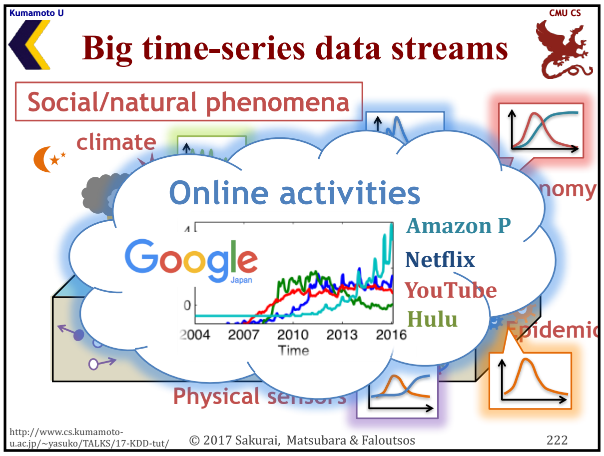
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 220



Big time-series data streams

Social/natural phenomena

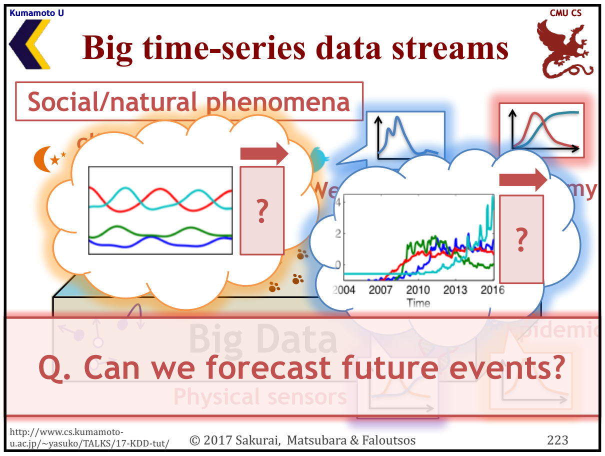
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 221



Big time-series data streams

Social/natural phenomena

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 222



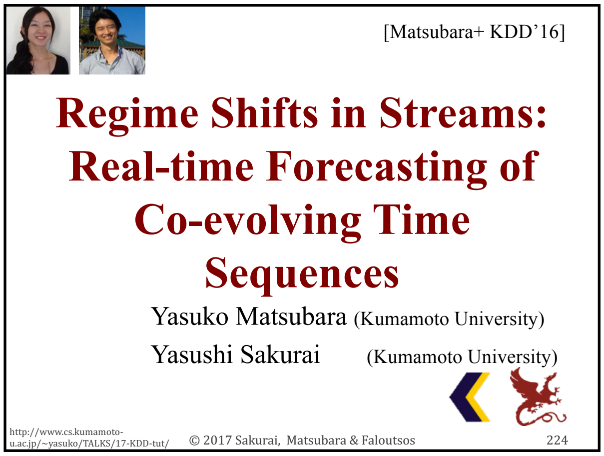
Big time-series data streams

Social/natural phenomena

Big Data

Physical sensors

Q. Can we forecast future events?

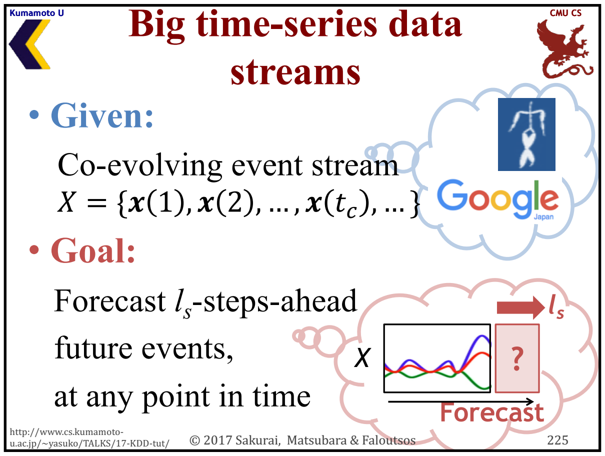


[Matsubara+ KDD'16]

Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences

Yasuko Matsubara (Kumamoto University)

Yasushi Sakurai (Kumamoto University)



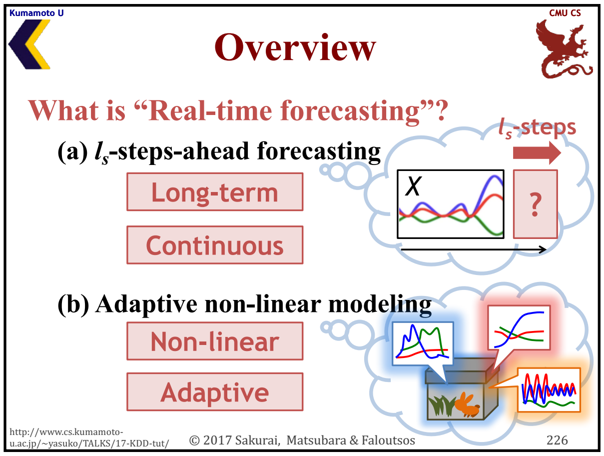
Big time-series data streams

- **Given:**

  Co-evolving event stream
  $X = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \dots, \boldsymbol{x}(t_c), \dots\}$

- **Goal:**

  Forecast l_s-steps-ahead
  future events,
  at any point in time



Overview

What is "Real-time forecasting"?

(a) l_s-steps-ahead forecasting

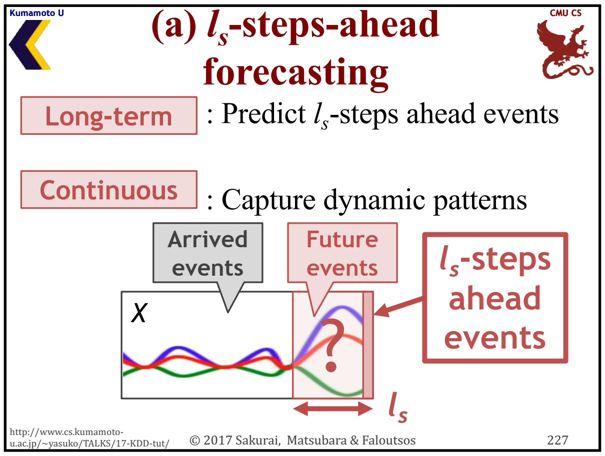
Long-term

Continuous

(b) Adaptive non-linear modeling

Non-linear

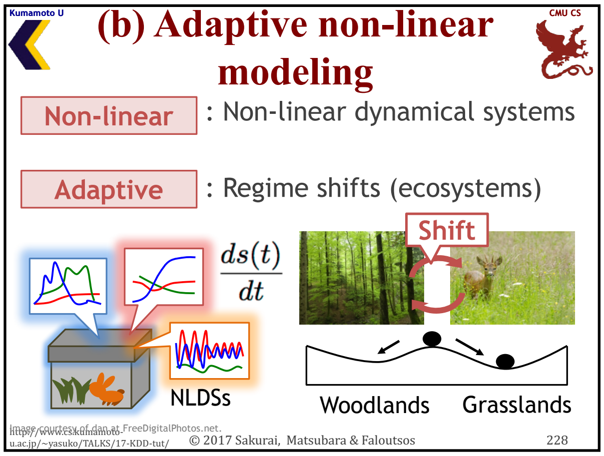
Adaptive



(a) l_s-steps-ahead forecasting

Long-term : Predict l_s-steps ahead events

Continuous : Capture dynamic patterns

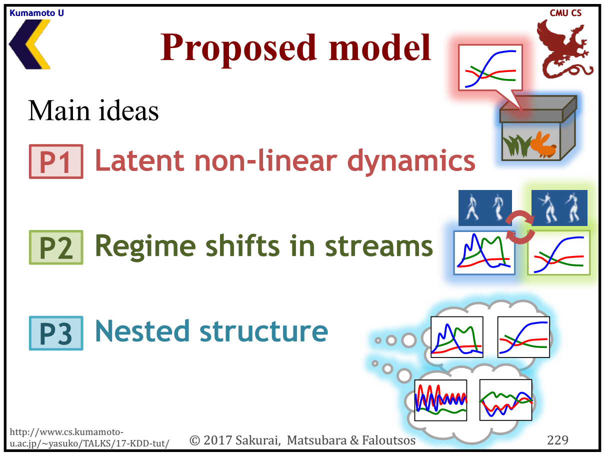


(b) Adaptive non-linear modeling

Non-linear : Non-linear dynamical systems

Adaptive : Regime shifts (ecosystems)

image courtesy of dan at FreeDigitalPhotos.net.



Kumamoto U CMU CS

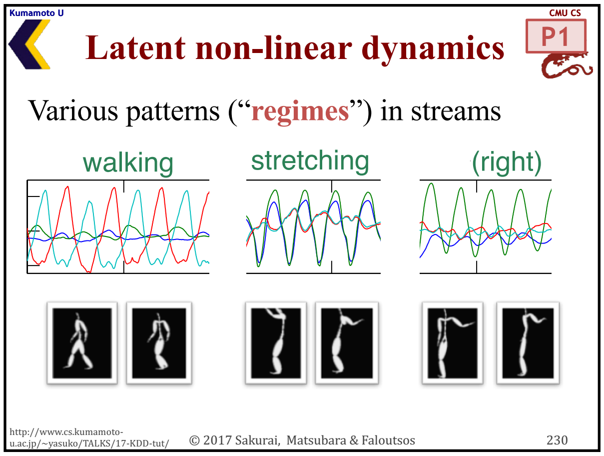
Proposed model

Main ideas

P1 Latent non-linear dynamics

P2 Regime shifts in streams

P3 Nested structure

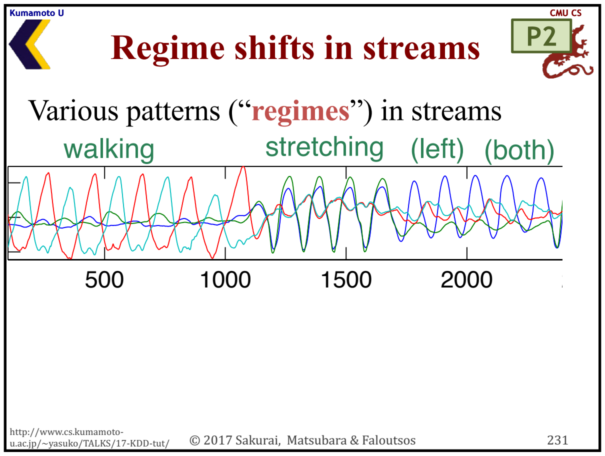


Kumamoto U CMU CS

Latent non-linear dynamics

P1

Various patterns ("**regimes**") in streams

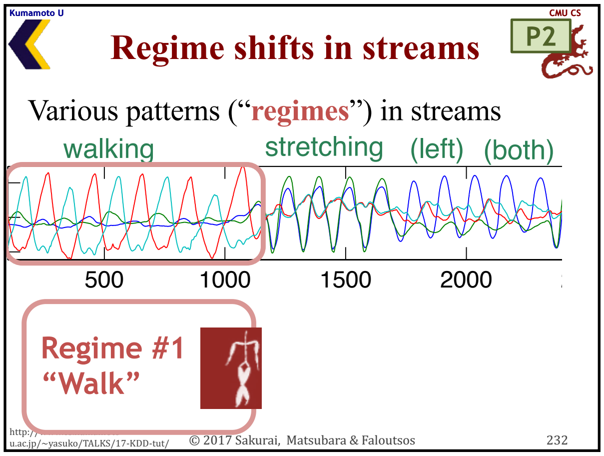


Kumamoto U CMU CS

Regime shifts in streams

P2

Various patterns ("**regimes**") in streams



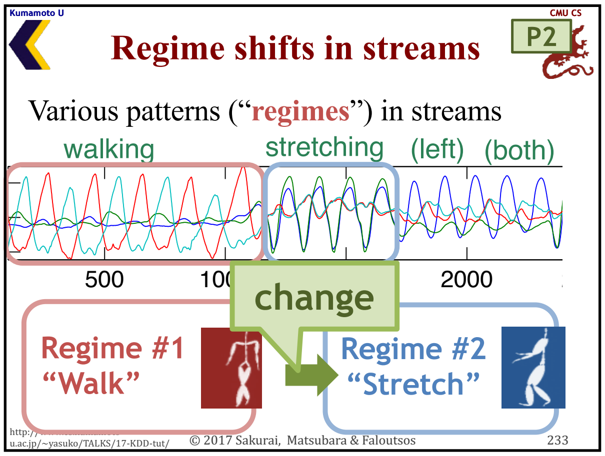
Kumamoto U CMU CS

Regime shifts in streams

P2

Various patterns ("**regimes**") in streams

Regime #1
"Walk"



Kumamoto U CMU CS

Regime shifts in streams

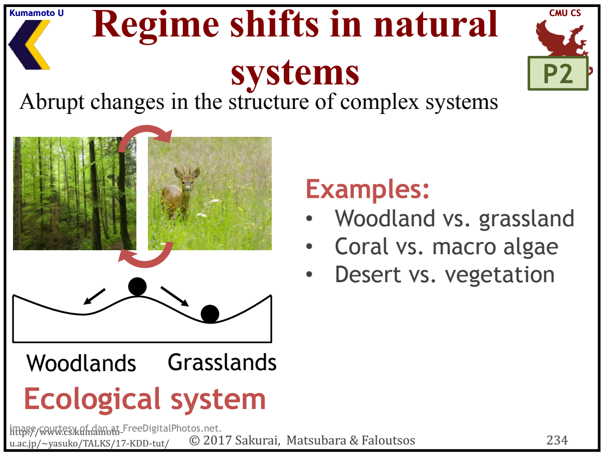
P2

Various patterns ("**regimes**") in streams

change

Regime #1
"Walk"

Regime #2
"Stretch"



Kumamoto U CMU CS

Regime shifts in natural systems

P2

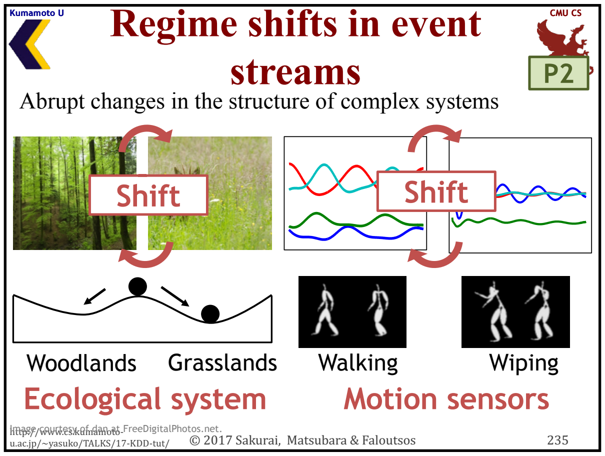
Abrupt changes in the structure of complex systems

Ecological system

Examples:

- Woodland vs. grassland
- Coral vs. macro algae
- Desert vs. vegetation

Image courtesy of dan at FreeDigitalPhotos.net.



Regime shifts in event streams

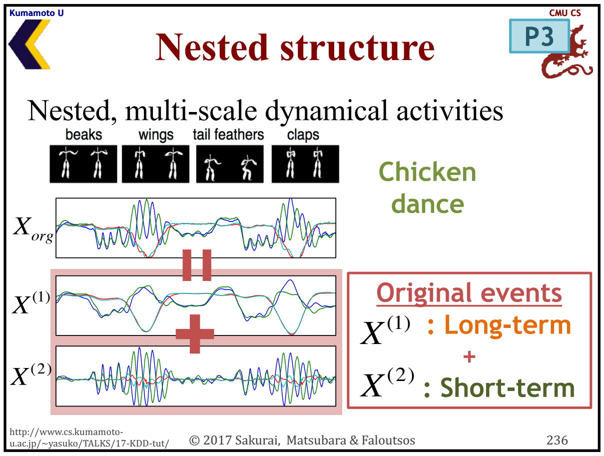
Abrupt changes in the structure of complex systems

Ecological system

Motion sensors

Image courtesy of dan at FreeDigitalPhotos.net.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 235



Nested structure

Nested, multi-scale dynamical activities

Chicken dance

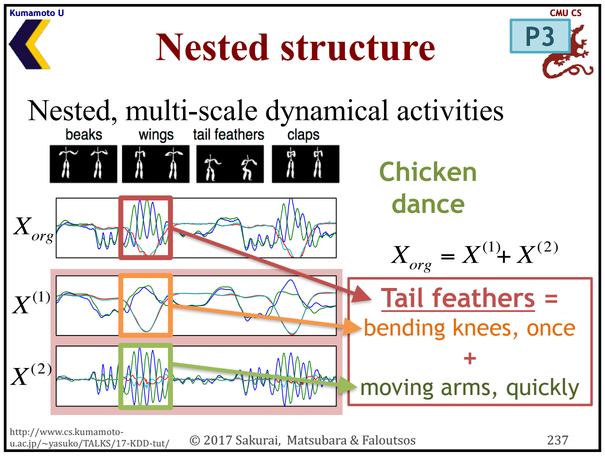
Original events

$X^{(1)}$: Long-term

+

$X^{(2)}$: Short-term

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 236



Nested structure

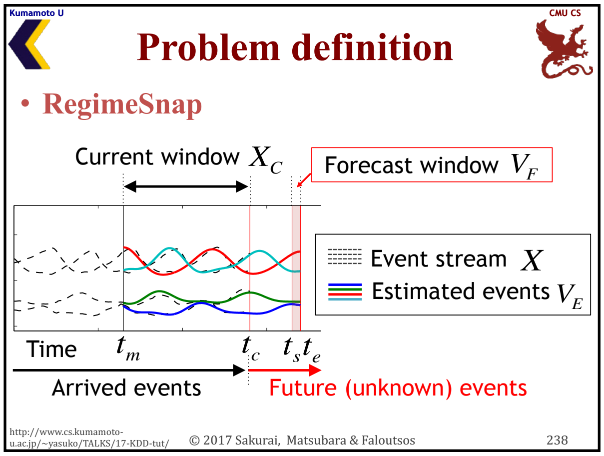
Nested, multi-scale dynamical activities

Chicken dance

$X_{org} = X^{(1)} + X^{(2)}$

Tail feathers = bending knees, once + moving arms, quickly

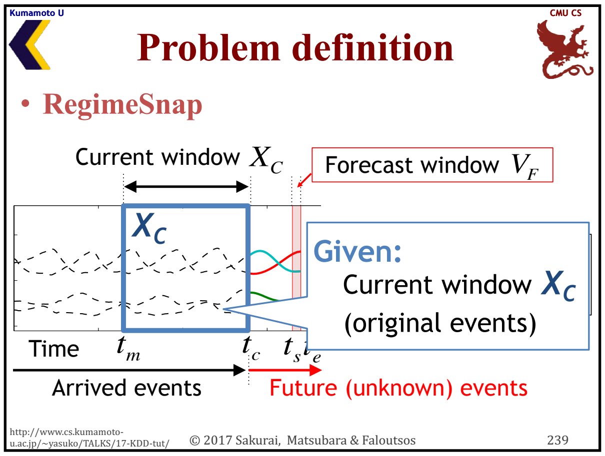
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 237



Problem definition

- RegimeSnap

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 238

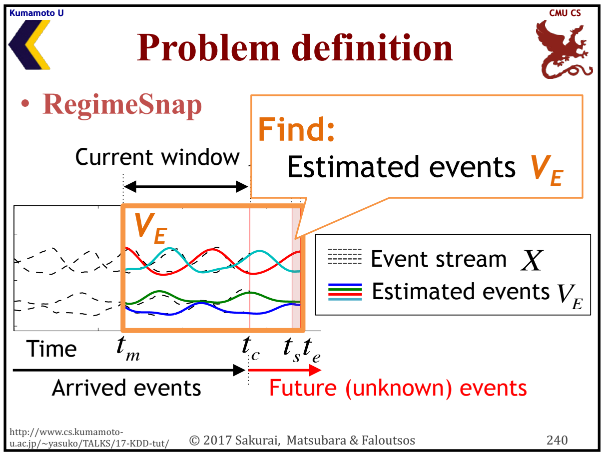


Problem definition

- RegimeSnap

Given: Current window X_C (original events)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 239

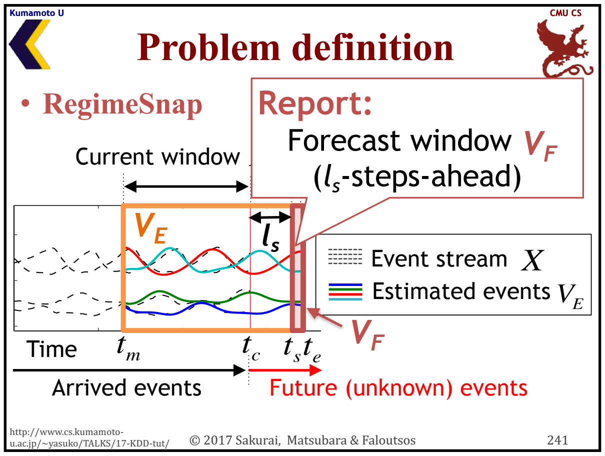


Problem definition

- RegimeSnap

Find: Estimated events V_E

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 240

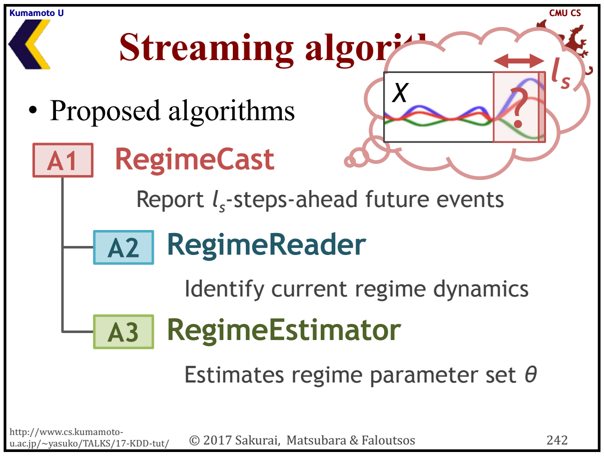


Problem definition

- **RegimeSnap**

Report:
Forecast window V_F
(l_s-steps-ahead)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 241



Streaming algori

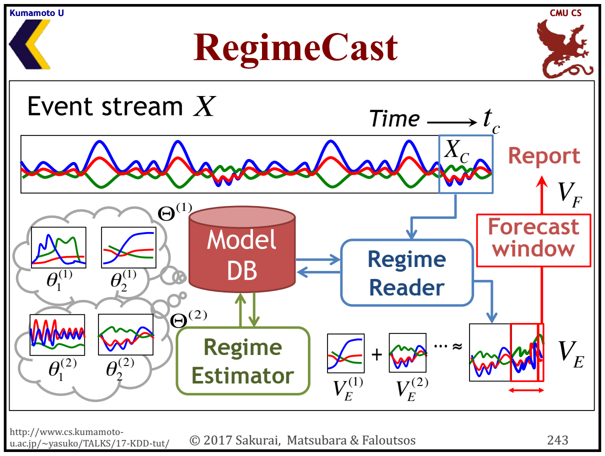
- Proposed algorithms

A1 RegimeCast
Report l_s-steps-ahead future events

A2 RegimeReader
Identify current regime dynamics

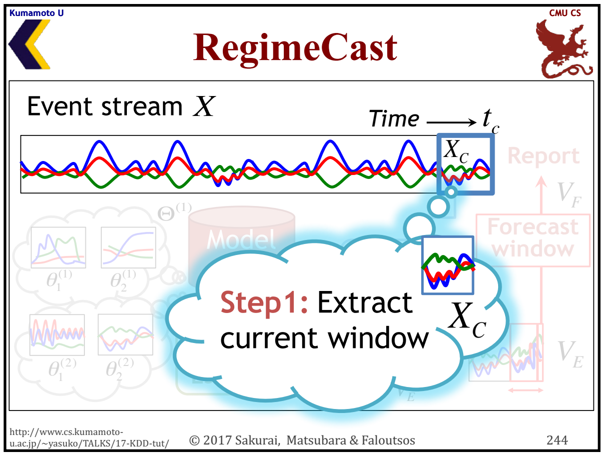
A3 RegimeEstimator
Estimates regime parameter set θ

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 242



RegimeCast

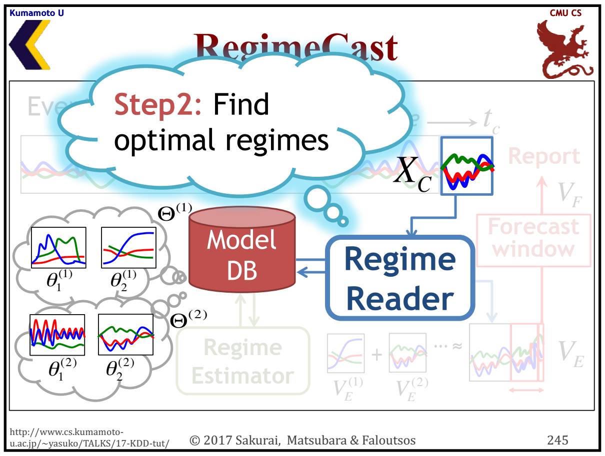
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 243



RegimeCast

Step1: Extract current window X_C

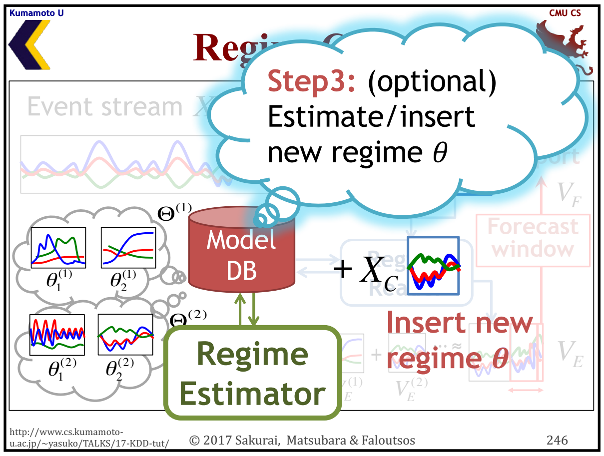
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 244



RegimeCast

Step2: Find optimal regimes

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 245

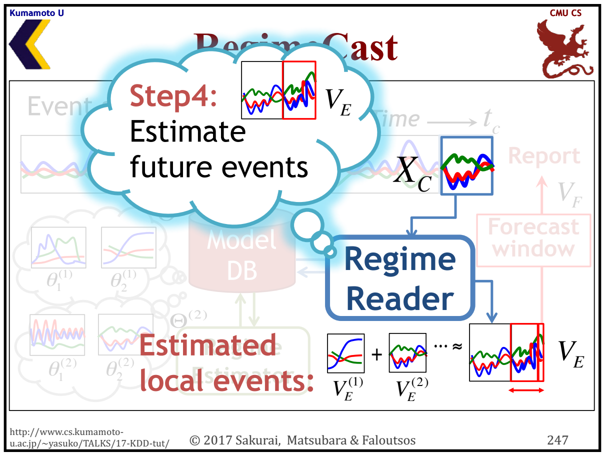


Regi

Step3: (optional) Estimate/insert new regime θ

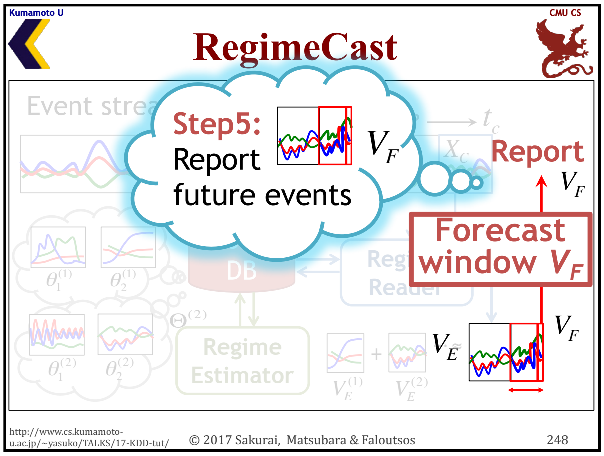
Insert new regime θ

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 246



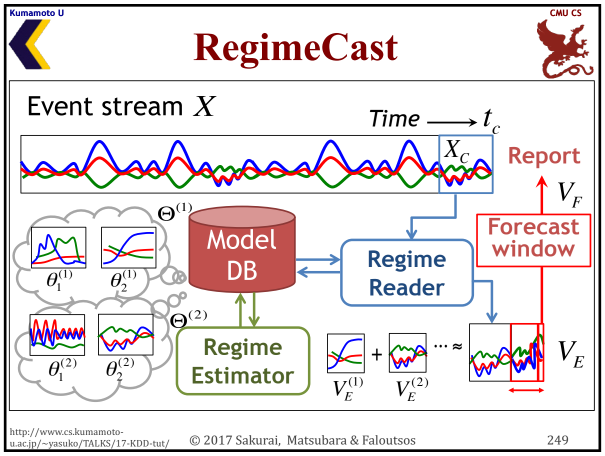
RegimeCast

Step4: Estimate future events V_E

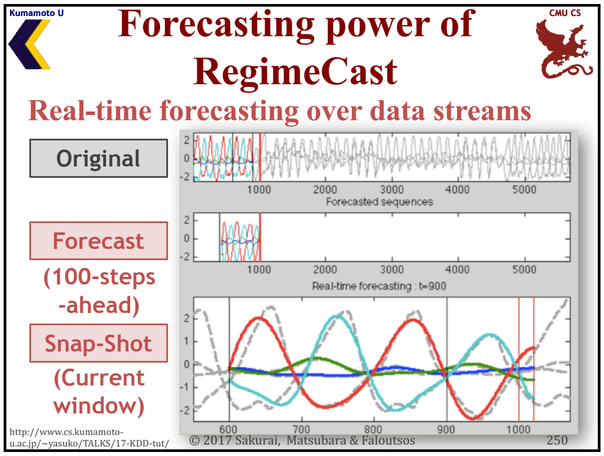


RegimeCast

Step5: Report V_F future events



RegimeCast



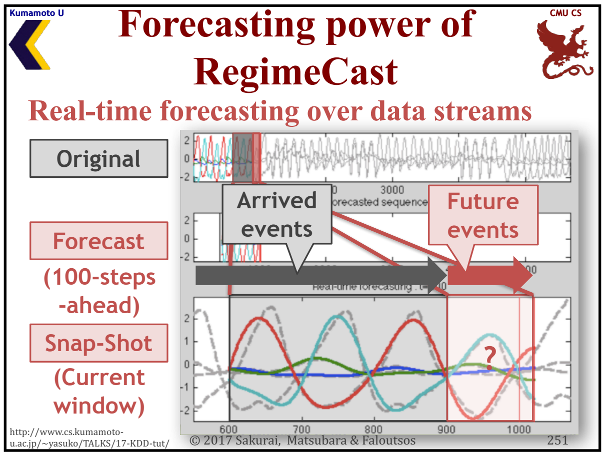
Forecasting power of RegimeCast

Real-time forecasting over data streams

Original

Forecast (100-steps -ahead)

Snap-Shot (Current window)



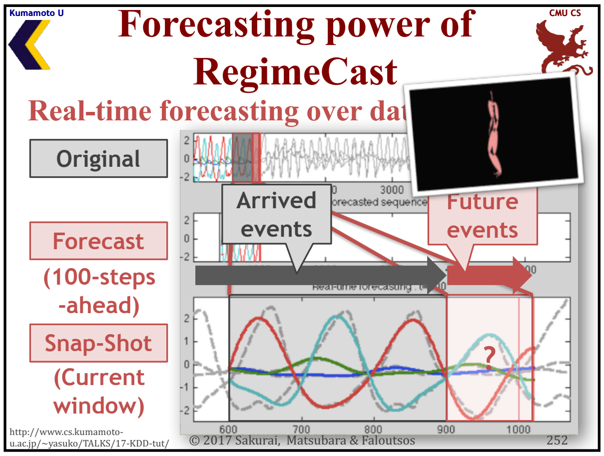
Forecasting power of RegimeCast

Real-time forecasting over data streams

Original

Forecast (100-steps -ahead)

Snap-Shot (Current window)



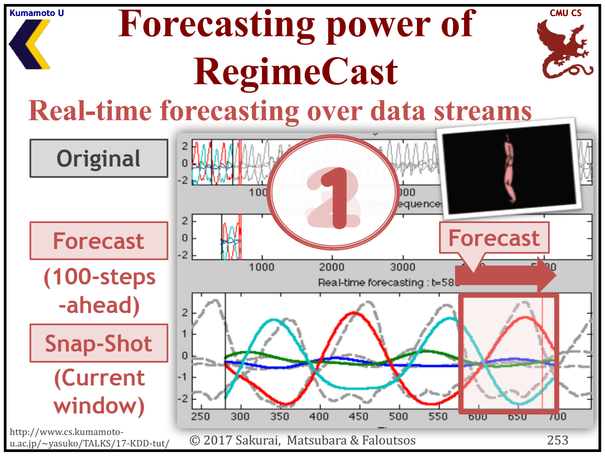
Forecasting power of RegimeCast

Real-time forecasting over dat

Original

Forecast (100-steps -ahead)

Snap-Shot (Current window)



Forecasting power of RegimeCast

Real-time forecasting over data streams

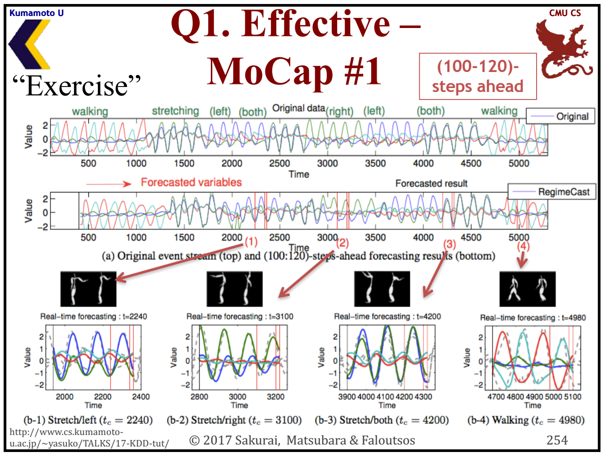
Original

Forecast (100-steps -ahead)

Snap-Shot (Current window)

Forecast

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 253



Q1. Effective – MoCap #1

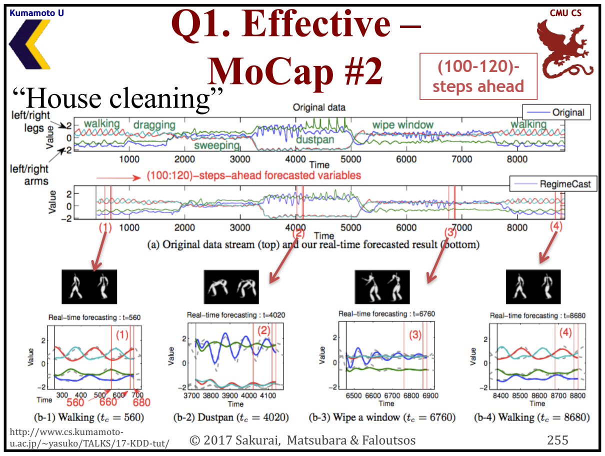
"Exercise"

(100-120)-steps ahead

(a) Original event stream (top) and (100:120)-steps-ahead forecasting results (bottom)

(b-1) Stretch/left ($t_c = 2240$) (b-2) Stretch/right ($t_c = 3100$) (b-3) Stretch/both ($t_c = 4200$) (b-4) Walking ($t_c = 4980$)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 254



Q1. Effective – MoCap #2

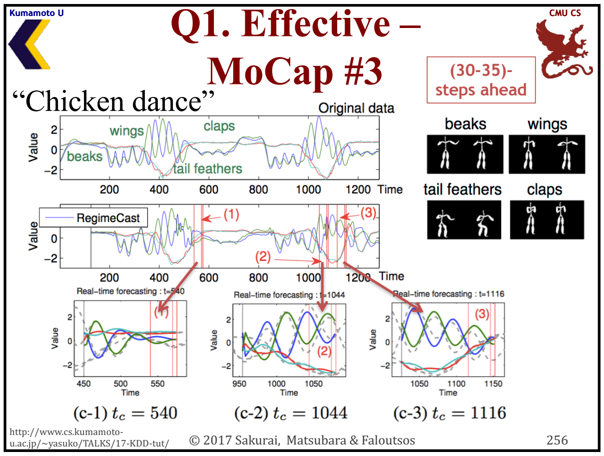
"House cleaning"

(100-120)-steps ahead

(a) Original data stream (top) and our real-time forecasted result (bottom)

(b-1) Walking ($t_c = 560$) (b-2) Dustpan ($t_c = 4020$) (b-3) Wipe a window ($t_c = 6760$) (b-4) Walking ($t_c = 8680$)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 255



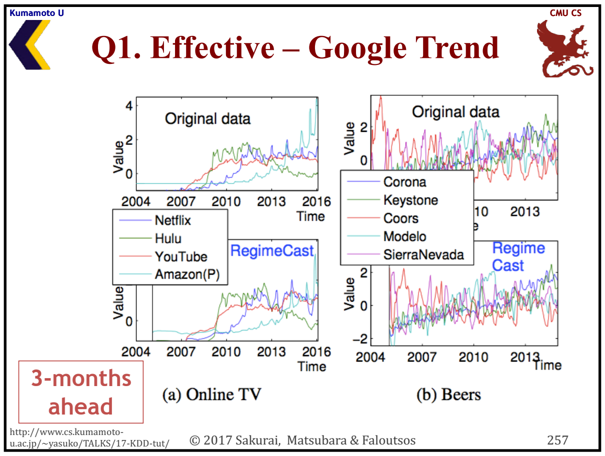
Q1. Effective – MoCap #3

"Chicken dance"

(30-35)-steps ahead

(c-1) $t_c = 540$ (c-2) $t_c = 1044$ (c-3) $t_c = 1116$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 256

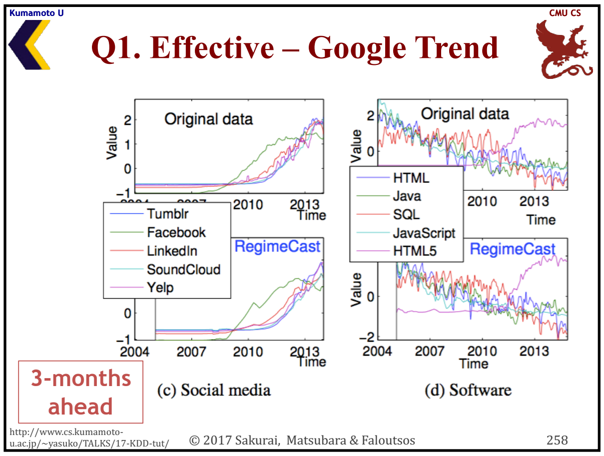


Q1. Effective – Google Trend

3-months ahead

(a) Online TV (b) Beers

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 257

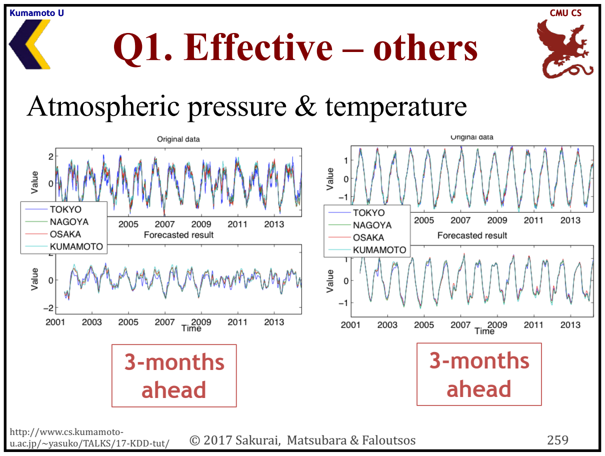


Q1. Effective – Google Trend

3-months ahead

(c) Social media (d) Software

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 258



Kumamoto U CMU CS

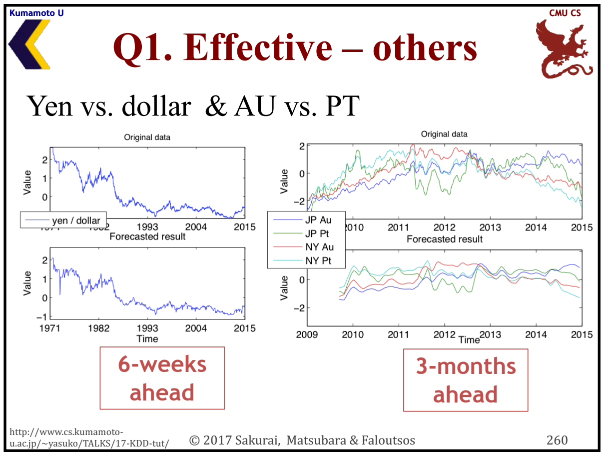
Q1. Effective – others

Atmospheric pressure & temperature

3-months ahead

3-months ahead

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 259



Kumamoto U CMU CS

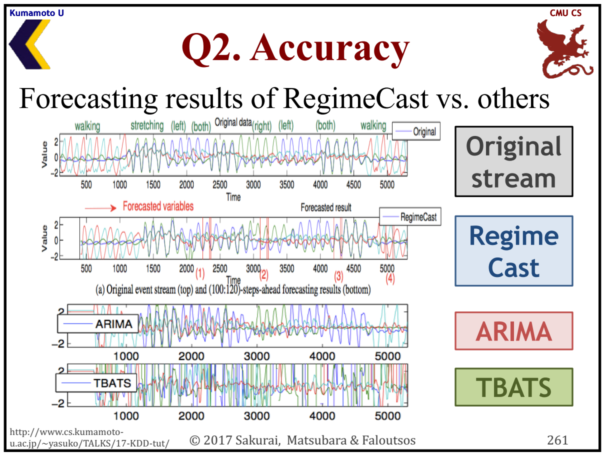
Q1. Effective – others

Yen vs. dollar & AU vs. PT

6-weeks ahead

3-months ahead

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 260



Kumamoto U CMU CS

Q2. Accuracy

Forecasting results of RegimeCast vs. others

(a) Original event stream (top) and (100:120)-steps-ahead forecasting results (bottom)

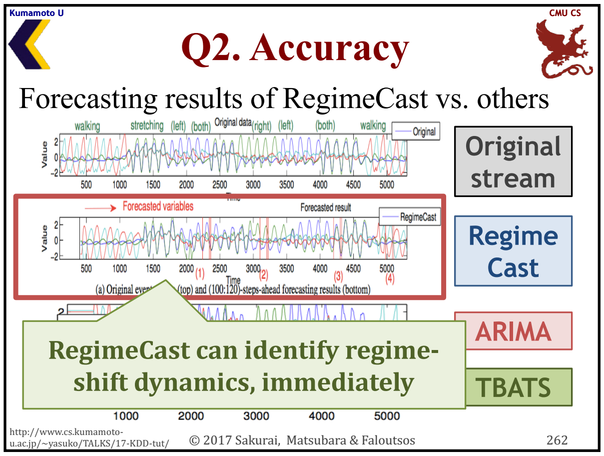
Original stream

Regime Cast

ARIMA

TBATS

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 261



Kumamoto U CMU CS

Q2. Accuracy

Forecasting results of RegimeCast vs. others

(a) Original event stream (top) and (100:120)-steps-ahead forecasting results (bottom)

Original stream

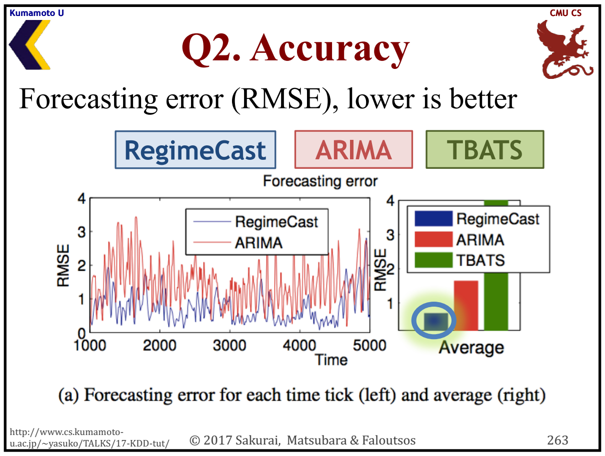
Regime Cast

ARIMA

TBATS

RegimeCast can identify regime-shift dynamics, immediately

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 262



Kumamoto U CMU CS

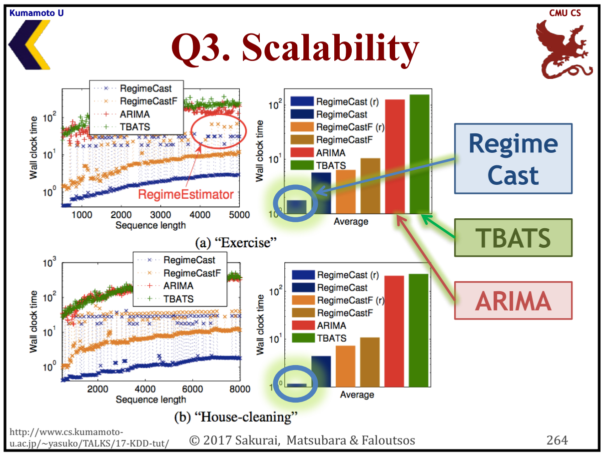
Q2. Accuracy

Forecasting error (RMSE), lower is better

RegimeCast ARIMA TBATS

(a) Forecasting error for each time tick (left) and average (right)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 263



Kumamoto U CMU CS

Q3. Scalability

(a) "Exercise"

(b) "House-cleaning"

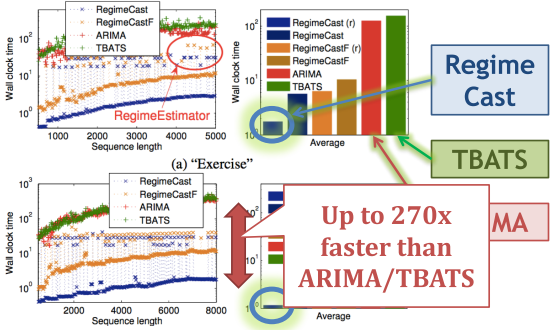
Regime Cast

TBATS

ARIMA

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 264

Q3. Scalability



(a) "Exercise"

(b) "House-cleaning"

Regime Cast

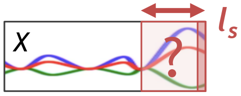
TBATS

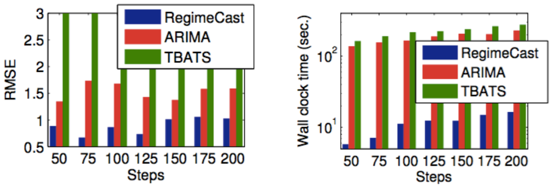
ARIMA

Up to 270x faster than ARIMA/TBATS

Q. Discussion

Q. How long ahead can it forecast?





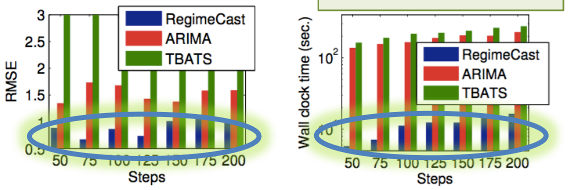
l_s-steps vs. error

l_s-steps vs. speed

Q. Discussion

Q. How long ahead can it forecast?

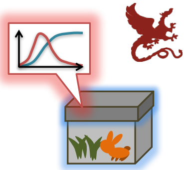
A. It can forecast future events for every step l_s



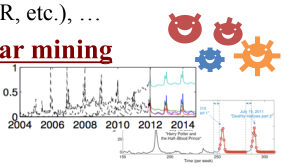
l_s-steps vs. error

l_s-steps vs. speed

Part 2 Conclusions



- ✓ **Why: "non-linear" modeling**
  - Black box: lag plots (k-NN search)
  - Grey-box: given a model
- ✓ **Fundamentals: popular non-linear models**
  - Logistic function, Lotka-Volterra, Competition, ...
  - Epidemics (SI, SIR, SEIR, etc.), …
- ✓ **Applications: non-linear mining**
  - Epidemics
  - Information diffusion
  - Online competition



References (1)

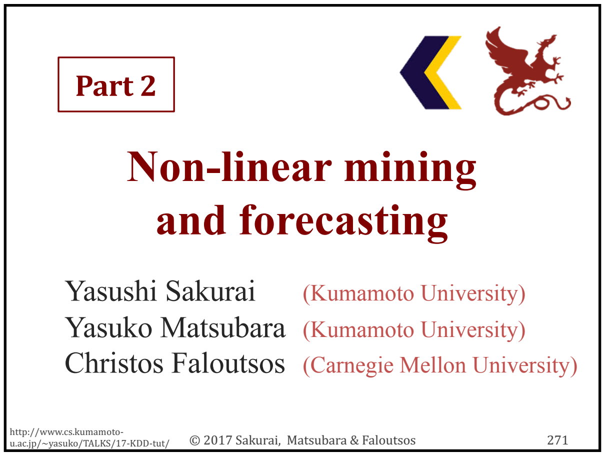
Fundamentals

- Non-linear forecasting
  - D. Chakrabarti and C. Faloutsos *F4: Large-Scale Automated Forecasting using Fractals* CIKM 2002, Washington DC, Nov. 2002.
  - Sauer, T. (1994). *Time series prediction using delay coordinate embedding*. (in book by Weigend and Gershenfeld, below) Addison-Wesley.
  - Takens, F. (1981). *Detecting strange attractors in fluid turbulence*. Dynamical Systems and Turbulence. Berlin: Springer-Verlag.
- Non-linear equations and modeling
  - F. Brauer and C. Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 40. Springer Verlag, New York, 2001.
  - R. M. Anderson and R. M. May. *Infectious Diseases of Humans Dynamics and Control*. Oxford University Press, 1992.
  - F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.
  - D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
  - R. M. Anderson and R. M. May. Infectious Diseases of Humans. Oxford University Press, 1991.
  - R. M. May. Qualitative stability in model ecosystems. *Ecology*, 54(3):638–641, 1973.
  - M. Nowak. *Evolutionary Dynamics*. Harvard University Press, 2006.
  - Schuster, H. G. and Wagner, P. A model for neuronal oscillations. Biol. Cybern., 1990.
- Others
  - A. G. Hawkes and D. Oakes. A cluster representation of a self-exciting process. J. Appl. Prob., 11:493–503, 1974.

References (2)

Applications

- Epidemics
  - Rohani, P., Earn, D. J. D., Finkenstadt, B. F. & Grenfell, B. T. Population dynamic interference among childhood diseases. Proc. R. Soc. Lond. B 265, 2033–2041 (1998).
  - Rohani, P., Green, C.J., Mantilla-Beniers, N.B. & Grenfell, B.T. Ecological Interference Among Fatal Infections. Nature 422: 885-888 (2003).
  - L.Stone,R.Olinky,andA.Huppert.Seasonaldynamicsofrecurrentepidemics. *Nature*, 446:533–536, March 2007.
  - Y. Matsubara, Y. Sakurai, W. G. van Panhuis, and C. Faloutsos. FUNNEL: automatic mining of spatially coevolving epidemics. In *KDD*, pages 105–114, 2014.
- Information diffusion
  - J. Leskovec, L. Backstrom, and J. M. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD*, pages 497–506, 2009.
  - J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *WSDM*, pages 177–186, 2011.
    J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *ICDM*, pages 599–608, 2010.
    R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. In *PNAS*, 2008.
  - F. Figueiredo, J. M. Almeida, Y. Matsubara, B. Ribeiro, and C. Faloutsos. Revisit behavior in social media: The phoenix-r model and discoveries. In *PKDD*, pages 386–401, 2014.
  - Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14, 2012.
- Online activities and competition
  - B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos. Winner takes all: competing viruses or ideas on fair-play networks. In *WWW*, pages 1037–1046, 2012.
  - A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos. Interacting viruses in networks: can both survive? In *KDD*, pages 426–434, 2012.
  - Y. Matsubara, Y. Sakurai, and C. Faloutsos. The web as a jungle: Non-linear dynamical systems for co-evolving online activities. In *WWW*, pages 721-731, 2015.
- Non-linear modeling for data streams
  - Y. Matsubara and Y. Sakurai. Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences. In *KDD*, pages 1045-1054, 2016.



Part 2

Non-linear mining and forecasting

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