Mining and Forecasting of 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 vs.

Kumamoto U CMU CS

Non-linear mining and forecasting

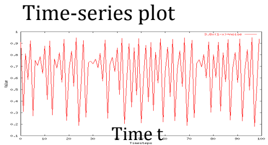
Q. What are "non-linear phenomena"?

Example: logistic parabola

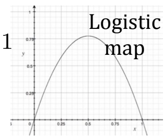
Models population of flies [R. May/1976]

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

Time-series plot



Logistic map



Kumamoto U CMU CS

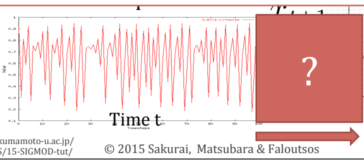
Non-linear mining and forecasting

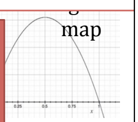
Q. What are "non-linear phenomena"?

Problem:

Given: a time series x_t

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



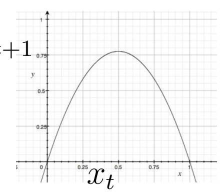


Kumamoto U CMU CS

How to forecast?

Solution 1

Linear equations, e.g., AR, ARIMA, …



Kumamoto U — CMU CS

How to forecast?

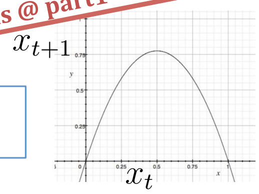
Solution 1

Linear equations, e.g., AR, ARIMA, …

Details @ part1

e.g., AR(1)

$x_{t+1} = a x_t + \epsilon$



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 7

Kumamoto U — CMU CS

How to forecast?

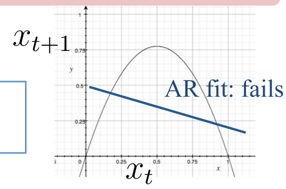
Solution 1

Linear equations, e.g., AR, ARIMA, …

but: linearity assumption

e.g., AR(1)

$x_{t+1} = a x_t + \epsilon$



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 8

Kumamoto U — CMU CS

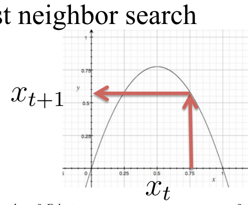
How to forecast?

Solution 2

“Delayed Coordinate Embedding”

= Lag Plots [Sauer92]

- Based on k-nearest neighbor search

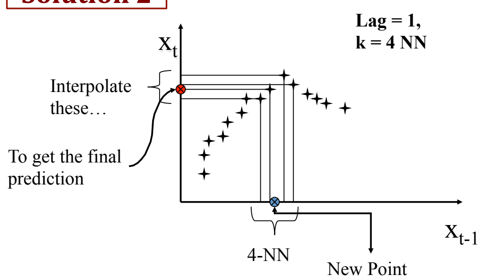


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 9

Kumamoto U — CMU CS

General Intuition (Lag Plot)

Solution 2



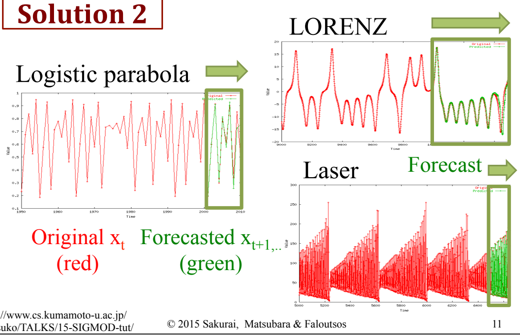
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 10

Kumamoto U — CMU CS

Forecasting results (Lag Plot)

[Chakrabarti+ CIKM'02]

Solution 2



Original x_t (red) Forecasted $x_{t+1,\ldots}$ (green)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 11

Kumamoto U — CMU CS

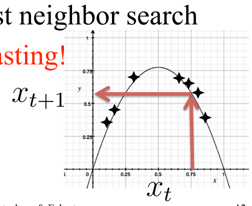
How to forecast?

Solution 2

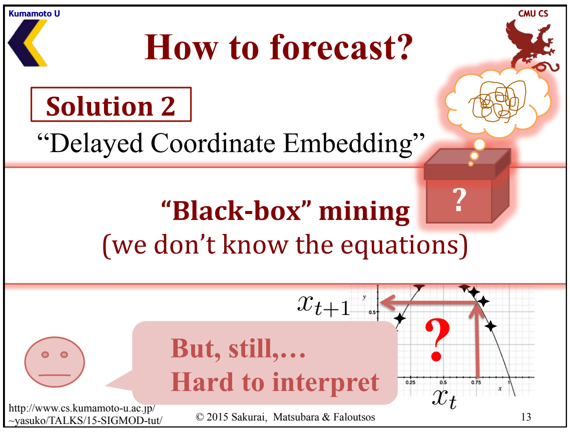
“Delayed Coordinate Embedding”

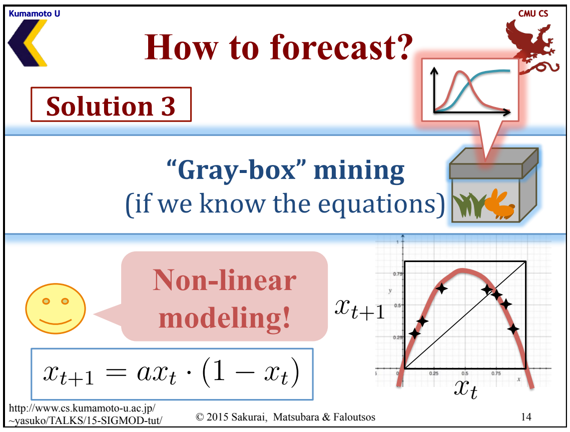
= Lag Plots [Sauer92]

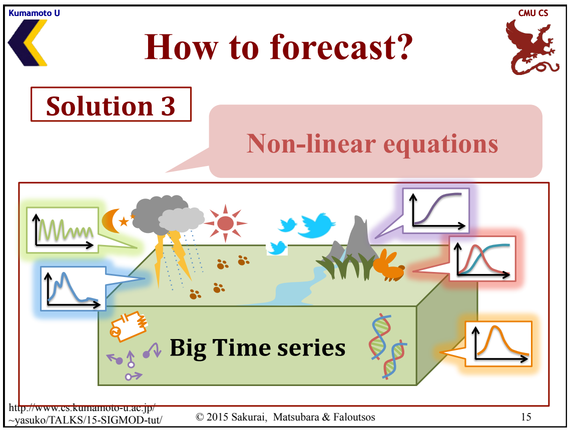
- Based on k-nearest neighbor search
- Non-linear Forecasting!

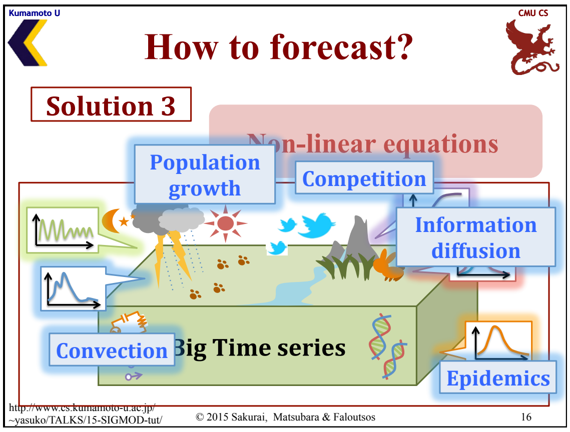


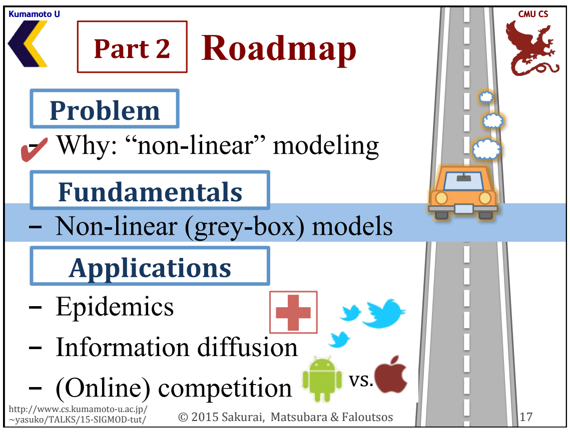
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 12

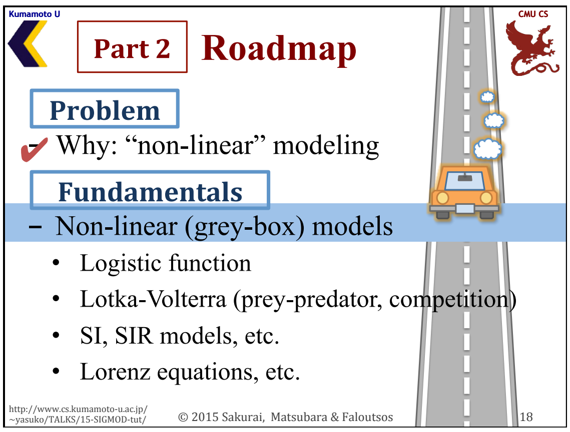

Kumamoto U
CMU CS
How to forecast?
Solution 2
"Delayed Coordinate Embedding"
"Black-box" mining
(we don't know the equations)
?
x_{t+1}
x_t
But, still,…
Hard to interpret
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
13

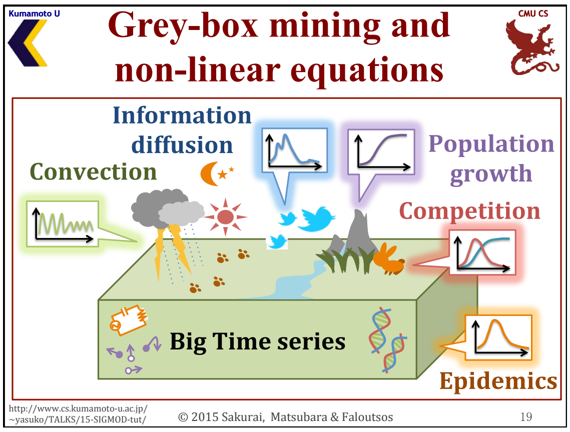

Kumamoto U
CMU CS
How to forecast?
Solution 3
"Gray-box" mining
(if we know the equations)
Non-linear modeling!
$x_{t+1} = a x_t \cdot (1 - x_t)$
x_{t+1}
x_t
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
14

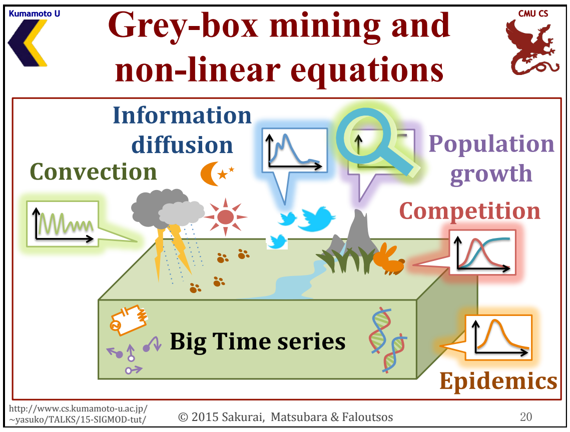

Kumamoto U
CMU CS
How to forecast?
Solution 3
Non-linear equations
Big Time series
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
15

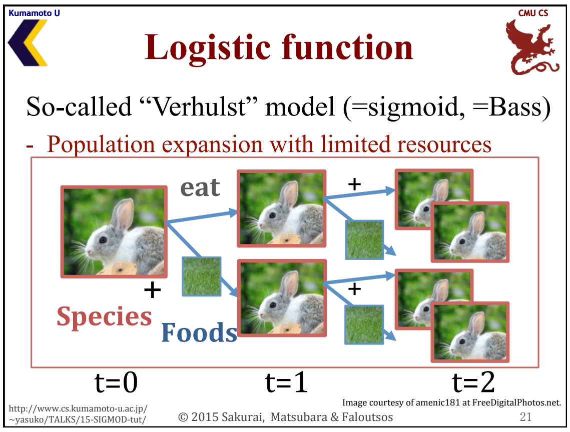

Kumamoto U
CMU CS
How to forecast?
Solution 3
Non-linear equations
Population growth
Competition
Information diffusion
Convection
Big Time series
Epidemics
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
16

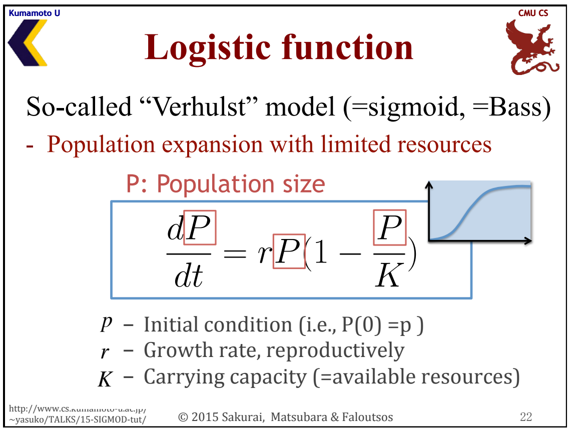

Kumamoto U
CMU CS
Part 2 Roadmap
Problem
Why: "non-linear" modeling
Fundamentals
– Non-linear (grey-box) models
Applications
– Epidemics
– Information diffusion
– (Online) competition
vs.
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
17

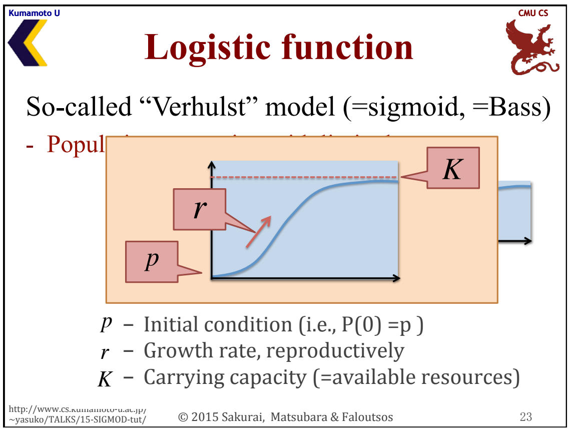

Kumamoto U
CMU CS
Part 2 Roadmap
Problem
Why: "non-linear" modeling
Fundamentals
– Non-linear (grey-box) models
• Logistic function
• Lotka-Volterra (prey-predator, competition)
• SI, SIR models, etc.
• Lorenz equations, etc.
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
18

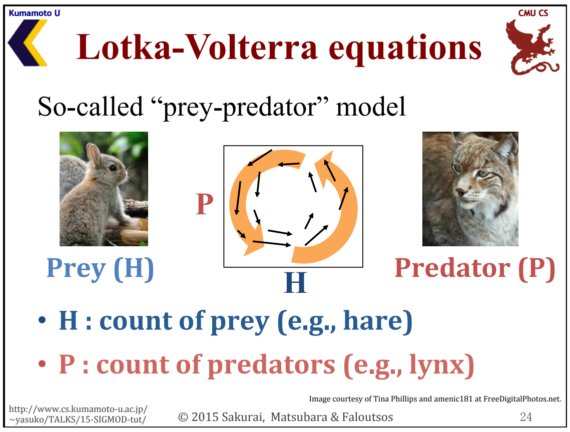

Kumamoto U
CMU CS
Grey-box mining and non-linear equations
Information diffusion
Convection
Population growth
Competition
Big Time series
Epidemics
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
19


Kumamoto U
CMU CS
Grey-box mining and non-linear equations
Information diffusion
Convection
Population growth
Competition
Big Time series
Epidemics
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
20


Kumamoto U
CMU CS
Logistic function
So-called "Verhulst" model (=sigmoid, =Bass)
- Population expansion with limited resources
eat
+
Species
Foods
t=0
t=1
t=2
Image courtesy of amenic181 at FreeDigitalPhotos.net.
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
21


Kumamoto U
CMU CS
Logistic function
So-called "Verhulst" model (=sigmoid, =Bass)
- Population expansion with limited resources
P: Population size
$\frac{dP}{dt} = rP(1 - \frac{P}{K})$
p – Initial condition (i.e., P(0) =p)
r – Growth rate, reproductively
K – Carrying capacity (=available resources)
© 2015 Sakurai, Matsubara & Faloutsos
22

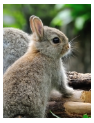

Kumamoto U
CMU CS
Logistic function
So-called "Verhulst" model (=sigmoid, =Bass)
K
r
p
p – Initial condition (i.e., P(0) =p)
r – Growth rate, reproductively
K – Carrying capacity (=available resources)
© 2015 Sakurai, Matsubara & Faloutsos
23


Kumamoto U
CMU CS
Lotka-Volterra equations
So-called "prey-predator" model
P
H
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/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
24

Kumamoto U | CMU CS

Lotka-Volterra equations

So-called "prey-predator" model



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

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

Prey (H) Predator (P)

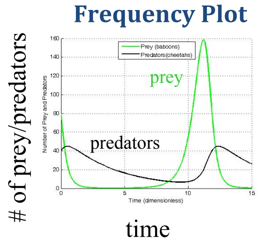
- H : count of prey (e.g., hare)
- P : count of predators (e.g., lynx)

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

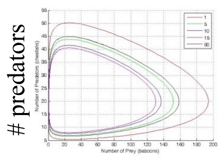
Kumamoto U | CMU CS

Solution to the Lotka-Volterra equations.

Frequency Plot



Phase Space Plot

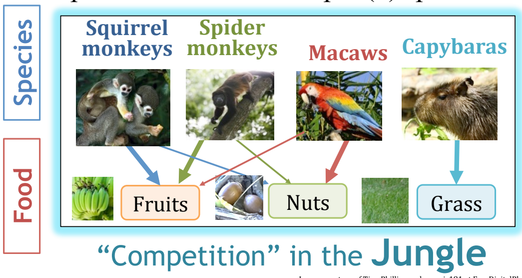


From Wikipedia

Kumamoto U | CMU CS

Extension: "Competitive" Lotka-Volterra equations

Competition between multiple (d) species



"Competition" in the Jungle

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

Details | CMU CS

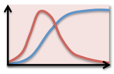
"Competitive" Lotka-Volterra equations

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

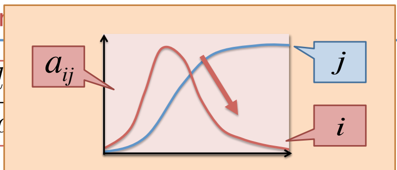


Details | CMU CS

"Competitive" Lotka-Volterra equations

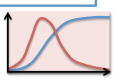
Competition between multiple (d) species

Popular j



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

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



Details | CMU CS

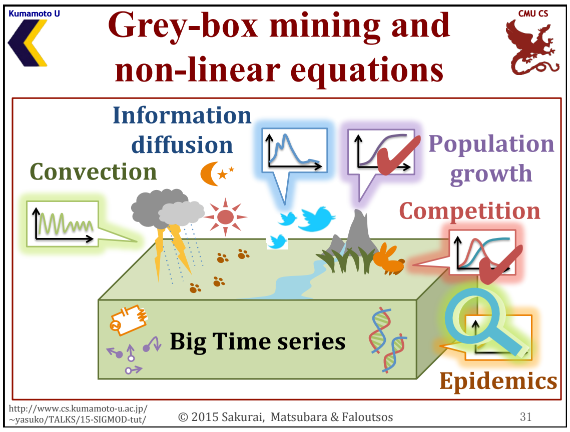
"Competitive" Lotka-Volterra equations

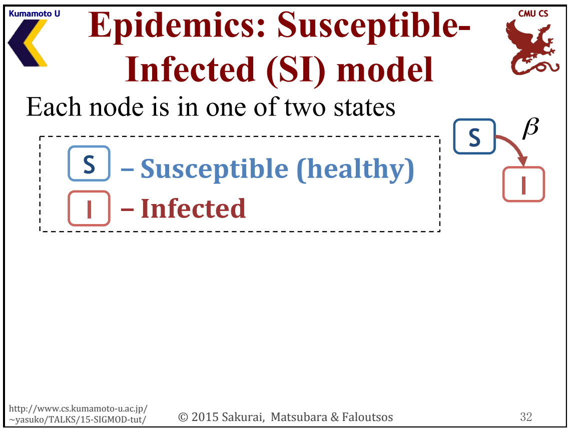
- Biological interaction
  - Table: Type of interaction

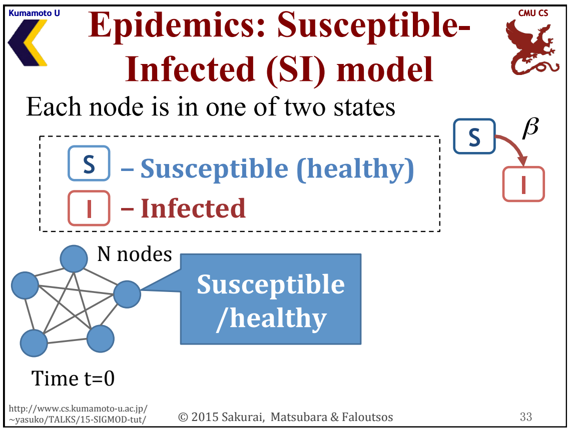
0 : no effect
- : detrimental
+ : beneficial

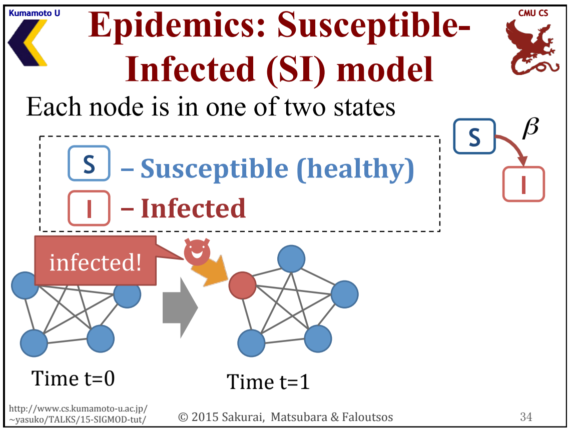
Species B

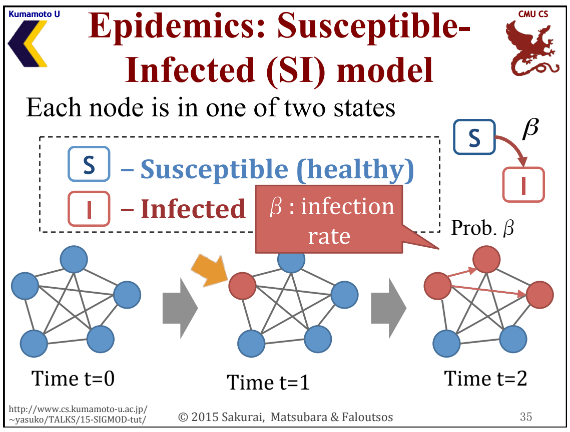
Species A	+	0	-
+	Mutualism		
0	Commensalism	Neutralism	
-	Antagonism	Amensalism	Competition

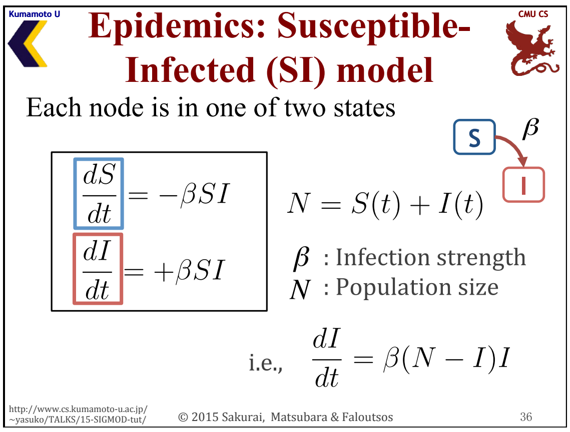

Kumamoto U
Grey-box mining and non-linear equations
CMU CS
Information diffusion
Convection
Population growth
Competition
Big Time series
Epidemics
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
31

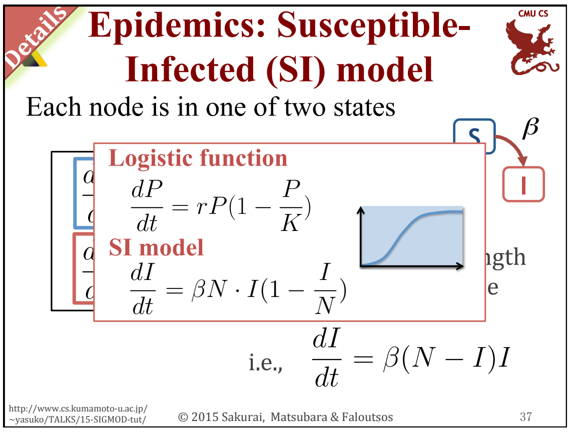

Kumamoto U
Epidemics: Susceptible-Infected (SI) model
CMU CS
Each node is in one of two states
S – Susceptible (healthy)
I – Infected
S
β
I
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
32

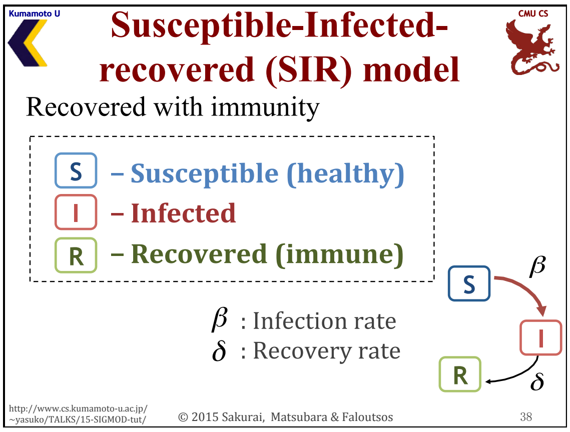

Kumamoto U
Epidemics: Susceptible-Infected (SI) model
CMU CS
Each node is in one of two states
S – Susceptible (healthy)
I – Infected
S
β
I
N nodes
Susceptible /healthy
Time t=0
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
33

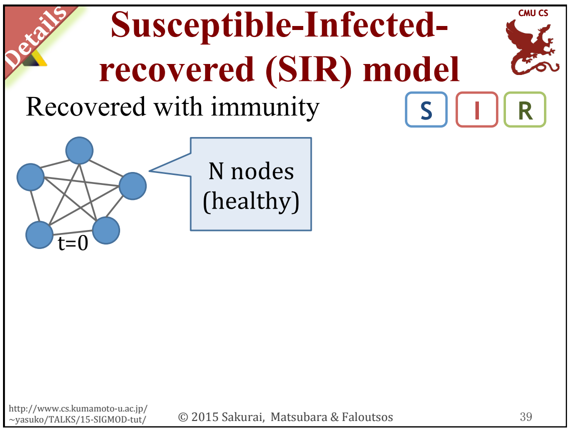

Kumamoto U
Epidemics: Susceptible-Infected (SI) model
CMU CS
Each node is in one of two states
S – Susceptible (healthy)
I – Infected
S
β
I
infected!
Time t=0
Time t=1
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
34

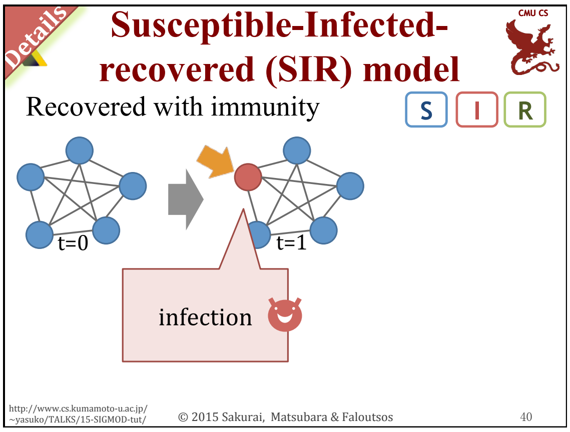

Kumamoto U
Epidemics: Susceptible-Infected (SI) model
CMU CS
Each node is in one of two states
S – Susceptible (healthy)
I – Infected
β : infection rate
S
β
I
Prob. β
Time t=0
Time t=1
Time t=2
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
35

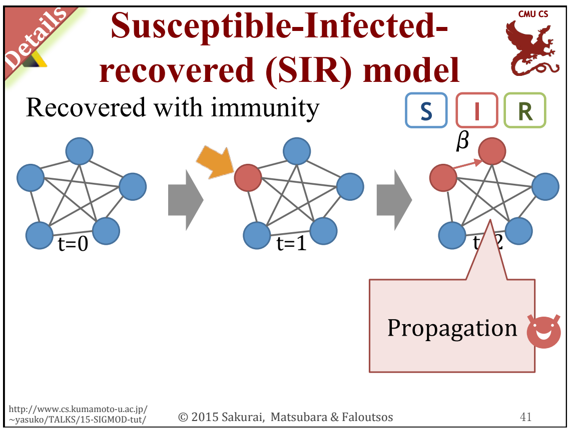

Kumamoto U
Epidemics: Susceptible-Infected (SI) model
CMU CS
Each node is in one of two states
S
β
I
$\frac{dS}{dt} = -\beta SI$
$\frac{dI}{dt} = +\beta SI$
$N = S(t) + I(t)$
β : Infection strength
N : Population size
i.e., $\frac{dI}{dt} = \beta(N - I)I$
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
36

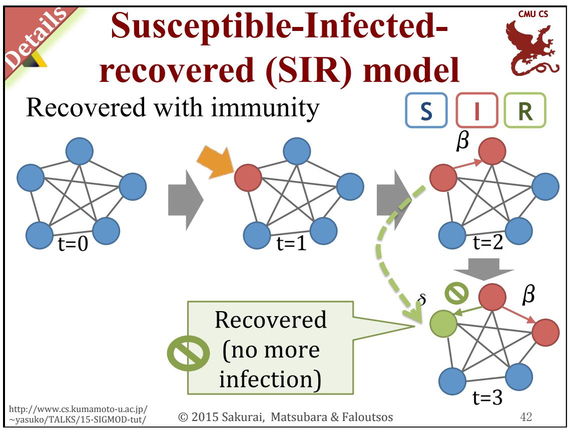

Details
Epidemics: Susceptible-Infected (SI) model
Each node is in one of two states
S
I
β
Logistic function
$\frac{dP}{dt} = rP(1 - \frac{P}{K})$
SI model
$\frac{dI}{dt} = \beta N \cdot I(1 - \frac{I}{N})$
i.e., $\frac{dI}{dt} = \beta(N - I)I$
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
37

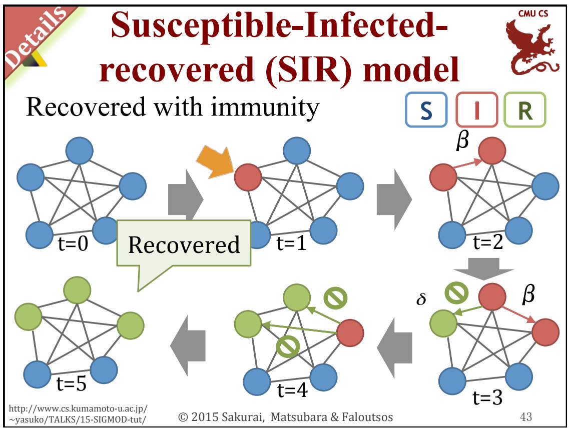

Kumamoto U
Susceptible-Infected-recovered (SIR) model
Recovered with immunity
S – Susceptible (healthy)
I – Infected
R – Recovered (immune)
β : Infection rate
δ : Recovery rate
S
I
R
β
δ
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
38

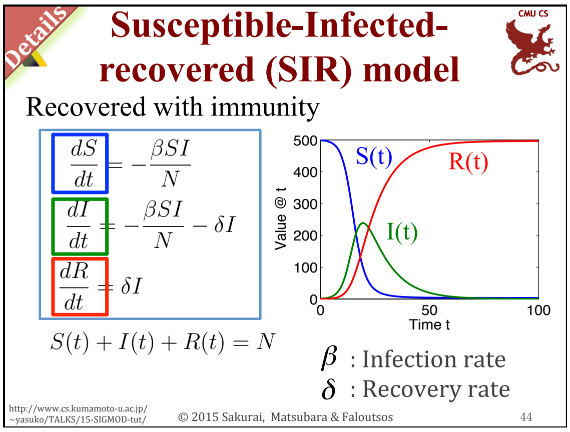

Details
Susceptible-Infected-recovered (SIR) model
Recovered with immunity
S I R
N nodes
(healthy)
t=0
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
39

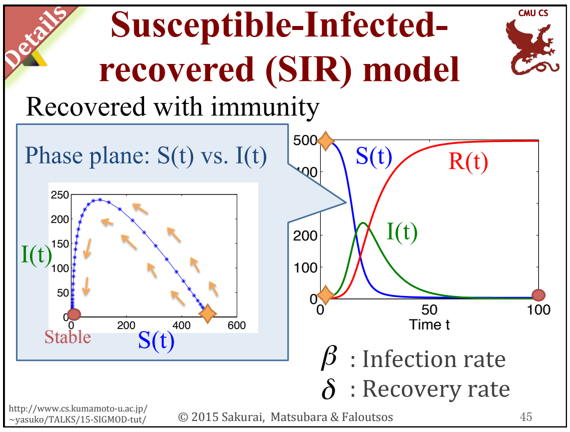

Details
Susceptible-Infected-recovered (SIR) model
Recovered with immunity
S I R
t=0
t=1
infection
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
40

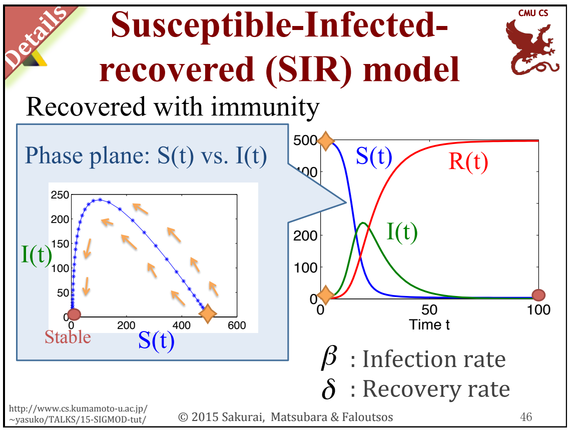

Details
Susceptible-Infected-recovered (SIR) model
Recovered with immunity
S I R
β
t=0
t=1
Propagation
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
41

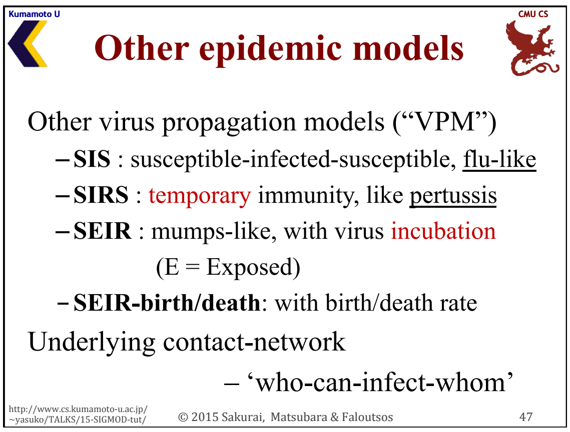

Details
Susceptible-Infected-recovered (SIR) model
Recovered with immunity
S I R
β
t=0
t=1
t=2
β
t=3
Recovered
(no more
infection)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
42

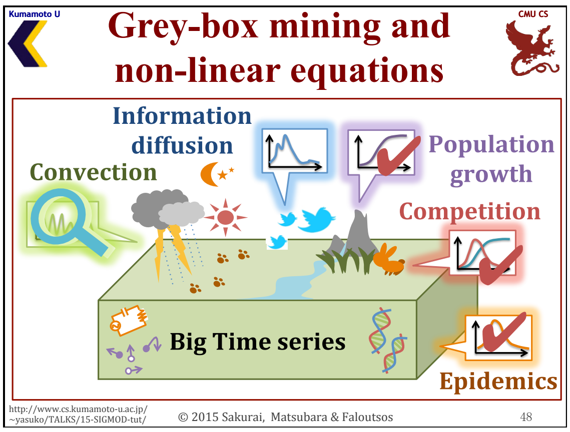

Details
Susceptible-Infected-recovered (SIR) model
CMU CS
Recovered with immunity
S
I
R
β
t=0
Recovered
t=1
t=2
δ
β
t=5
t=4
t=3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
43


Details
Susceptible-Infected-recovered (SIR) model
CMU CS
Recovered with immunity
$\frac{dS}{dt} = -\frac{\beta SI}{N}$
$\frac{dI}{dt} = \frac{\beta SI}{N} - \delta I$
$\frac{dR}{dt} = \delta I$
$S(t) + I(t) + R(t) = N$
S(t)
R(t)
I(t)
Value @ t
Time t
β : Infection rate
δ : Recovery rate
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
44


Details
Susceptible-Infected-recovered (SIR) model
CMU CS
Recovered with immunity
Phase plane: S(t) vs. I(t)
I(t)
Stable
S(t)
S(t)
R(t)
I(t)
Time t
β : Infection rate
δ : Recovery rate
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
45


Details
Susceptible-Infected-recovered (SIR) model
CMU CS
Recovered with immunity
Phase plane: S(t) vs. I(t)
I(t)
Stable
S(t)
S(t)
R(t)
I(t)
Time t
β : Infection rate
δ : Recovery rate
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
46


Kumamoto U
Other epidemic models
CMU CS
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'
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
47


Kumamoto U
Grey-box mining and non-linear equations
CMU CS
Information diffusion
Population growth
Convection
Competition
Big Time series
Epidemics
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
48

Other non-linear models

LORENZ: eqs. for atmospheric convection

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

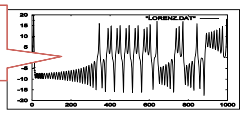
- x: convective intensity
- y: temperature difference between ascending and descending currents
- z: difference in vertical temperature profile from linearity

Other non-linear models

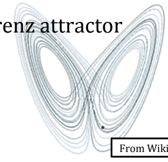
LORENZ: eqs. for atmospheric convection

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

Butterfly effect (chaos)



Lorenz attractor

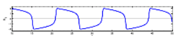


From Wikipedia

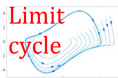
Other non-linear models

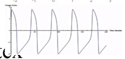
From Wikipedia

- 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
- …
- …

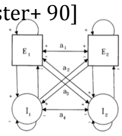


Limit cycle





[Schuster+ 90]



Part 2 Roadmap

Problem

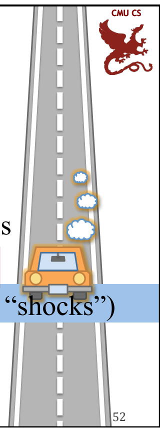
✔ Why: "non-linear" modeling

Fundamentals

✔ Non-linear ("gray-box") models

Applications

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



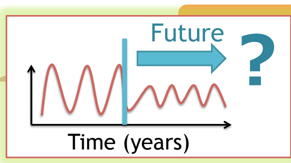
Mining and forecasting of co-evolving epidemics

Flu

Measles

Mumps

Mining and forecasting of co-evolving epidemics



Q. Can we forecast future epidemics?

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]

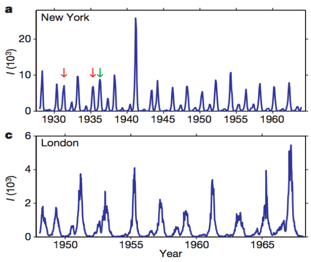
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Time series of reported measles cases

New York

London



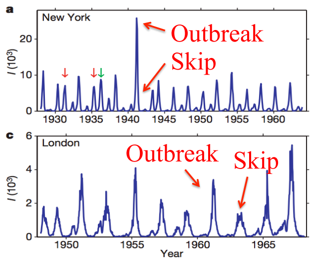
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Time series of reported measles cases

New York

London

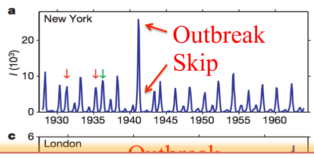


Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

- Time series of reported measles cases

New York



Q. Outbreak vs. skip?

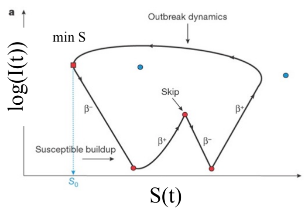
Details

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))



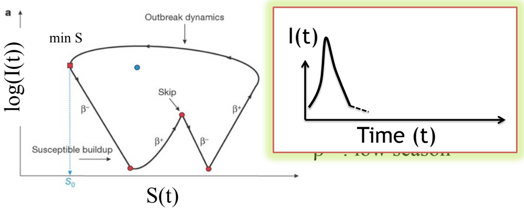
Contact rate
β+ : high season
β - : low season

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))

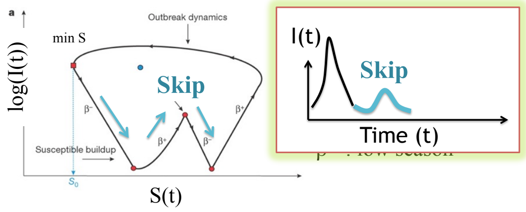


Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

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

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

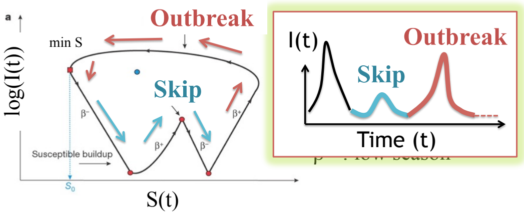


Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

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

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



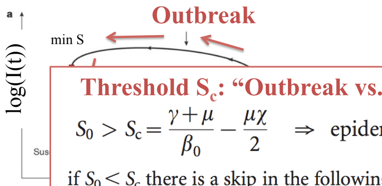
Recurrent epidemics: Outbreak or skip?

[Stone+ Nature'07]

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

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

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

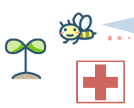


Threshold S_c: "Outbreak vs. Skip"

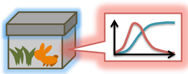
$$S_0 > S_c = \frac{\gamma+\mu}{\beta_0} - \frac{\mu\chi}{2} \quad \Rightarrow \quad \text{epidemic}$$

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

Epidemics - roadmap



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

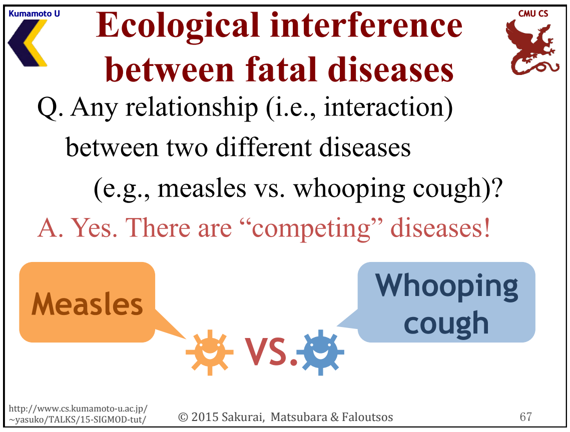


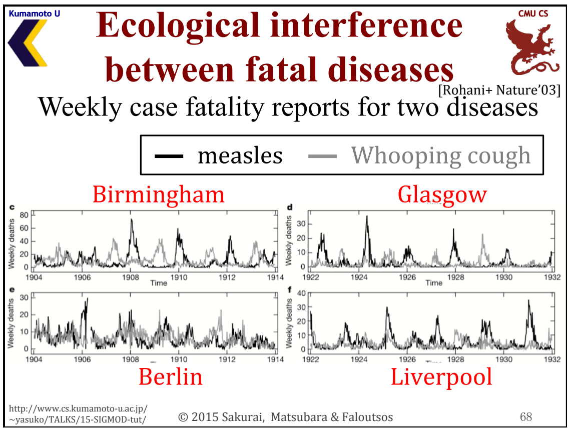
Solutions

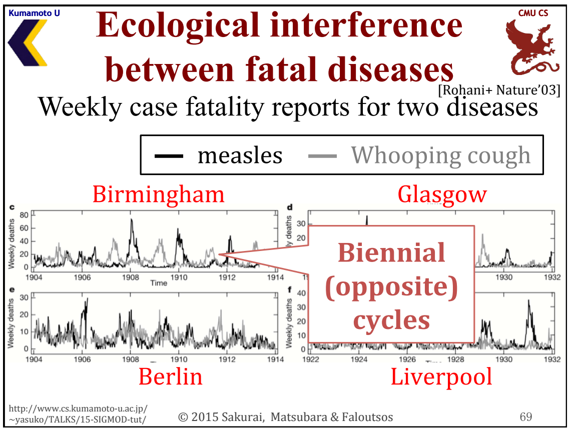
- Outbreak vs. Skips [Stone+ Nature'07]
- **Interaction between diseases [Rohani+ Nature'03]**
- FUNNEL [Matsubara+ KDD'14]

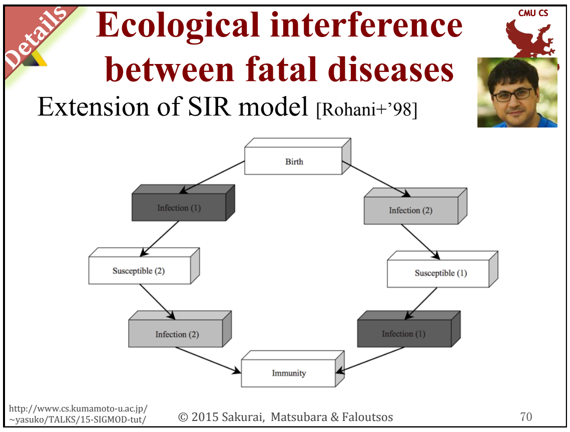
Ecological interference between fatal diseases

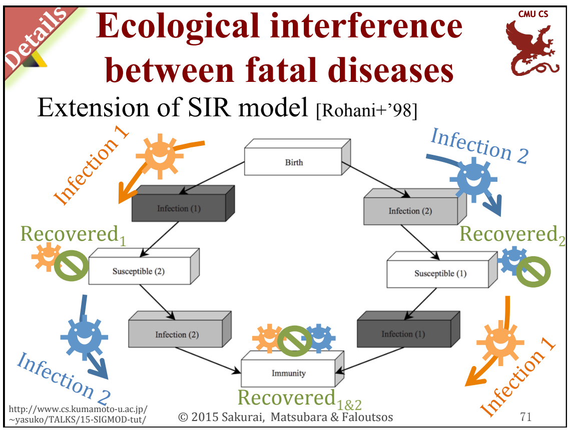
Q. Any relationship (i.e., interaction) between two different diseases (e.g., measles vs. whooping cough)?

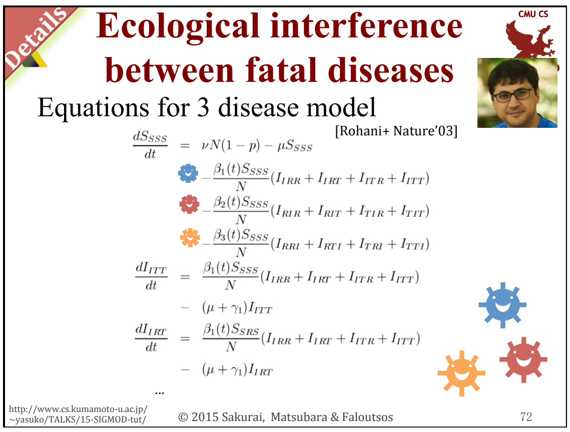

Kumamoto U
CMU CS
Ecological interference between fatal diseases
Q. Any relationship (i.e., interaction) between two different diseases (e.g., measles vs. whooping cough)?
A. Yes. There are "competing" diseases!
Measles
Whooping cough
VS.
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
67


Kumamoto U
CMU CS
Ecological interference between fatal diseases
[Rohani+ Nature'03]
Weekly case fatality reports for two diseases
measles
Whooping cough
Birmingham
Glasgow
Berlin
Liverpool
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
68


Kumamoto U
CMU CS
Ecological interference between fatal diseases
[Rohani+ Nature'03]
Weekly case fatality reports for two diseases
measles
Whooping cough
Birmingham
Glasgow
Biennial (opposite) cycles
Berlin
Liverpool
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
69

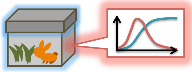

Details
CMU CS
Ecological interference between fatal diseases
Extension of SIR model [Rohani+'98]
Birth
Infection (1)
Infection (2)
Susceptible (2)
Susceptible (1)
Infection (2)
Infection (1)
Immunity
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
70


Details
CMU CS
Ecological interference between fatal diseases
Extension of SIR model [Rohani+'98]
Infection 1
Infection 2
Birth
Recovered1
Recovered2
Infection (1)
Infection (2)
Susceptible (2)
Susceptible (1)
Infection (2)
Infection (1)
Immunity
Infection 2
Recovered1&2
Infection 1
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
71


Details
CMU CS
Ecological interference between fatal diseases
Equations for 3 disease model
[Rohani+ Nature'03]
$\frac{dS_{SSS}}{dt} = \nu N(1-p) - \mu S_{SSS}$
$- \frac{\beta_1(t)S_{SSS}}{N}(I_{IRR} + I_{IRT} + I_{ITR} + I_{ITT})$
$- \frac{\beta_2(t)S_{SSS}}{N}(I_{RIR} + I_{RIT} + I_{TIR} + I_{TIT})$
$- \frac{\beta_3(t)S_{SSS}}{N}(I_{RRI} + I_{RTI} + I_{TRI} + I_{TTI})$
$\frac{dI_{ITT}}{dt} = \frac{\beta_1(t)S_{SSS}}{N}(I_{IRR} + I_{IRT} + I_{ITR} + I_{ITT})$
$- (\mu + \gamma_1)I_{ITT}$
$\frac{dI_{IRT}}{dt} = \frac{\beta_1(t)S_{SRS}}{N}(I_{IRR} + I_{IRT} + I_{ITR} + I_{ITT})$
$- (\mu + \gamma_1)I_{IRT}$
...
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
72

Epidemics - roadmap

Non-linear (gray-box) modeling!



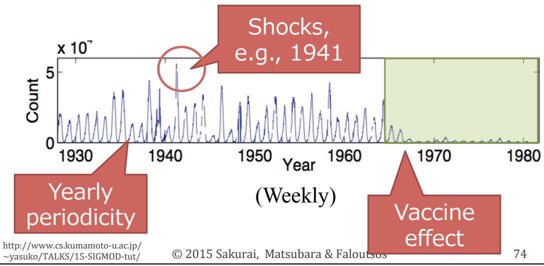
Solutions

- 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/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 73

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

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

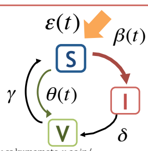


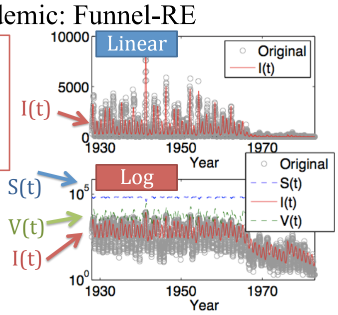
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 74

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





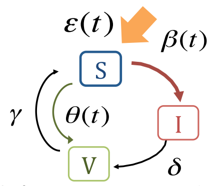
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 75

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

With a single epidemic: Funnel-RE

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

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



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 76

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

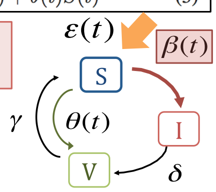
With a single epidemic: Funnel-RE

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

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

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

$$P_p = 52$$



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 77

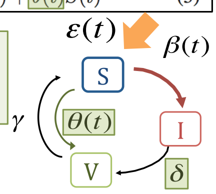
FUNNEL [Matsubara+ KDD'14] with a single epidemic

With a single epidemic: Funnel-RE

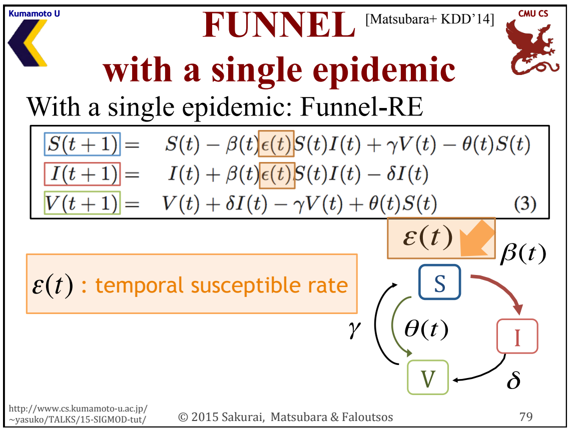
$$\begin{aligned} S(t+1) &= S(t) - \beta(t)\epsilon(t)S(t)I(t) + \gamma V(t) - \boxed{\theta(t)}S(t) \\ I(t+1) &= I(t) + \beta(t)\epsilon(t)S(t)I(t) - \boxed{\delta}I(t) \\ V(t+1) &= V(t) + \boxed{\delta}I(t) - \gamma V(t) + \boxed{\theta(t)}S(t) \qquad (3) \end{aligned}$$

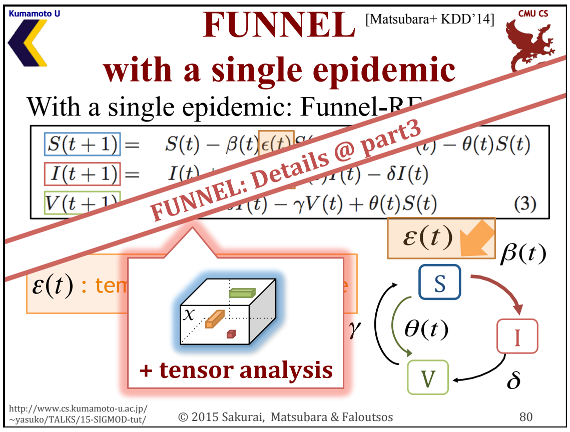
δ : healing rate
$\theta(t)$: disease reduction effect

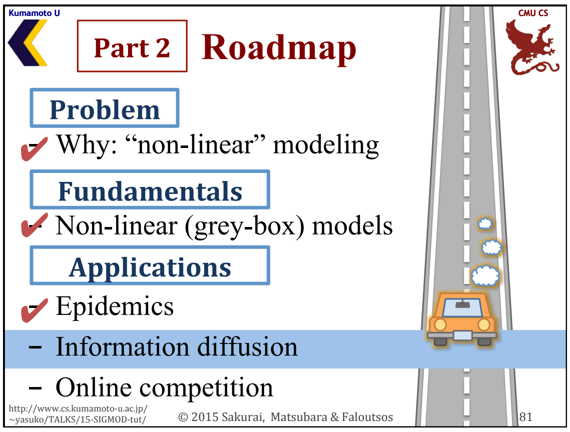
$$\theta(t) = \begin{cases} 0 & (t < t_\theta) \\ \theta_0 & (t \geq t_\theta) \end{cases}$$

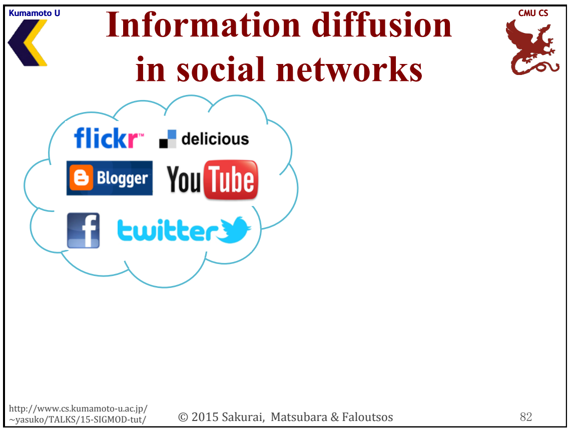


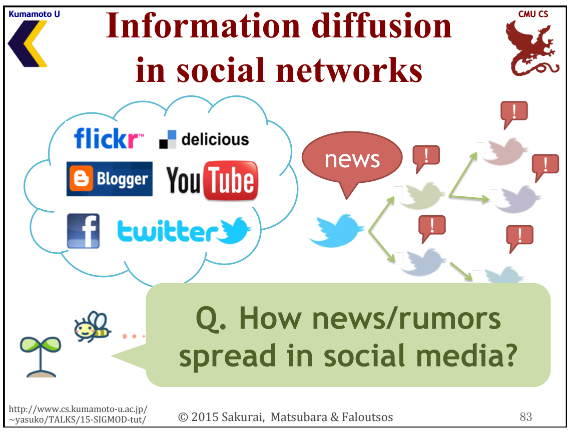
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 78

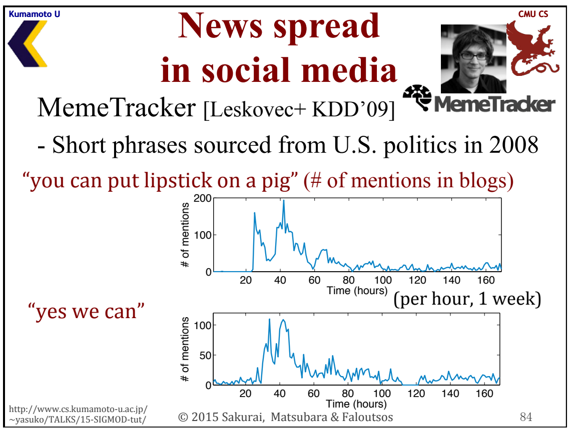

Kumamoto U
FUNNEL [Matsubara+ KDD'14]
with a single epidemic
CMU CS
With a single epidemic: Funnel-RE
$S(t+1) = S(t) - \beta(t)\epsilon(t)S(t)I(t) + \gamma V(t) - \theta(t)S(t)$
$I(t+1) = I(t) + \beta(t)\epsilon(t)S(t)I(t) - \delta I(t)$
$V(t+1) = V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t)$ (3)
$\varepsilon(t)$
$\varepsilon(t)$: temporal susceptible rate
$\beta(t)$
S
I
V
$\theta(t)$
γ
δ
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
79


Kumamoto U
FUNNEL [Matsubara+ KDD'14]
with a single epidemic
CMU CS
With a single epidemic: Funnel-RE
FUNNEL: Details @ part3
+ tensor analysis
$\varepsilon(t)$
$\beta(t)$
S
I
V
$\theta(t)$
γ
δ
(3)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
80


Kumamoto U
Part 2 Roadmap
CMU CS
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/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
81


Kumamoto U
Information diffusion
in social networks
CMU CS
flickr
delicious
Blogger
YouTube
twitter
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
82


Kumamoto U
Information diffusion
in social networks
CMU CS
flickr
delicious
Blogger
YouTube
twitter
news
Q. How news/rumors
spread in social media?
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
83

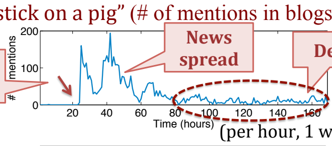

Kumamoto U
News spread
in social media
CMU CS
MemeTracker [Leskovec+ KDD'09]
MemeTracker
- Short phrases sourced from U.S. politics in 2008
"you can put lipstick on a pig" (# of mentions in blogs)
# of mentions
200
100
0
20 40 60 80 100 120 140 160
Time (hours)
(per hour, 1 week)
"yes we can"
# of mentions
100
50
0
20 40 60 80 100 120 140 160
Time (hours)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
84

News spread in social media

MemeTracker [Leskovec+ KDD'09]

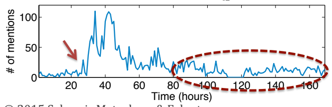
- Short phrases sourced from U.S. politics in 2008

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



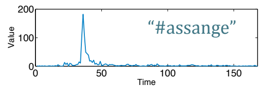
(per hour, 1 week)

"yes we can"

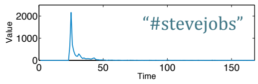


News spread in social media

- Twitter (# of hashtags per hour)



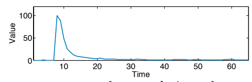
(per hour, 1week)



(per hour, 1 week)

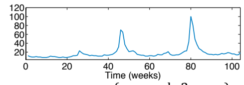
- Google trend (# of queries per week)

"tsunami" (in 2005)



(per week, 1 year)

"harry potter" (2010 - 2011)

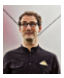


(per week, 2 years)

News spread in social media

Q. How many patterns are there?

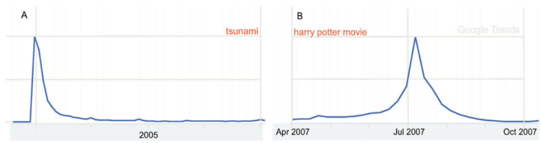
- Four classes on YouTube, etc. [Crane et al. PNAS'08]
- Six classes on Social media [Yang et al. WSDM'11]



News spread in social media

[Crane et al. PNAS'08]

- The volume of Google searches

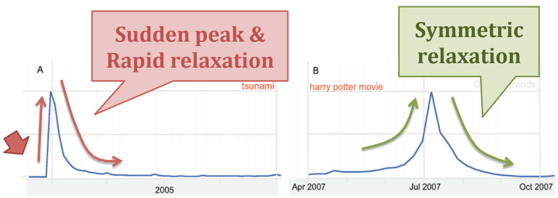


"Tsunami" "Harry potter movie"

News spread in social media

[Crane et al. PNAS'08]

- The volume of Google searches



"Tsunami" (Exogenous)

"Harry potter movie" (Endogenous)

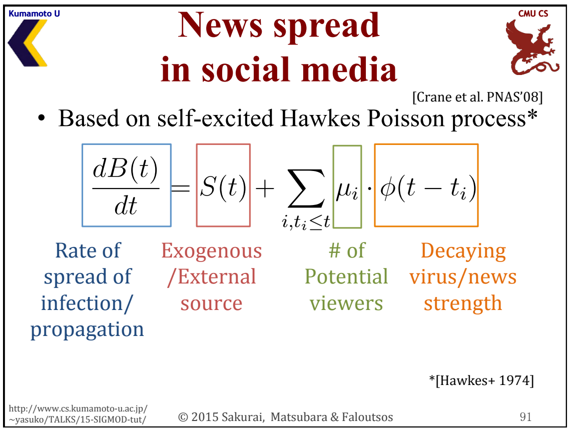
News spread in social media

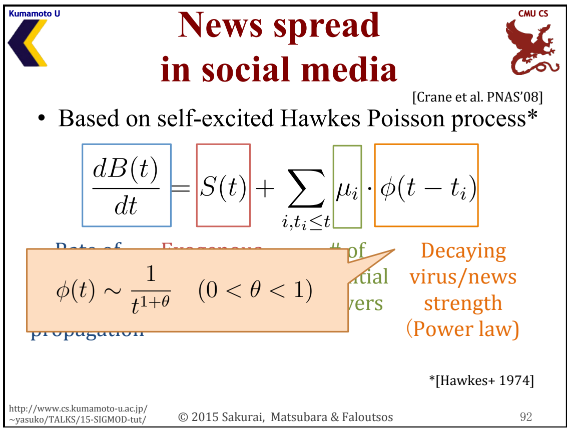
[Crane et al. PNAS'08]

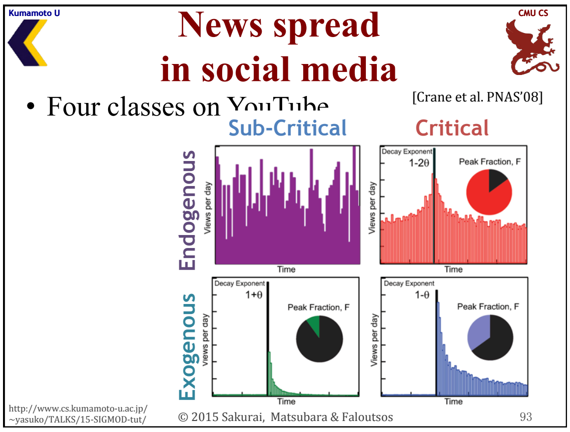
- Based on self-excited Hawkes Poisson process*

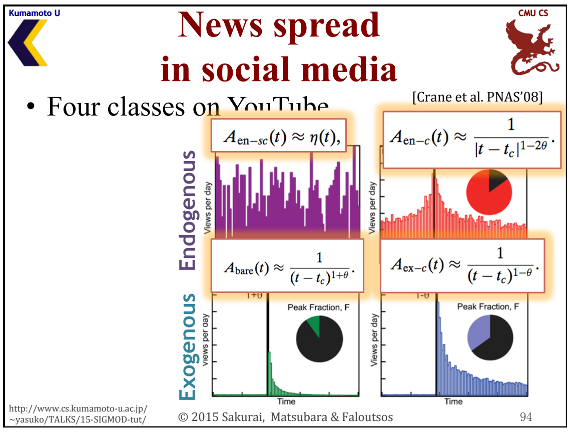
$$\frac{dB(t)}{dt} = S(t) + \sum_{i, t_i \le t} \mu_i \cdot \phi(t - t_i)$$

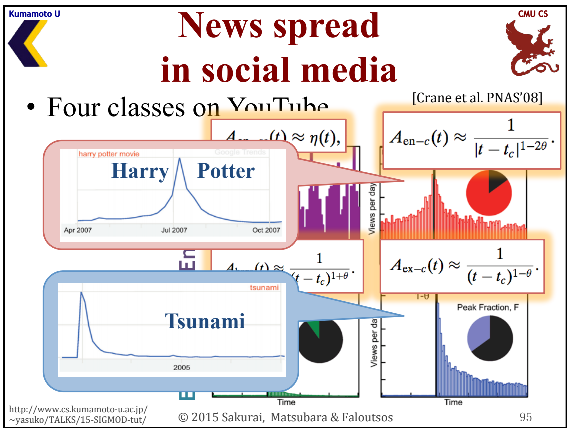
*[Hawkes+ 1974]

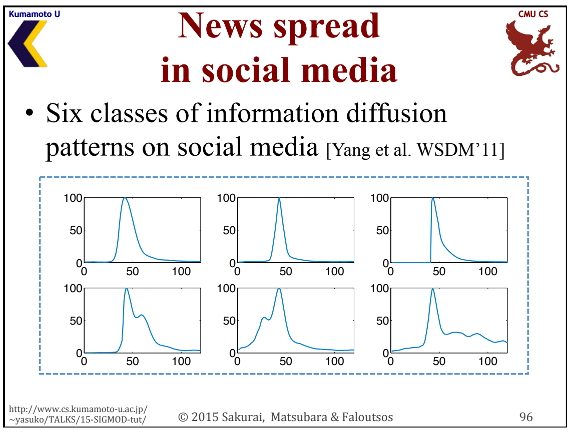

Kumamoto U
CMU CS
News spread in social media
[Crane et al. PNAS'08]
• Based on self-excited Hawkes Poisson process*
$\frac{dB(t)}{dt} = S(t) + \sum_{i,t_i \le t} \mu_i \cdot \phi(t - t_i)$
Rate of spread of infection/ propagation
Exogenous /External source
# of Potential viewers
Decaying virus/news strength
*[Hawkes+ 1974]
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
91


Kumamoto U
CMU CS
News spread in social media
[Crane et al. PNAS'08]
• Based on self-excited Hawkes Poisson process*
$\frac{dB(t)}{dt} = S(t) + \sum_{i,t_i \le t} \mu_i \cdot \phi(t - t_i)$
$\phi(t) \sim \frac{1}{t^{1+\theta}} \quad (0 < \theta < 1)$
Decaying virus/news strength (Power law)
*[Hawkes+ 1974]
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
92


Kumamoto U
CMU CS
News spread in social media
[Crane et al. PNAS'08]
• Four classes on YouTube
Sub-Critical
Critical
Endogenous
Exogenous
Decay Exponent 1-2θ
Decay Exponent 1+θ
Decay Exponent 1-θ
Peak Fraction, F
Views per day
Time
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
93


Kumamoto U
CMU CS
News spread in social media
[Crane et al. PNAS'08]
• Four classes on YouTube
$A_{en-sc}(t) \approx \eta(t),$
$A_{en-c}(t) \approx \frac{1}{|t-t_c|^{1-2\theta}}.$
$A_{bare}(t) \approx \frac{1}{(t-t_c)^{1+\theta}}.$
$A_{ex-c}(t) \approx \frac{1}{(t-t_c)^{1-\theta}}.$
Endogenous
Exogenous
Peak Fraction, F
Views per day
Time
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
94

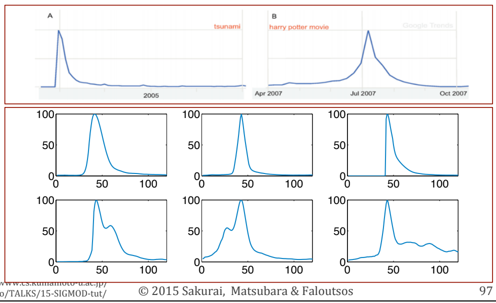

Kumamoto U
CMU CS
News spread in social media
[Crane et al. PNAS'08]
• Four classes on YouTube
$A_{en-c}(t) \approx \frac{1}{|t-t_c|^{1-2\theta}}.$
$A_{ex-c}(t) \approx \frac{1}{(t-t_c)^{1-\theta}}.$
Harry Potter
Apr 2007
Jul 2007
Oct 2007
Tsunami
2005
Peak Fraction, F
Time
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
95


Kumamoto U
CMU CS
News spread in social media
• Six classes of information diffusion patterns on social media [Yang et al. WSDM'11]
100
50
0
50
100
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
96

Kumamoto U | CMU CS

News spread in social media

Q. How many patterns are there, after all?



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 97

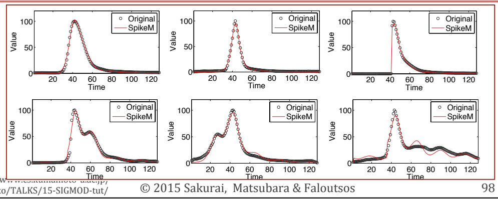
Kumamoto U | CMU CS

News spread in social media

A. Our answer is “ONE”!

A single non-linear model !

“SpikeM”



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 98

[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)

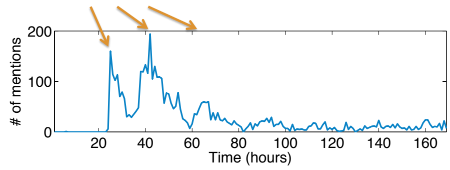
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 99

Kumamoto U | CMU CS

Rise and fall patterns in social media

SpikeM captures 3 properties of real spike

1. periodicities



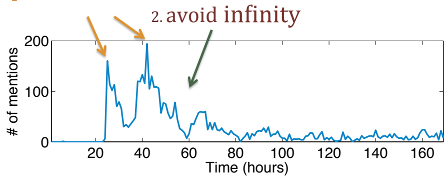
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 100

Kumamoto U | CMU CS

Rise and fall patterns in social media

SpikeM captures 3 properties of real spike

1. periodicities
2. avoid infinity



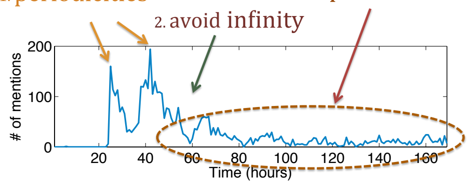
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 101

Kumamoto U | CMU CS

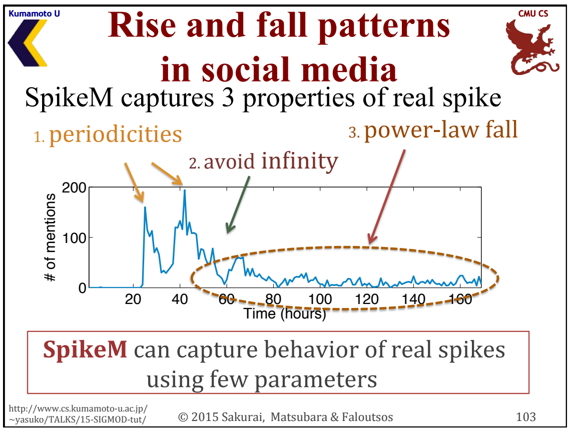
Rise and fall patterns in social media

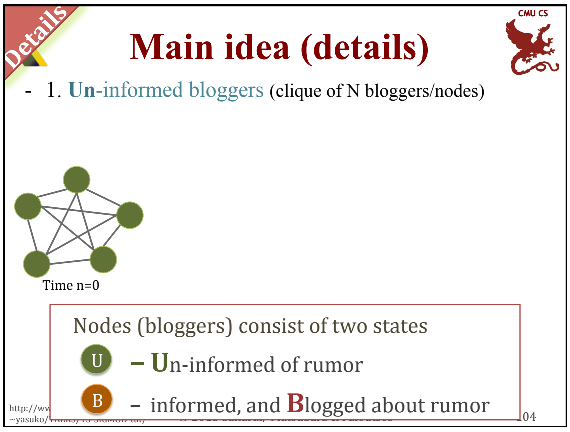
SpikeM captures 3 properties of real spike

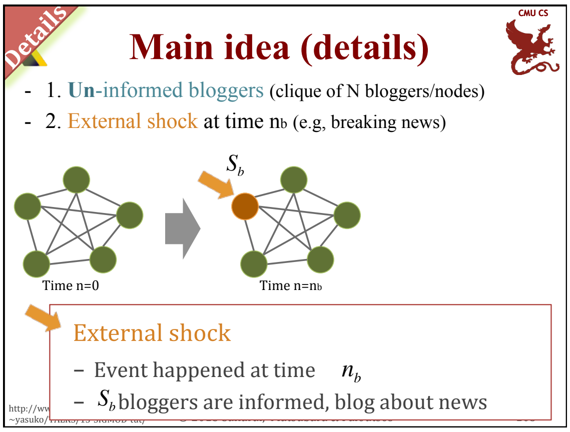
1. periodicities
2. avoid infinity
3. power-law fall

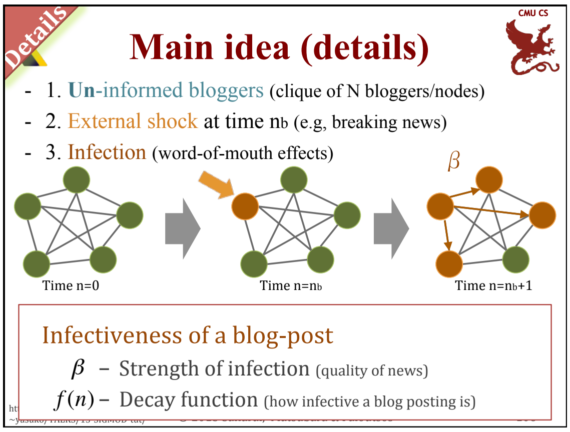


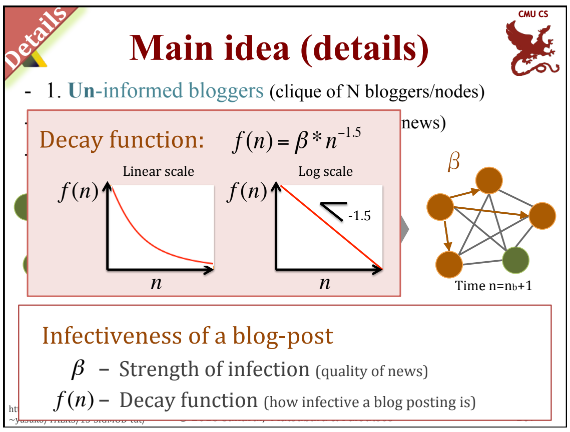
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 102

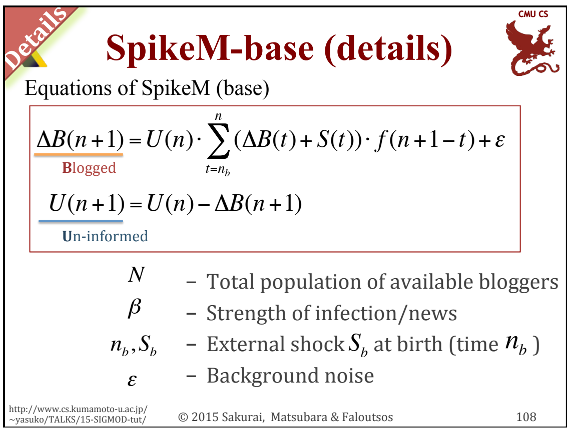

Kumamoto U
CMU CS
Rise and fall patterns in social media
SpikeM captures 3 properties of real spike
1. periodicities
2. avoid infinity
3. power-law fall
# of mentions
200
100
0
20 40 60 80 100 120 140 160
Time (hours)
SpikeM can capture behavior of real spikes using few parameters
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
103

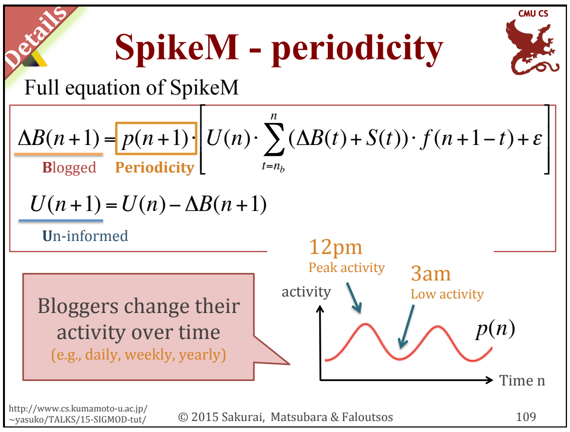

Details
CMU CS
Main idea (details)
- 1. Un-informed bloggers (clique of N bloggers/nodes)
Time n=0
Nodes (bloggers) consist of two states
U – Un-informed of rumor
B – informed, and Blogged about rumor
104

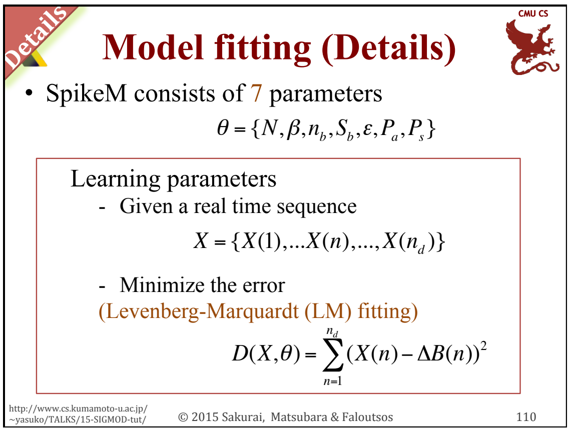

Details
CMU CS
Main idea (details)
- 1. Un-informed bloggers (clique of N bloggers/nodes)
- 2. External shock at time nb (e.g, breaking news)
S_b
Time n=0
Time n=nb
External shock
– Event happened at time n_b
– S_b bloggers are informed, blog about news

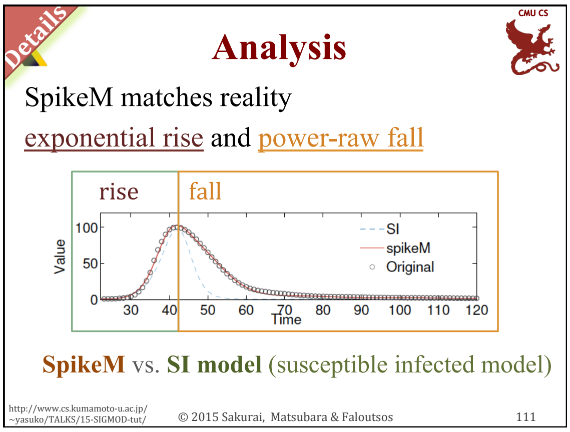

Details
CMU CS
Main idea (details)
- 1. Un-informed bloggers (clique of N bloggers/nodes)
- 2. External shock at time nb (e.g, breaking news)
- 3. Infection (word-of-mouth effects)
β
Time n=0
Time n=nb
Time n=nb+1
Infectiveness of a blog-post
β – Strength of infection (quality of news)
$f(n)$ – Decay function (how infective a blog posting is)

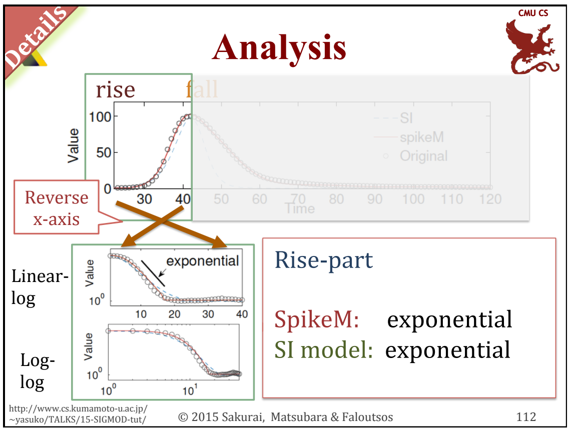

Details
CMU CS
Main idea (details)
- 1. Un-informed bloggers (clique of N bloggers/nodes)
news)
Decay function: $f(n) = \beta * n^{-1.5}$
Linear scale
Log scale
$f(n)$
n
-1.5
β
Time n=nb+1
Infectiveness of a blog-post
β – Strength of infection (quality of news)
$f(n)$ – Decay function (how infective a blog posting is)

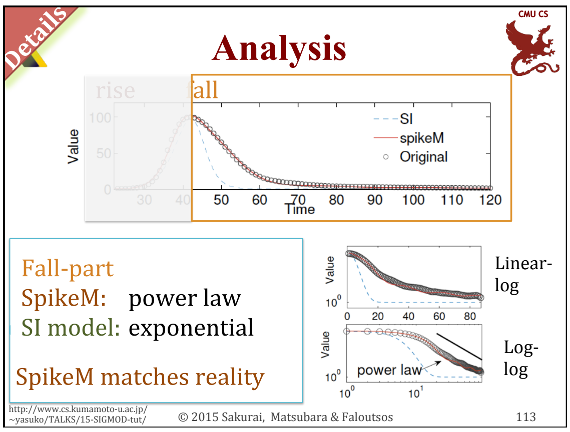

Details
CMU CS
SpikeM-base (details)
Equations of SpikeM (base)
$\Delta B(n+1) = U(n) \cdot \sum_{t=n_b}^{n} (\Delta B(t) + S(t)) \cdot f(n+1-t) + \varepsilon$
Blogged
$U(n+1) = U(n) - \Delta B(n+1)$
Un-informed
N – Total population of available bloggers
β – Strength of infection/news
n_b, S_b – External shock S_b at birth (time n_b)
ε – Background noise
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
108

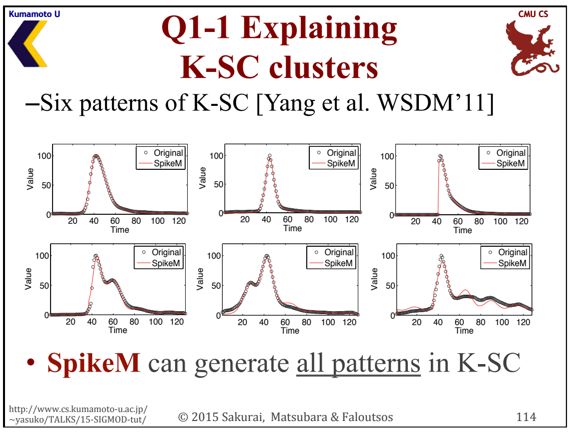

Details
SpikeM - periodicity
Full equation of SpikeM
$\Delta B(n+1) = p(n+1) \cdot \left[U(n) \cdot \sum_{t=n_b}^{n} (\Delta B(t) + S(t)) \cdot f(n+1-t) + \varepsilon \right]$
Blogged
Periodicity
$U(n+1) = U(n) - \Delta B(n+1)$
Un-informed
12pm
Peak activity
3am
Low activity
activity
$p(n)$
Time n
Bloggers change their activity over time
(e.g., daily, weekly, yearly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
109


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/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
110


Details
Analysis
SpikeM matches reality
exponential rise and power-raw fall
rise
fall
SI
spikeM
Original
Value
Time
SpikeM vs. SI model (susceptible infected model)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
111

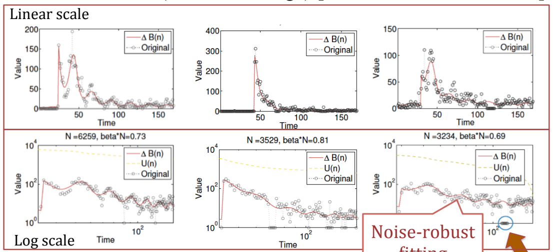

Details
Analysis
rise
fall
Value
Time
Reverse x-axis
Linear-log
Log-log
exponential
Rise-part
SpikeM: exponential
SI model: exponential
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
112


Details
Analysis
rise
fall
SI
spikeM
Original
Value
Time
Fall-part
SpikeM: power law
SI model: exponential
SpikeM matches reality
Linear-log
Log-log
power law
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
113


Kumamoto U
Q1-1 Explaining K-SC clusters
–Six patterns of K-SC [Yang et al. WSDM'11]
Original
SpikeM
Value
Time
• SpikeM can generate all patterns in K-SC
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
114

Q1-2 Matching MemeTracker patterns

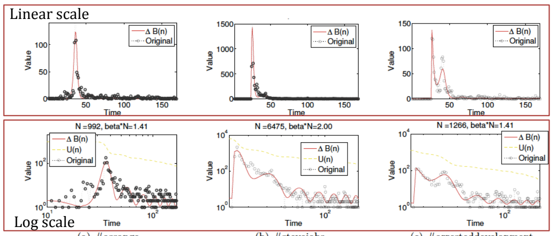
MemeTracker (memes in blogs) [Leskovec et al. KDD'09]



SpikeM can fit various patterns in blog

Q1-3 Matching Twitter data

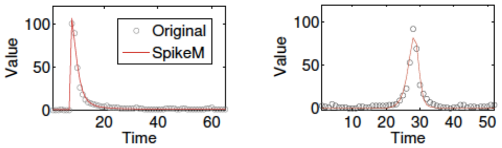
Twitter data (hashtags)



It can generate various patterns in social media

Q1-4 Matching Google trend data

Volume of searches for queries on Google

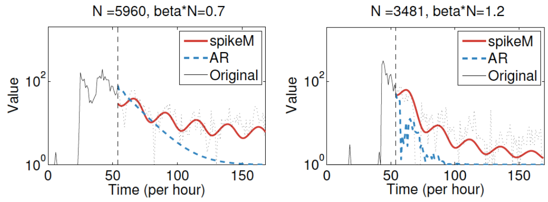


(a) "tsunami" (2005) (b) "Harry Potter" (2007)

SpikeM can capture various patterns

Q2 Tail-part forecasts

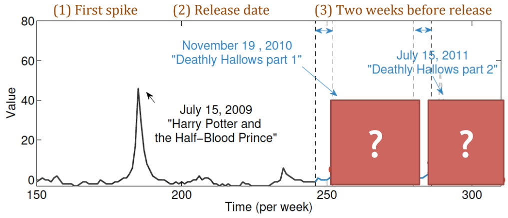
- Given a first part of the spike
  - forecast the tail part



SpikeM can capture tail part (AR: fail)

A1. "What-if" forecasting

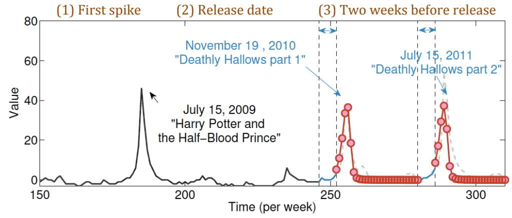
Forecast not only tail-part, but also **rise-part**!



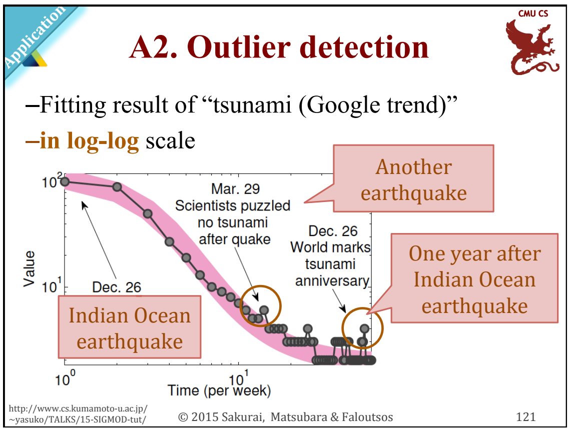
e.g., given (1) first spike,
(2) release date of two sequel movies
(3) access volume before the release date

A1. "What-if" forecasting

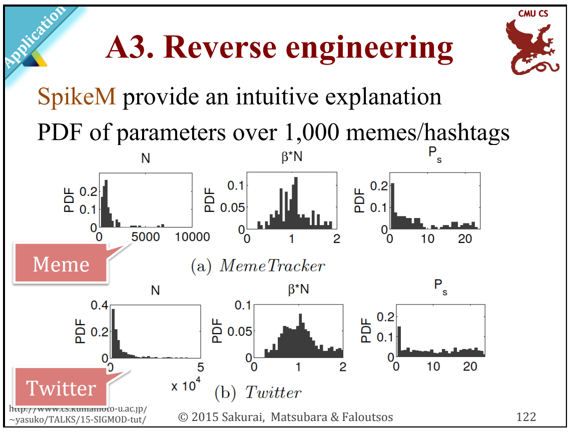
Forecast not only tail-part, but also **rise-part**!



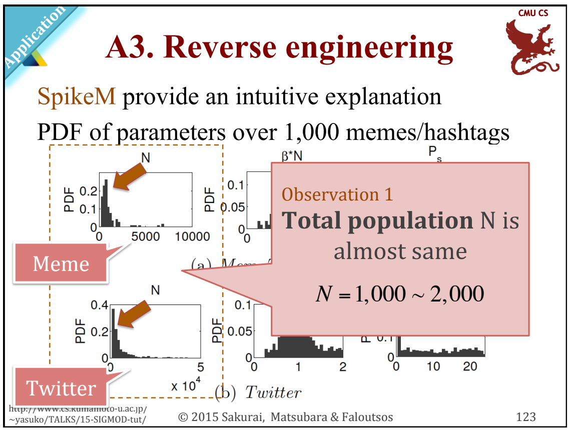
SpikeM can forecast upcoming spikes!



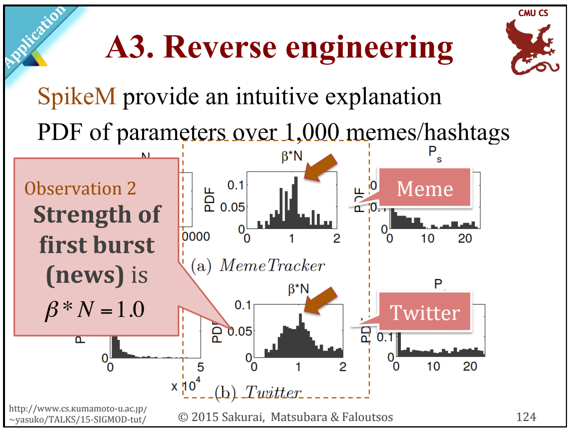
Application
A2. Outlier detection
–Fitting result of "tsunami (Google trend)"
–in log-log scale
Another earthquake
Mar. 29 Scientists puzzled no tsunami after quake
Dec. 26 World marks tsunami anniversary
One year after Indian Ocean earthquake
Dec. 26
Indian Ocean earthquake
Value
Time (per week)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
121



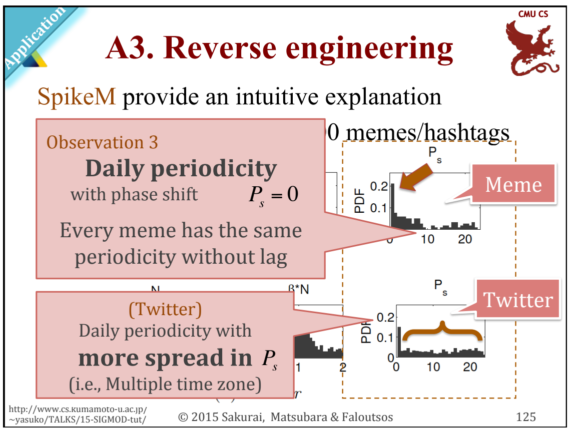
Application
A3. Reverse engineering
SpikeM provide an intuitive explanation
PDF of parameters over 1,000 memes/hashtags
N
β*N
Ps
Meme
(a) MemeTracker
Twitter
(b) Twitter
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
122



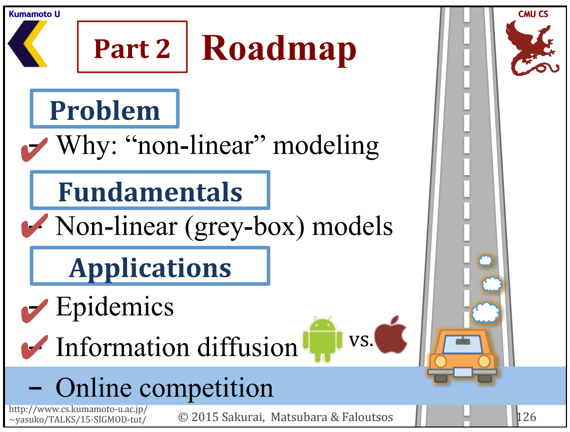
Application
A3. Reverse engineering
SpikeM provide an intuitive explanation
PDF of parameters over 1,000 memes/hashtags
Observation 1
Total population N is almost same
N = 1,000 ~ 2,000
Meme
Twitter
(b) Twitter
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
123



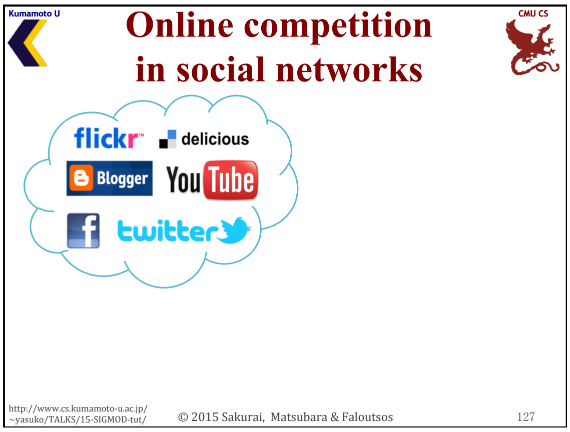
Application
A3. Reverse engineering
SpikeM provide an intuitive explanation
PDF of parameters over 1,000 memes/hashtags
Observation 2
Strength of first burst (news) is
β * N = 1.0
Meme
Twitter
(a) MemeTracker
(b) Twitter
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
124

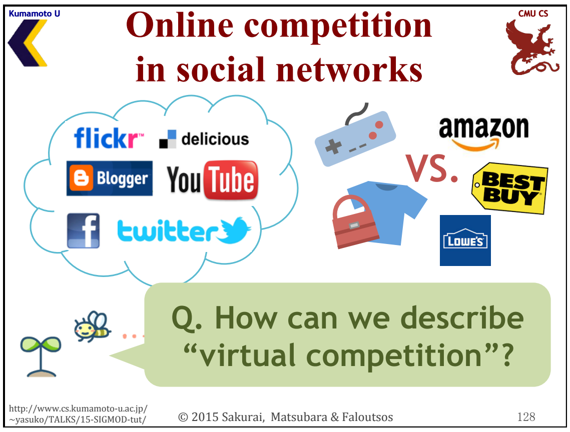


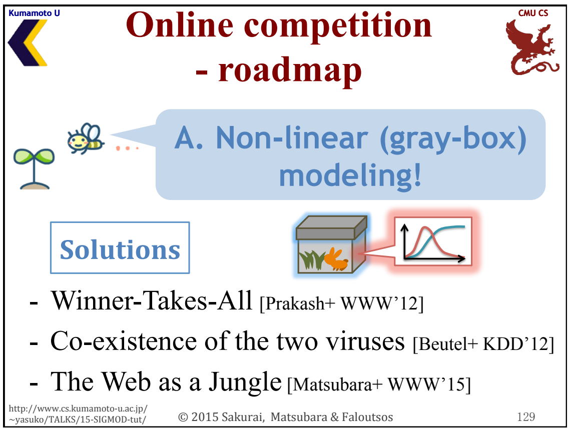
Application
A3. Reverse engineering
SpikeM provide an intuitive explanation
Observation 3
Daily periodicity
with phase shift Ps = 0
Every meme has the same periodicity without lag
Meme
(Twitter)
Daily periodicity with more spread in Ps
(i.e., Multiple time zone)
Twitter
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
125

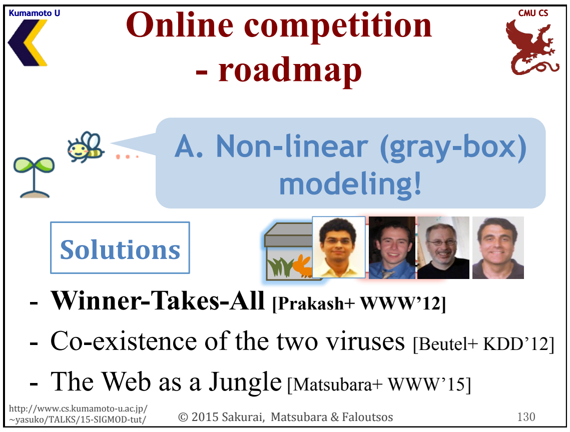


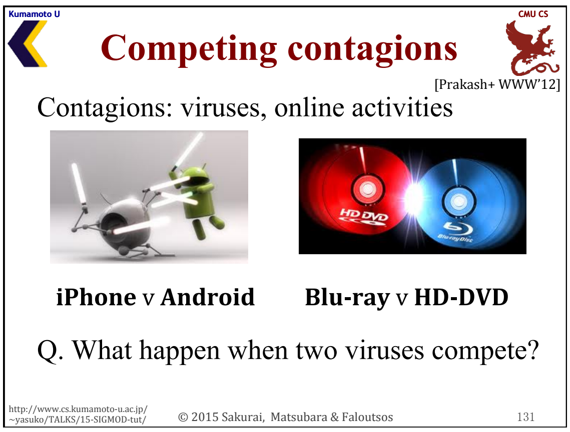
Kumamoto U
Part 2
Roadmap
Problem
Why: "non-linear" modeling
Fundamentals
Non-linear (grey-box) models
Applications
Epidemics
Information diffusion vs.
– Online competition
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
126

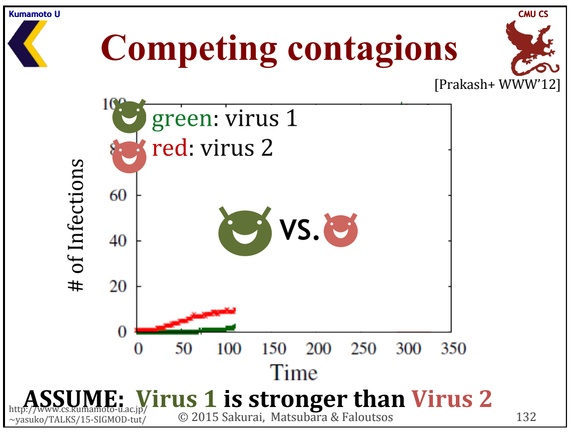

Kumamoto U
CMU CS
Online competition in social networks
flickr
delicious
Blogger
YouTube
twitter
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
127

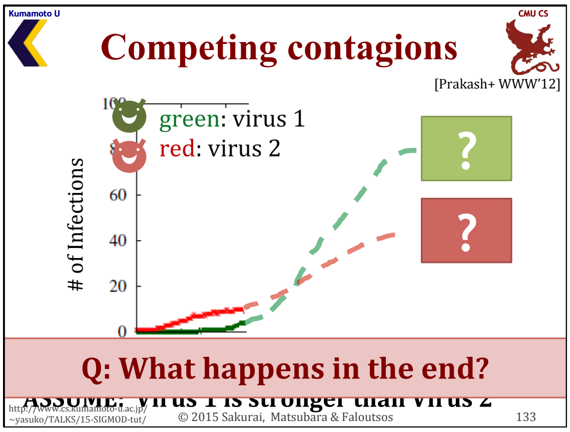

Kumamoto U
CMU CS
Online competition in social networks
flickr
delicious
Blogger
YouTube
twitter
amazon
VS.
BEST BUY
LOWE'S
Q. How can we describe "virtual competition"?
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
128

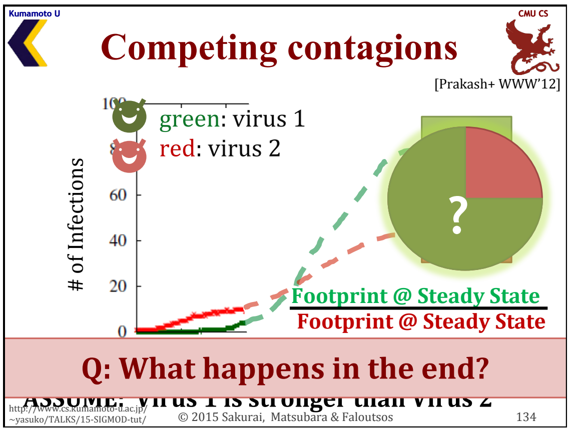

Kumamoto U
CMU CS
Online competition - roadmap
A. Non-linear (gray-box) modeling!
Solutions
- Winner-Takes-All [Prakash+ WWW'12]
- Co-existence of the two viruses [Beutel+ KDD'12]
- The Web as a Jungle [Matsubara+ WWW'15]
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
129

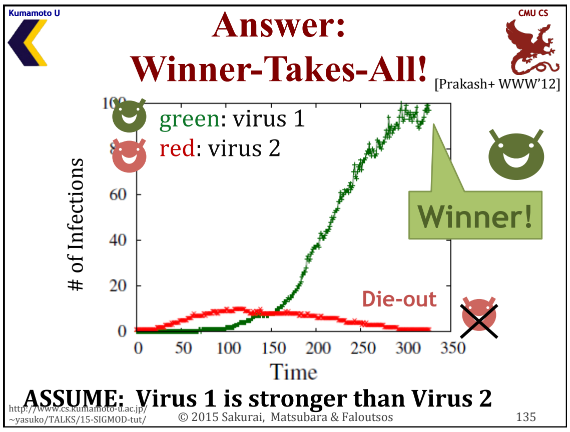

Kumamoto U
CMU CS
Online competition - roadmap
A. Non-linear (gray-box) modeling!
Solutions
- Winner-Takes-All [Prakash+ WWW'12]
- Co-existence of the two viruses [Beutel+ KDD'12]
- The Web as a Jungle [Matsubara+ WWW'15]
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
130

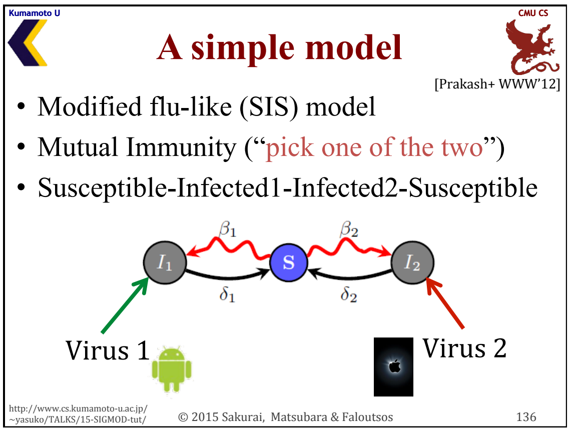

Kumamoto U
CMU CS
Competing contagions
[Prakash+ WWW'12]
Contagions: viruses, online activities
HD DVD
iPhone v Android
Blu-ray v HD-DVD
Q. What happen when two viruses compete?
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
131

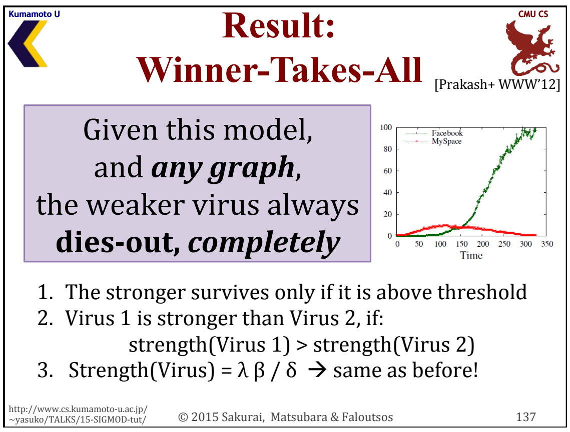

Kumamoto U
CMU CS
Competing contagions
[Prakash+ WWW'12]
green: virus 1
red: virus 2
VS.
# of Infections
0
20
40
60
0
50
100
150
200
250
300
350
Time
ASSUME: Virus 1 is stronger than Virus 2
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
132

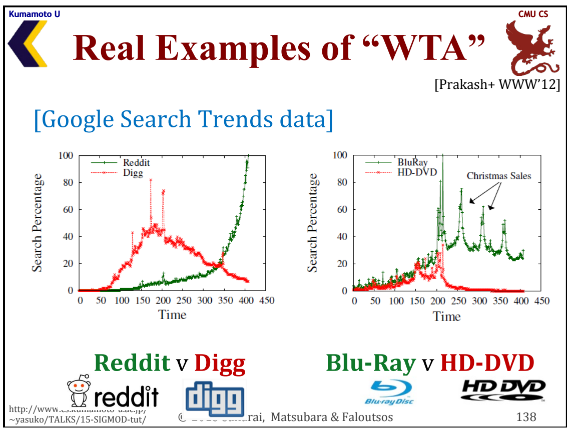

Competing contagions
[Prakash+ WWW'12]
green: virus 1
red: virus 2
# of Infections
?
?
Q: What happens in the end?
© 2015 Sakurai, Matsubara & Faloutsos
133


Competing contagions
[Prakash+ WWW'12]
green: virus 1
red: virus 2
# of Infections
?
Footprint @ Steady State
Footprint @ Steady State
Q: What happens in the end?
© 2015 Sakurai, Matsubara & Faloutsos
134


Answer:
Winner-Takes-All!
[Prakash+ WWW'12]
green: virus 1
red: virus 2
# of Infections
Winner!
Die-out
Time
ASSUME: Virus 1 is stronger than Virus 2
© 2015 Sakurai, Matsubara & Faloutsos
135


A simple model
[Prakash+ WWW'12]
• Modified flu-like (SIS) model
• Mutual Immunity ("pick one of the two")
• Susceptible-Infected1-Infected2-Susceptible
Virus 1
Virus 2
© 2015 Sakurai, Matsubara & Faloutsos
136

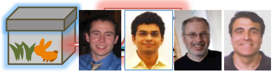

Result:
Winner-Takes-All
[Prakash+ WWW'12]
Given this model, and any graph, the weaker virus always dies-out, completely
1. The stronger survives only if it is above threshold
2. Virus 1 is stronger than Virus 2, if: strength(Virus 1) > strength(Virus 2)
3. Strength(Virus) = λ β / δ → same as before!
© 2015 Sakurai, Matsubara & Faloutsos
137


Real Examples of "WTA"
[Prakash+ WWW'12]
[Google Search Trends data]
Search Percentage
Time
Christmas Sales
Reddit v Digg
Blu-Ray v HD-DVD
138

Online competition in social networks

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

Solutions

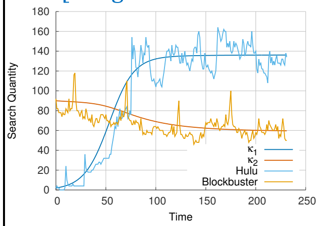


- Winner-Takes-All [Prakash+ WWW'12]
- **Co-existence of the two viruses [Beutel+ KDD'12]**
- The Web as a Jungle [Matsubara+ WWW'15]

Interacting Viruses: Can Both Survive?

Real example of "co-existence"

[Google Search Trends data]

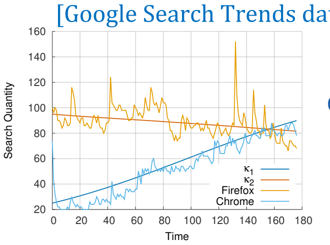


Hulu v Blockbuster

Interacting Viruses: Can Both Survive?

Real example of "co-existence"

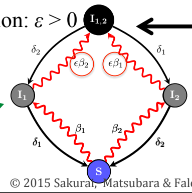
[Google Search Trends data]



Chrome v Firefox

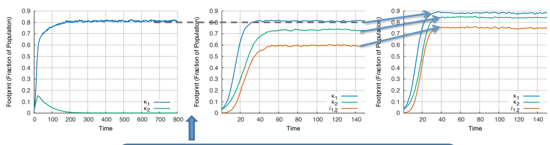
A simple model: $SI_{1|2}S$

- 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: $\varepsilon > 0$



Question: What happens in the end?

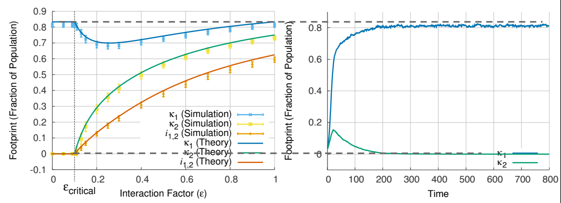
ε = 0	ε = 1	ε = 2
Winner takes all	Co-exist independently	Viruses cooperate



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

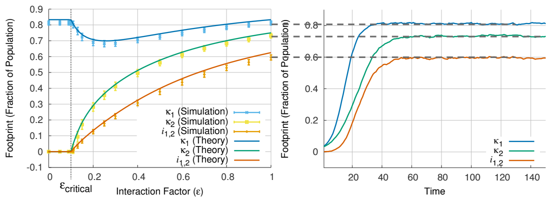
ASSUME: Virus 1 is stronger than Virus 2

Answer: Yes! There is a phase transition



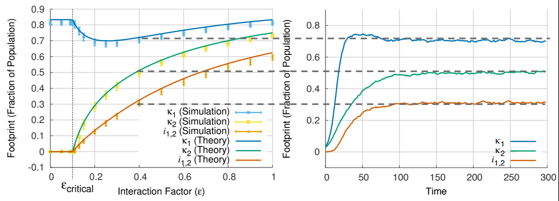
ASSUME: Virus 1 is stronger than Virus 2

Answer: Yes! There is a phase transition



ASSUME: Virus 1 is stronger than Virus 2

Answer: Yes! There is a phase transition



ASSUME: Virus 1 is stronger than Virus 2

Result: Viruses can Co-exist

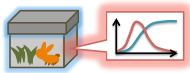
Given this model and a fully connected graph, there exists an $\varepsilon_{critical}$ such that for $\varepsilon \geq \varepsilon_{critical}$, there is a fixed point where both viruses survive.

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) $\sigma = N\beta / \delta$

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]

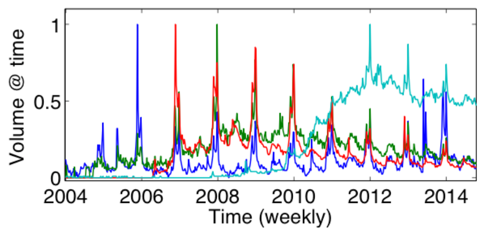
The Web as a Jungle: Non-Linear Dynamical Systems for Co-evolving Online Activities

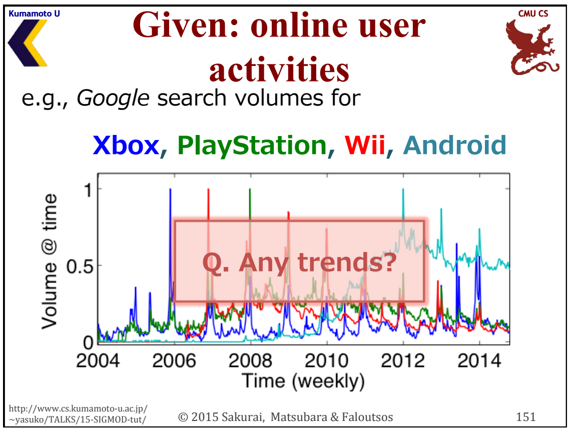
Yasuko Matsubara (Kumamoto University)
Yasushi Sakurai (Kumamoto University)
Christos Faloutsos (CMU)

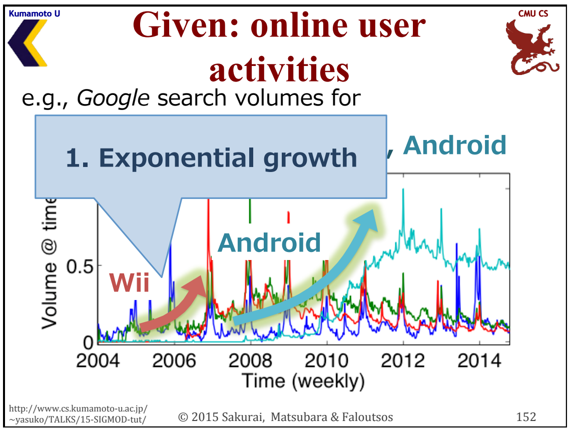
Given: online user activities

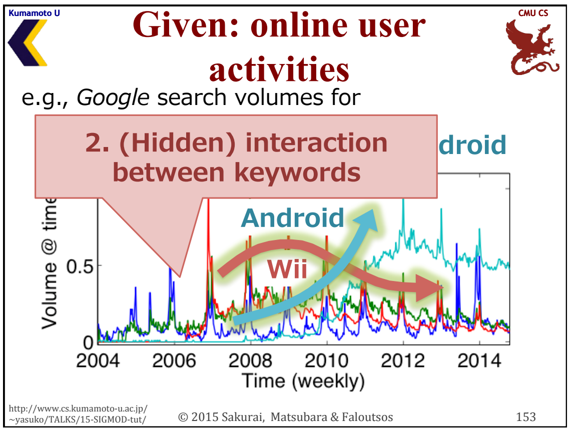
e.g., *Google* search volumes for

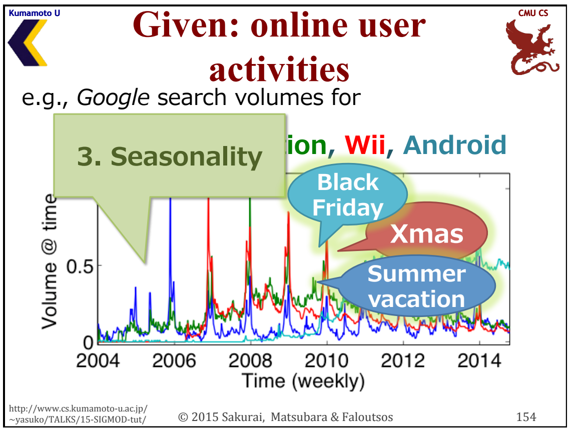
Xbox, PlayStation, Wii, Android

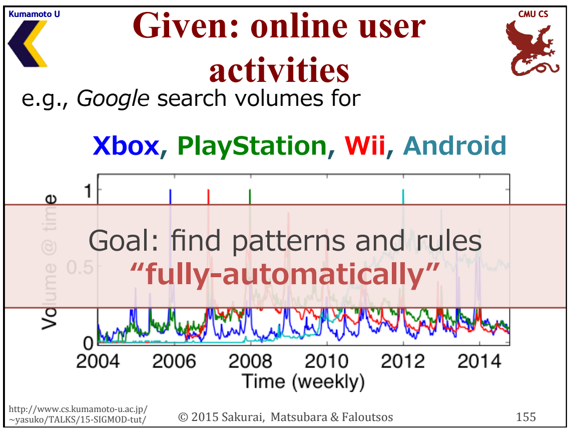


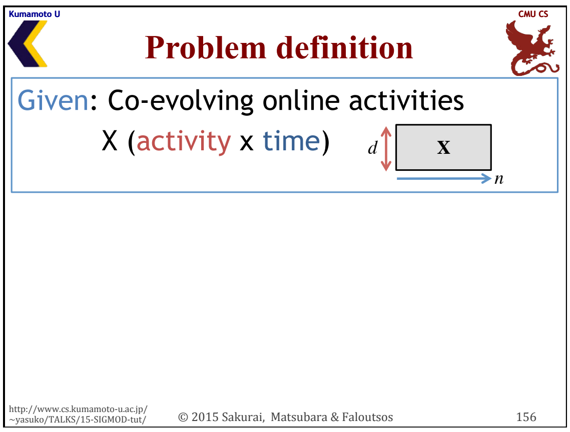

Kumamoto U
CMU CS
Given: online user activities
e.g., Google search volumes for
Xbox, PlayStation, Wii, Android
Q. Any trends?
Volume @ time
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
151

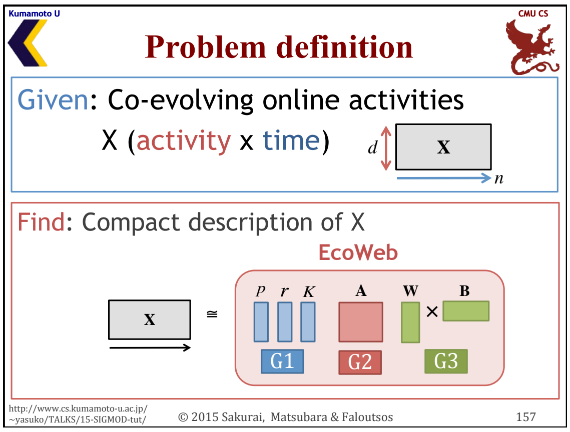

Kumamoto U
CMU CS
Given: online user activities
e.g., Google search volumes for
1. Exponential growth
, Android
Android
Wii
Volume @ time
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
152

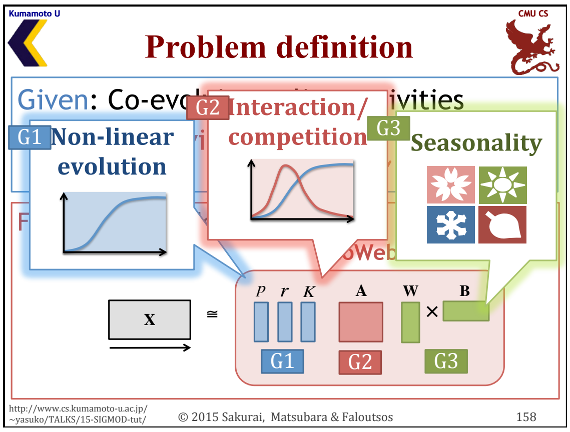

Kumamoto U
CMU CS
Given: online user activities
e.g., Google search volumes for
2. (Hidden) interaction between keywords
droid
Android
Wii
Volume @ time
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
153

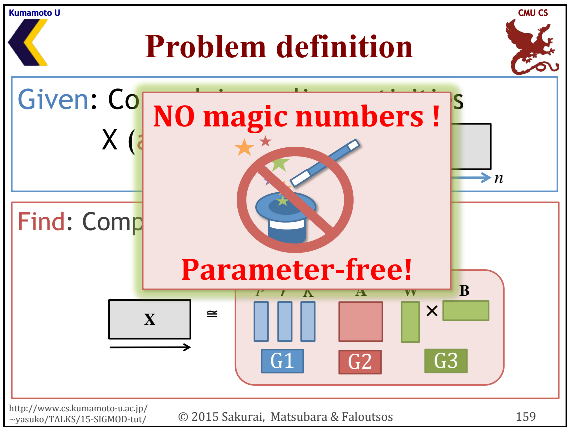

Kumamoto U
CMU CS
Given: online user activities
e.g., Google search volumes for
3. Seasonality
ion, Wii, Android
Black Friday
Xmas
Summer vacation
Volume @ time
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
154

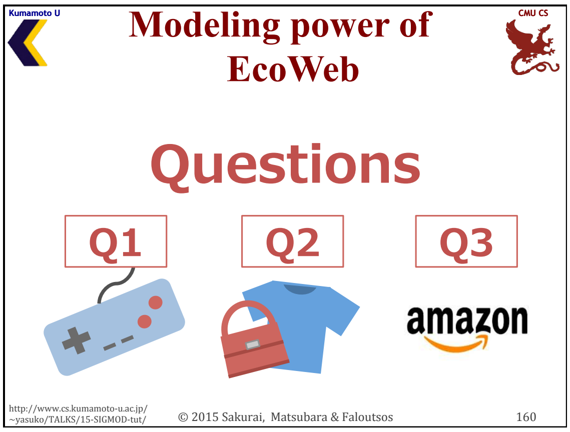

Kumamoto U
CMU CS
Given: online user activities
e.g., Google search volumes for
Xbox, PlayStation, Wii, Android
Goal: find patterns and rules
"fully-automatically"
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
155

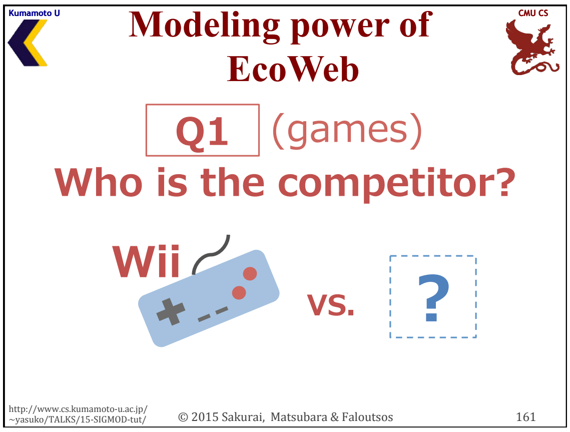

Kumamoto U
CMU CS
Problem definition
Given: Co-evolving online activities
X (activity x time)
d
X
n
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
156

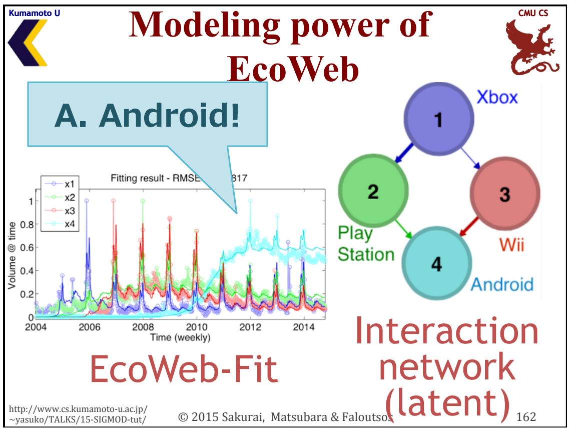

Kumamoto U
CMU CS
Problem definition
Given: Co-evolving online activities
X (activity x time)
d
X
n
Find: Compact description of X
EcoWeb
X
≅
p r K
A
W
B
×
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
157

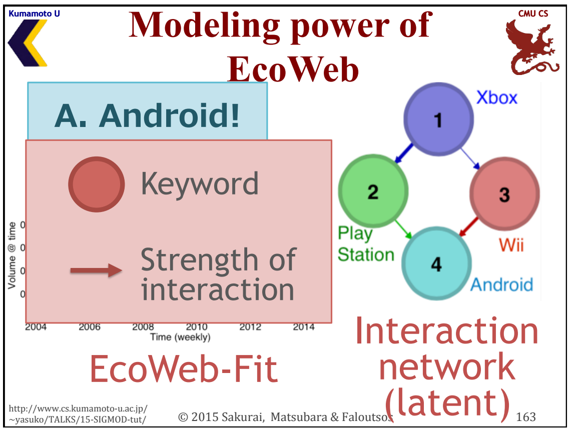

Kumamoto U
CMU CS
Problem definition
Given: Co-evo
ivities
G1 Non-linear evolution
G2 Interaction/ competition
G3 Seasonality
Web
X
≅
p r K
A
W
B
×
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
158

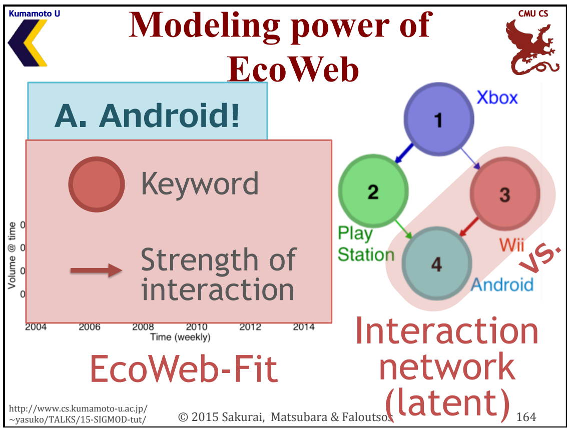

Kumamoto U
CMU CS
Problem definition
Given: Co
X (a
n
Find: Comp
NO magic numbers !
Parameter-free!
X
≅
B
×
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
159

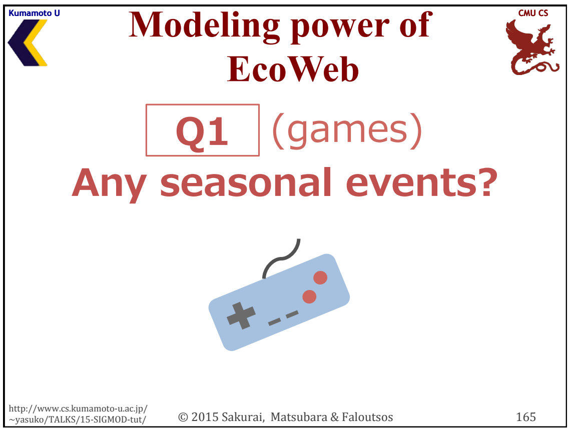

Kumamoto U
CMU CS
Modeling power of EcoWeb
Questions
Q1
Q2
Q3
amazon
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
160

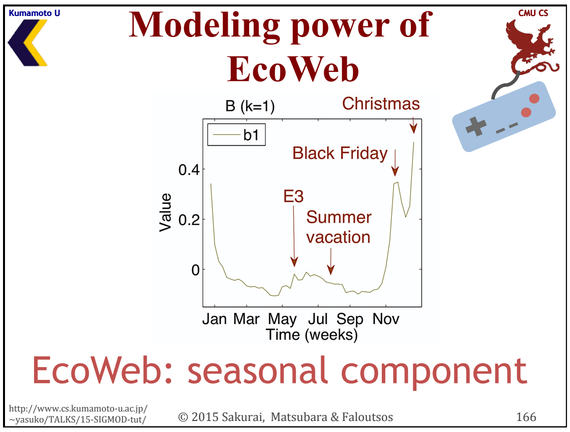

Kumamoto U
CMU CS
Modeling power of EcoWeb
Q1 (games)
Who is the competitor?
Wii
vs.
?
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
161

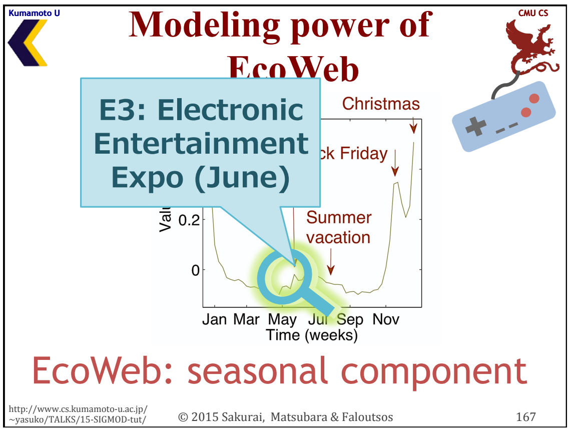

Kumamoto U
CMU CS
Modeling power of EcoWeb
A. Android!
Fitting result - RMSE
x1
x2
x3
x4
Volume @ time
Time (weekly)
EcoWeb-Fit
1 Xbox
2 Play Station
3 Wii
4 Android
Interaction network (latent)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
162

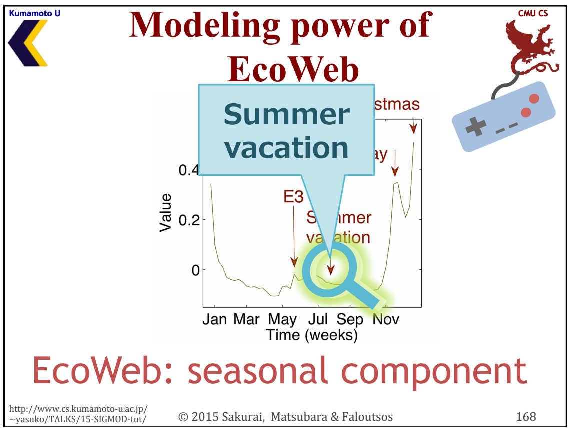

Kumamoto U
Modeling power of EcoWeb
CMU CS
A. Android!
Keyword
Strength of interaction
Xbox
Play Station
Wii
Android
EcoWeb-Fit
Interaction network (latent)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
163

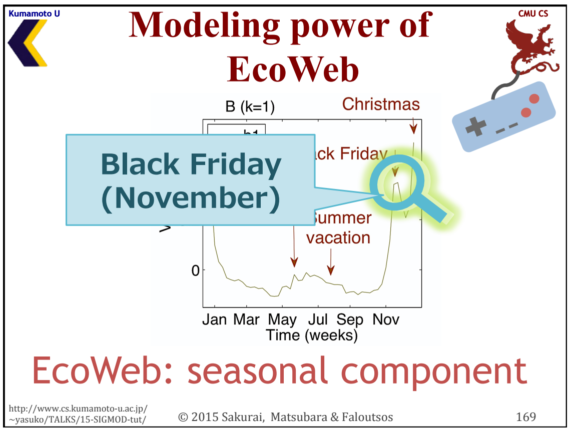

Kumamoto U
Modeling power of EcoWeb
CMU CS
A. Android!
Keyword
Strength of interaction
Xbox
Play Station
Wii
Android
VS.
EcoWeb-Fit
Interaction network (latent)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
164

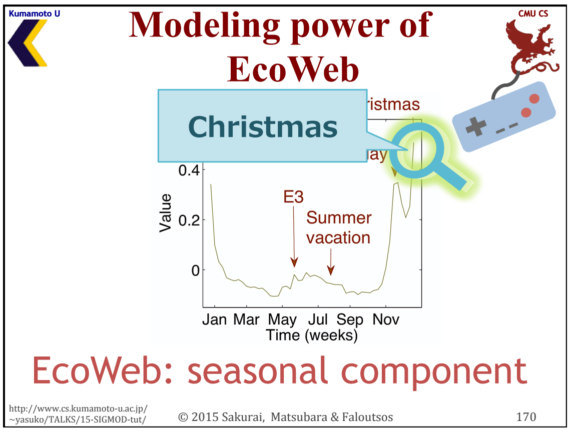

Kumamoto U
Modeling power of EcoWeb
CMU CS
Q1 (games)
Any seasonal events?
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
165

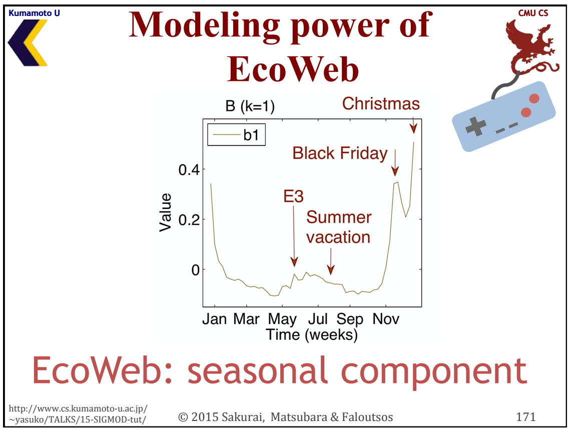

Kumamoto U
Modeling power of EcoWeb
CMU CS
B (k=1)
b1
Christmas
Black Friday
E3
Summer vacation
Value
Jan Mar May Jul Sep Nov
Time (weeks)
EcoWeb: seasonal component
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
166

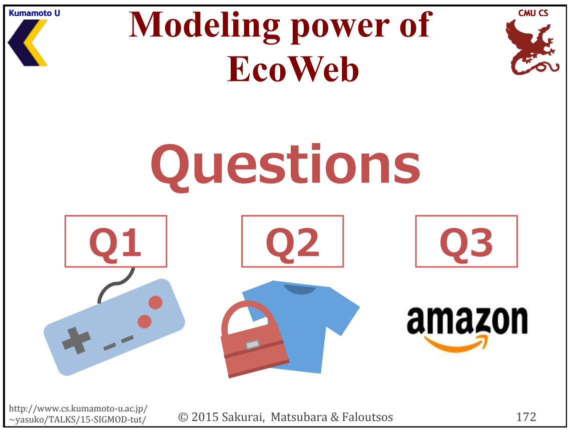

Kumamoto U
Modeling power of EcoWeb
CMU CS
E3: Electronic Entertainment Expo (June)
Christmas
Summer vacation
Jan Mar May Jul Sep Nov
Time (weeks)
EcoWeb: seasonal component
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
167

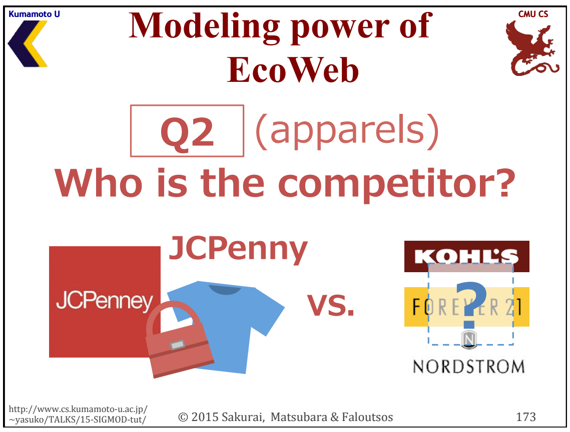

Kumamoto U
Modeling power of EcoWeb
CMU CS
Summer vacation
E3
Value
Jan Mar May Jul Sep Nov
Time (weeks)
EcoWeb: seasonal component
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
168

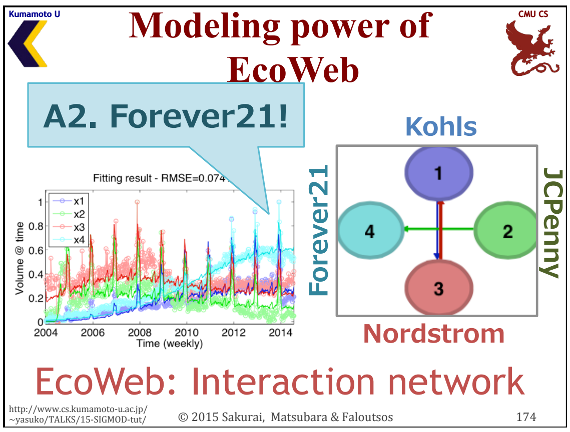

Modeling power of EcoWeb
Black Friday (November)
B (k=1)
Christmas
Summer vacation
Jan Mar May Jul Sep Nov
Time (weeks)
EcoWeb: seasonal component
© 2015 Sakurai, Matsubara & Faloutsos
169

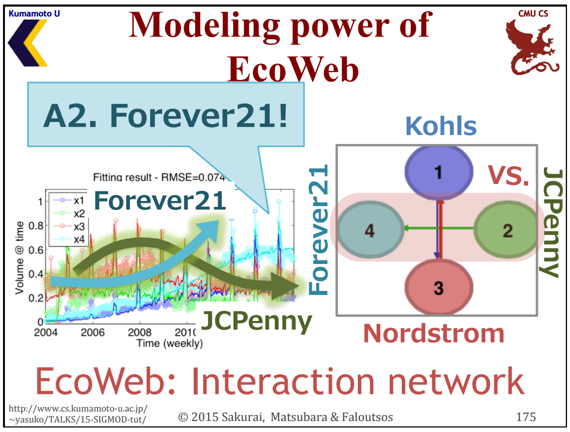

Modeling power of EcoWeb
Christmas
E3
Summer vacation
Value
Jan Mar May Jul Sep Nov
Time (weeks)
EcoWeb: seasonal component
© 2015 Sakurai, Matsubara & Faloutsos
170

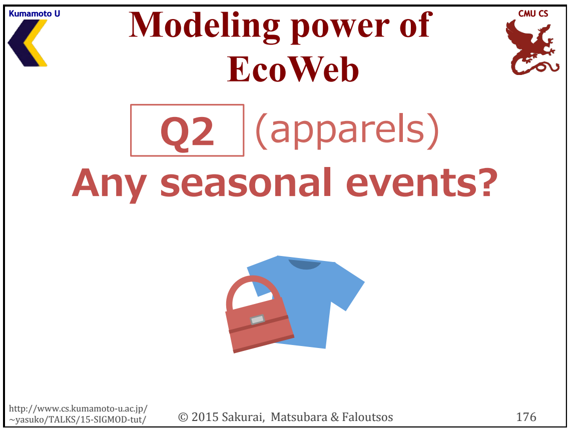

Modeling power of EcoWeb
B (k=1)
Christmas
b1
Black Friday
E3
Summer vacation
Value
Jan Mar May Jul Sep Nov
Time (weeks)
EcoWeb: seasonal component
© 2015 Sakurai, Matsubara & Faloutsos
171

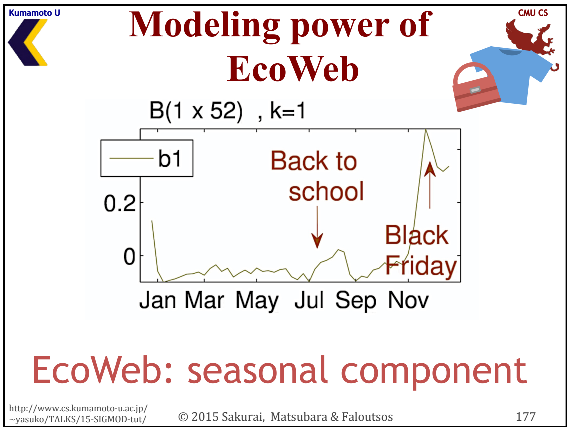

Modeling power of EcoWeb
Questions
Q1
Q2
Q3
amazon
© 2015 Sakurai, Matsubara & Faloutsos
172

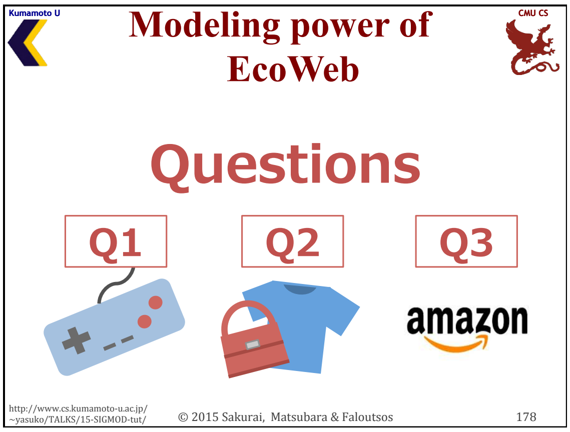

Modeling power of EcoWeb
Q2 (apparels)
Who is the competitor?
JCPenny
JCPenney
vs.
KOHL'S
NORDSTROM
© 2015 Sakurai, Matsubara & Faloutsos
173


Modeling power of EcoWeb
A2. Forever21!
Fitting result - RMSE=0.074
Volume @ time
Time (weekly)
Kohls
Forever21
JCPenny
Nordstrom
EcoWeb: Interaction network
© 2015 Sakurai, Matsubara & Faloutsos
174

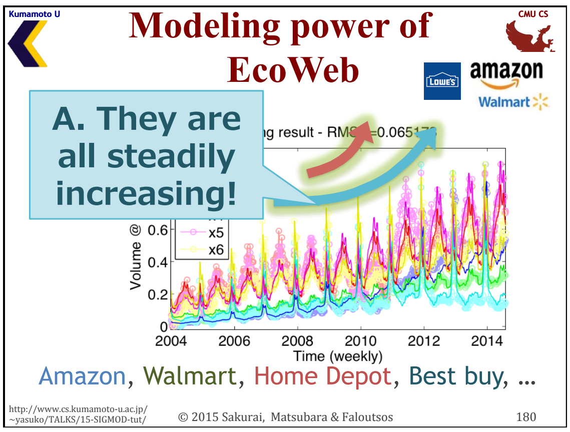

Kumamoto U
CMU CS
Modeling power of EcoWeb
A2. Forever21!
Fitting result - RMSE=0.074
Forever21
Kohls
vs.
Forever21
JCPenny
JCPenny
Nordstrom
Volume @ time
Time (weekly)
EcoWeb: Interaction network
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
175

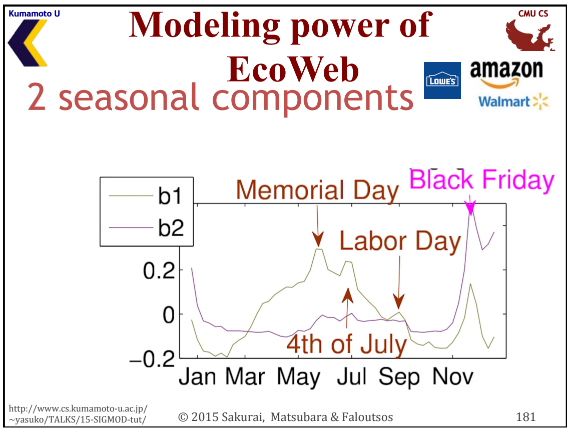

Kumamoto U
CMU CS
Modeling power of EcoWeb
Q2 (apparels)
Any seasonal events?
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
176

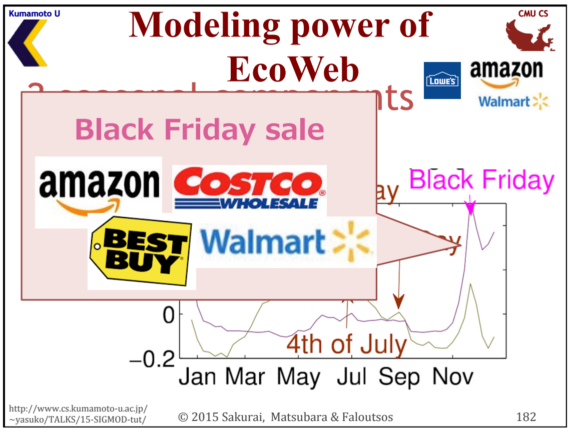

Kumamoto U
CMU CS
Modeling power of EcoWeb
B(1 x 52) , k=1
b1
Back to school
Black Friday
Jan Mar May Jul Sep Nov
EcoWeb: seasonal component
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
177

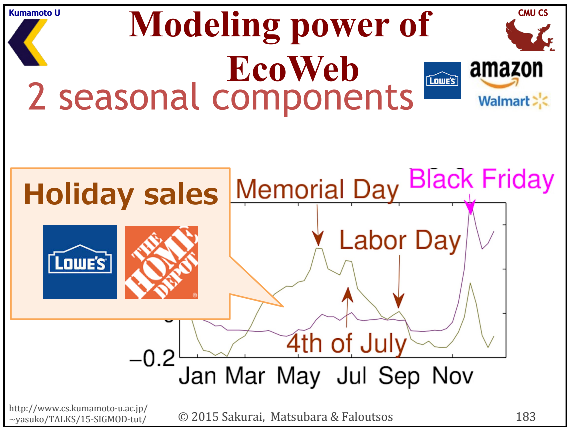

Kumamoto U
CMU CS
Modeling power of EcoWeb
Questions
Q1
Q2
Q3
amazon
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
178

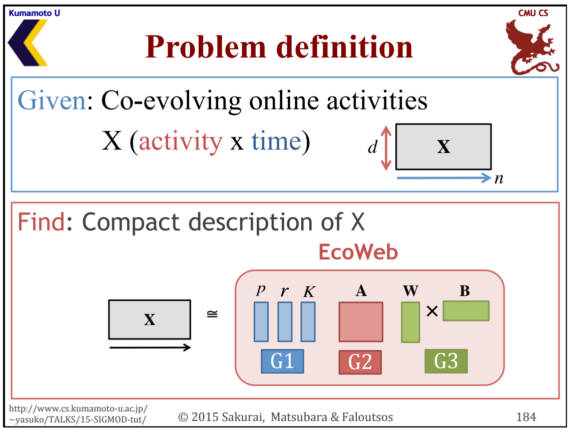

Kumamoto U
CMU CS
Modeling power of EcoWeb
Q3 (retails)
Any patterns/trends?
amazon
LOWE'S
COSTCO WHOLESALE
BEST BUY
Walmart
THE HOME DEPOT
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
179

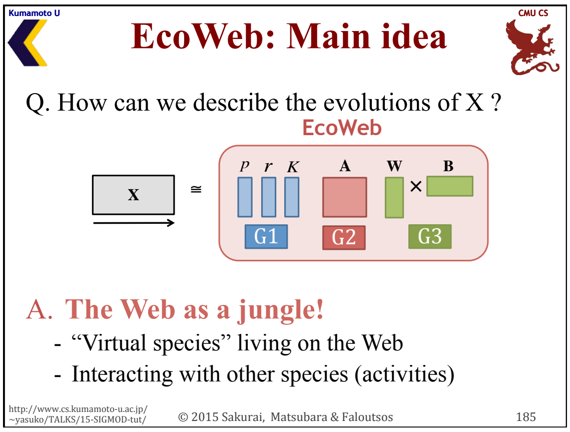

Kumamoto U
CMU CS
Modeling power of EcoWeb
LOWE'S
amazon
Walmart
A. They are all steadily increasing!
Volume @
Time (weekly)
Amazon, Walmart, Home Depot, Best buy, ...
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
180

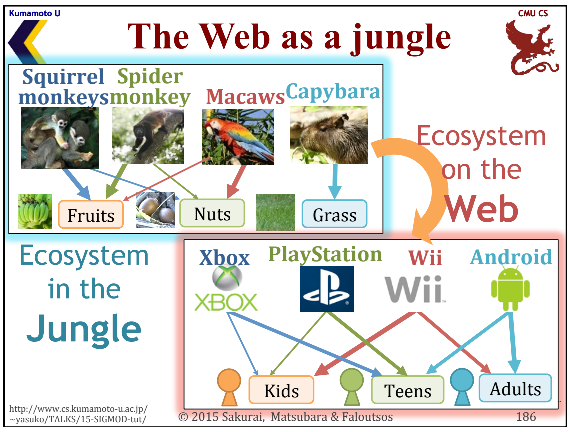

Kumamoto U
CMU CS
Modeling power of EcoWeb
2 seasonal components
amazon
Walmart
b1
b2
Memorial Day
Black Friday
Labor Day
4th of July
0.2
0
−0.2
Jan Mar May Jul Sep Nov
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
181

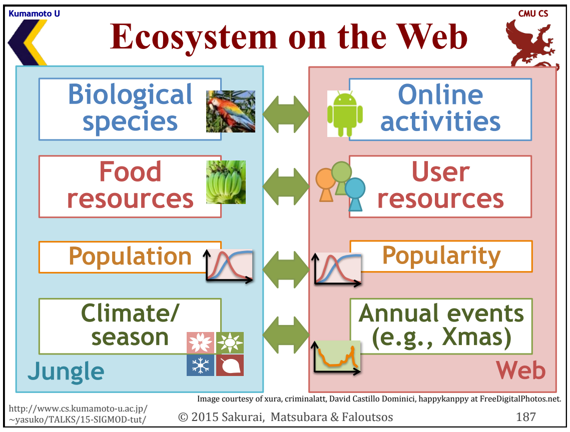

Kumamoto U
CMU CS
Modeling power of EcoWeb
amazon
Walmart
Black Friday sale
amazon
COSTCO WHOLESALE
BEST BUY
Walmart
Black Friday
0
−0.2
4th of July
Jan Mar May Jul Sep Nov
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
182


Kumamoto U
CMU CS
Modeling power of EcoWeb
2 seasonal components
amazon
Walmart
Holiday sales
LOWE'S
THE HOME DEPOT
Memorial Day
Black Friday
Labor Day
4th of July
−0.2
Jan Mar May Jul Sep Nov
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
183

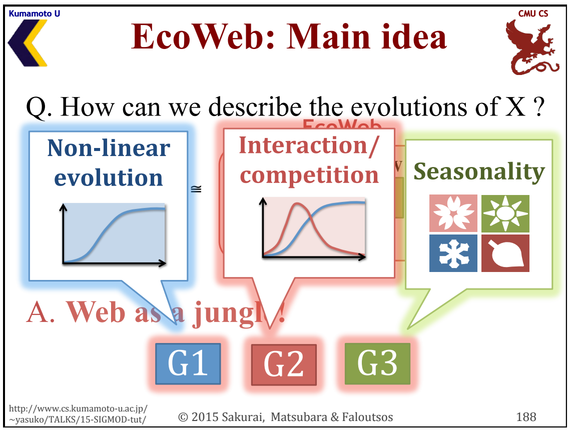

Kumamoto U
CMU CS
Problem definition
Given: Co-evolving online activities
X (activity x time)
d
X
n
Find: Compact description of X
EcoWeb
X
≅
P r K
A
W
B
×
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
184


Kumamoto U
CMU CS
EcoWeb: Main idea
Q. How can we describe the evolutions of X ?
EcoWeb
X
≅
P r K
A
W
B
×
G1
G2
G3
A. The Web as a jungle!
- "Virtual species" living on the Web
- Interacting with other species (activities)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
185

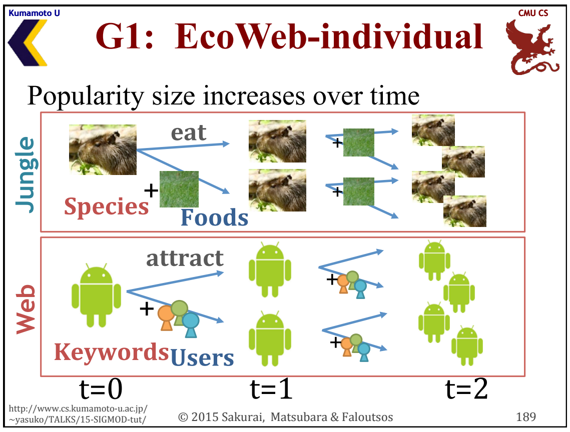

Kumamoto U
CMU CS
The Web as a jungle
Squirrel monkeys
Spider monkey
Macaws
Capybara
Fruits
Nuts
Grass
Ecosystem on the Web
Ecosystem in the Jungle
Xbox
XBOX
PlayStation
Wii
Android
Kids
Teens
Adults
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
186



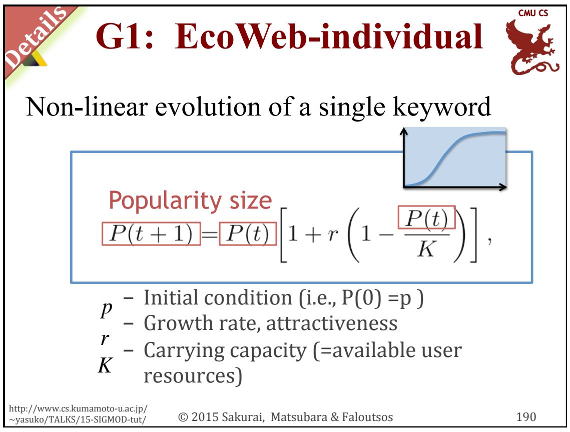
Kumamoto U
CMU CS
Ecosystem on the Web
Biological species
Online activities
Food resources
User resources
Population
Popularity
Climate/ season
Annual events (e.g., Xmas)
Jungle
Web
Image courtesy of xura, criminalatt, David Castillo Dominici, happykanppy at FreeDigitalPhotos.net.
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
187



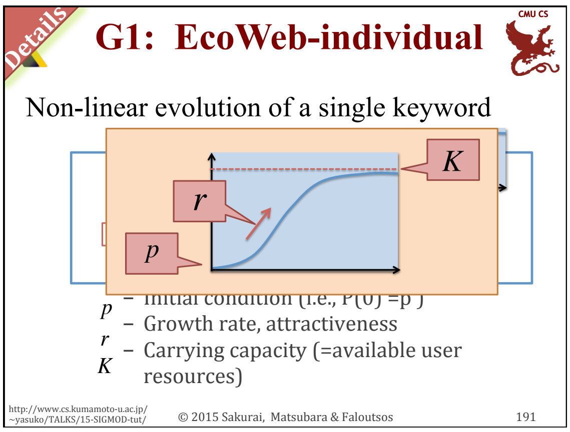
Kumamoto U
CMU CS
EcoWeb: Main idea
Q. How can we describe the evolutions of X ?
Non-linear evolution
Interaction/ competition
Seasonality
A. Web as a jungle!
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
188



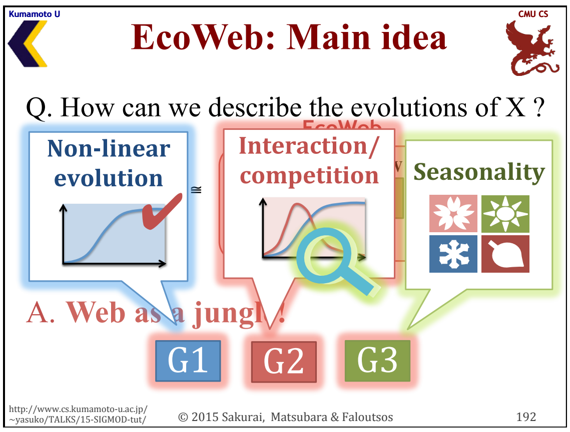
Kumamoto U
CMU CS
G1: EcoWeb-individual
Popularity size increases over time
Jungle
eat
Species
Foods
Web
attract
Keywords
Users
t=0
t=1
t=2
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
189



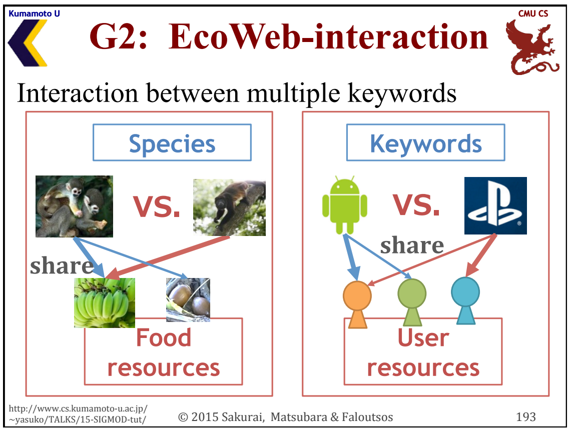
Details
CMU CS
G1: EcoWeb-individual
Non-linear evolution of a single keyword
Popularity size
$P(t+1) = P(t)\left[1 + r\left(1 - \frac{P(t)}{K}\right)\right],$
p – Initial condition (i.e., P(0) =p)
r – Growth rate, attractiveness
K – Carrying capacity (=available user resources)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
190

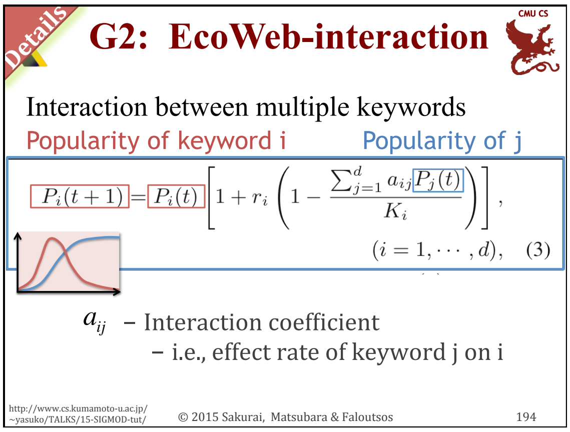


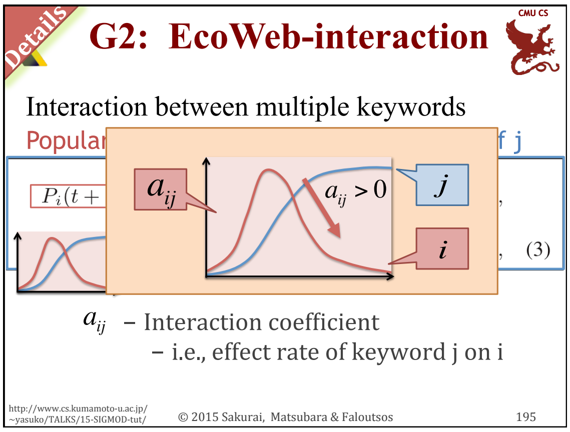
Details
CMU CS
G1: EcoWeb-individual
Non-linear evolution of a single keyword
K
r
p
p – Initial condition (i.e., P(0) =p)
r – Growth rate, attractiveness
K – Carrying capacity (=available user resources)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
191

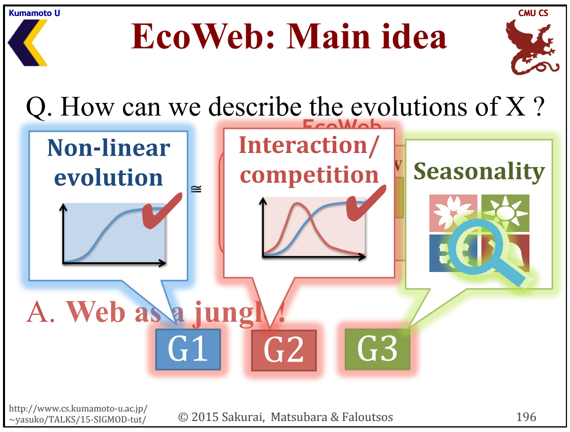


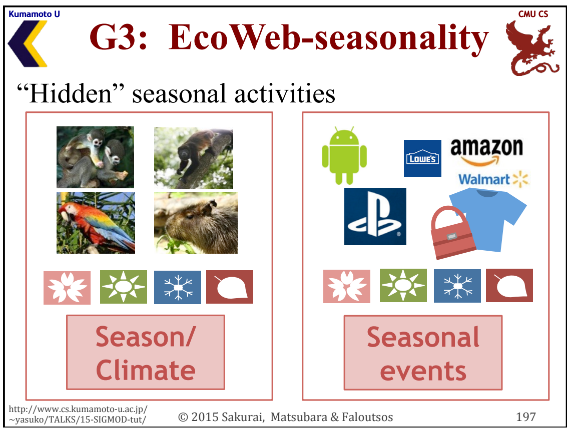
Kumamoto U
CMU CS
EcoWeb: Main idea
Q. How can we describe the evolutions of X ?
Non-linear evolution
Interaction/ competition
Seasonality
A. Web as a jungle!
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
192

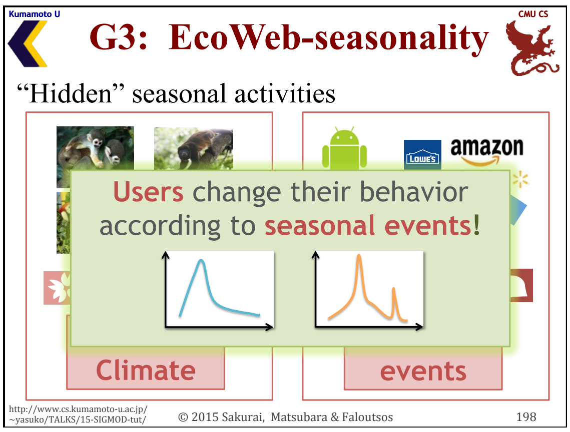

Kumamoto U
CMU CS
G2: EcoWeb-interaction
Interaction between multiple keywords
Species
vs.
share
Food resources
Keywords
vs.
share
User resources
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
193

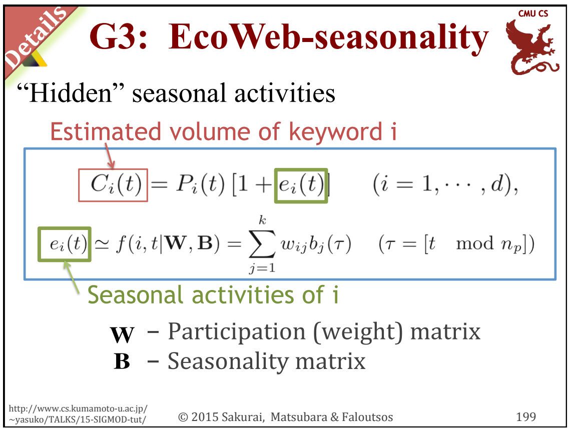

Details
CMU CS
G2: EcoWeb-interaction
Interaction between multiple keywords
Popularity of keyword i
Popularity of j
$P_i(t+1) = P_i(t)\left[1 + r_i\left(1 - \frac{\sum_{j=1}^{d} a_{ij} P_j(t)}{K_i}\right)\right],$
$(i = 1, \cdots, d),$ (3)
a_{ij} – Interaction coefficient
– i.e., effect rate of keyword j on i
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
194

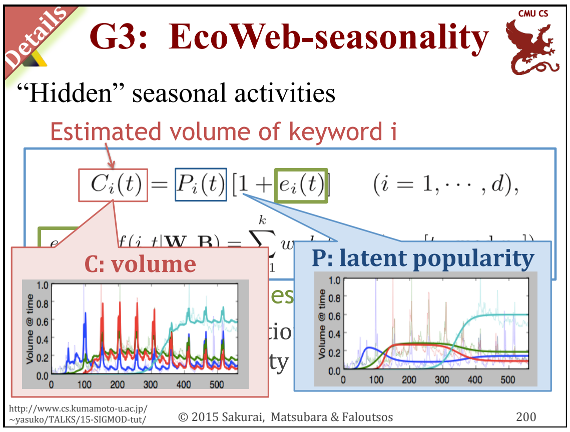

Details
CMU CS
G2: EcoWeb-interaction
Interaction between multiple keywords
Popular
f j
$P_i(t+$
a_{ij}
$a_{ij} > 0$
j
i
(3)
a_{ij} – Interaction coefficient
– i.e., effect rate of keyword j on i
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
195

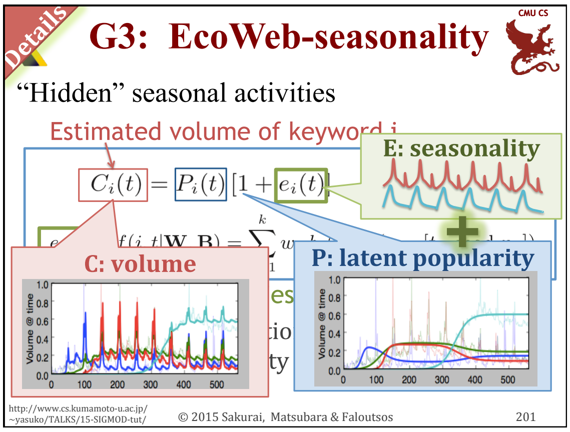

Kumamoto U
CMU CS
EcoWeb: Main idea
Q. How can we describe the evolutions of X ?
Non-linear evolution
Interaction/ competition
Seasonality
A. Web as a jungle!
G1
G2
G3
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
196

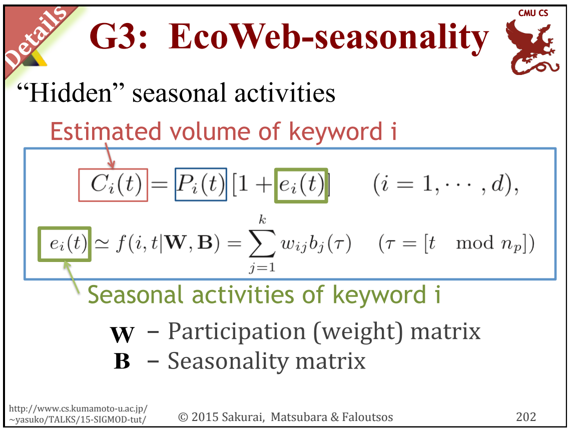

Kumamoto U
CMU CS
G3: EcoWeb-seasonality
"Hidden" seasonal activities
amazon
Walmart
Season/ Climate
Seasonal events
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
197

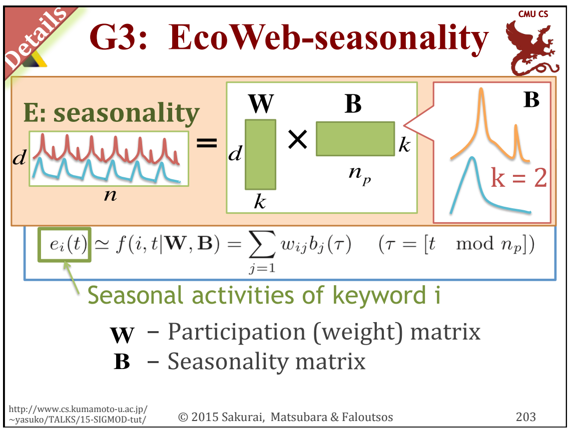

Kumamoto U
CMU CS
G3: EcoWeb-seasonality
"Hidden" seasonal activities
amazon
Users change their behavior according to seasonal events!
Climate
events
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
198

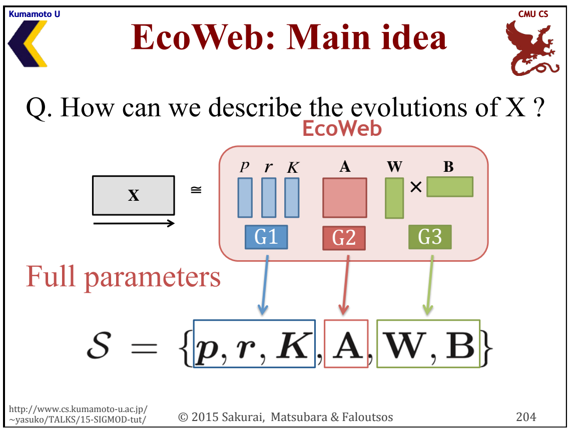

Details
G3: EcoWeb-seasonality
CMU CS
"Hidden" seasonal activities
Estimated volume of keyword i
$C_i(t) = P_i(t)\,[1 + e_i(t)]$ $(i = 1, \cdots, d)$,
$e_i(t) \simeq f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{j=1}^{k} w_{ij} b_j(\tau)$ $(\tau = [t \mod n_p])$
Seasonal activities of i
W – Participation (weight) matrix
B – Seasonality matrix
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
199

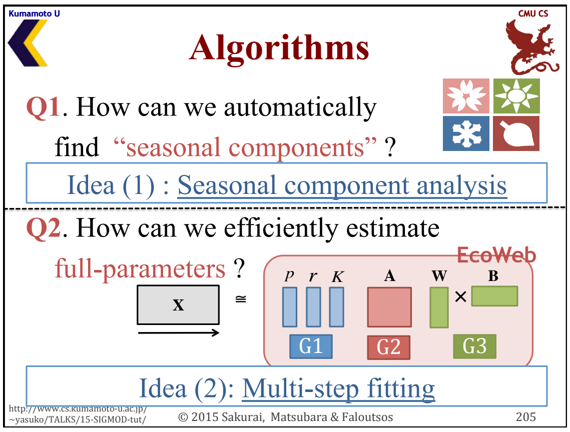

Details
G3: EcoWeb-seasonality
CMU CS
"Hidden" seasonal activities
Estimated volume of keyword i
$C_i(t) = P_i(t)\,[1 + e_i(t)]$ $(i = 1, \cdots, d)$,
C: volume
P: latent popularity
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
200

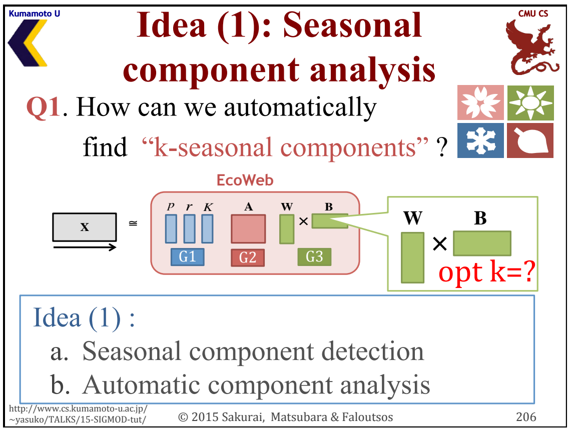

Details
G3: EcoWeb-seasonality
CMU CS
"Hidden" seasonal activities
Estimated volume of keyword i
$C_i(t) = P_i(t)\,[1 + e_i(t)]$
E: seasonality
C: volume
P: latent popularity
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
201

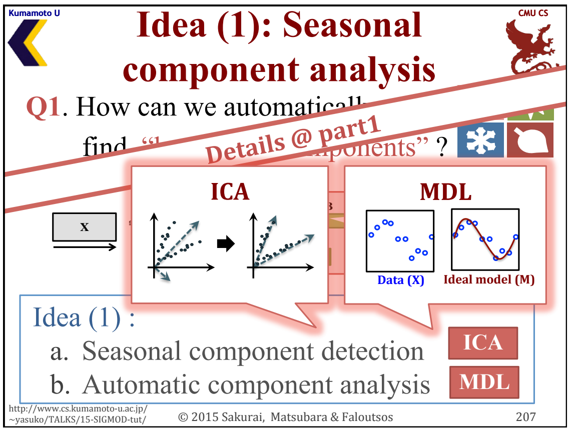

Details
G3: EcoWeb-seasonality
CMU CS
"Hidden" seasonal activities
Estimated volume of keyword i
$C_i(t) = P_i(t)\,[1 + e_i(t)]$ $(i = 1, \cdots, d)$,
$e_i(t) \simeq f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{j=1}^{k} w_{ij} b_j(\tau)$ $(\tau = [t \mod n_p])$
Seasonal activities of keyword i
W – Participation (weight) matrix
B – Seasonality matrix
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
202

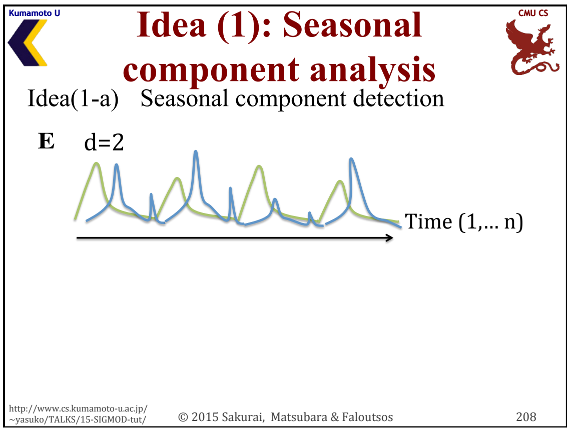

Details
G3: EcoWeb-seasonality
CMU CS
E: seasonality
d
n
=
W
d
k
×
B
k
n_p
B
k = 2
$e_i(t) \simeq f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{j=1}^{k} w_{ij} b_j(\tau)$ $(\tau = [t \mod n_p])$
Seasonal activities of keyword i
W – Participation (weight) matrix
B – Seasonality matrix
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
203

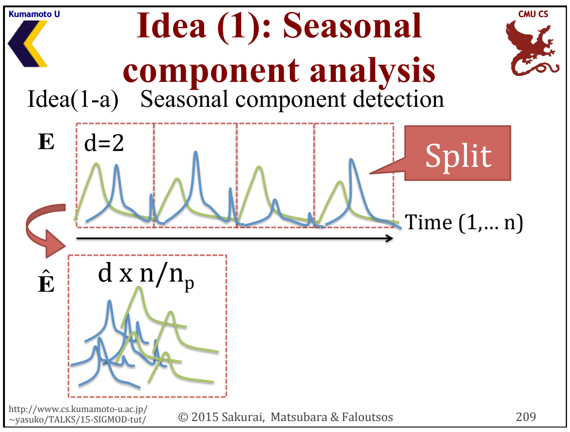

Kumamoto U
EcoWeb: Main idea
CMU CS
Q. How can we describe the evolutions of X ?
EcoWeb
X
≅
p r K
A
W
×
B
G1
G2
G3
Full parameters
$\mathcal{S} = \{\mathbf{p}, \mathbf{r}, \mathbf{K}, \mathbf{A}, \mathbf{W}, \mathbf{B}\}$
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
204

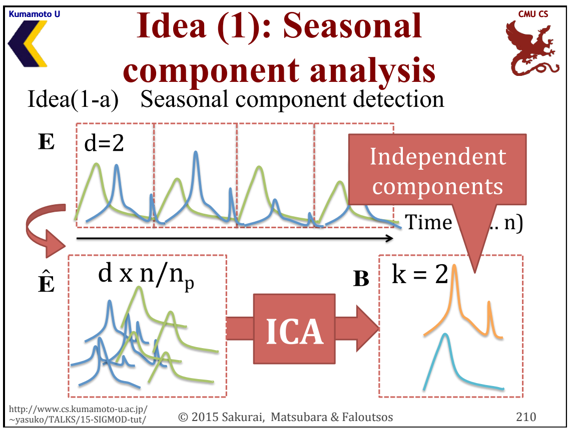

Kumamoto U
CMU CS
Algorithms
Q1. How can we automatically find "seasonal components" ?
Idea (1) : Seasonal component analysis
Q2. How can we efficiently estimate full-parameters ?
EcoWeb
X
p r K
A
W
B
G1
G2
G3
Idea (2): Multi-step fitting
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
205


Kumamoto U
CMU CS
Idea (1): Seasonal component analysis
Q1. How can we automatically find "k-seasonal components" ?
EcoWeb
X
p r K
A
W
B
G1
G2
G3
W
B
opt k=?
Idea (1) :
a. Seasonal component detection
b. Automatic component analysis
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
206


Kumamoto U
CMU CS
Idea (1): Seasonal component analysis
Q1. How can we automaticall
Details @ part1
ICA
MDL
Data (X)
Ideal model (M)
Idea (1) :
a. Seasonal component detection
b. Automatic component analysis
ICA
MDL
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
207


Kumamoto U
CMU CS
Idea (1): Seasonal component analysis
Idea(1-a) Seasonal component detection
E
d=2
Time (1,... n)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
208


Kumamoto U
CMU CS
Idea (1): Seasonal component analysis
Idea(1-a) Seasonal component detection
E
d=2
Split
Time (1,... n)
Ê
d x n/n_p
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
209

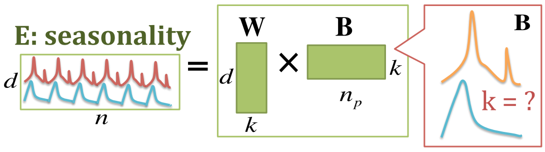

Kumamoto U
CMU CS
Idea (1): Seasonal component analysis
Idea(1-a) Seasonal component detection
E
d=2
Independent components
Time
. n)
Ê
d x n/n_p
ICA
B
k = 2
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
210

Idea (1): Seasonal component analysis

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

d: dimension



opt k=?

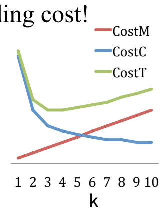
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 211

Idea (1): Seasonal component analysis

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

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

Model cost   Coding cost



Good compression ⟲ Good description

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 212

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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 213

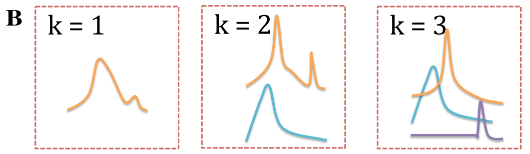
Idea (1): Seasonal component analysis

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

d: dimension

W × B opt k=?



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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 214

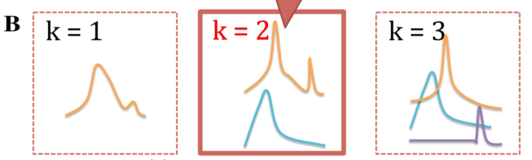
Idea (1): Seasonal component analysis

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

W × B opt k=?

Optimal k

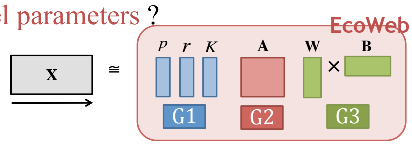


Cost(1) = $$ Cost(2) = $ Cost(3) = $$$

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 215

Idea (2): EcoWeb-Fit

Q2. How can we efficiently estimate model parameters ?



Idea (2): Multi-step fitting

a. **StepFit** (sub)

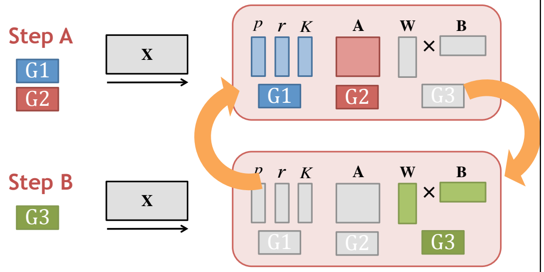
b. **EcoWeb-Fit** (full)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 216

Kumamoto U | CMU CS

Idea (2): EcoWeb-Fit

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

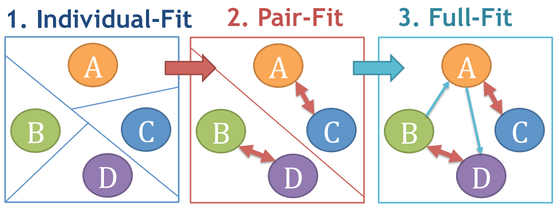


Kumamoto U | CMU CS

Idea (2): EcoWeb-Fit

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

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



EcoWeb-Fit updates parameters, separately

Kumamoto U | CMU CS

Experiments

We answer the following questions…

Q1. Effectiveness
How successful is it in spotting patterns?

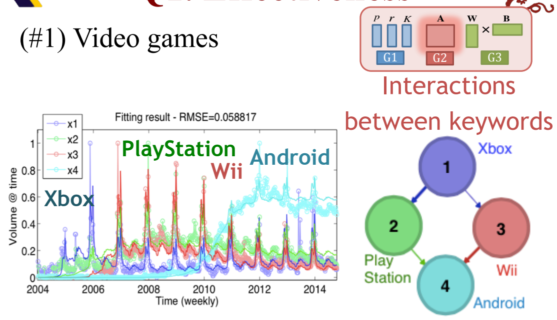
Q2. Accuracy
How well does it match the data?

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

Kumamoto U | CMU CS

Q1. Effectiveness

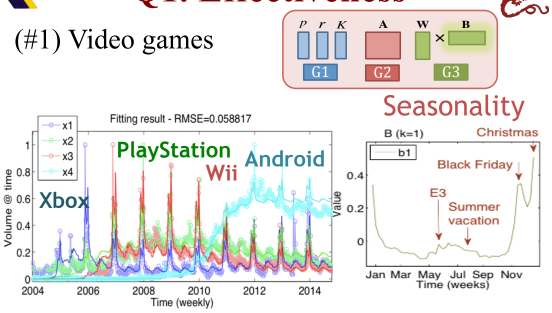
(#1) Video games



Kumamoto U | CMU CS

Q1. Effectiveness

(#1) Video games

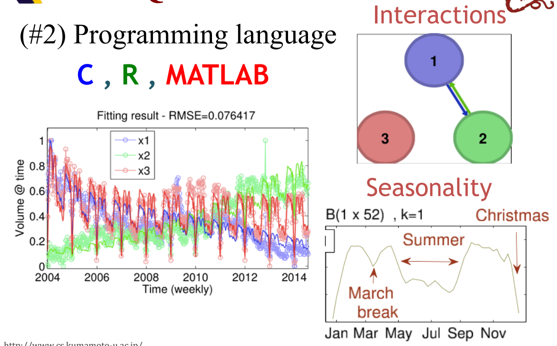


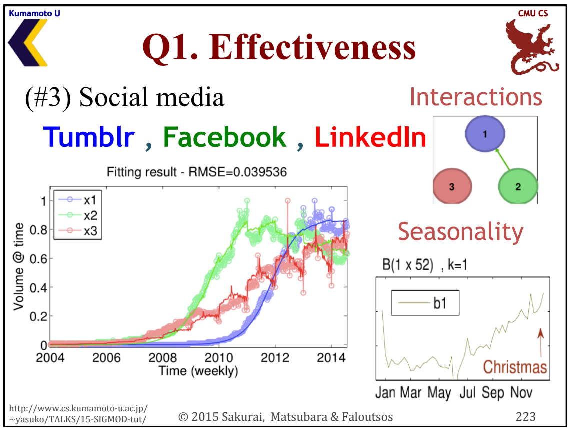
Kumamoto U | CMU CS

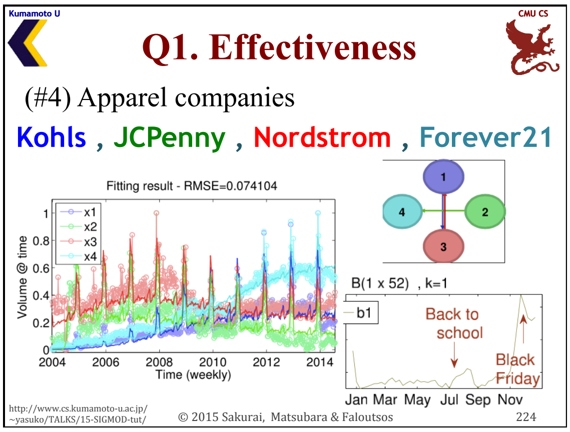
Q1. Effectiveness

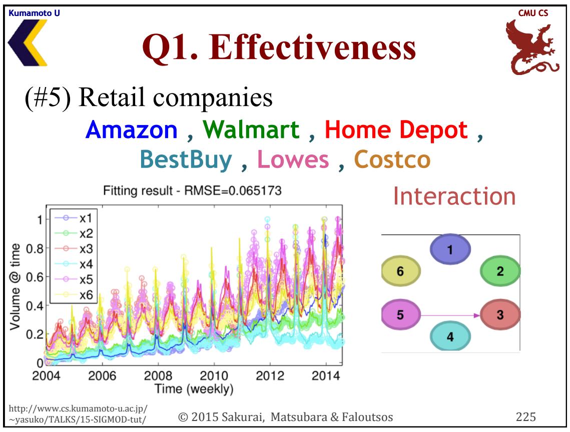
(#2) Programming language

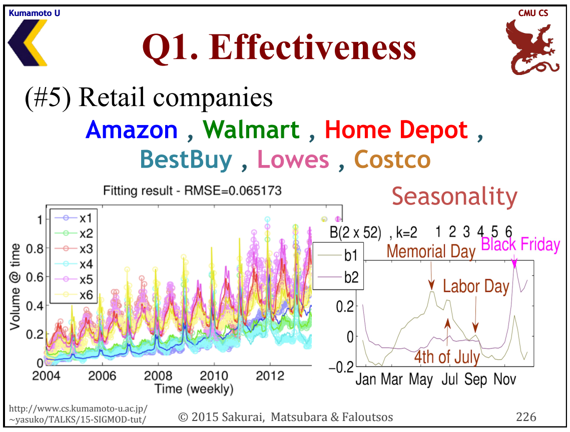
C , R , MATLAB

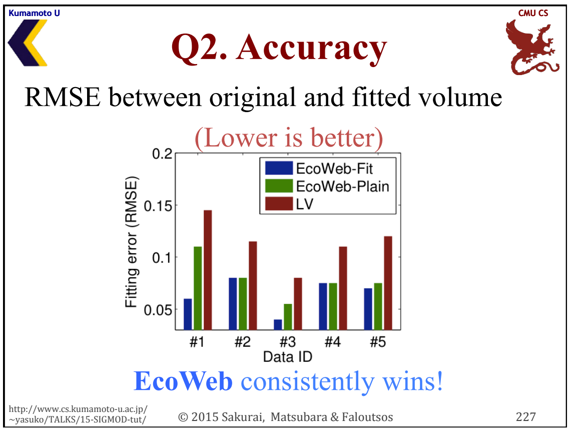


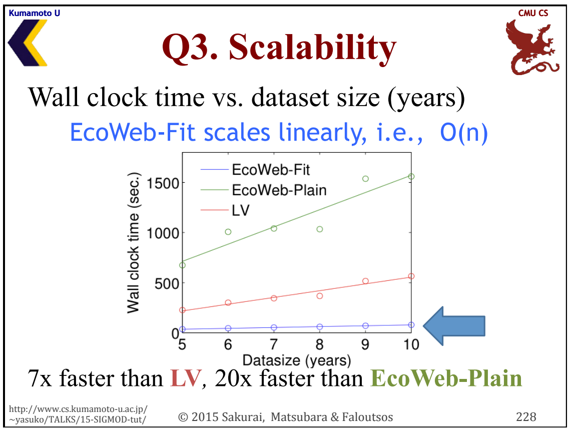

Kumamoto U
CMU CS
Q1. Effectiveness
(#3) Social media
Tumblr , Facebook , LinkedIn
Interactions
Fitting result - RMSE=0.039536
Volume @ time
Time (weekly)
Seasonality
B(1 x 52) , k=1
Christmas
Jan Mar May Jul Sep Nov
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
223


Kumamoto U
CMU CS
Q1. Effectiveness
(#4) Apparel companies
Kohls , JCPenny , Nordstrom , Forever21
Fitting result - RMSE=0.074104
Volume @ time
Time (weekly)
B(1 x 52) , k=1
Back to school
Black Friday
Jan Mar May Jul Sep Nov
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
224


Kumamoto U
CMU CS
Q1. Effectiveness
(#5) Retail companies
Amazon , Walmart , Home Depot ,
BestBuy , Lowes , Costco
Fitting result - RMSE=0.065173
Interaction
Volume @ time
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
225


Kumamoto U
CMU CS
Q1. Effectiveness
(#5) Retail companies
Amazon , Walmart , Home Depot ,
BestBuy , Lowes , Costco
Fitting result - RMSE=0.065173
Seasonality
B(2 x 52) , k=2
Memorial Day
Black Friday
Labor Day
4th of July
Jan Mar May Jul Sep Nov
Volume @ time
Time (weekly)
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
226


Kumamoto U
CMU CS
Q2. Accuracy
RMSE between original and fitted volume
(Lower is better)
EcoWeb-Fit
EcoWeb-Plain
LV
Fitting error (RMSE)
Data ID
EcoWeb consistently wins!
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
227

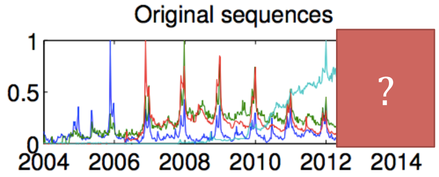

Kumamoto U
CMU CS
Q3. Scalability
Wall clock time vs. dataset size (years)
EcoWeb-Fit scales linearly, i.e., O(n)
EcoWeb-Fit
EcoWeb-Plain
LV
Wall clock time (sec.)
Datasize (years)
7x faster than LV, 20x faster than EcoWeb-Plain
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
228

EcoWeb at work - forecasting

Forecasting future activities

Train: 2/3 sequences

Forecast: 1/3 following years

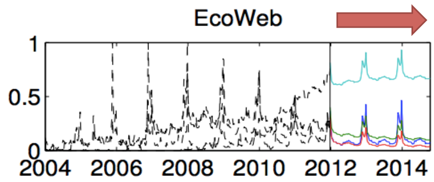


EcoWeb at work - forecasting

Forecasting future activities

Train: 2/3 sequences

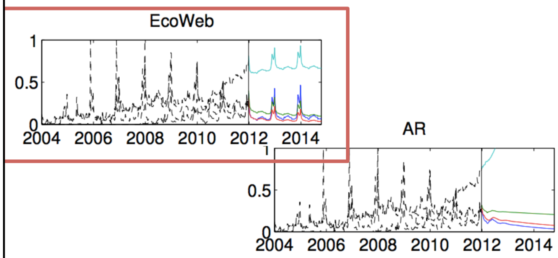
Forecast: 1/3 following years



EcoWeb can capture future patterns

EcoWeb at work - forecasting

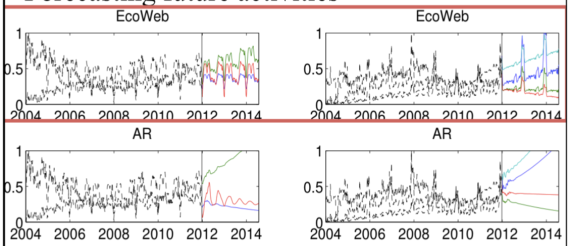
Forecasting future activities



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!

Part 2 Roadmap

Problem

✔ Why: "non-linear" modeling

Fundamentals

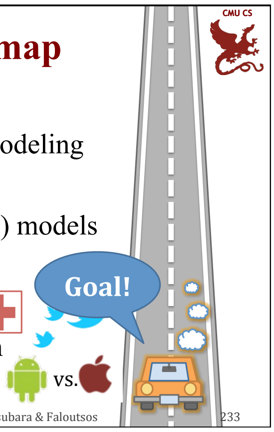
✔ Non-linear (grey-box) models

Applications

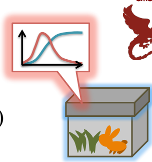
✔ Epidemics

✔ Information diffusion

✔ Online competition



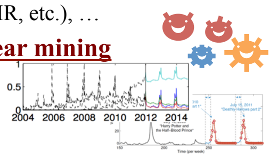
Part 2 Conclusions



✔ **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



Kumamoto U CMU CS

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.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ 235

Kumamoto U CMU CS

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*, 2015.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ 236

Part 2

Non-linear mining and forecasting

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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ 237

