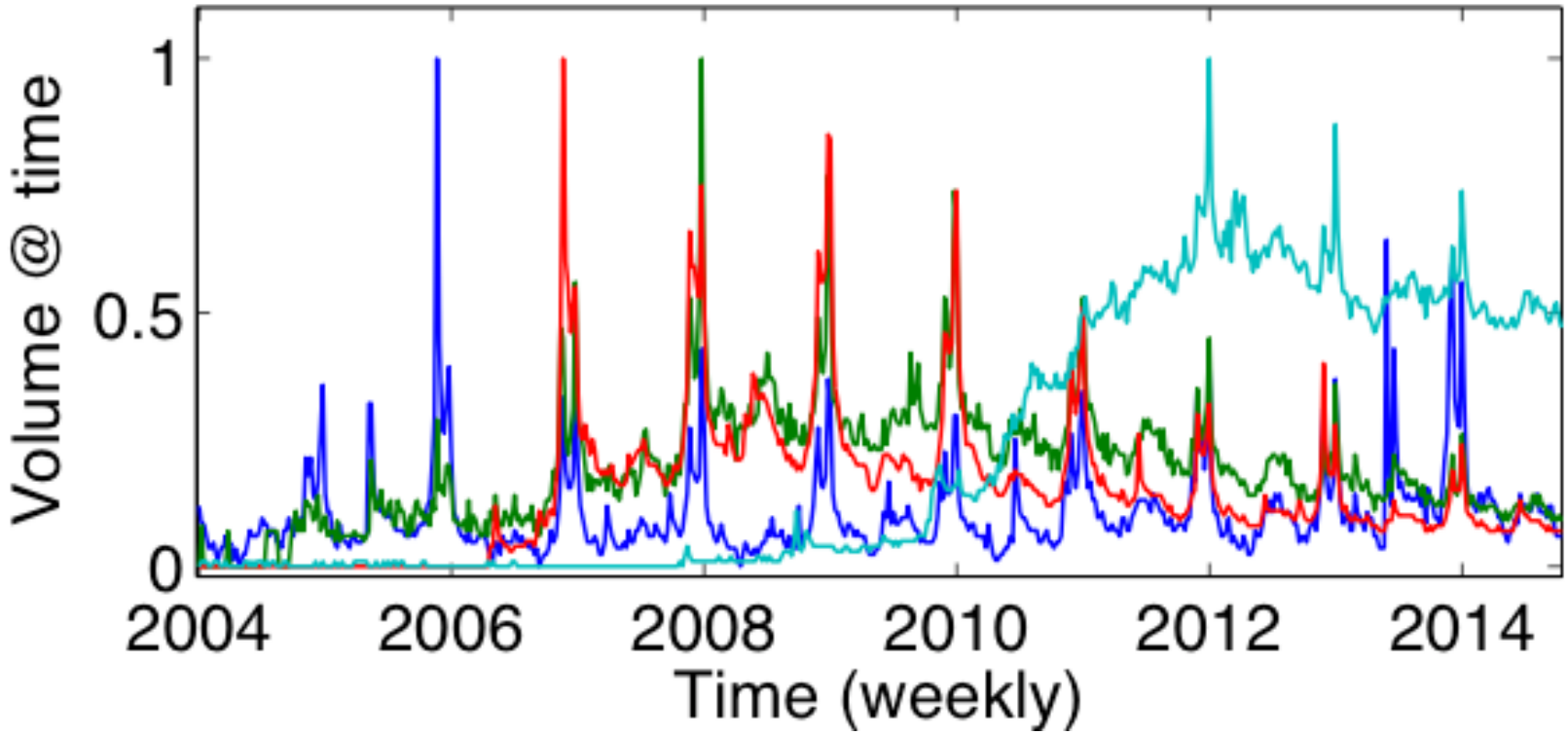
The Web as a Jungle: Non-Linear Dynamical Systems for Co-evolving Online Activities

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

Given: online user activities

e.g., *Google* search volumes for

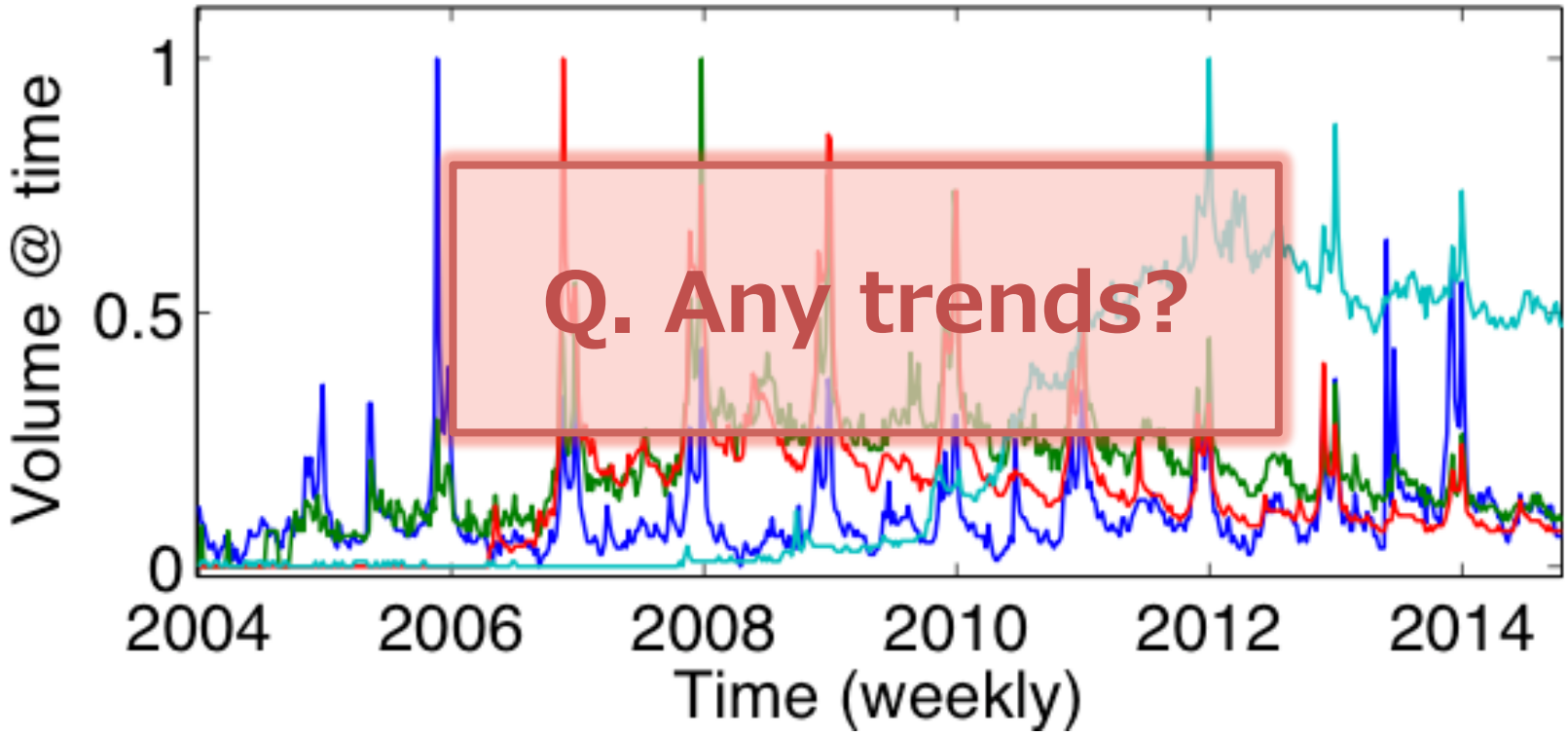
Xbox, **PlayStation**, **Wii**, **Android**



Given: online user activities

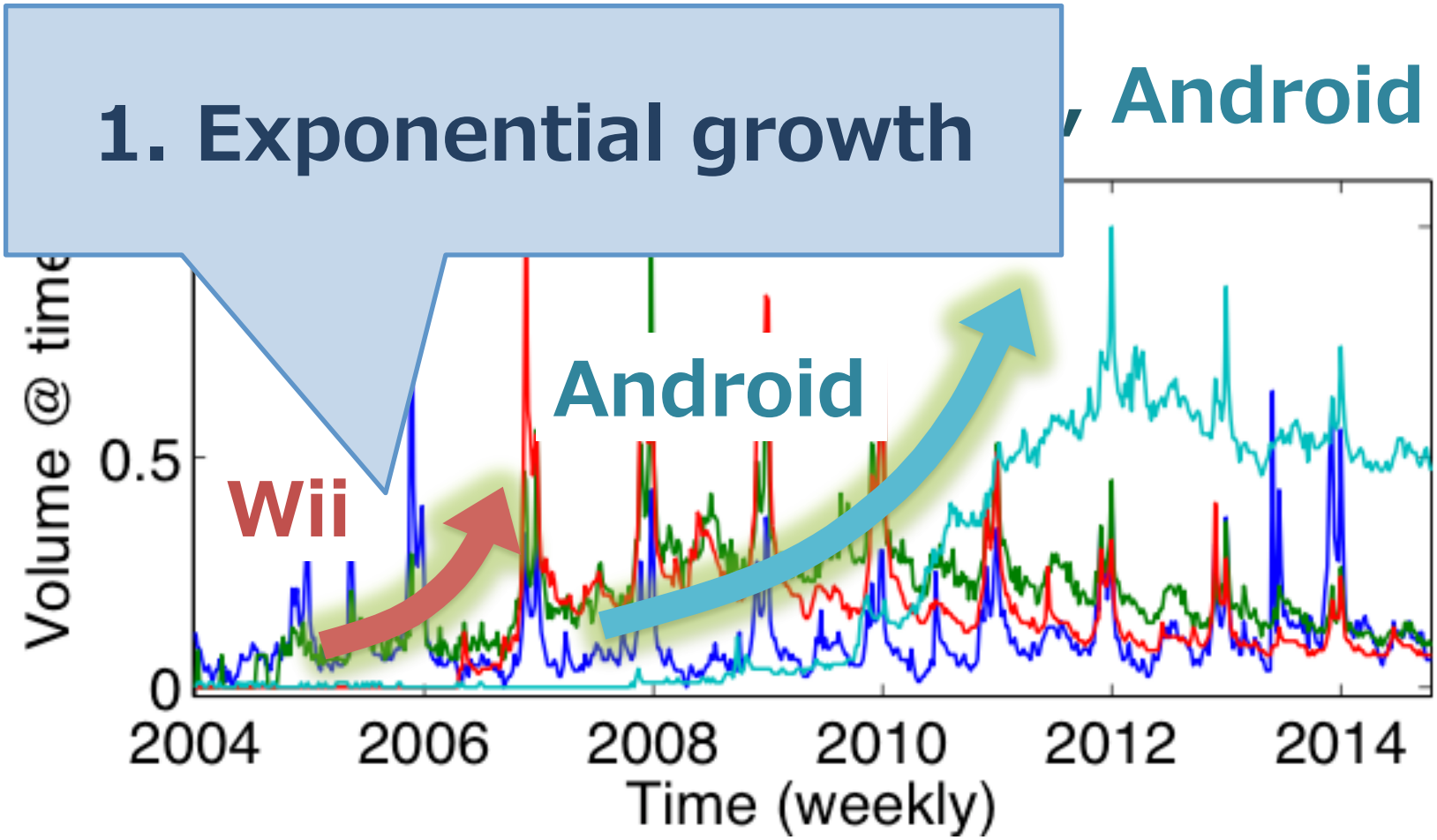
e.g., *Google* search volumes for

Xbox, **PlayStation**, **Wii**, **Android**



Given: online user activities

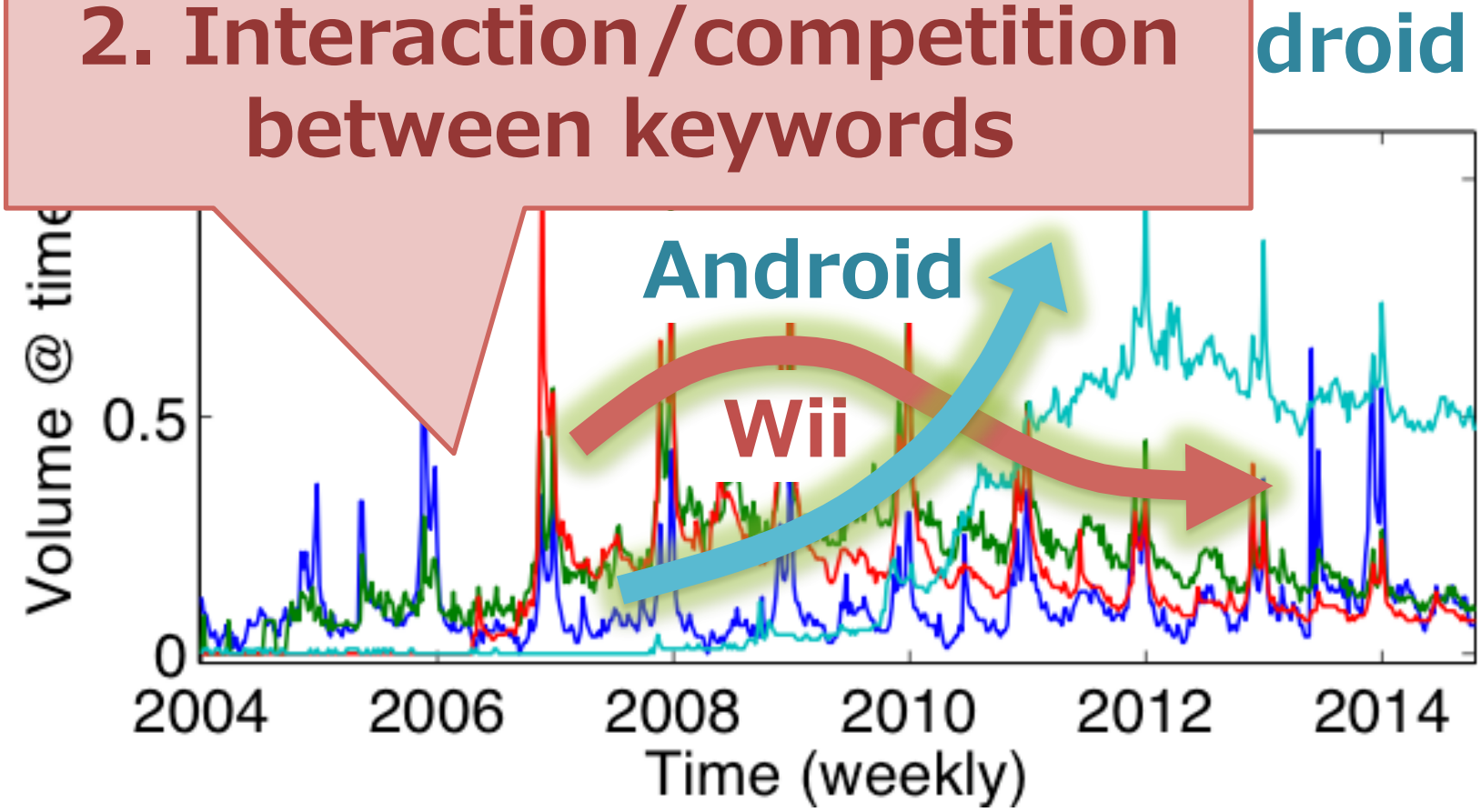
e.g., *Google* search volumes for



Given: online user activities

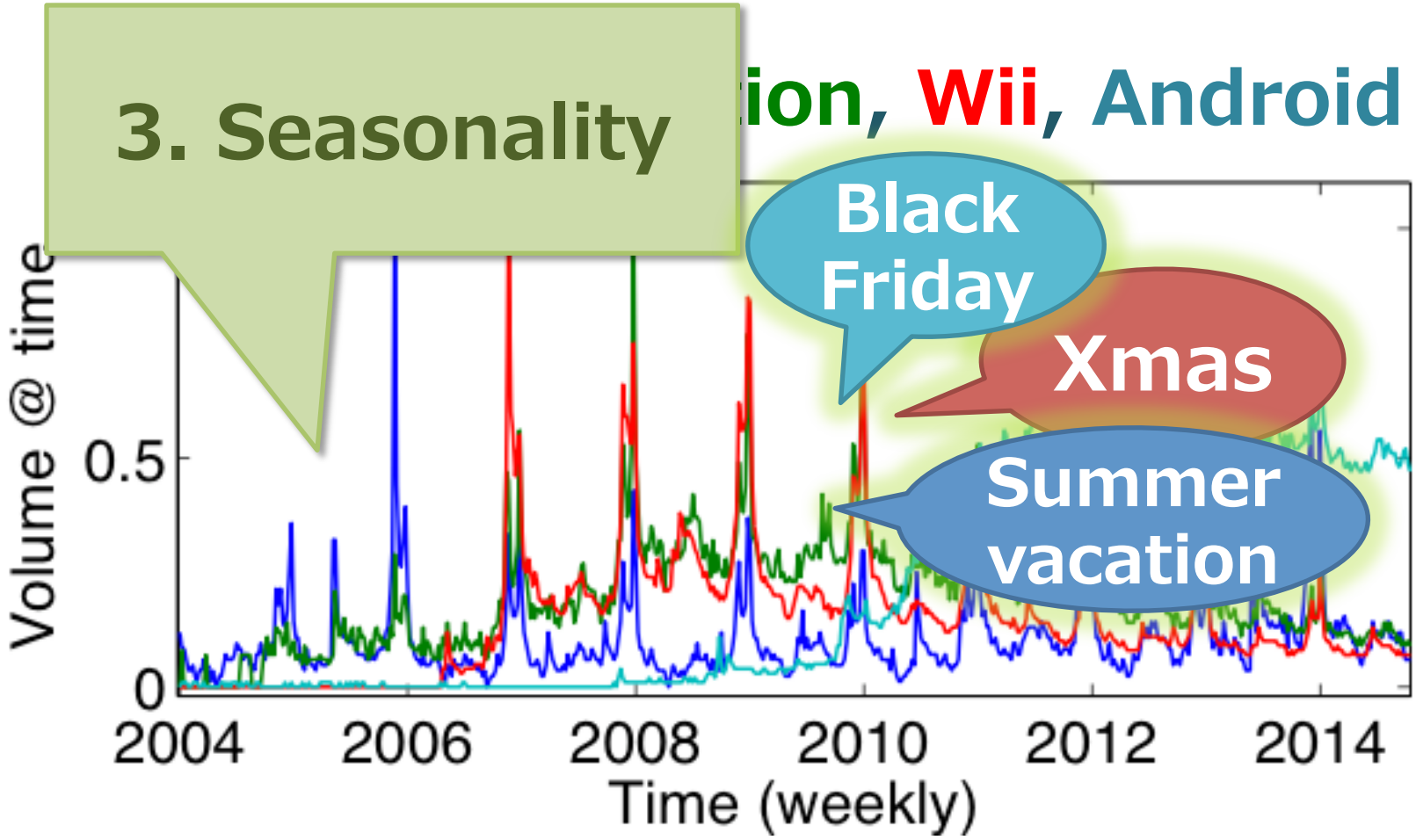
e.g., *Google* search volumes for

2. Interaction/competition between keywords



Given: online user activities

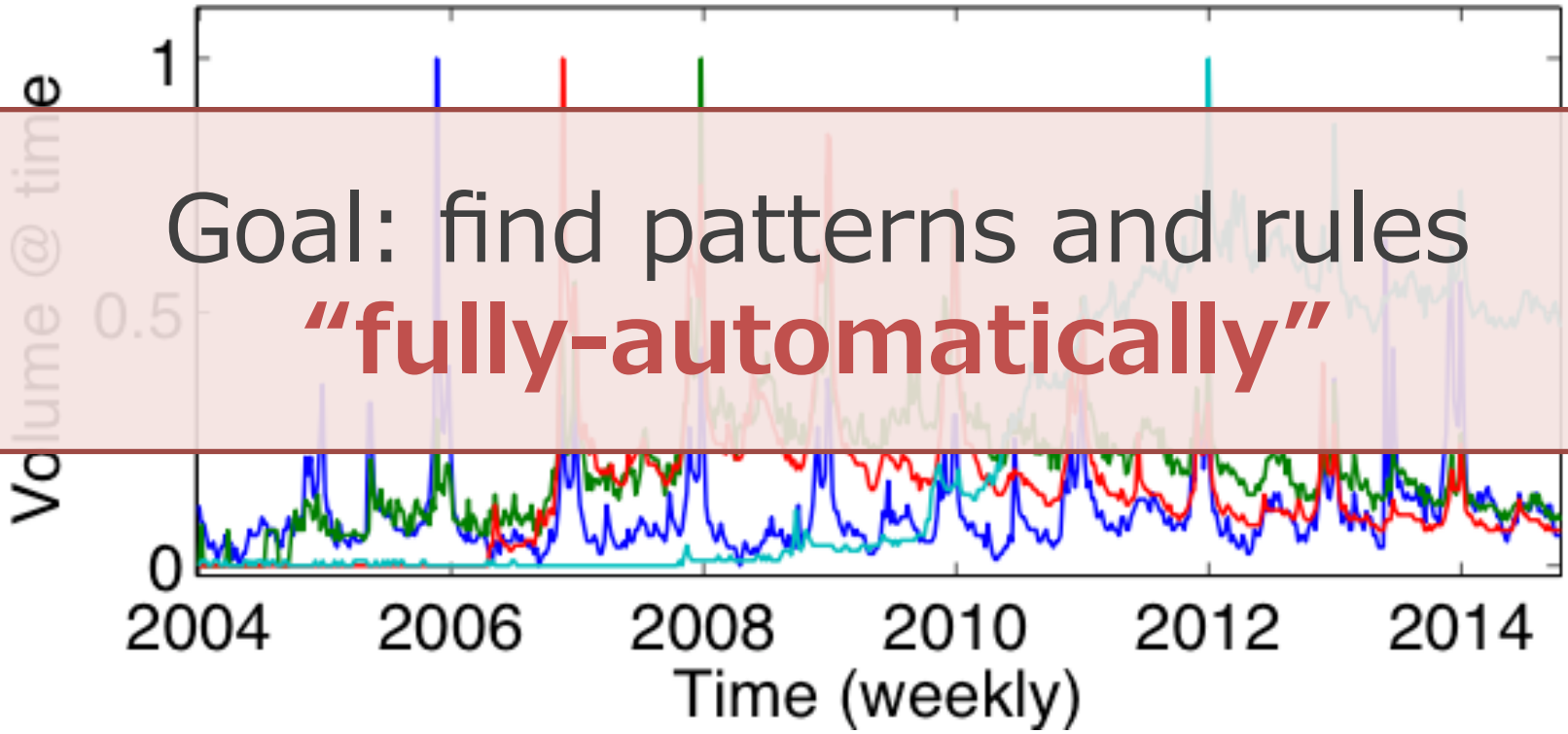
e.g., *Google* search volumes for



Given: online user activities

e.g., *Google* search volumes for

Xbox, **PlayStation**, **Wii**, **Android**

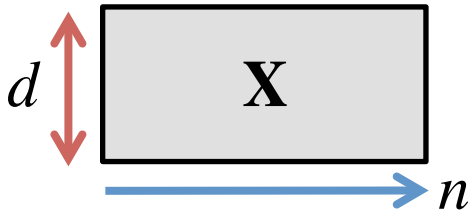


Goal: find patterns and rules
"fully-automatically"

Problem definition

Given: Co-evolving online activities

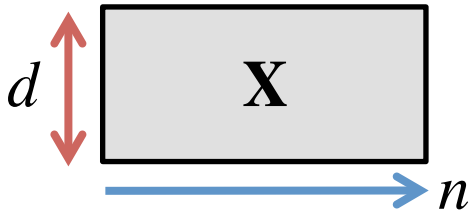
X (activity x time)



Problem definition

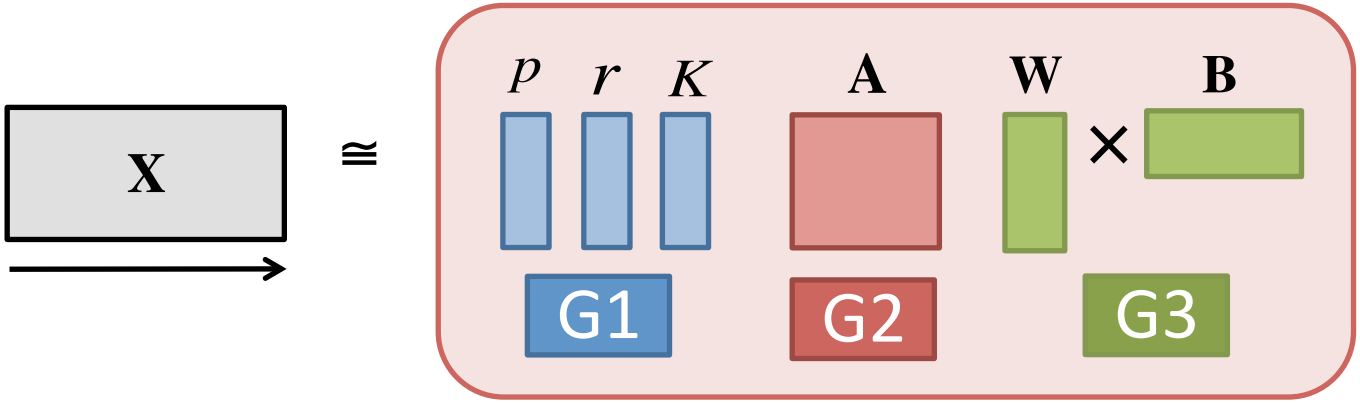
Given: Co-evolving online activities

X (activity x time)



Find: Compact description of X

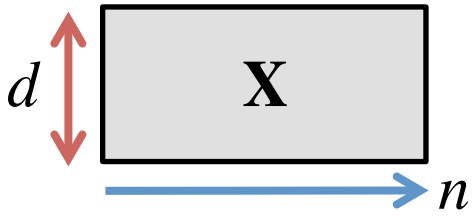
EcoWeb



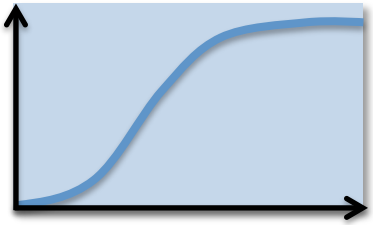
Problem definition

Given: Co-evolving online activities

x time)

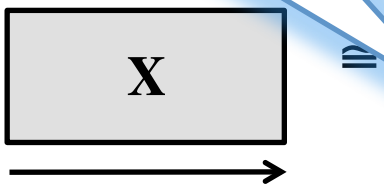


G1 **Non-linear evolution**

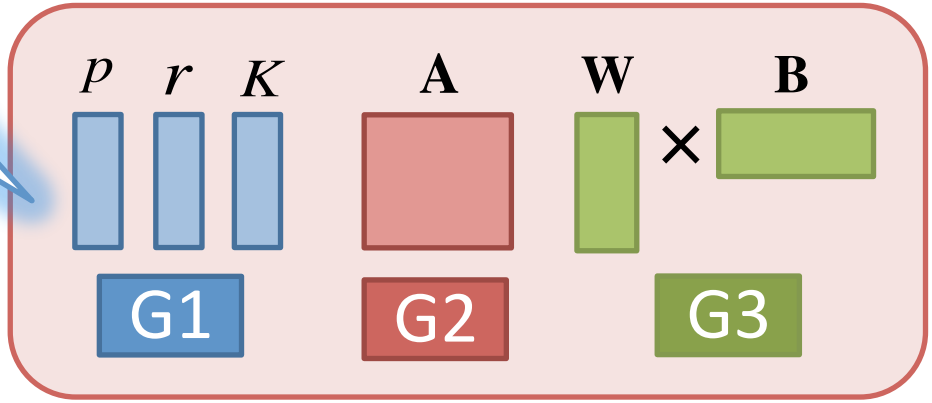


Fi scription of X

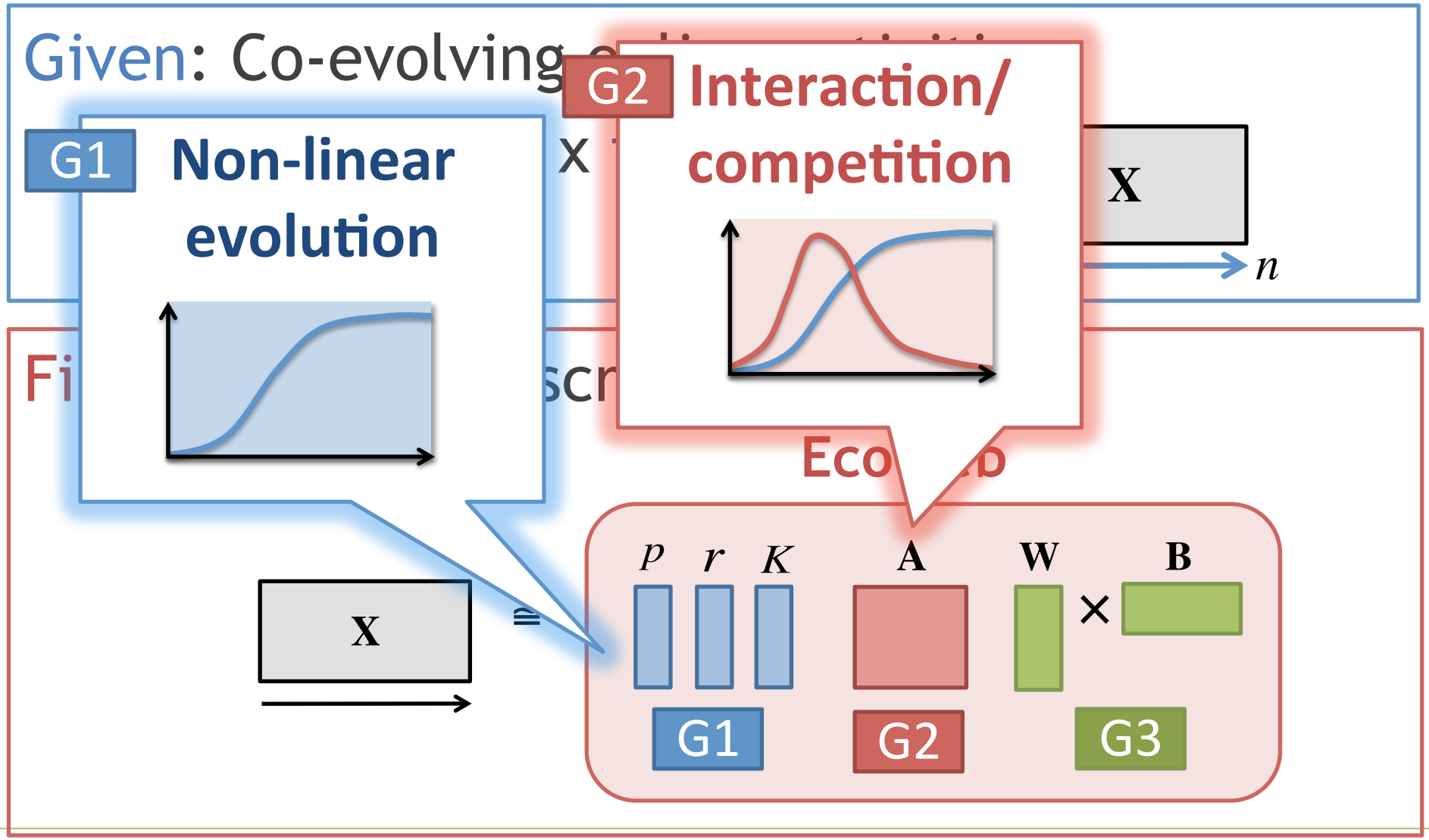
EcoWeb



≅



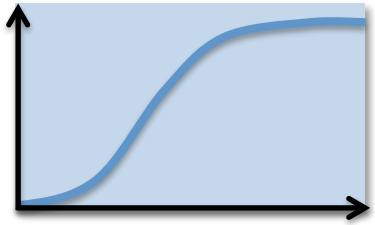
Problem definition



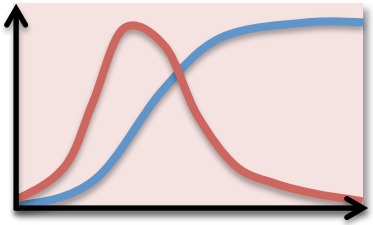
Problem definition

Given: Co-evolvi… …es

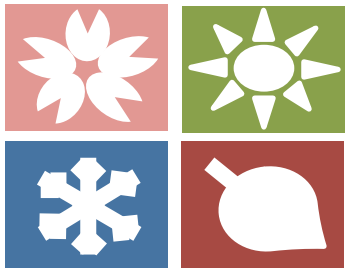
G1 **Non-linear evolution**



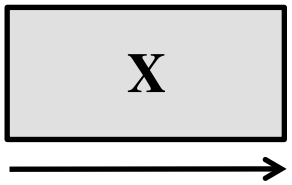
G2 **Interaction/ competition**



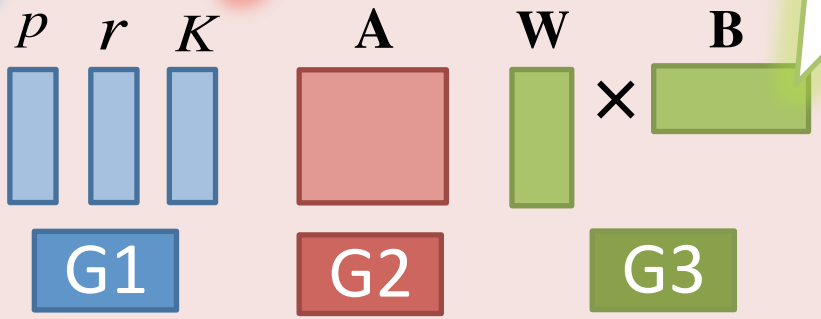
G3 **Seasonality**



F… …oWeb



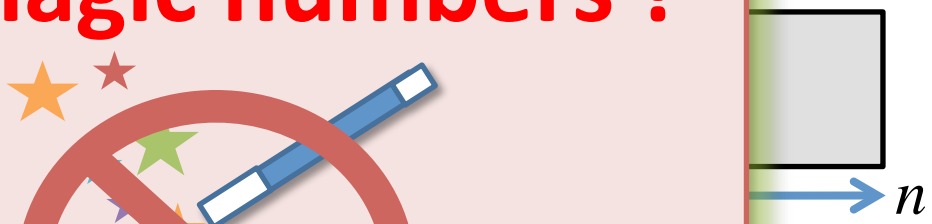
≅



Problem definition

Given: Co-e

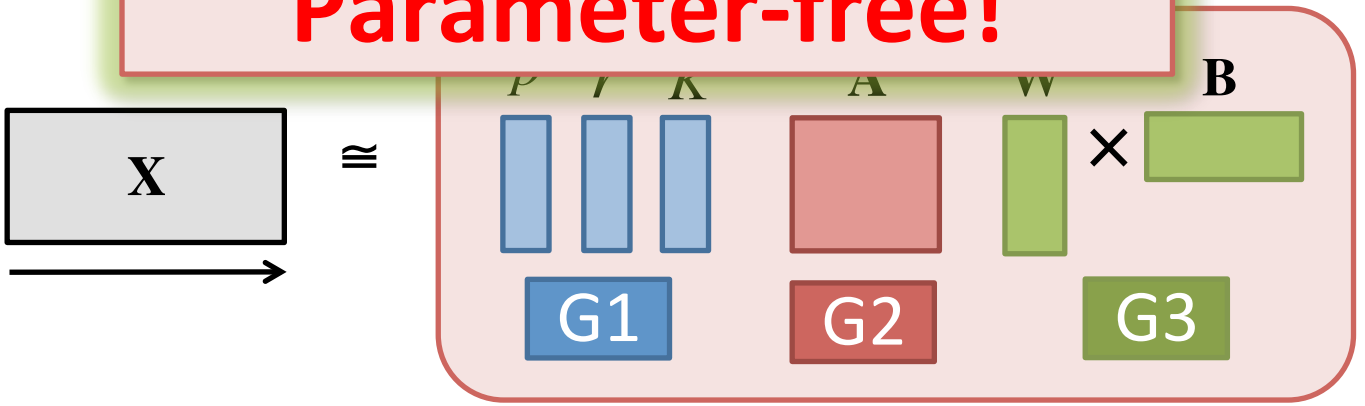
X (a



Find: Comp

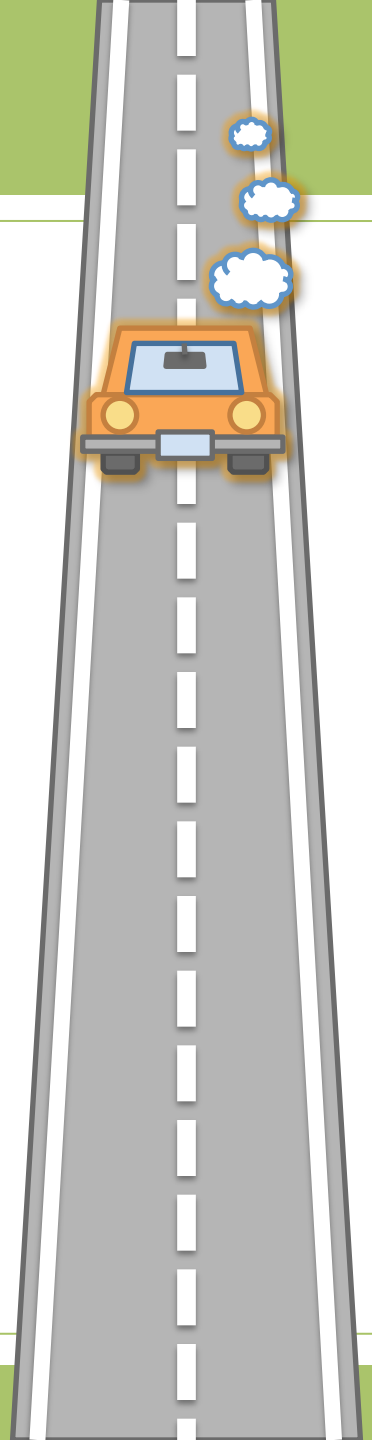
NO magic numbers !

Parameter-free!



Roadmap

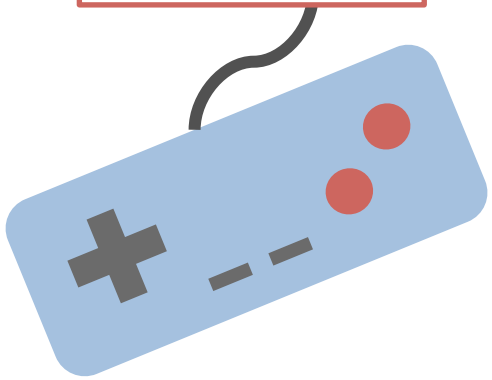
- ✔ Motivation
- Modeling power of EcoWeb
- Overview
- Proposed model
- Algorithm
- Experiments
- EcoWeb – at work
- Conclusions



Modeling power of EcoWeb

Questions

Q1



Q2

Q3

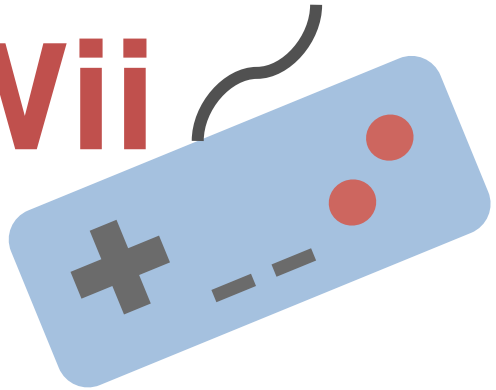
amazon

Modeling power of EcoWeb

Q1 (games)

Who is the competitor?

Wii

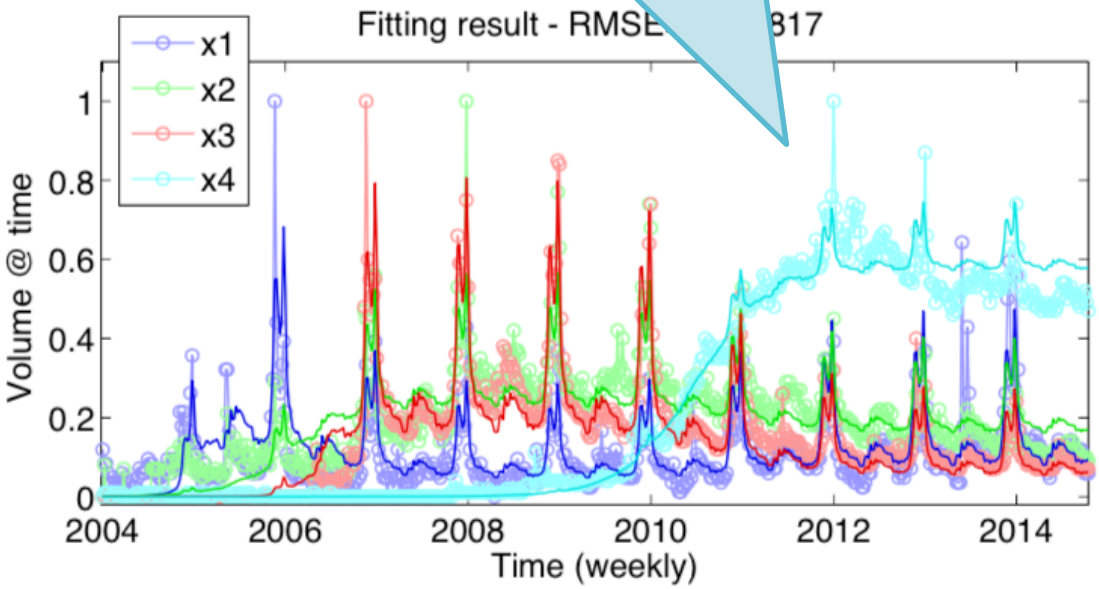


vs.

?

Modeling power of EcoWeb

A. Android!



EcoWeb-Fit

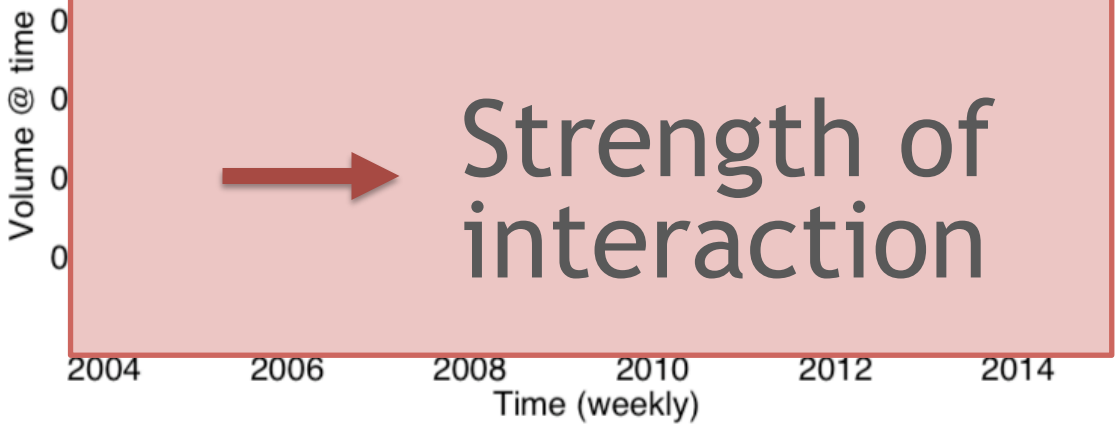
Interaction network (latent)

Modeling power of EcoWeb

A. Android!

Keyword

Strength of interaction



EcoWeb-Fit

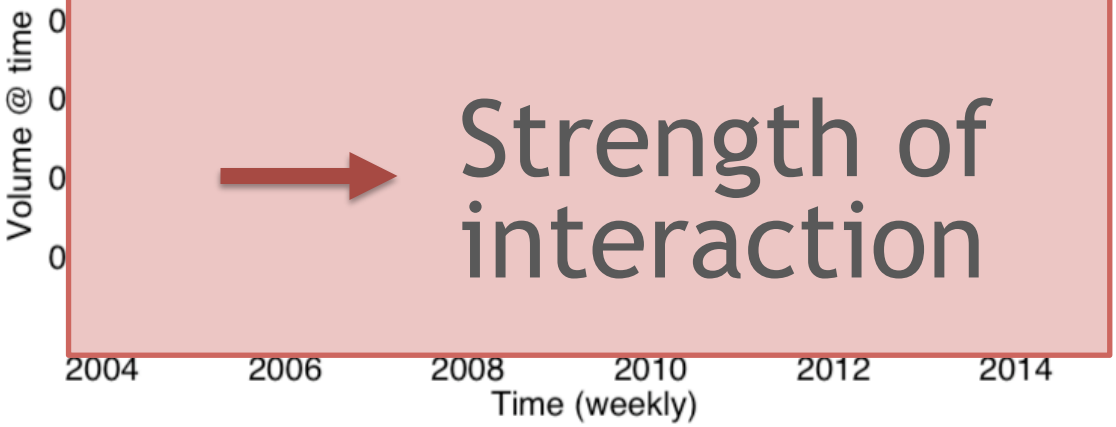
Interaction network (latent)

Modeling power of EcoWeb

A. Android!

Keyword

Strength of interaction



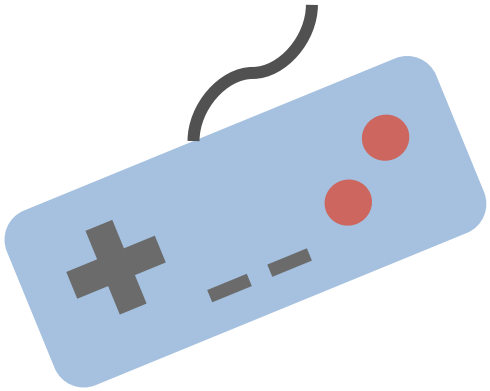
EcoWeb-Fit

Interaction network (latent)

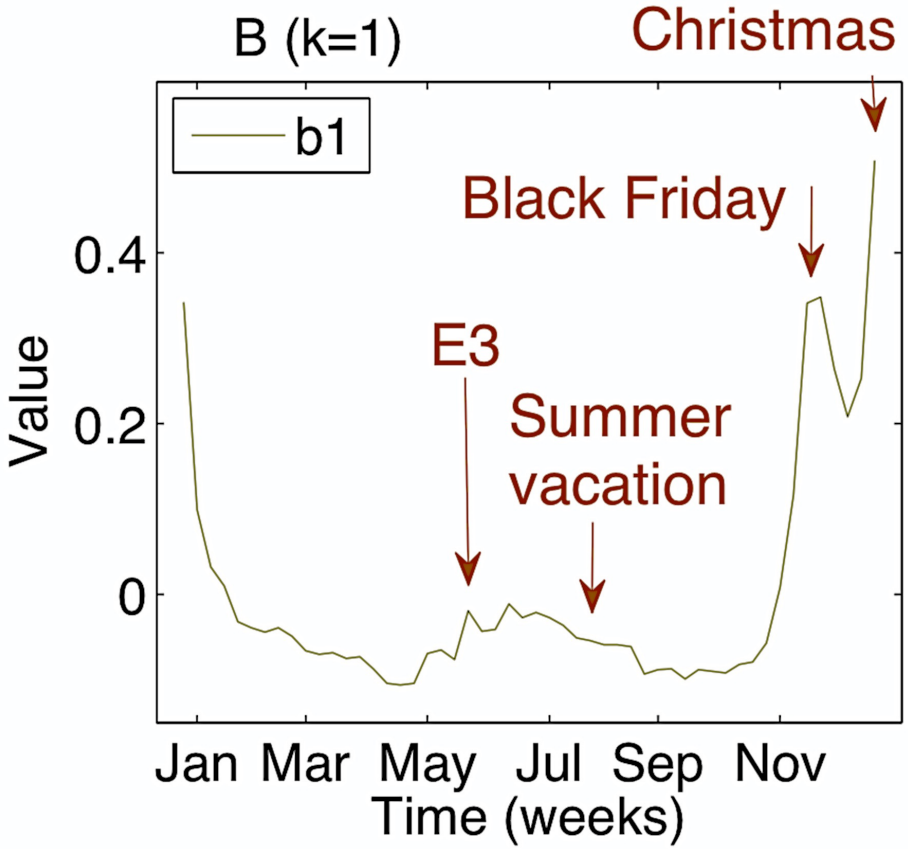
Modeling power of EcoWeb

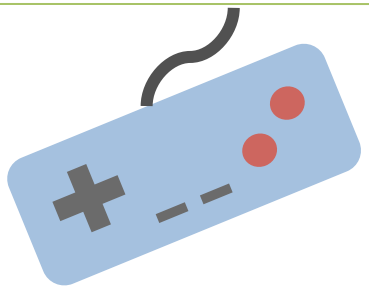
Q1 (games)

Any seasonal events?



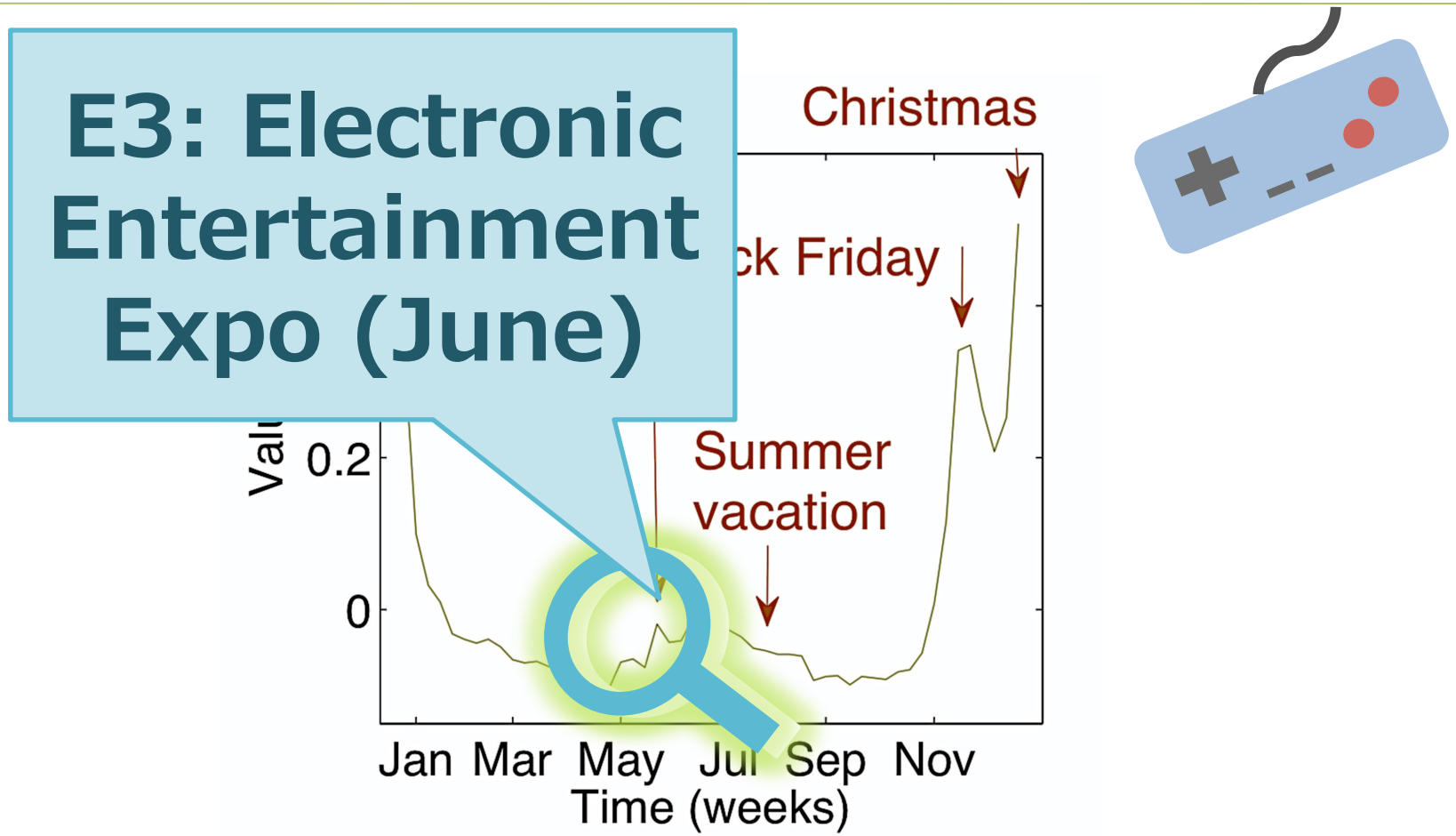
Modeling power of EcoWeb





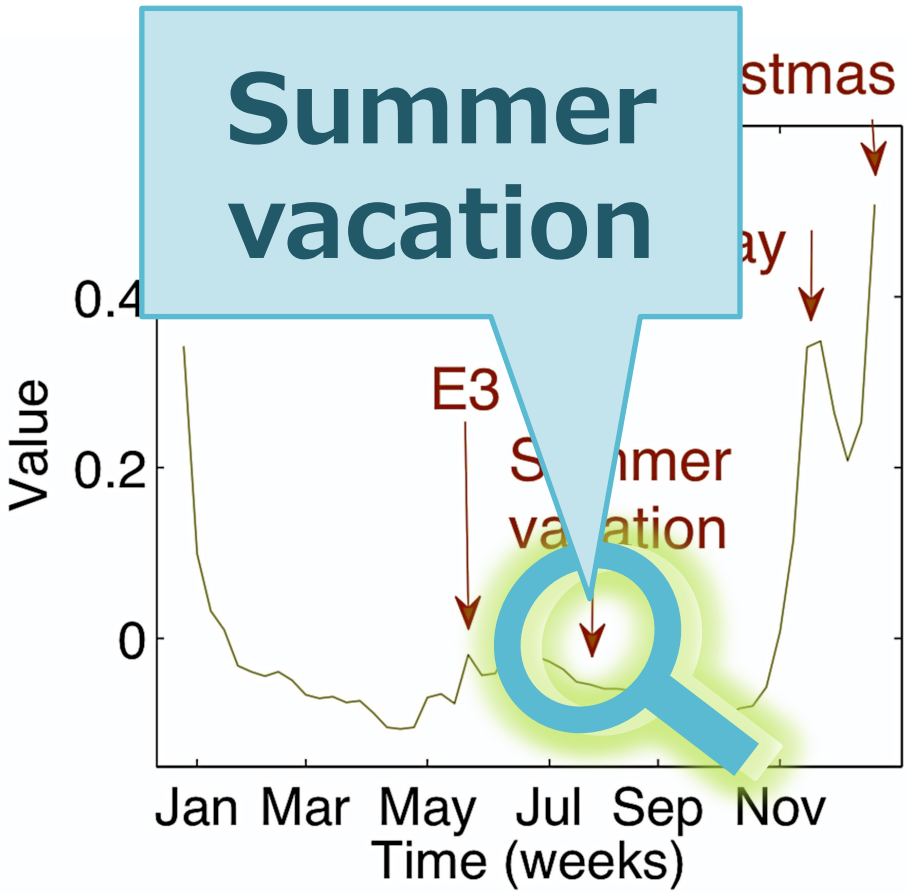
EcoWeb: seasonal component

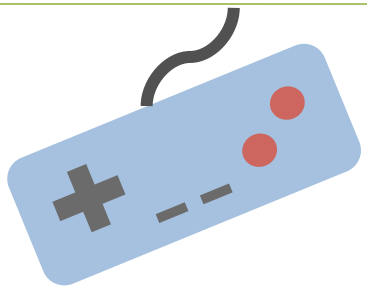
Modeling power of EcoWeb



EcoWeb: seasonal component

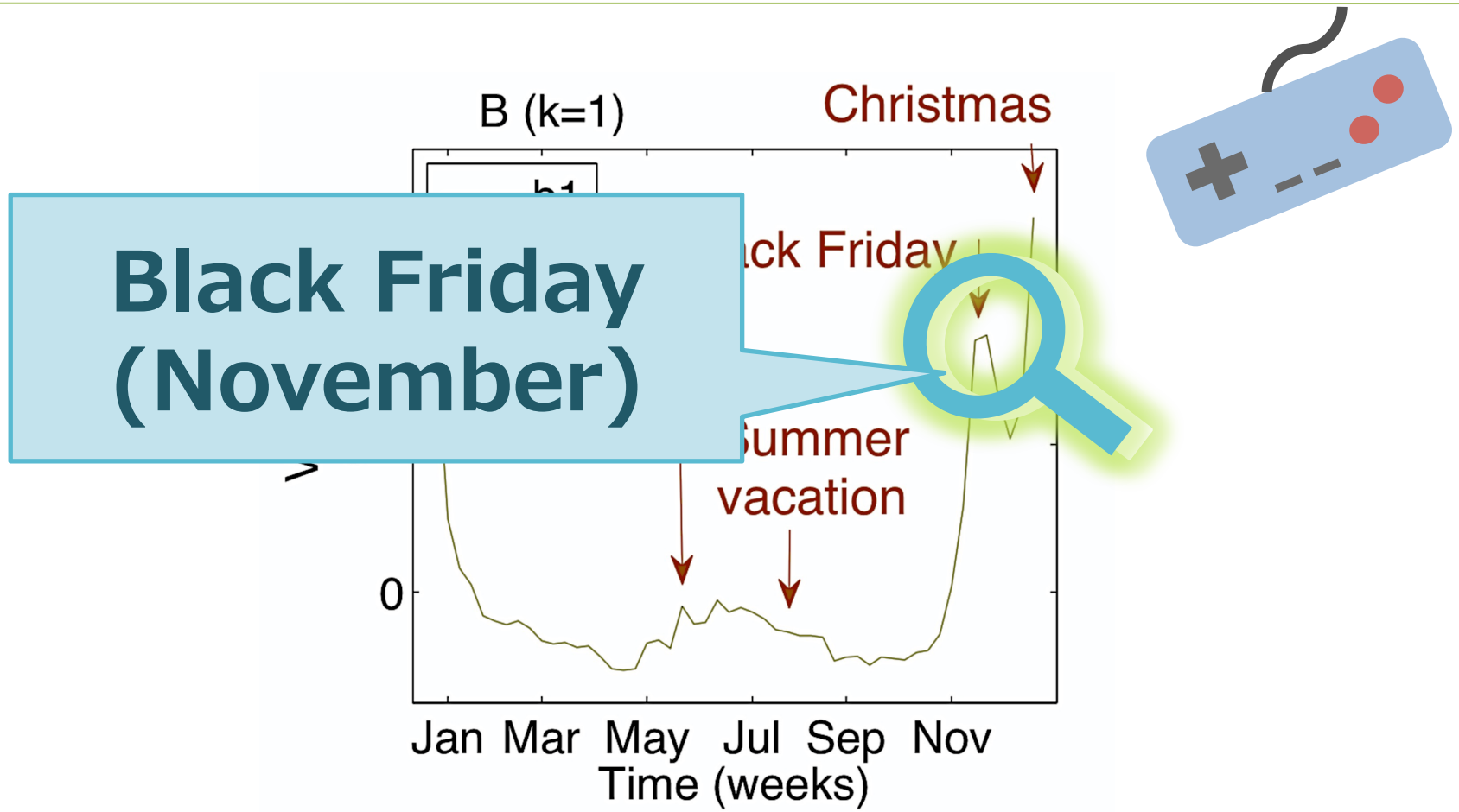
Modeling power of EcoWeb





EcoWeb: seasonal component

Modeling power of EcoWeb

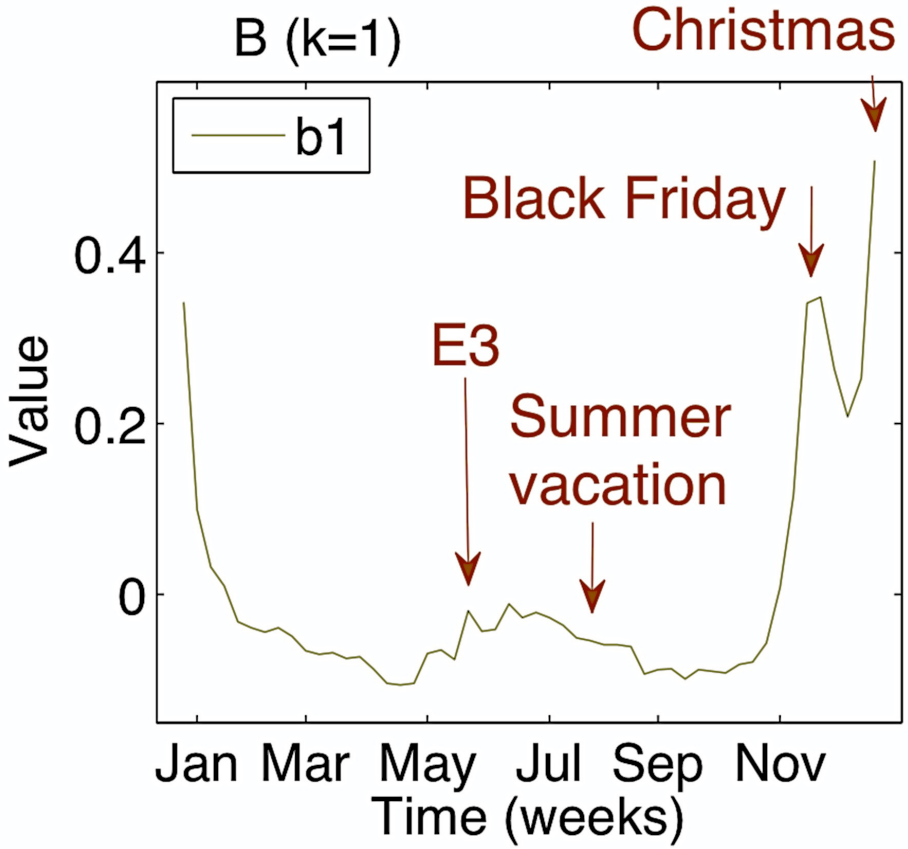


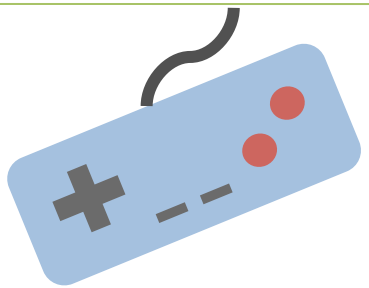
EcoWeb: seasonal component

Modeling power of EcoWeb

EcoWeb: seasonal component

Modeling power of EcoWeb



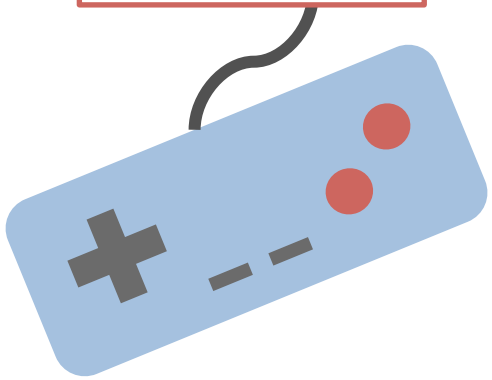


EcoWeb: seasonal component

Modeling power of EcoWeb

Questions

Q1



Q2

Q3

amazon

Modeling power of EcoWeb

Q2 (apparels)

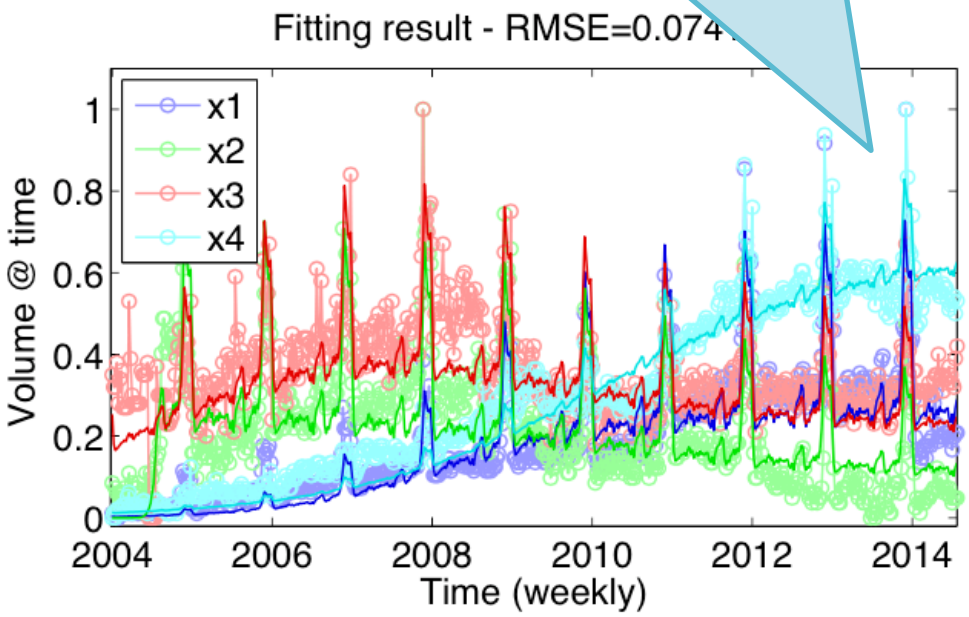
Who is the competitor?

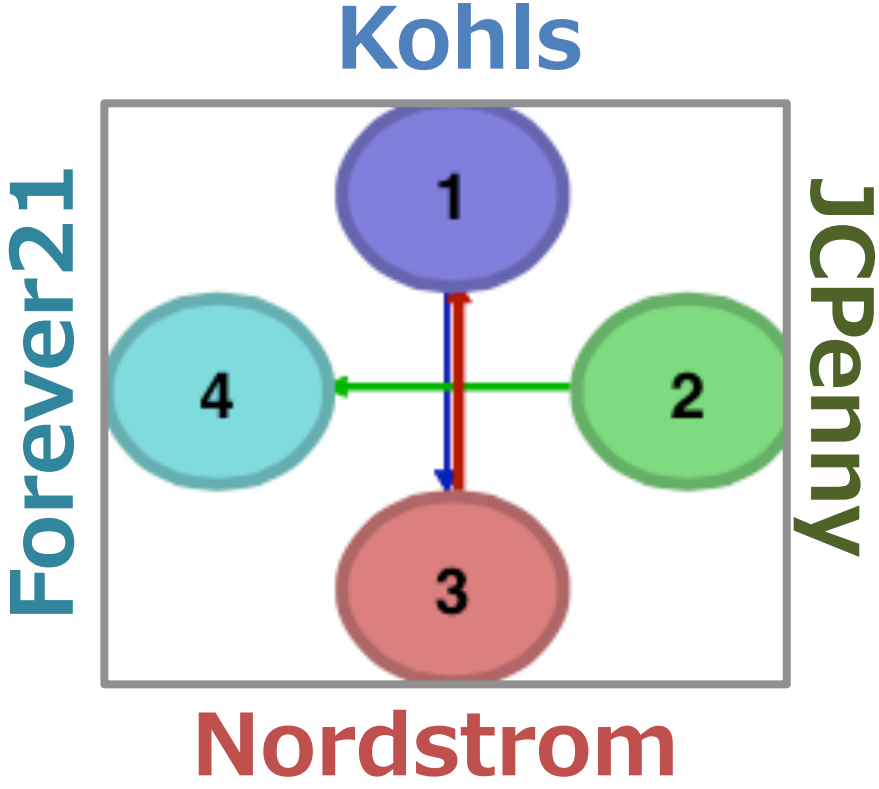
JCPenny

vs.

Modeling power of EcoWeb

A2. Forever21!

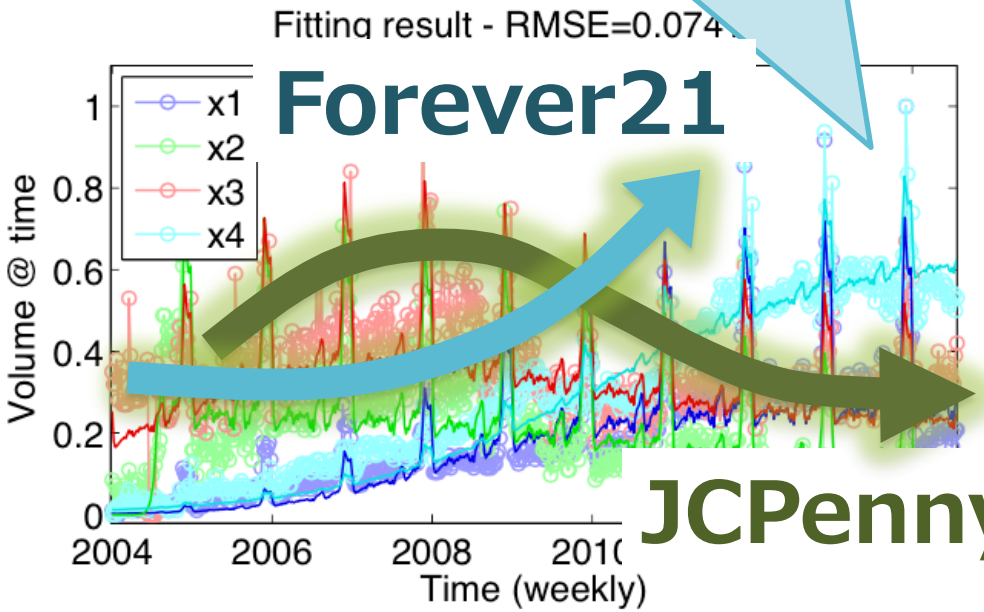


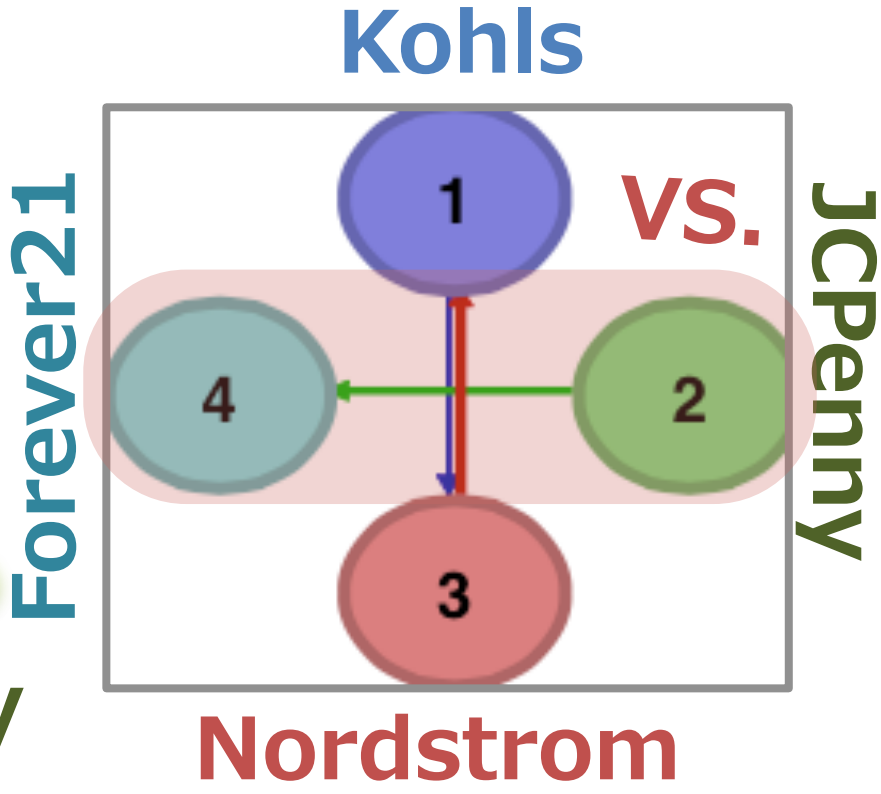


EcoWeb: Interaction network

Modeling power of EcoWeb

A2. Forever21!





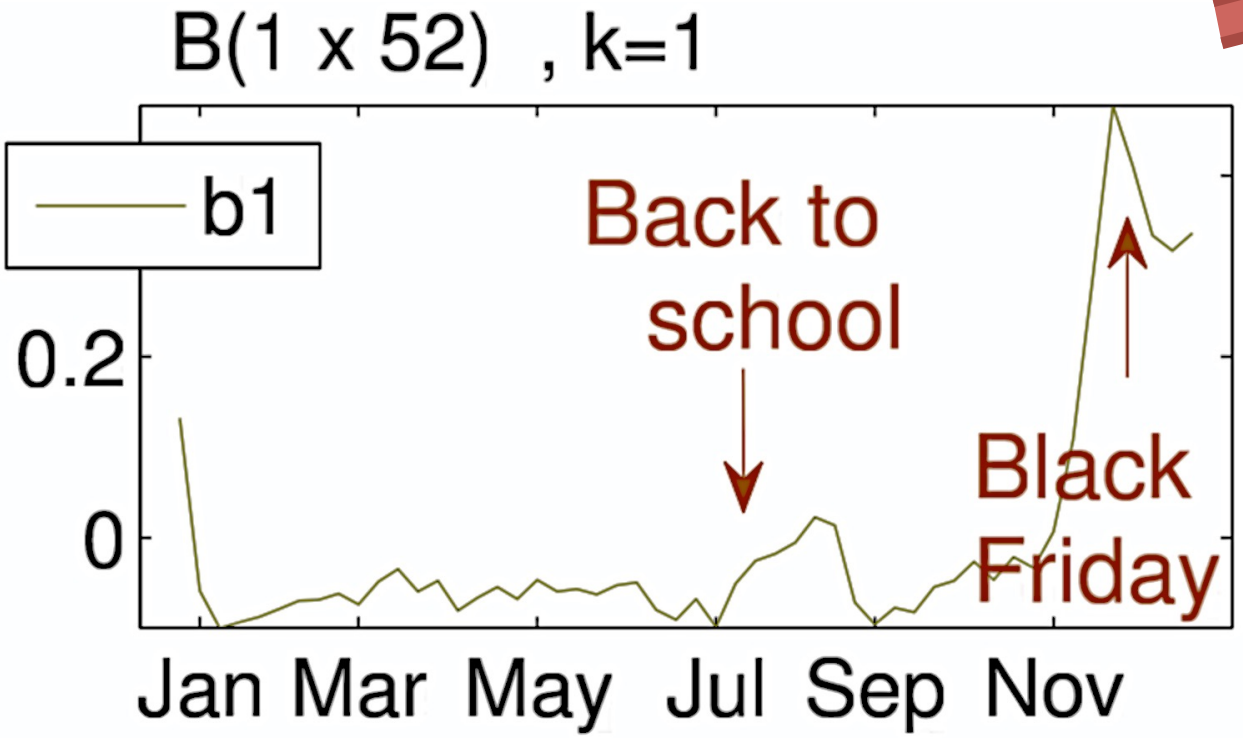
EcoWeb: Interaction network

Modeling power of EcoWeb

Q2 (apparels)

Any seasonal events?

Modeling power of EcoWeb

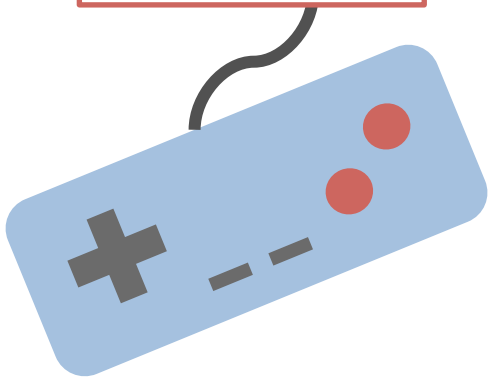


EcoWeb: seasonal component

Modeling power of EcoWeb

Questions

Q1



Q2

Q3

amazon

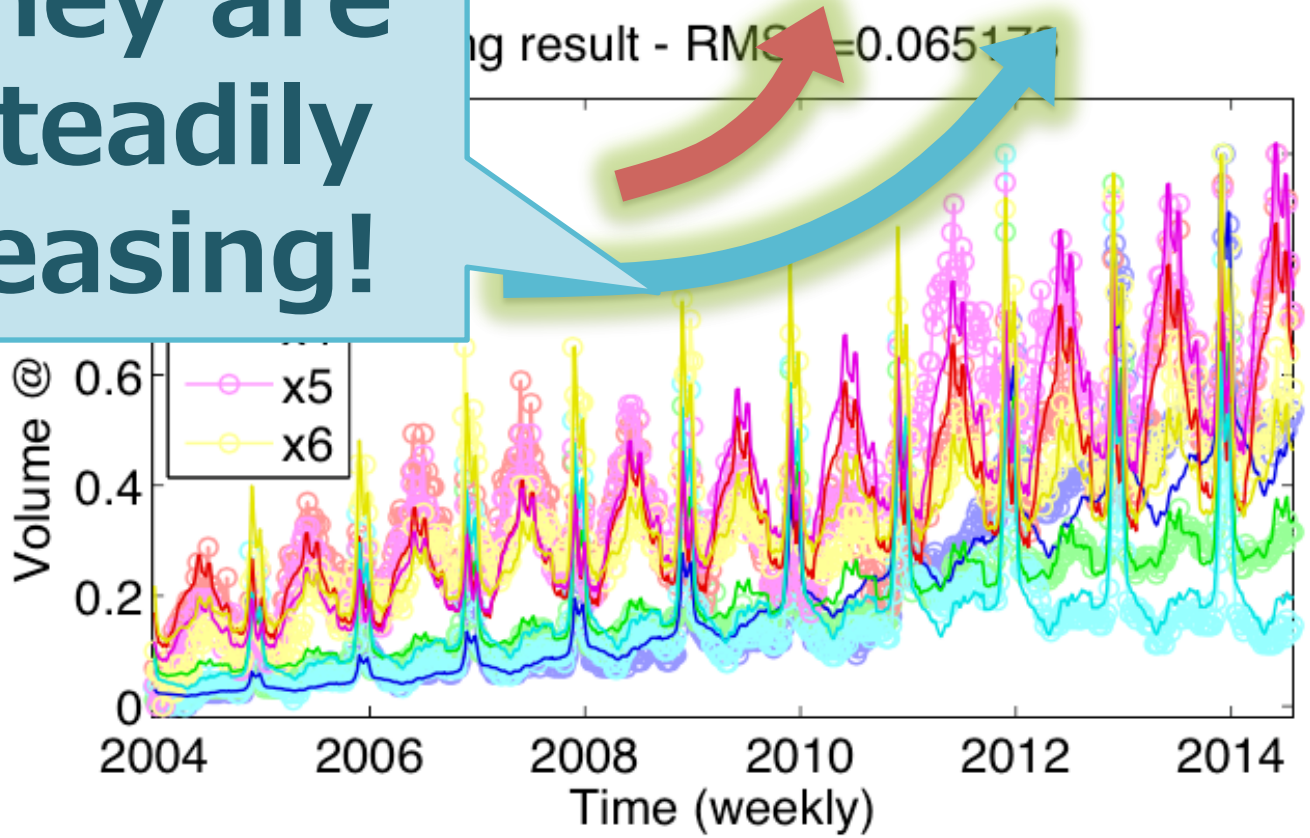
Modeling power of EcoWeb

Q3 (retails)

Any patterns/trends?

Modeling power of EcoWeb

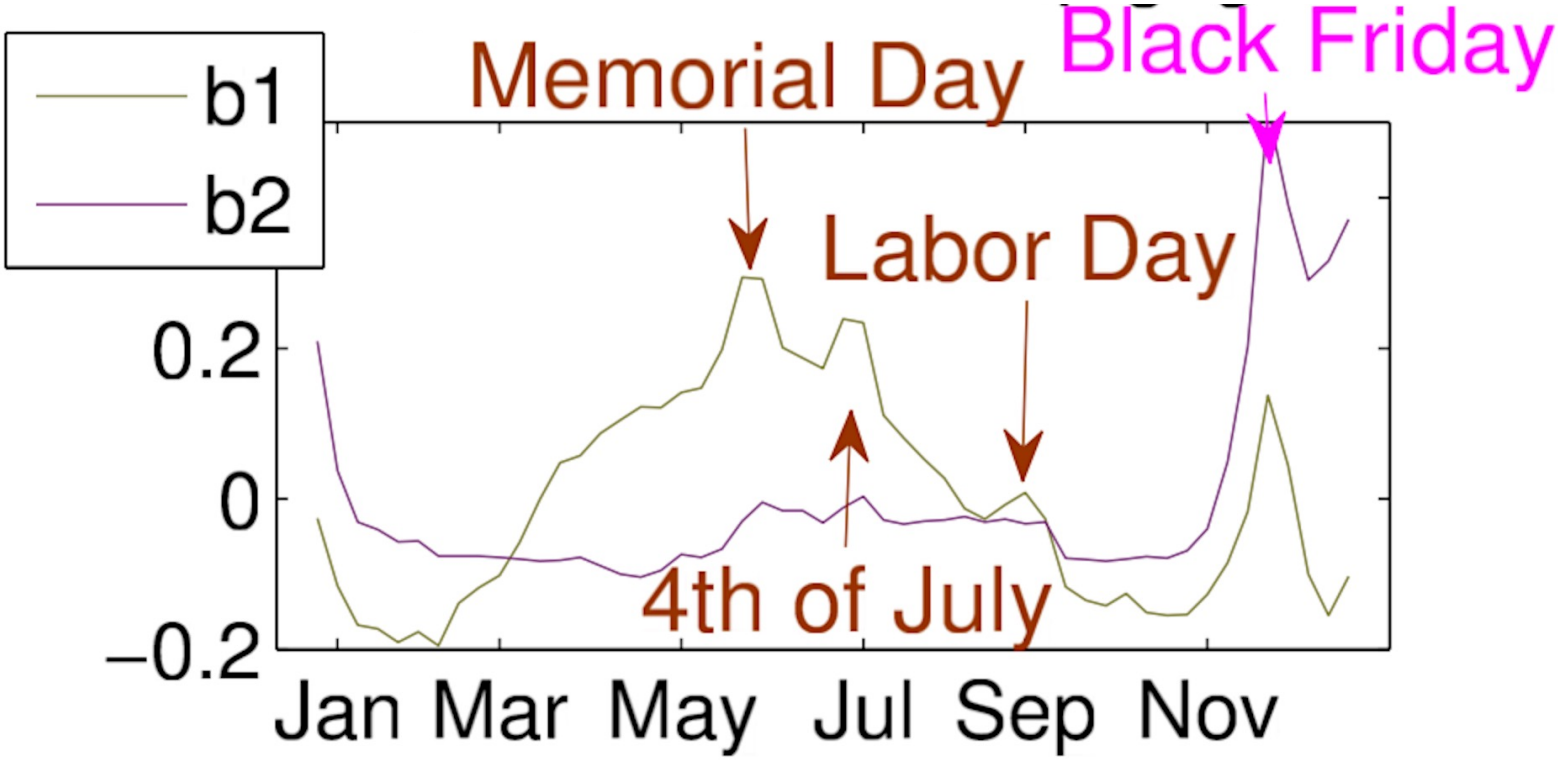
A. They are all steadily increasing!



Amazon, Walmart, Home Depot, Best buy, …

Modeling power of EcoWeb

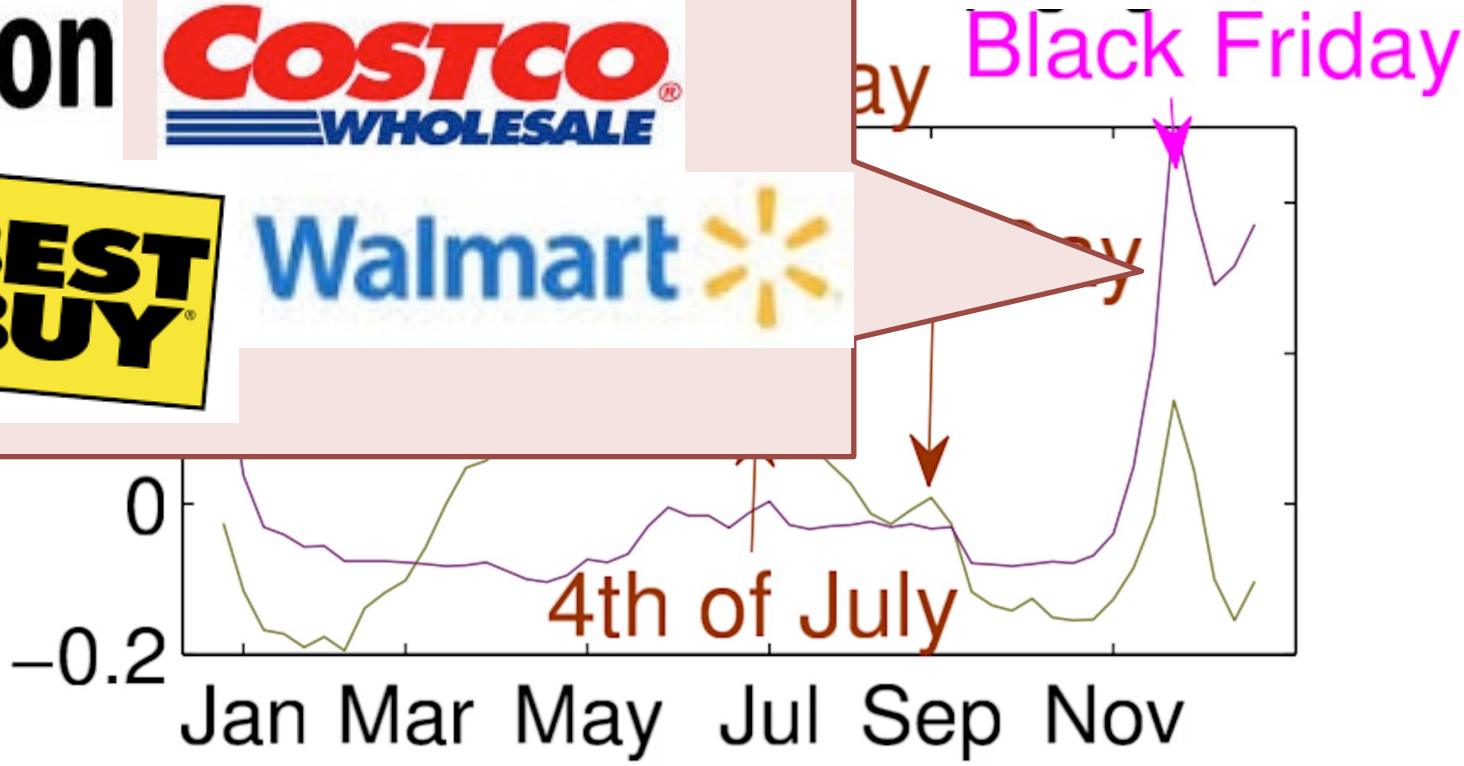
2 seasonal components



Modeling power of EcoWeb

2 seasonal components

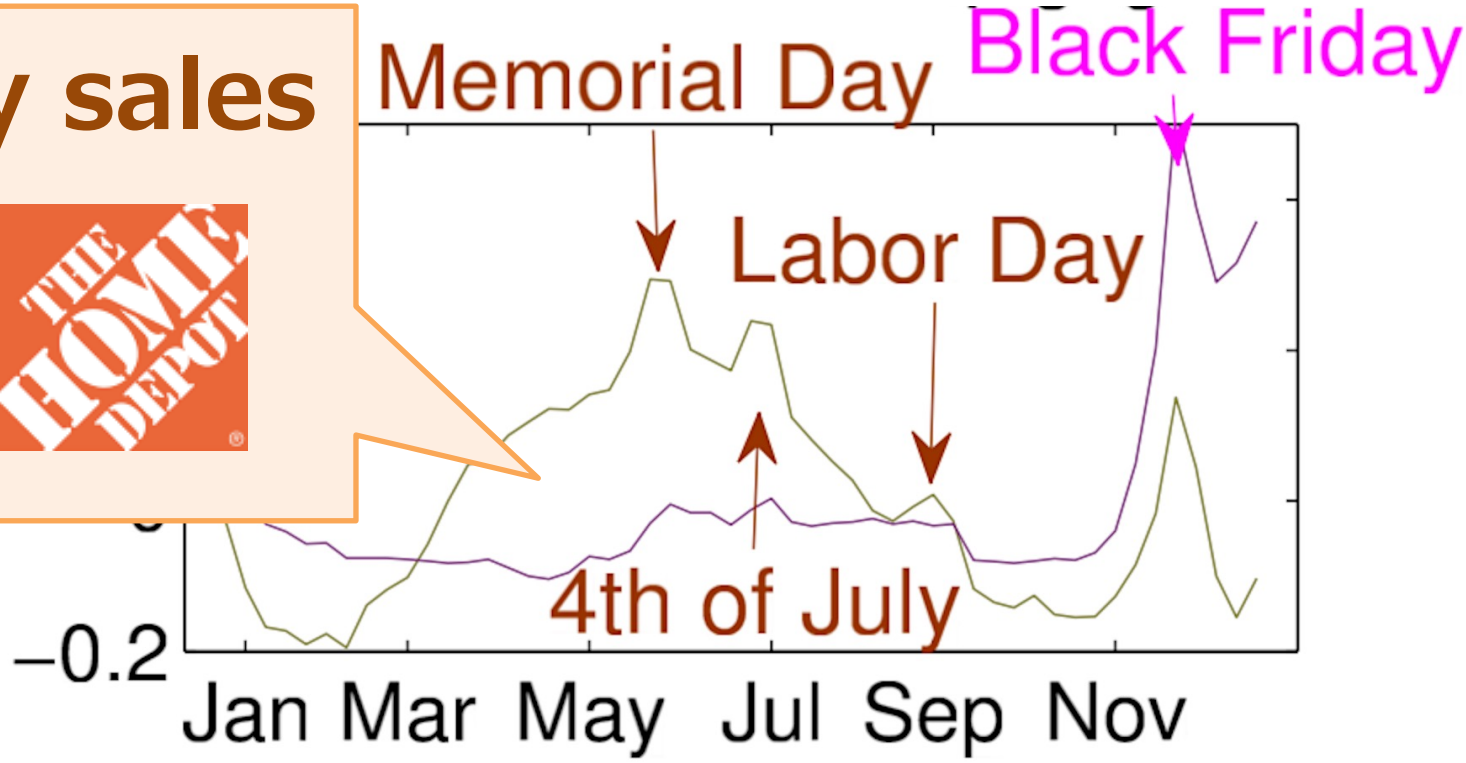
Black Friday sale



Modeling power of EcoWeb

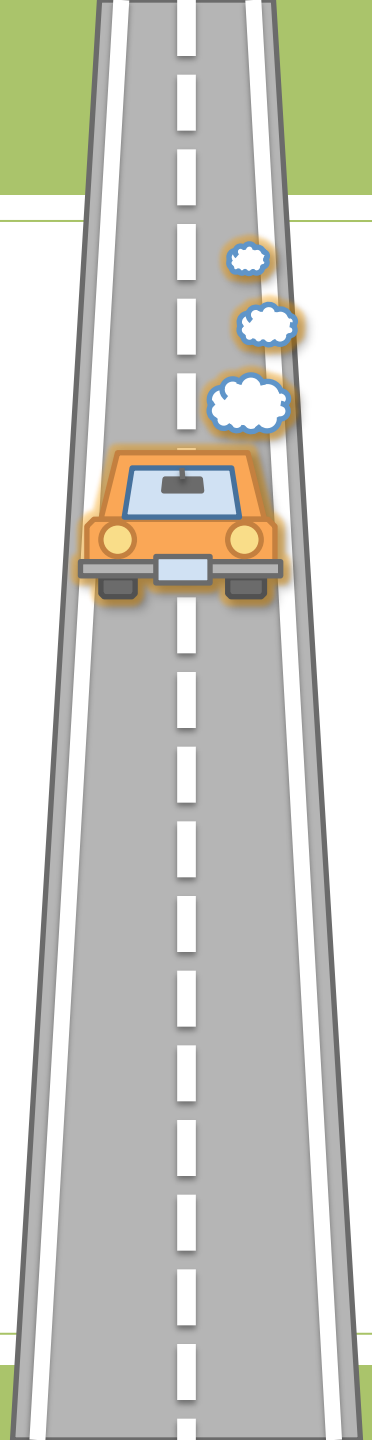
2 seasonal components

Holiday sales



Roadmap

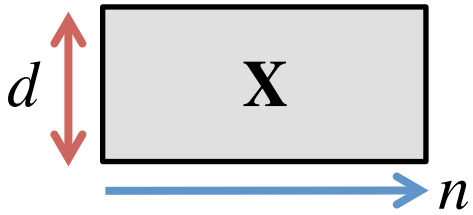
- ✔ Motivation
- ✔ Modeling power of EcoWeb
  - Overview
  - Proposed model
  - Algorithm
  - Experiments
  - EcoWeb – at work
  - Conclusions



Problem definition

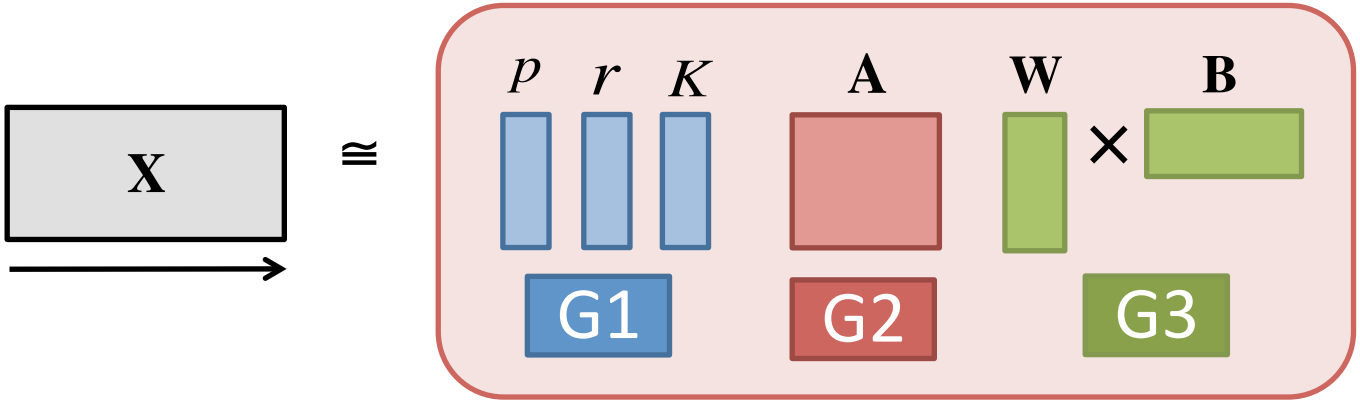
Given: Co-evolving online activities

X (activity x time)



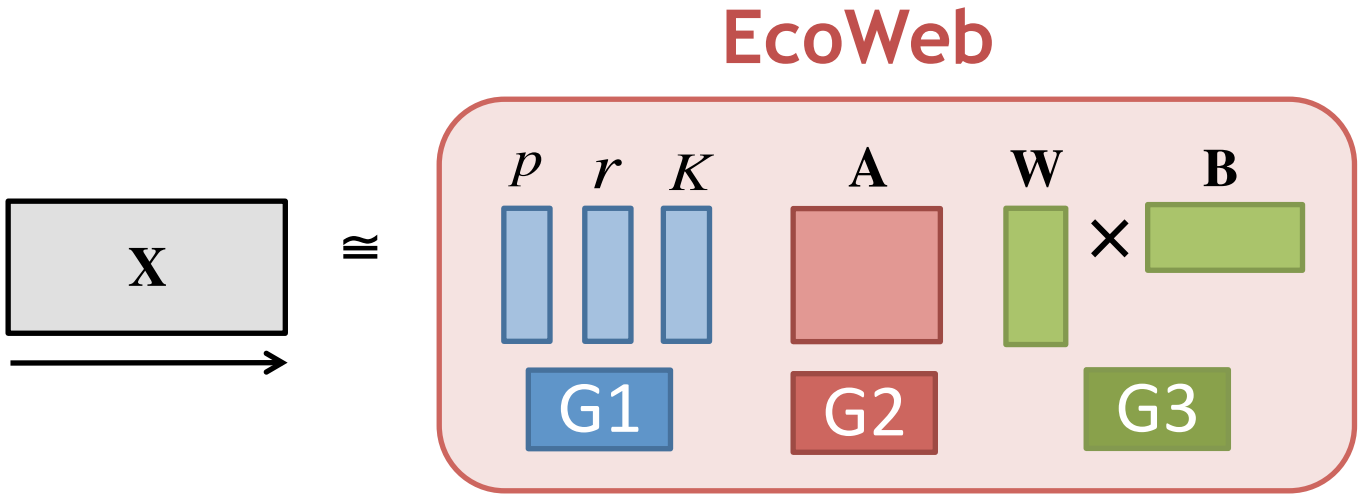
Find: Compact description of X

EcoWeb



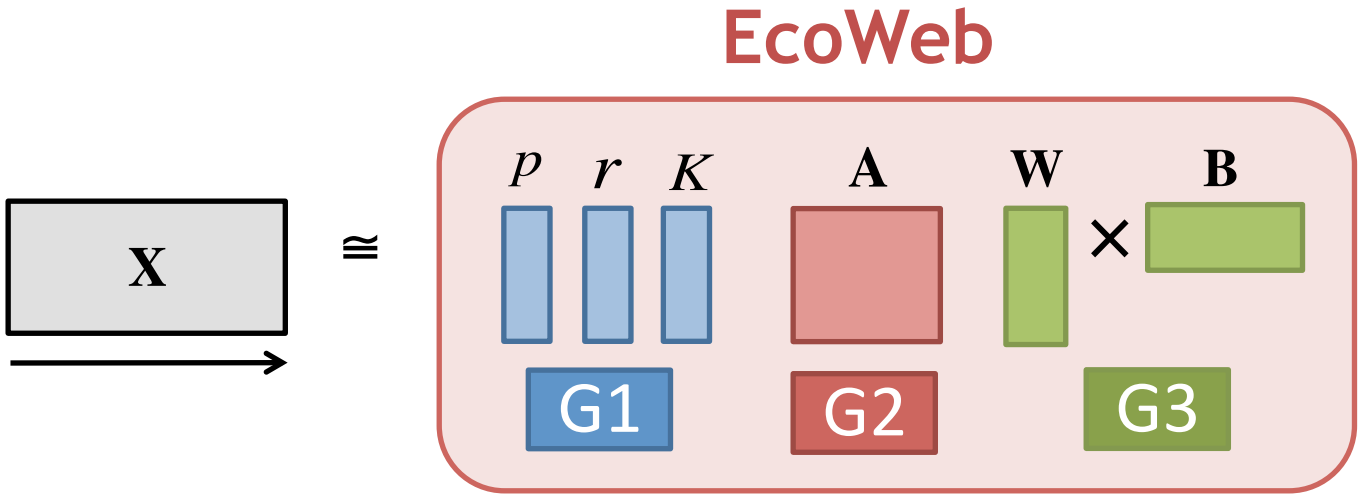
EcoWeb: Main idea

Q. How can we describe the evolutions of X ?



EcoWeb: Main idea

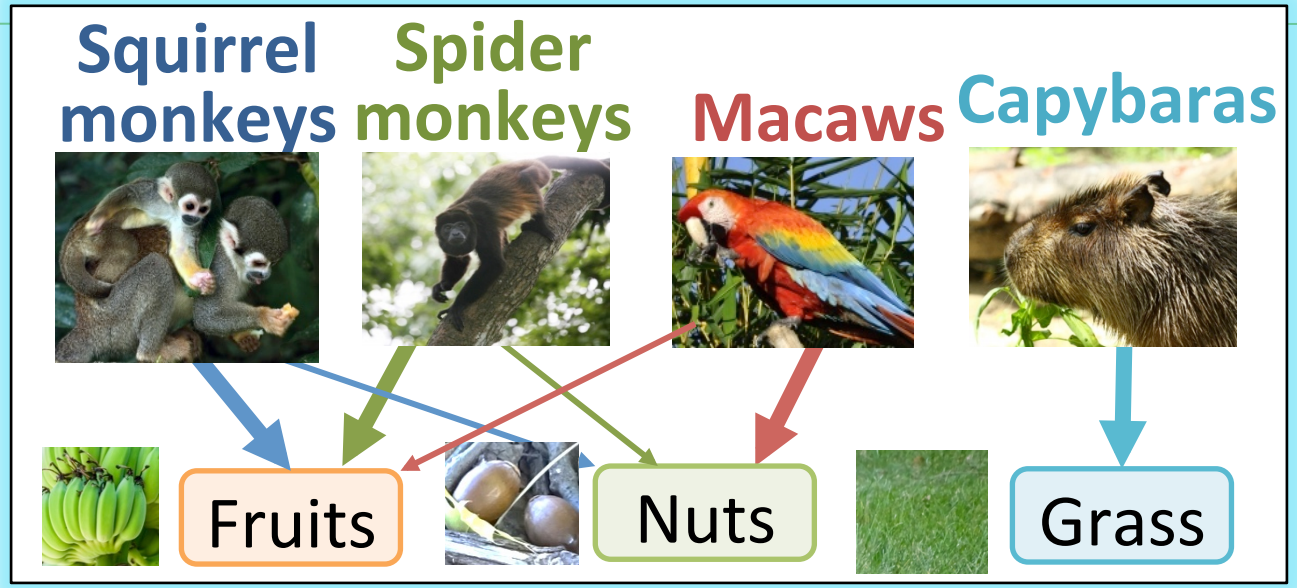
Q. How can we describe the evolutions of X ?



A. The Web as a jungle!

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

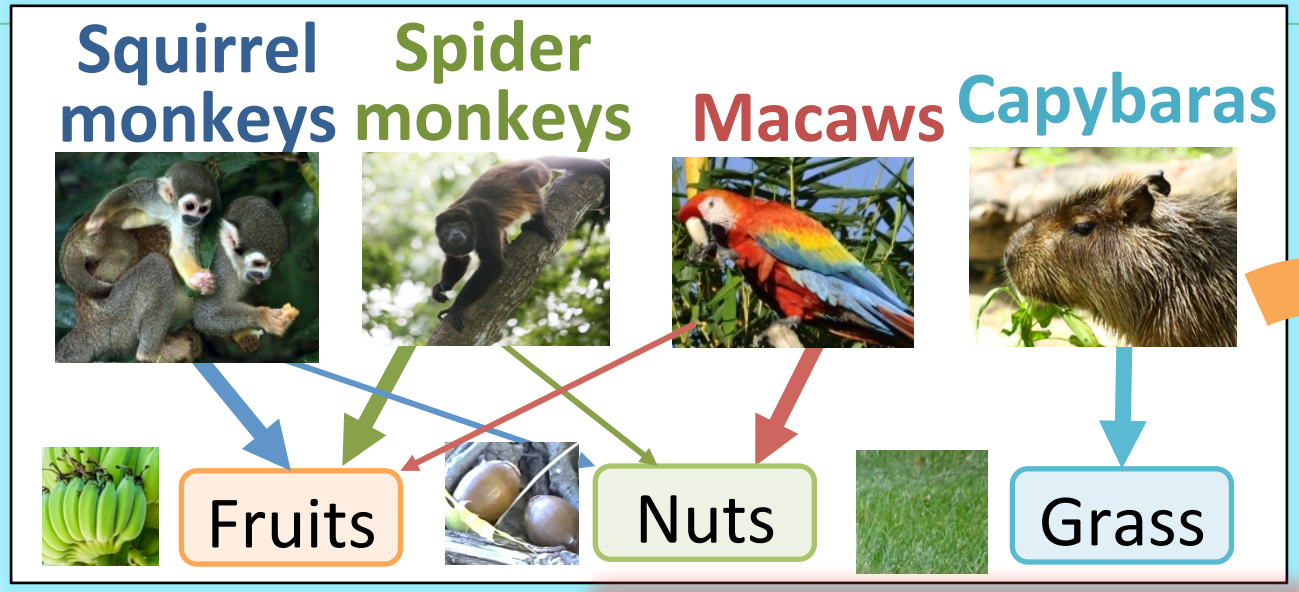
Main idea: the Web as a jungle



Ecosystem in the **Jungle**

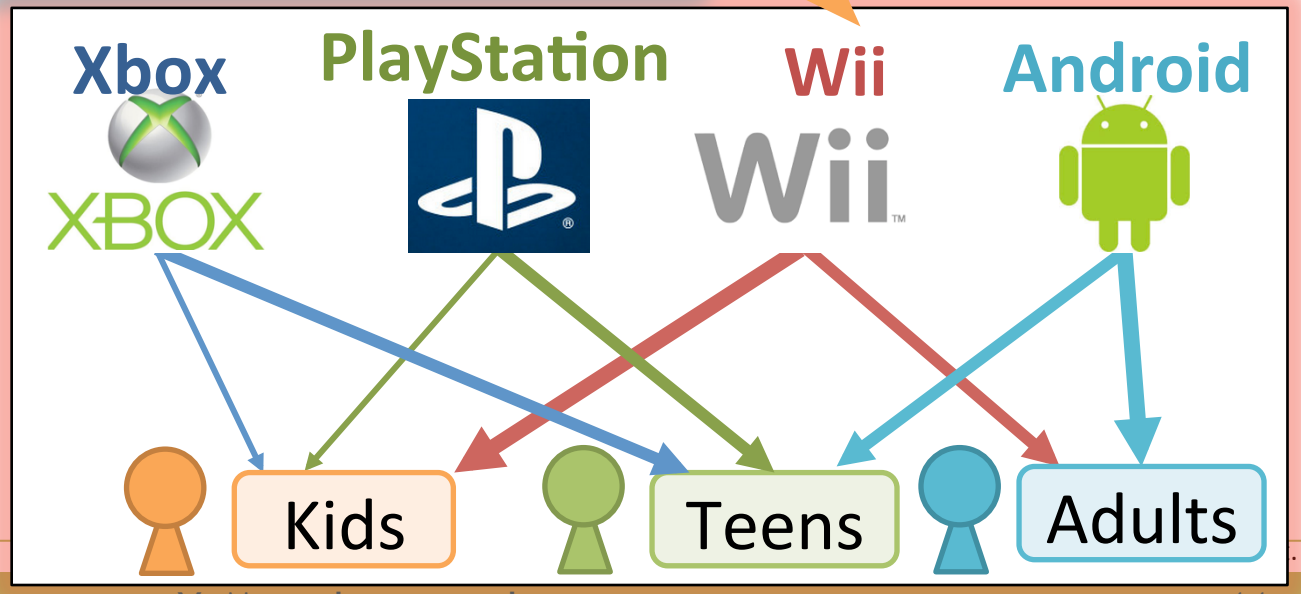
Image courtesy of xura, criminalatt, David Castillo Dominici, happykanppy at FreeDigitalPhotos.net.

Main idea: the Web as a jungle



Ecosystem in the **Jungle**

Ecosystem on the **Web**



Analogies: ecosystem on the Web

Biological species

Online activities

Jungle

Web

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

Analogies: ecosystem on the Web

Biological species

Food resources

Online activities

User resources

Jungle

Web

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

Analogies: ecosystem on the Web

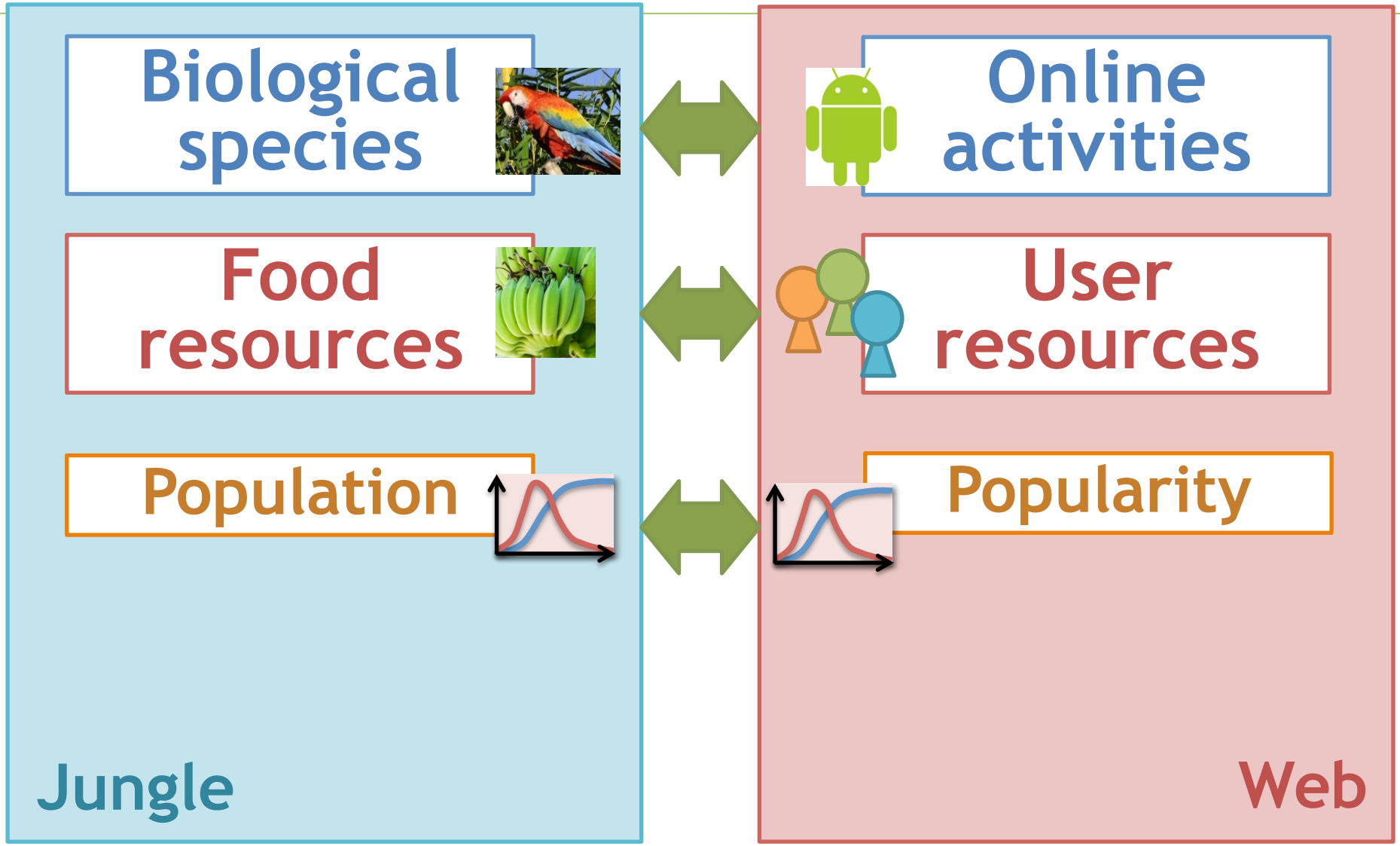


Image courtesy of xura, criminalatt, David Castillo Dominici, happykanppy at FreeDigitalPhotos.net.

Analogies: ecosystem on the Web

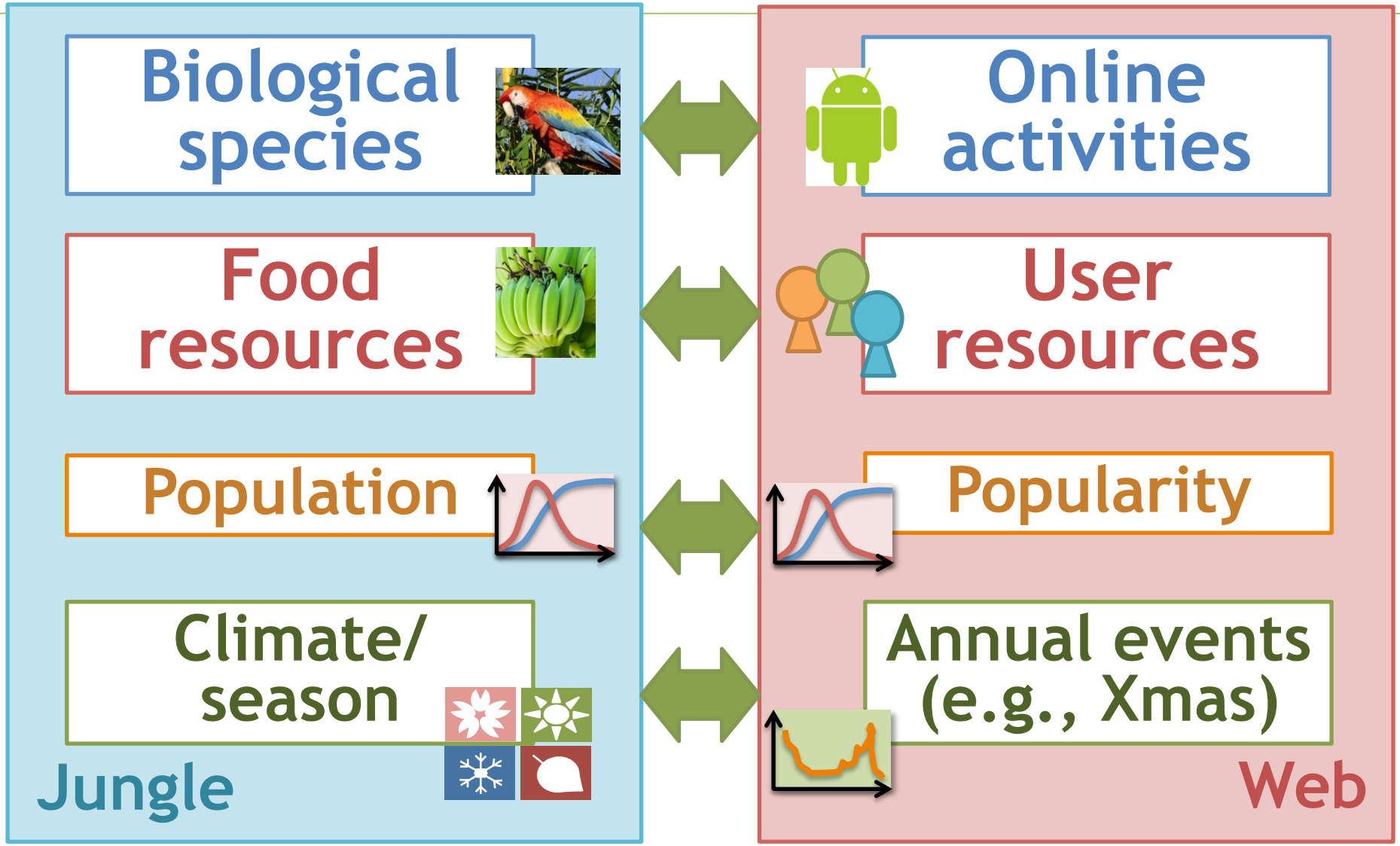
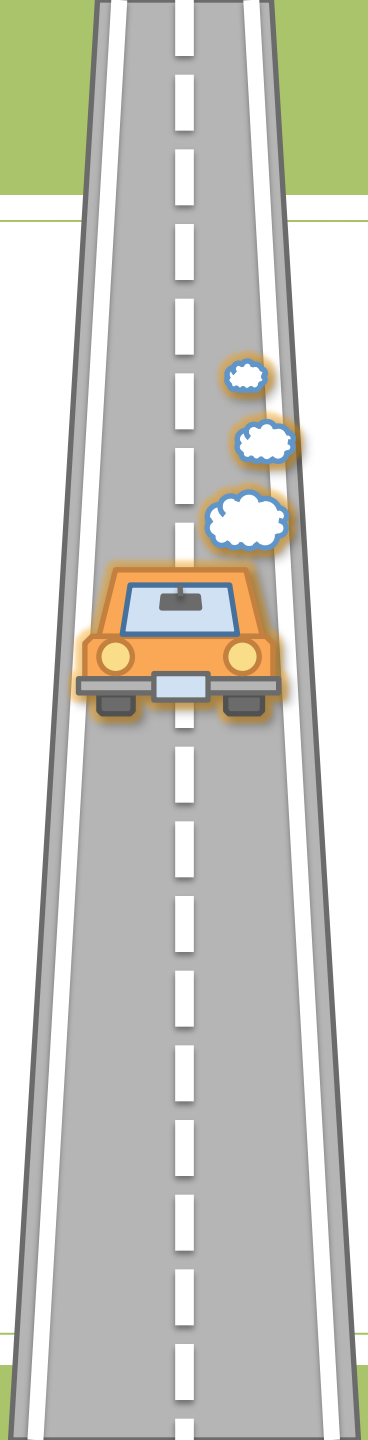


Image courtesy of xura, criminalatt, David Castillo Dominici, happykanppy at FreeDigitalPhotos.net.

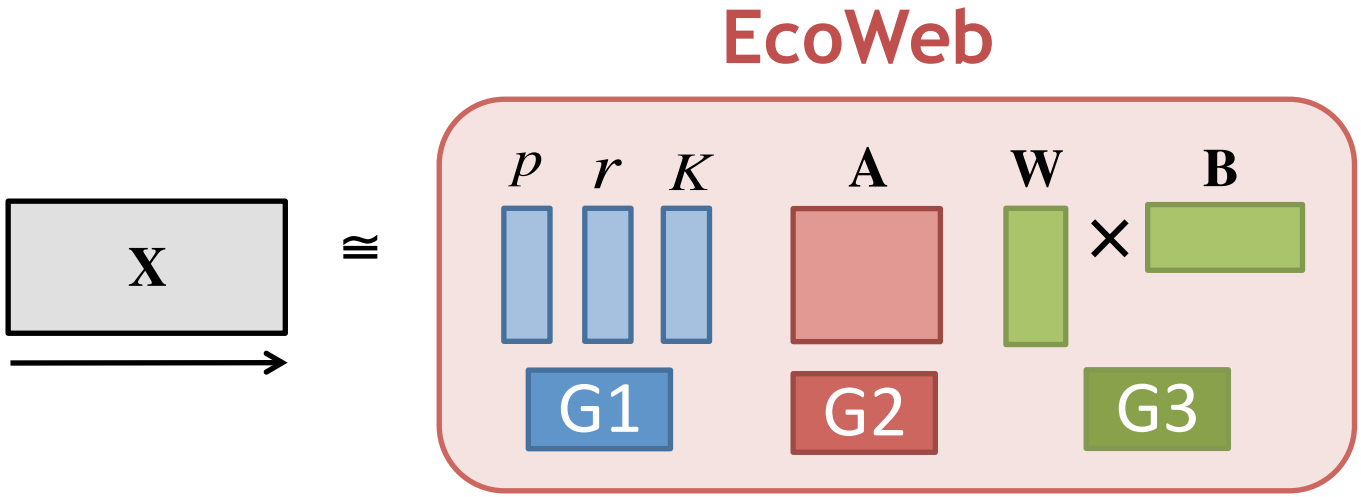
Roadmap

- ✔ Motivation
- ✔ Modeling power of EcoWeb
- ✔ Overview
- Proposed model
- Algorithm
- Experiments
- EcoWeb – at work
- Conclusions



EcoWeb: Main idea

Q. How can we describe the evolutions of X ?



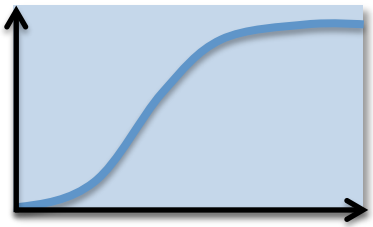
A. Web as a jungle!

G1 G2 G3

EcoWeb: Main idea

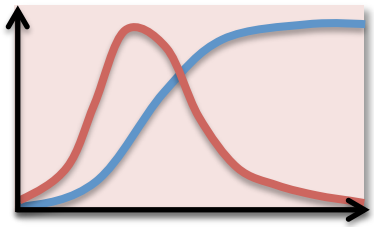
Q. How can we describe the evolutions of X ?

Non-linear evolution

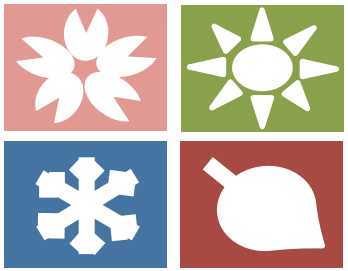


≅

Interaction/ competition



Seasonality

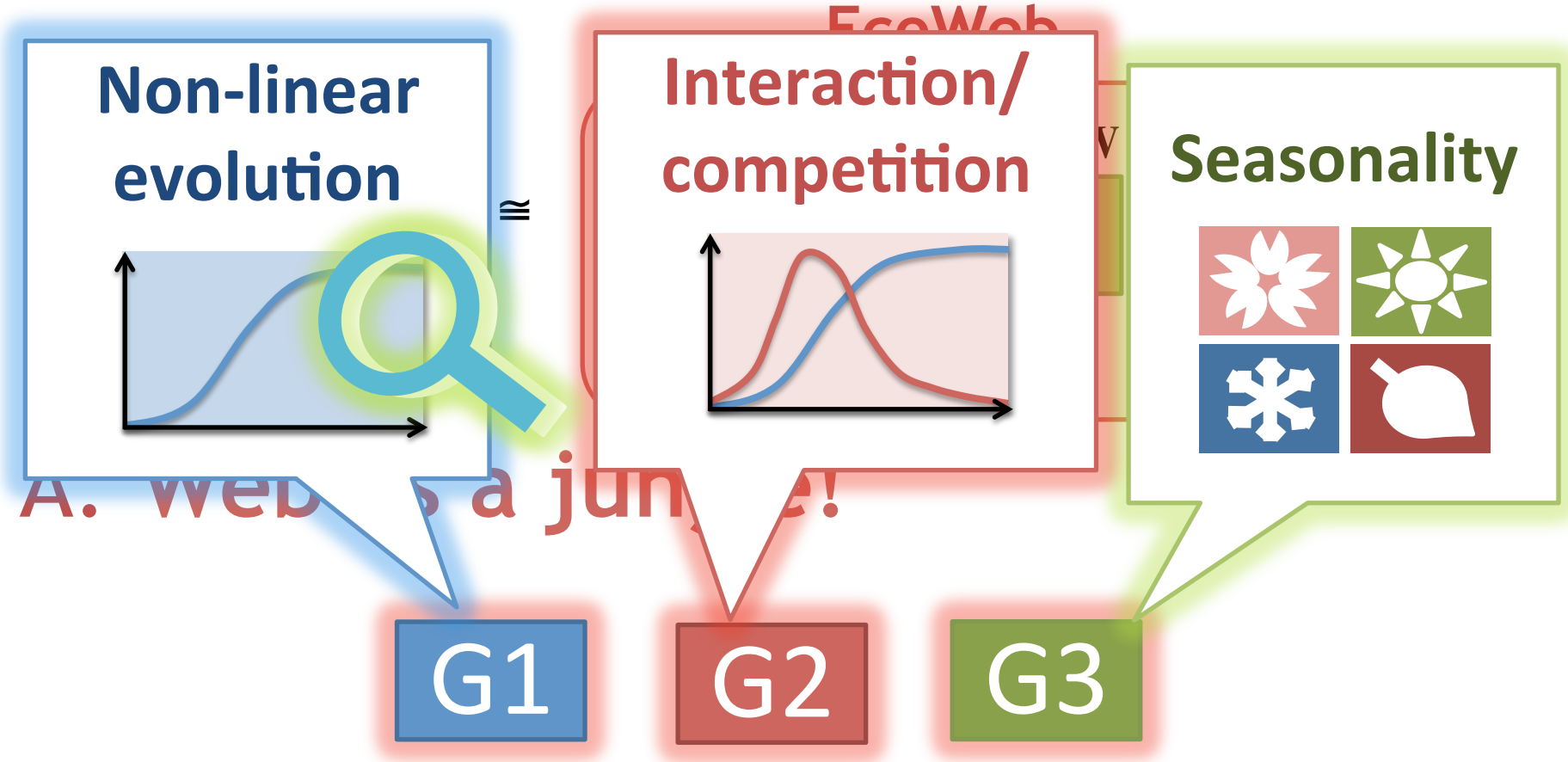


A. Web is a jungle!

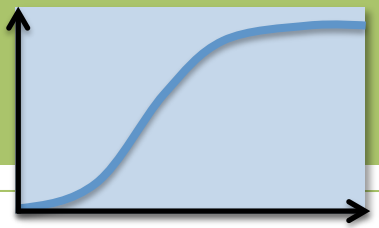
G1 G2 G3

EcoWeb: Main idea

Q. How can we describe the evolutions of X ?



G1 EcoWeb-individual



Non-linear evolution of a single keyword

Jungle

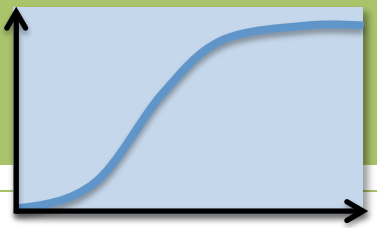
Species

Web

Keywords

t=0 t=1 t=2

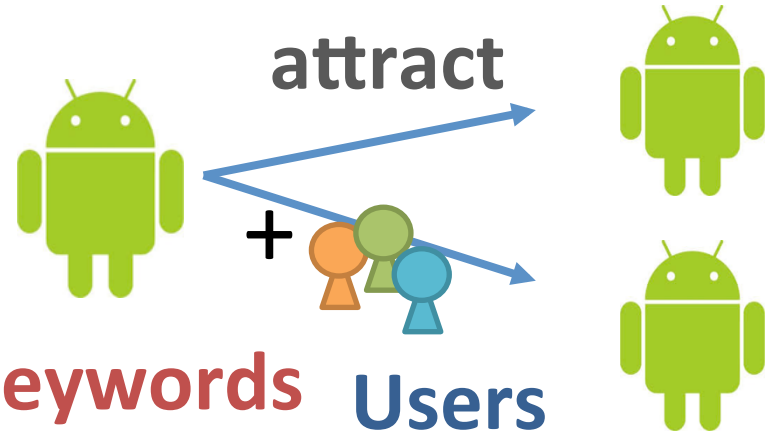
G1 EcoWeb-individual



Non-linear evolution of a single keyword

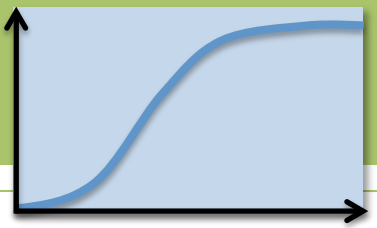
Jungle

Web

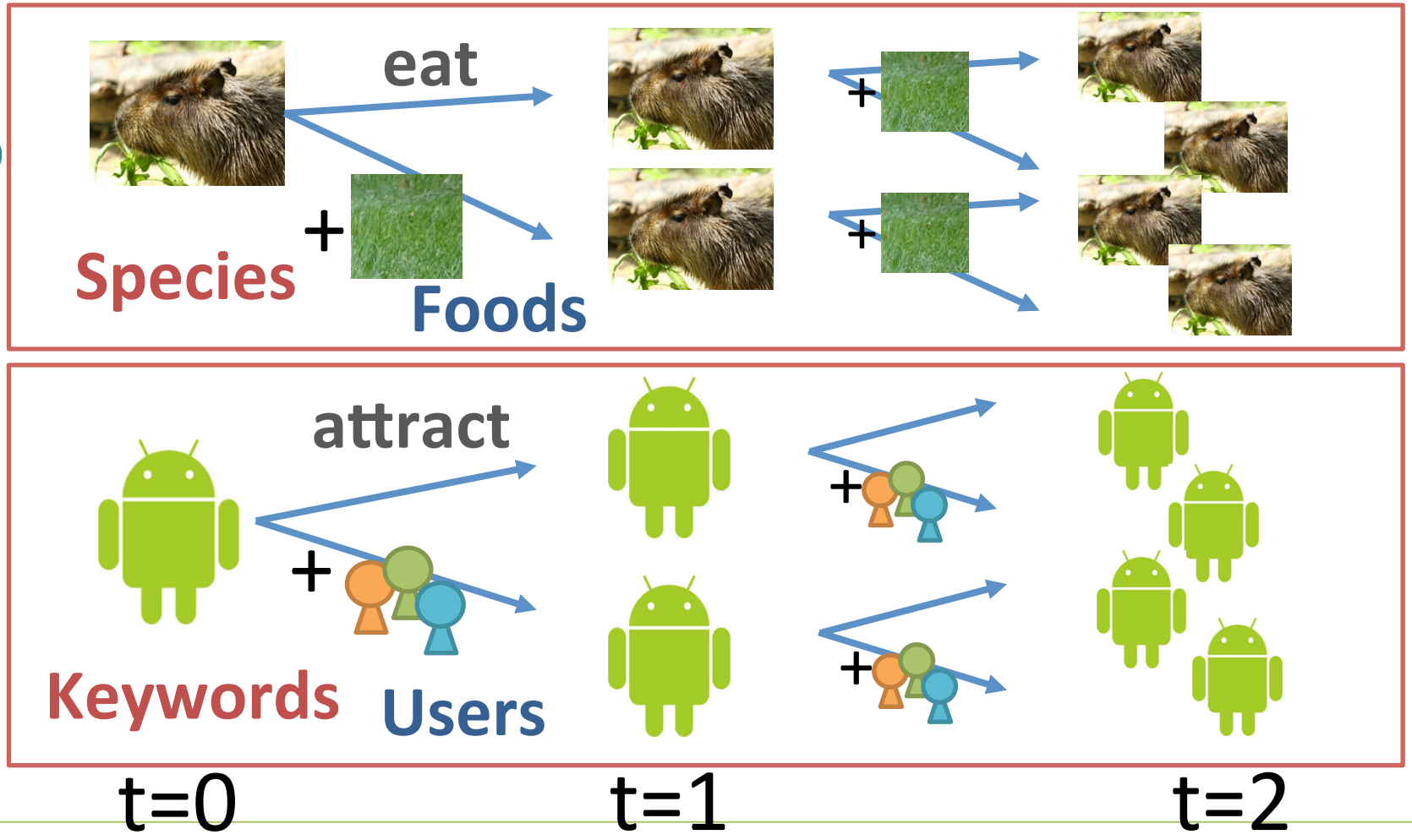


t=0 t=1 t=2

G1 EcoWeb-individual



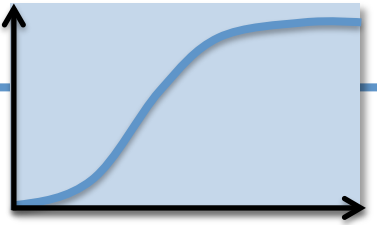
Popularity size increases over time



G1 EcoWeb-individual

Details

Non-linear evolution of a single keyword



Popularity size

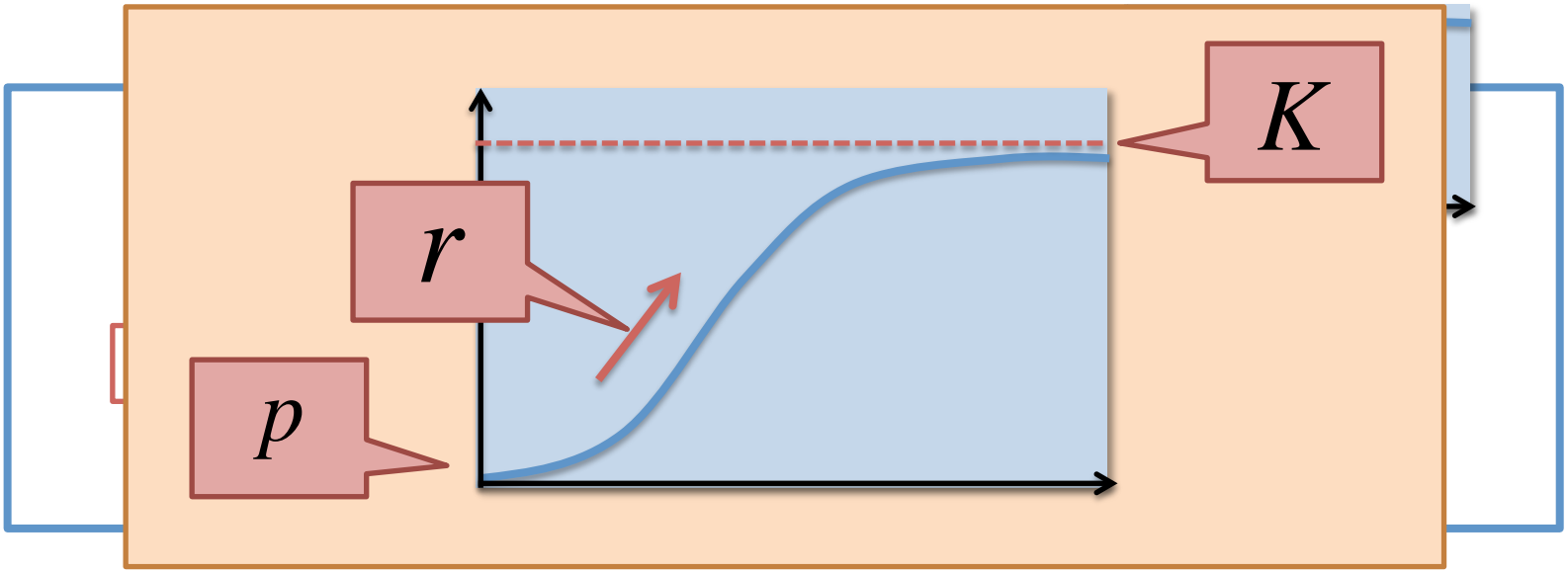
$$P(t+1) = P(t)\left[1 + r\left(1 - \frac{P(t)}{K}\right)\right],$$

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

G1 EcoWeb-individual

Details

Non-linear evolution of a single keyword



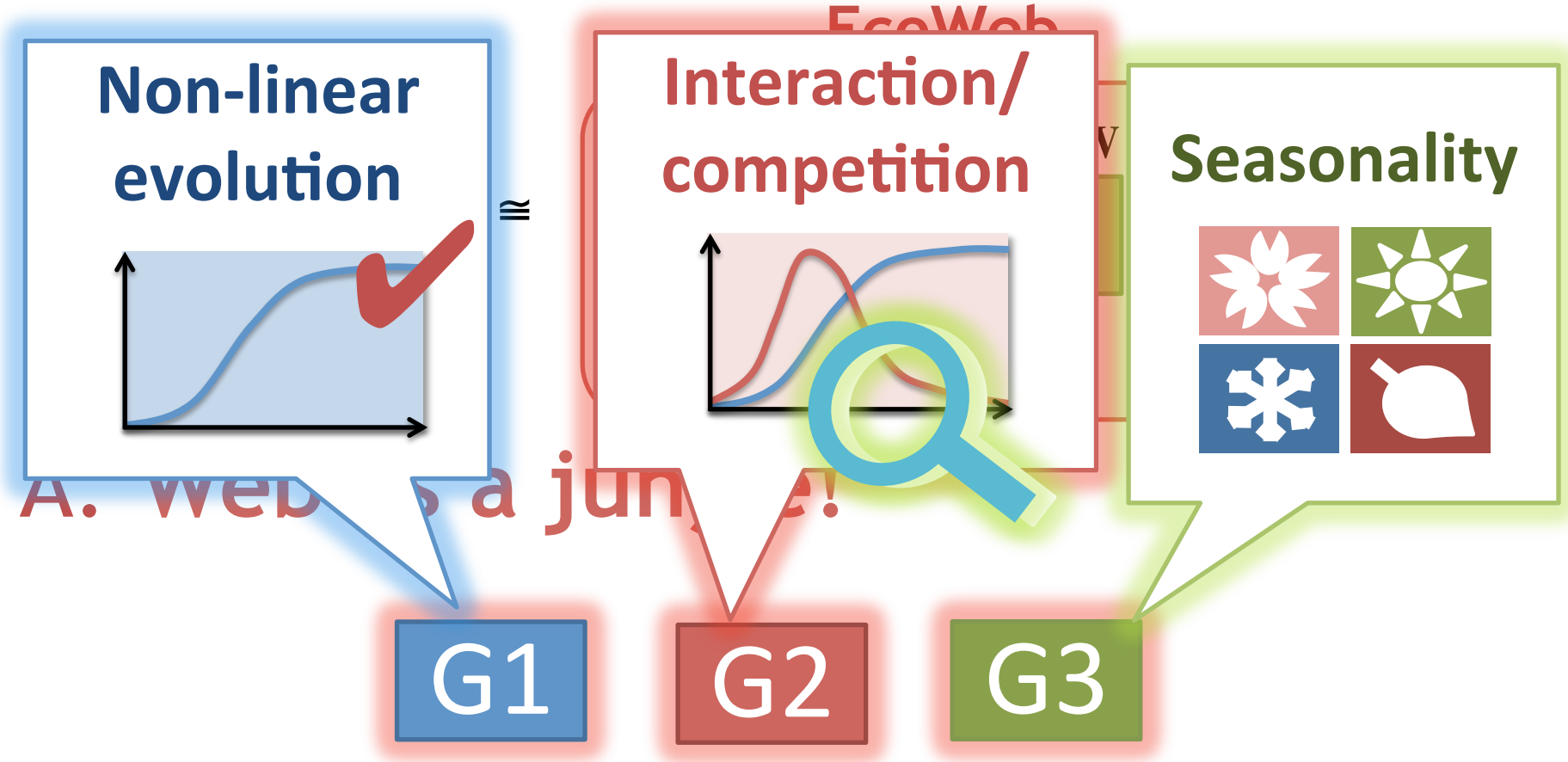
p – Initial condition (i.e., P(0) =p)

r – Growth rate, attractiveness

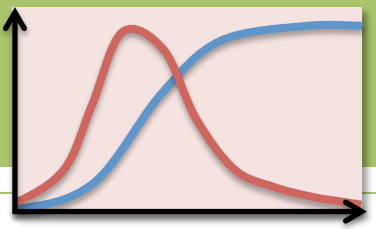
K – Carrying capacity (=available user resources)

EcoWeb: Main idea

Q. How can we describe the evolutions of X ?

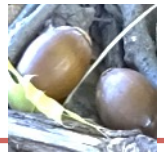


G2 EcoWeb-interaction



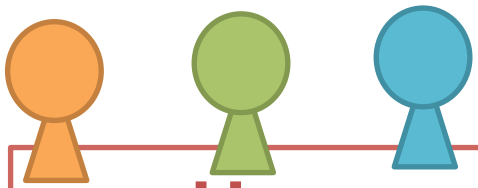
Interaction between multiple keywords

Species



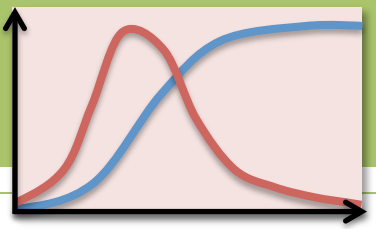
Food resources

Keywords

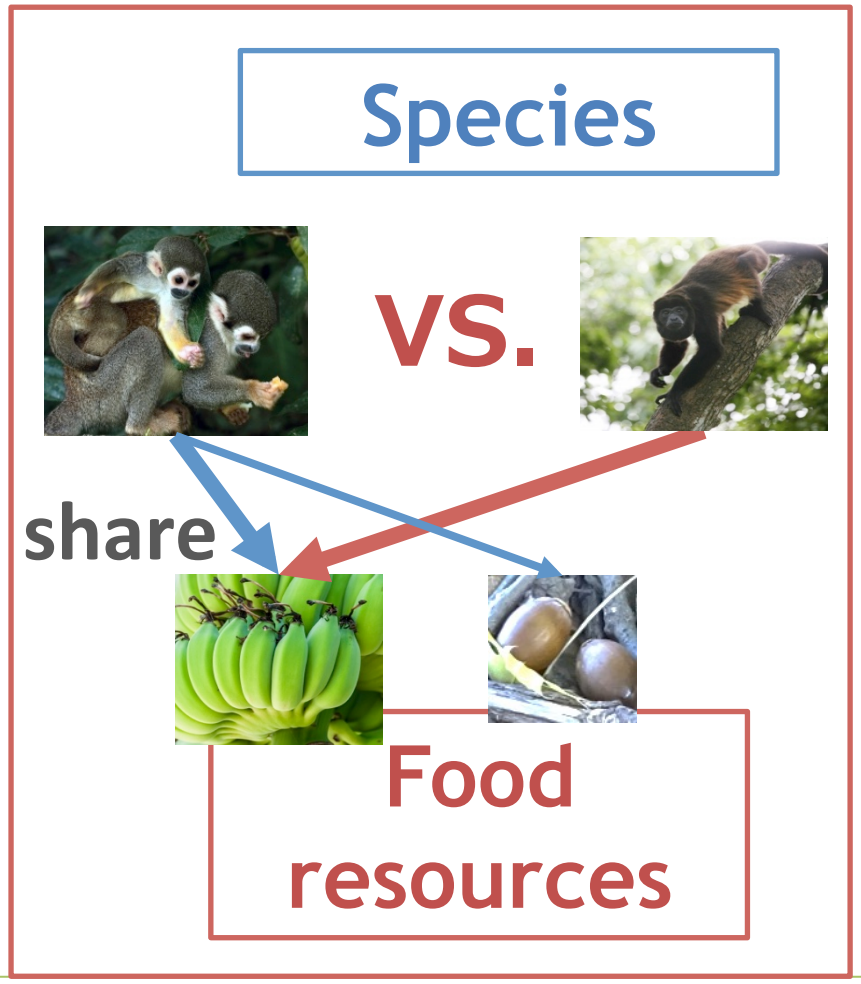


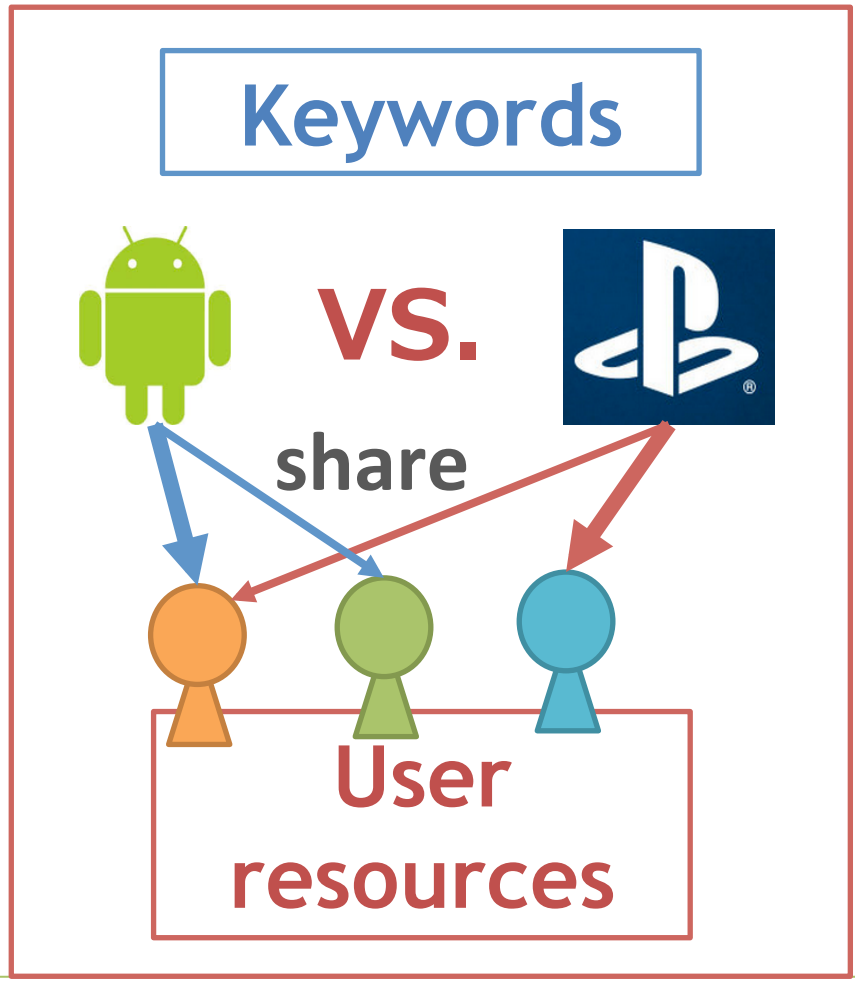
User resources

G2 EcoWeb-interaction



Interaction between multiple keywords





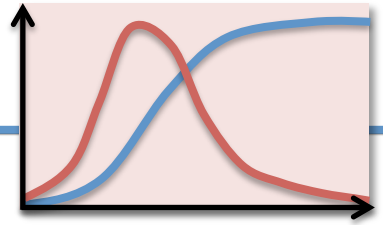
G2 EcoWeb-interaction

Details

Interaction between multiple keywords

Popularity of keyword i Popularity of j

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



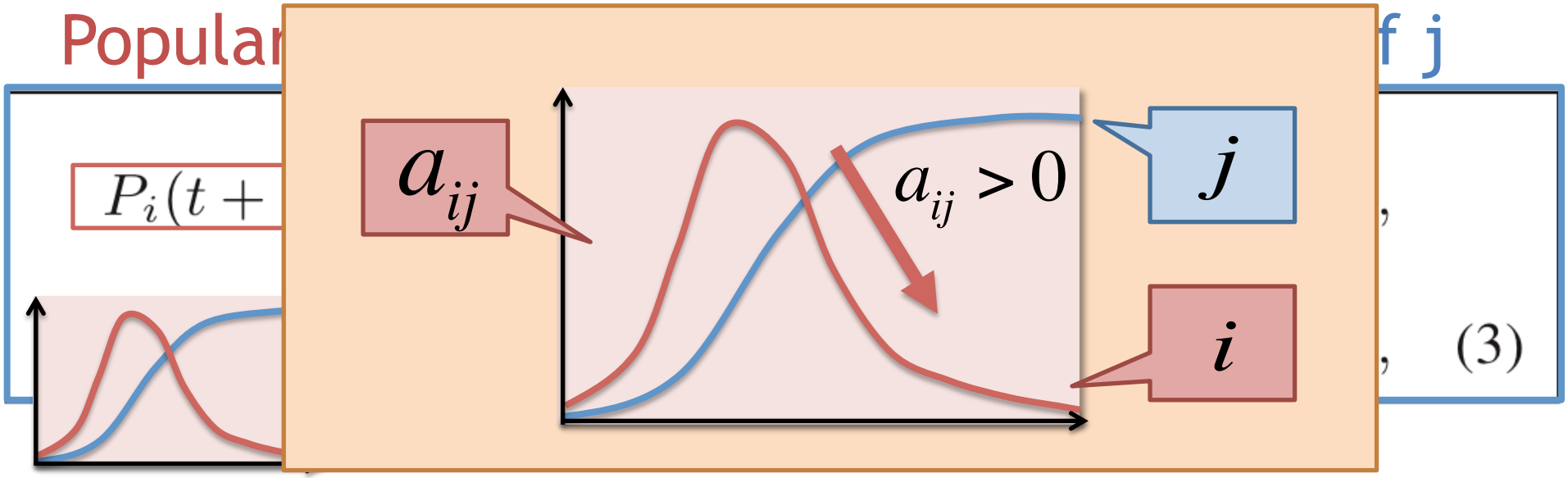
a_{ij} – Interaction coefficient

– i.e., effect rate of keyword j on i

G2 EcoWeb-interaction

Details

Interaction between multiple keywords

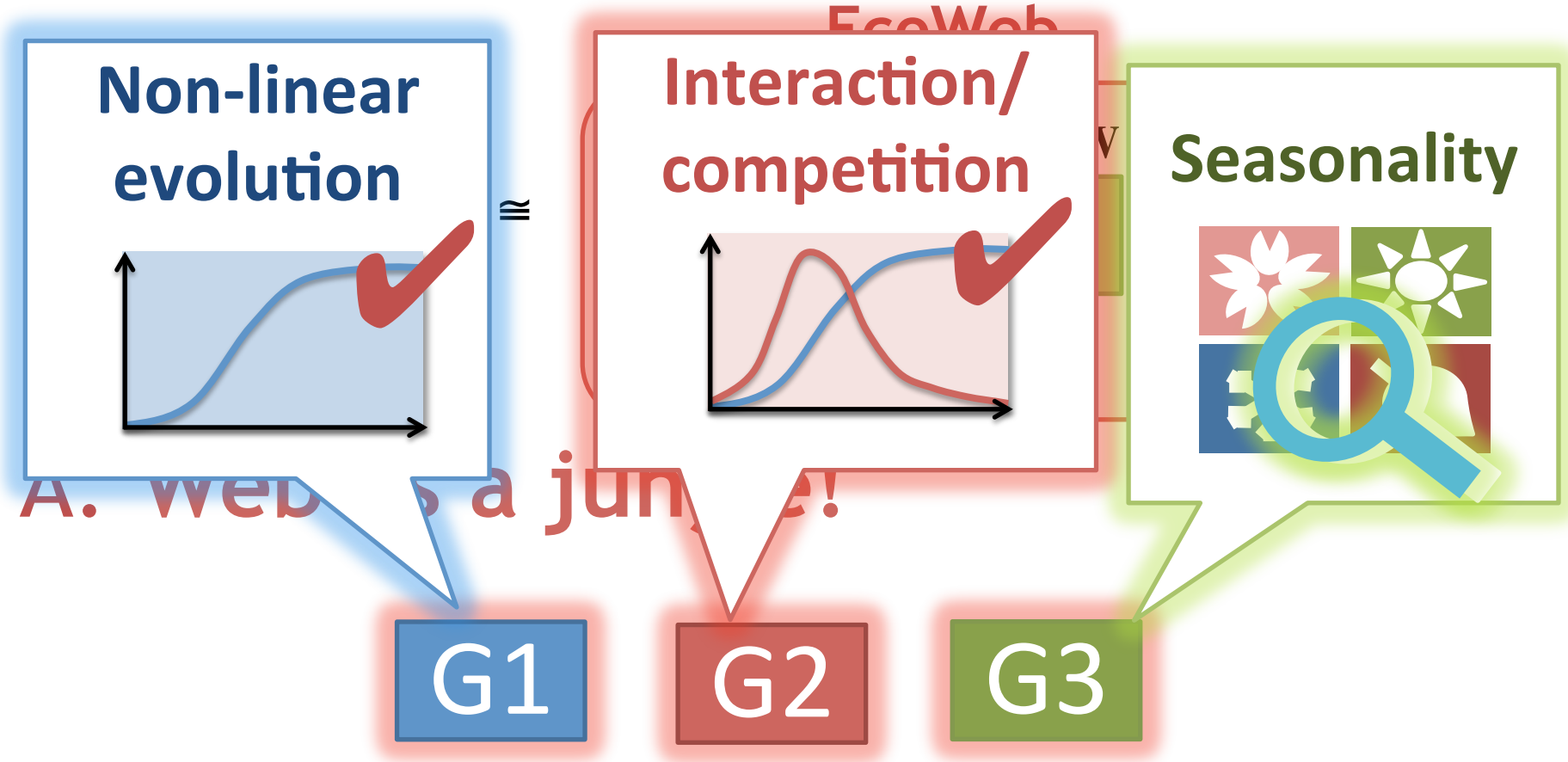


a_{ij} – Interaction coefficient

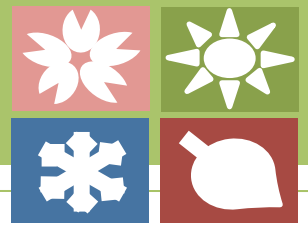
– i.e., effect rate of keyword j on i

EcoWeb: Main idea

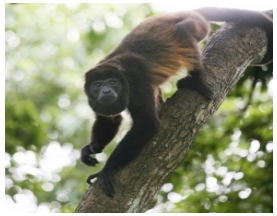
Q. How can we describe the evolutions of X ?



G3 EcoWeb-seasonality



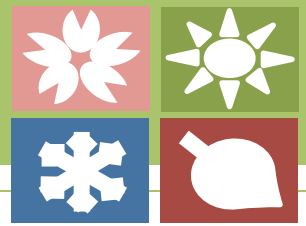
“Hidden” seasonal activities

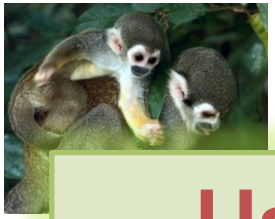
Season/ Climate

Seasonal events

G3 EcoWeb-seasonality



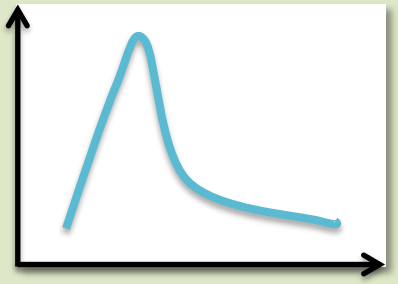
"Hidden" seasonal activities

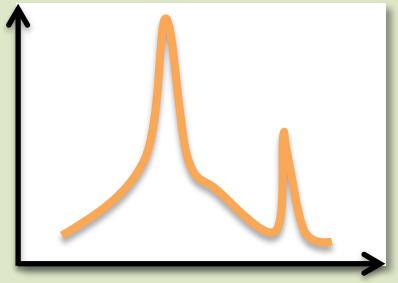


LOWE'S

amazon

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

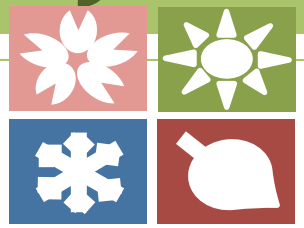




Climate

events

G3 EcoWeb-seasonality



Details

"Hidden" seasonal activities

Estimated volume of keyword i

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

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

Seasonal activities of i

$\mathbf{W}$ – Participation (weight) matrix

$\mathbf{B}$ – Seasonality matrix

G3 EcoWeb-seasonality

Details

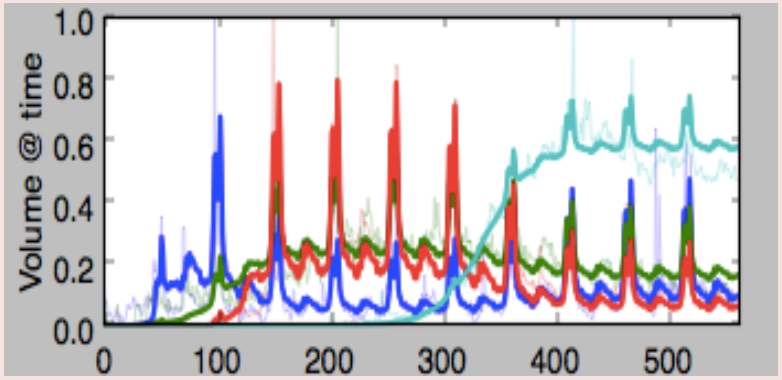
"Hidden" seasonal activities

Estimated volume of keyword i

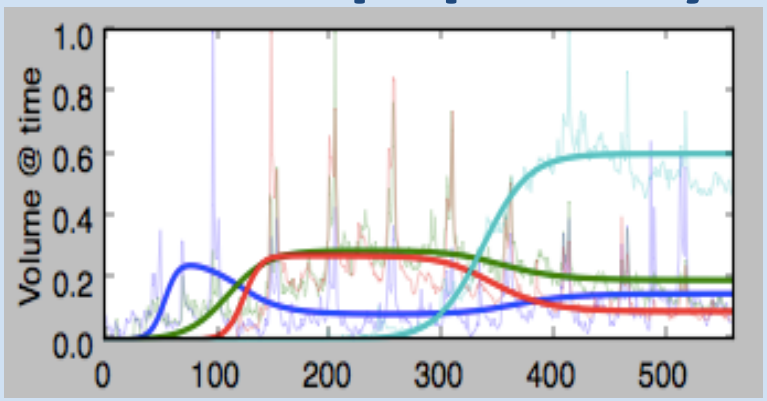
$$C_i(t) = P_i(t)\,[1 + e_i(t)] \qquad (i = 1, \cdots, d),$$

$$f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{1}^{k} w$$

C: volume



P: latent popularity



G3 EcoWeb-seasonality

Details

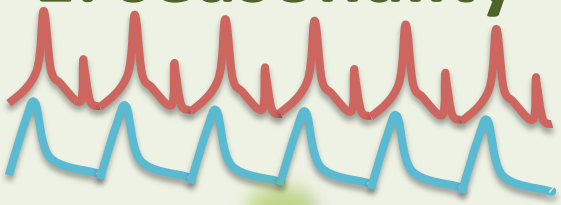
"Hidden" seasonal activities

Estimated volume of keyword i

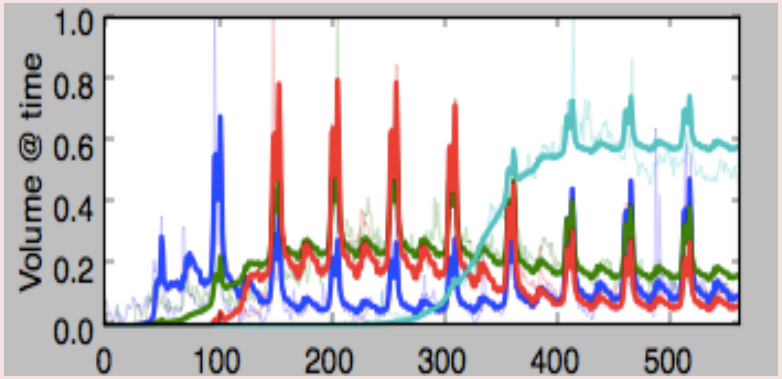
$$C_i(t) = P_i(t)\,[1 + e_i(t)]$$

$$f(i, t|\mathbf{W}, \mathbf{B}) = \sum_{1}^{k} w$$

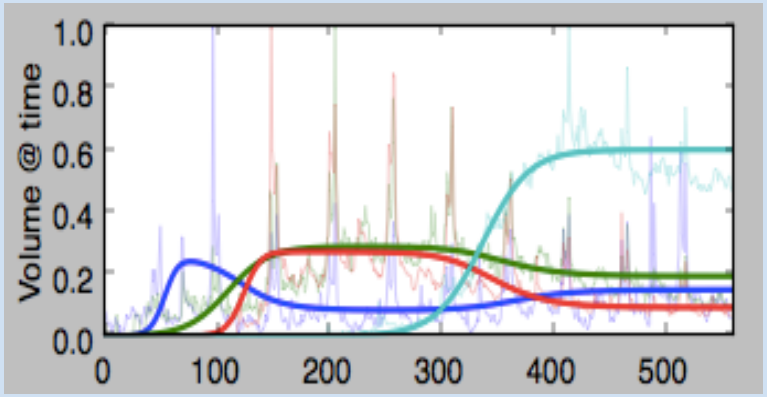
E: seasonality



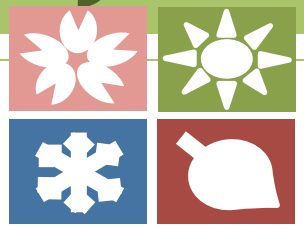
C: volume



P: latent popularity



G3 EcoWeb-seasonality



Details

"Hidden" seasonal activities

Estimated volume of keyword i

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

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

Seasonal activities of keyword i

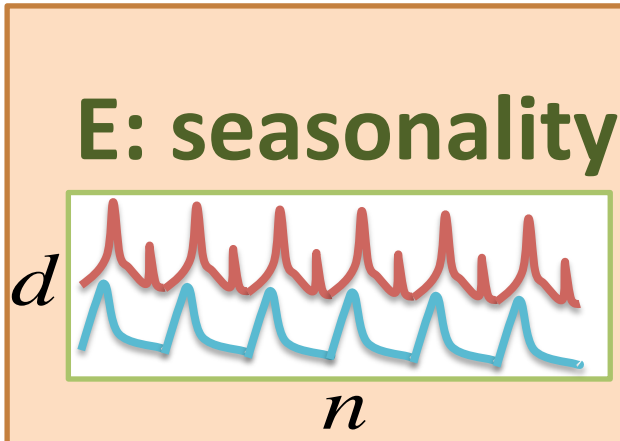
$\mathbf{W}$ – Participation (weight) matrix

$\mathbf{B}$ – Seasonality matrix

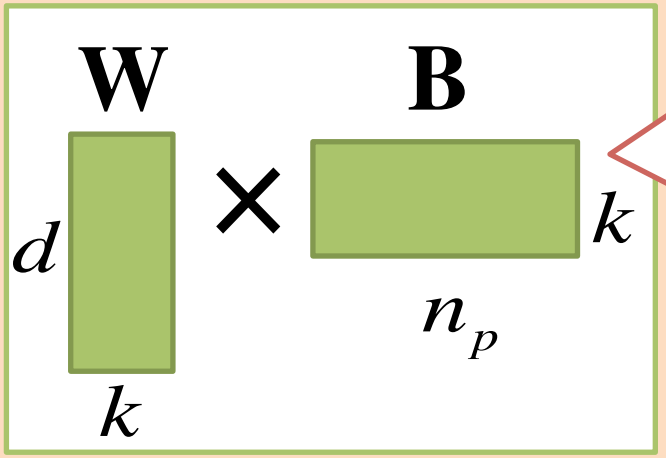
G3 EcoWeb-seasonality

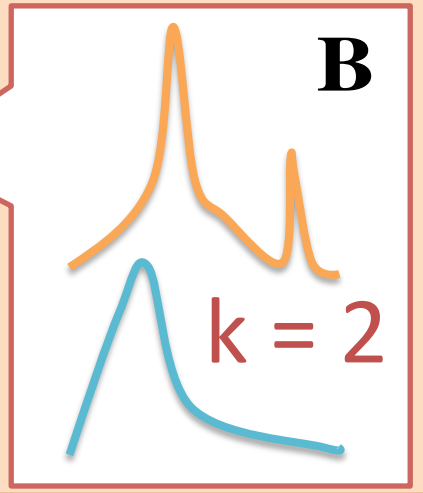
Details

E: seasonality



=





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

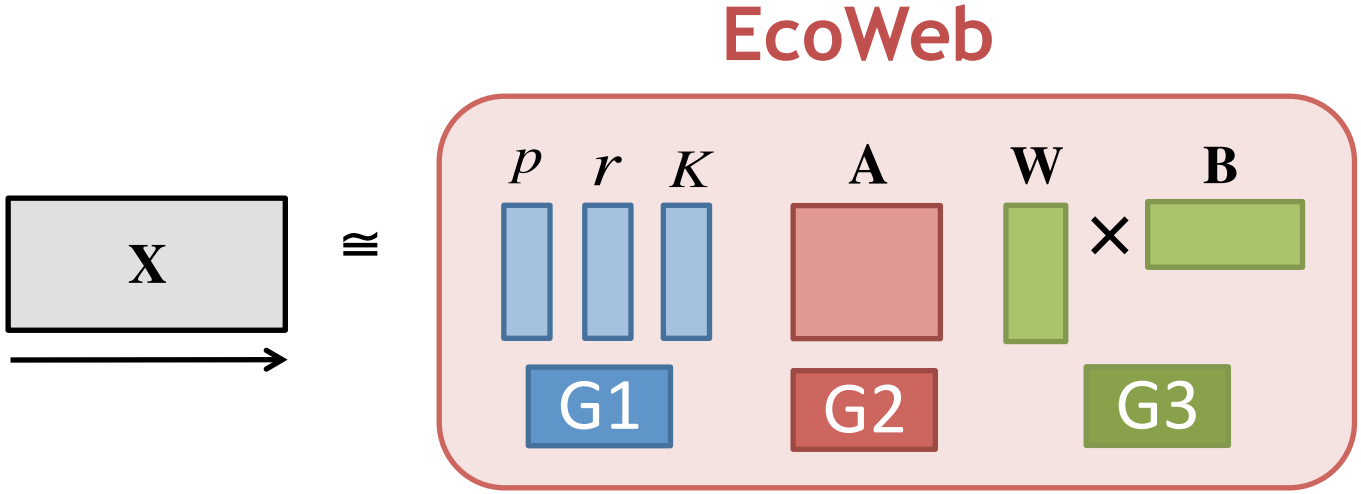
Seasonal activities of keyword i

W – Participation (weight) matrix

B – Seasonality matrix

EcoWeb: Main idea

Q. How can we describe the evolutions of X ?

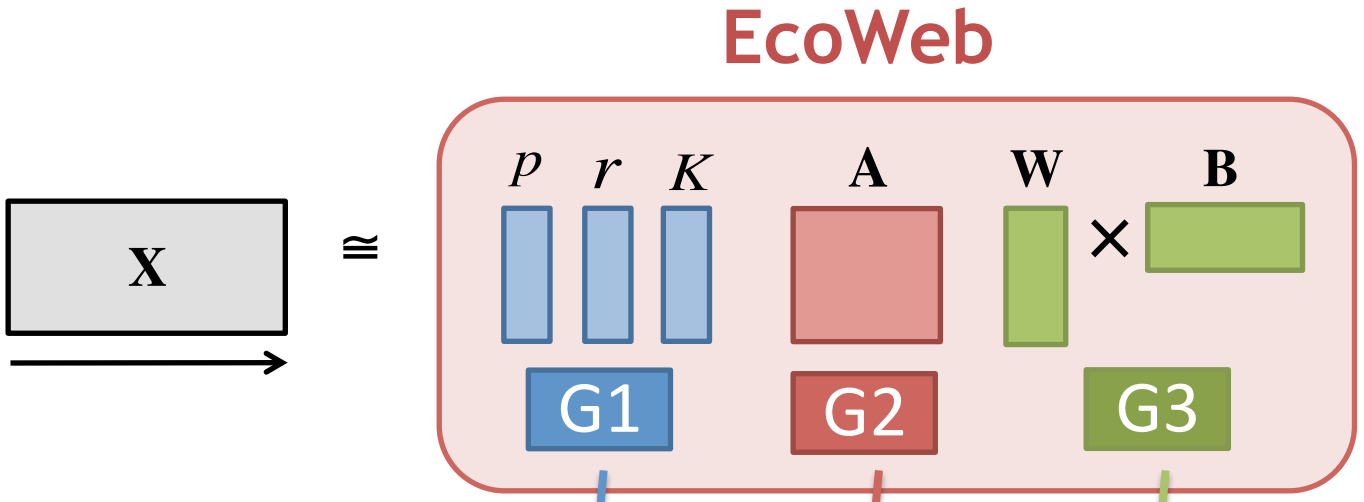


Full parameters (6)

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

EcoWeb: Main idea

Q. How can we describe the evolutions of X ?

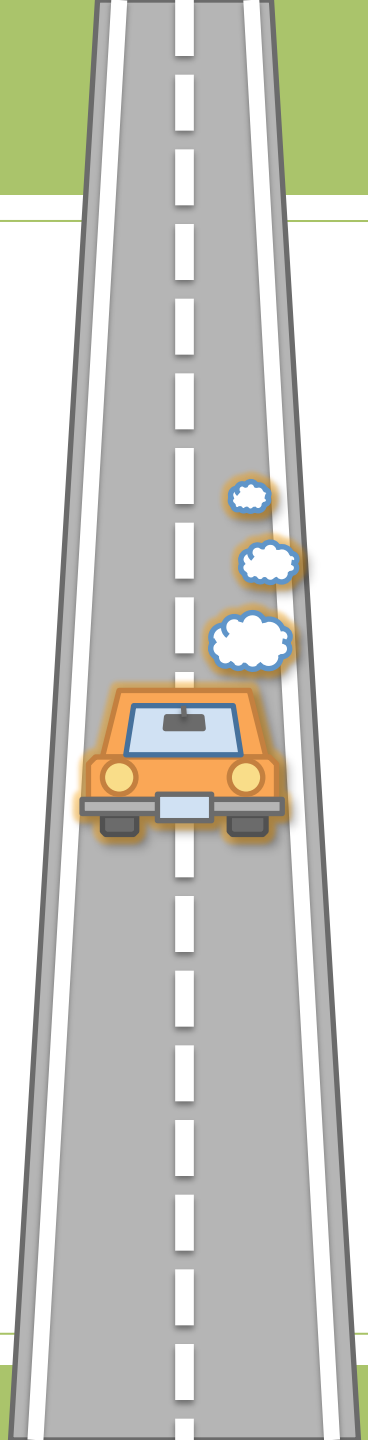


Full parameters

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

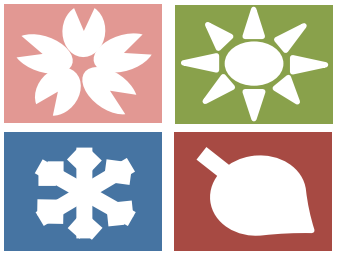
Roadmap

- ✔ Motivation
- ✔ Modeling power of EcoWeb
- ✔ Overview
- ✔ Proposed model
  - Algorithm
  - Experiments
  - EcoWeb – at work
  - Conclusions

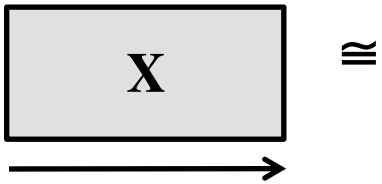


Challenges

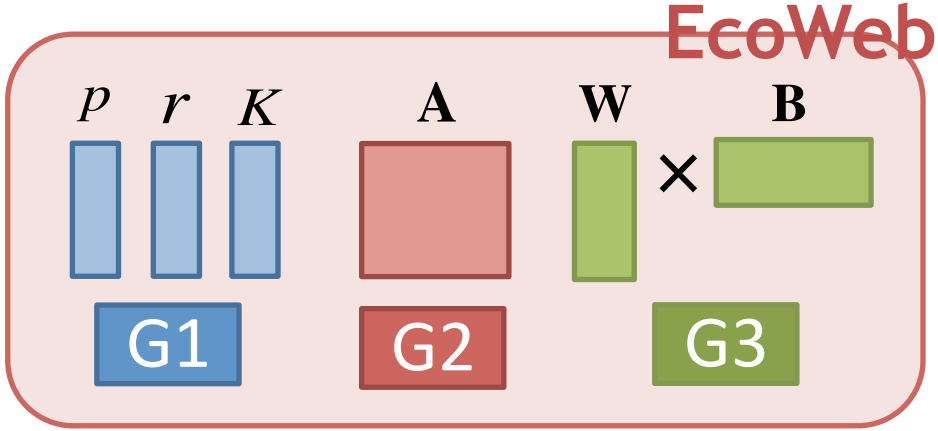
Q1. How can we automatically find "seasonal components" ?



Q2. How can we efficiently estimate full-parameters ?

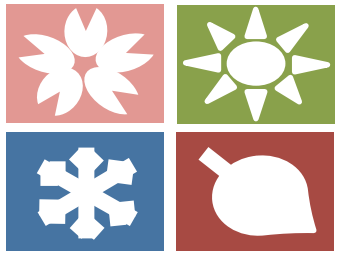


≅



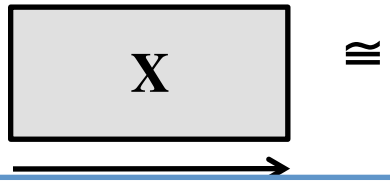
Challenges

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

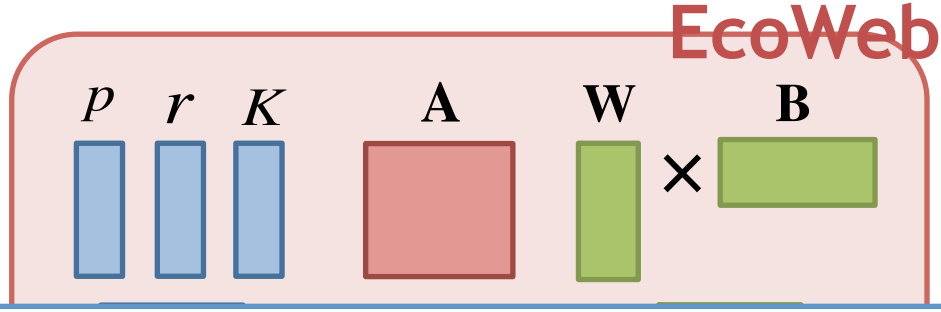


Idea (1) : Seasonal component analysis

Q2. How can we efficiently estimate full-parameters ?



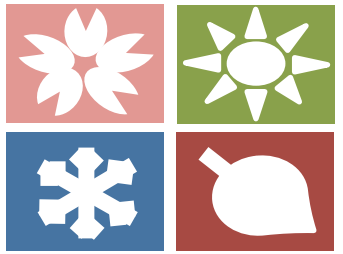
≅



Idea (2): Multi-step fitting

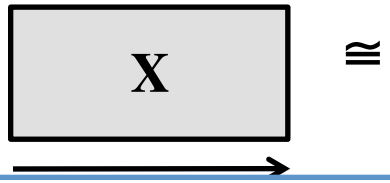
Challenges

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

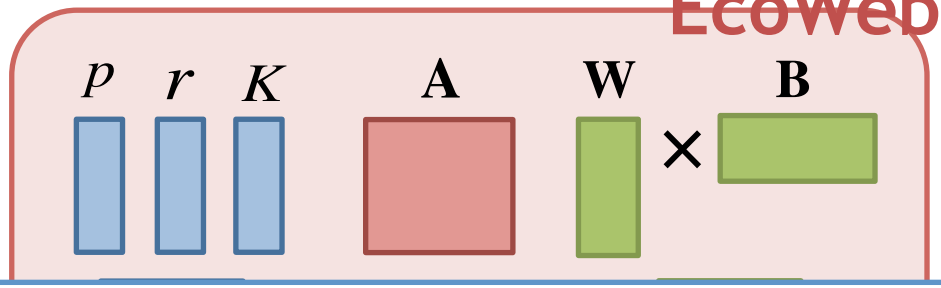


Idea (1) : Seasonal component analysis

Q2. How can we efficiently estimate full-parameters ?



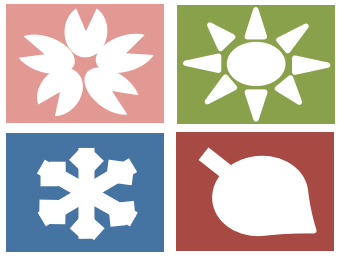
≅

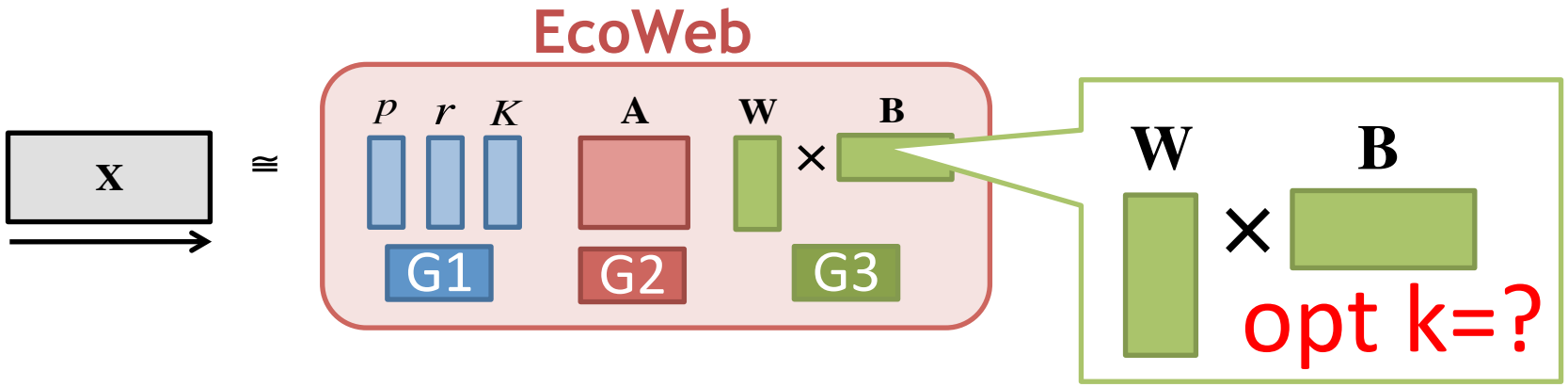


Idea (2): Multi-step fitting

Idea (1): Seasonal component analysis

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





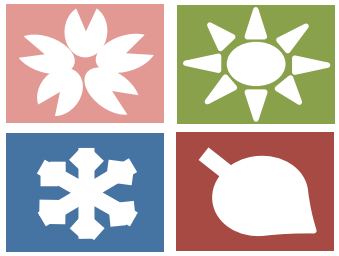
Idea (1) :

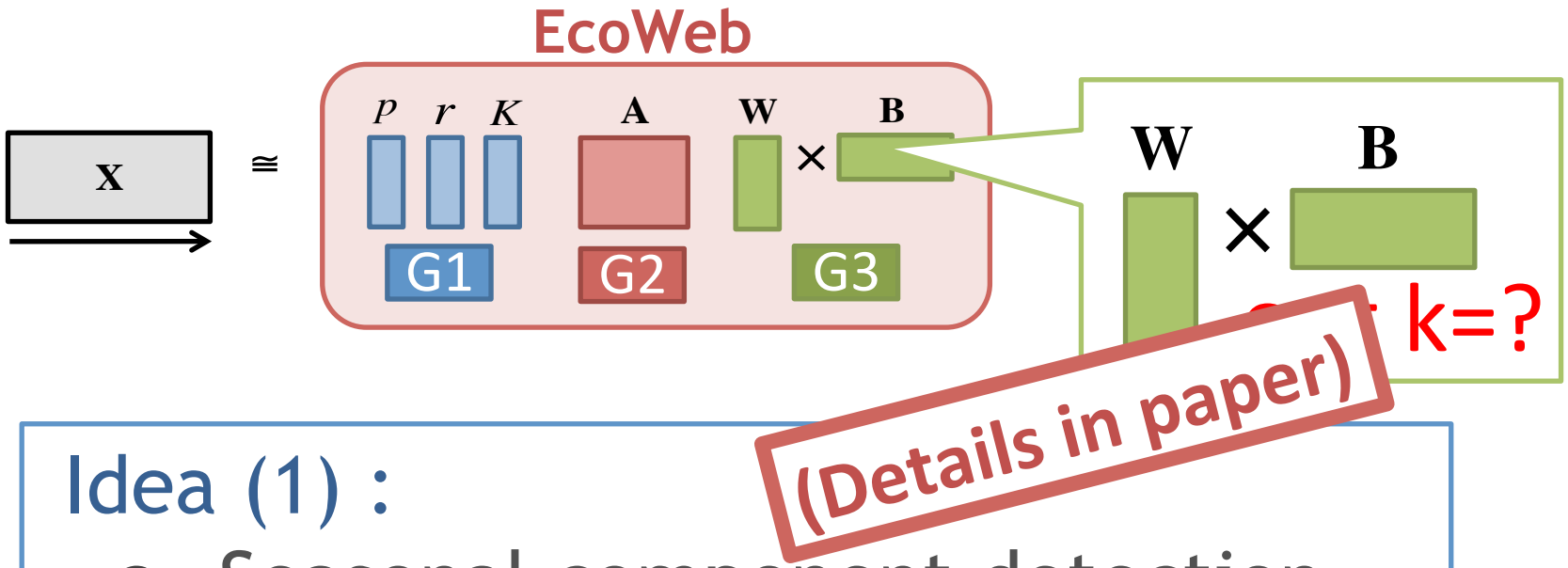
a. Seasonal component detection

b. Automatic component analysis

Idea (1): Seasonal component analysis

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





Idea (1) :

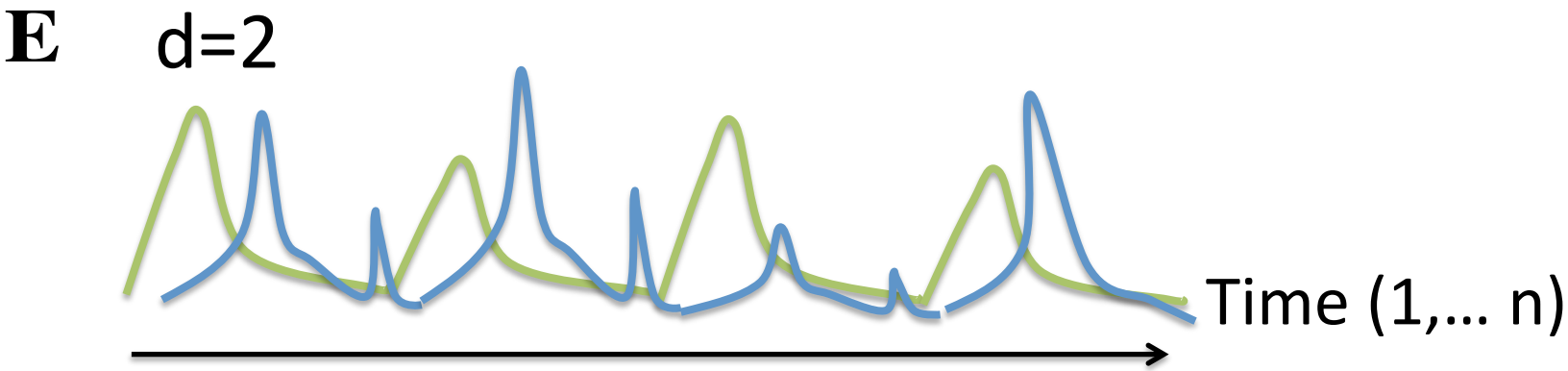
a. Seasonal component detection

b. Automatic component analysis

Idea (1): Seasonal component analysis

Details

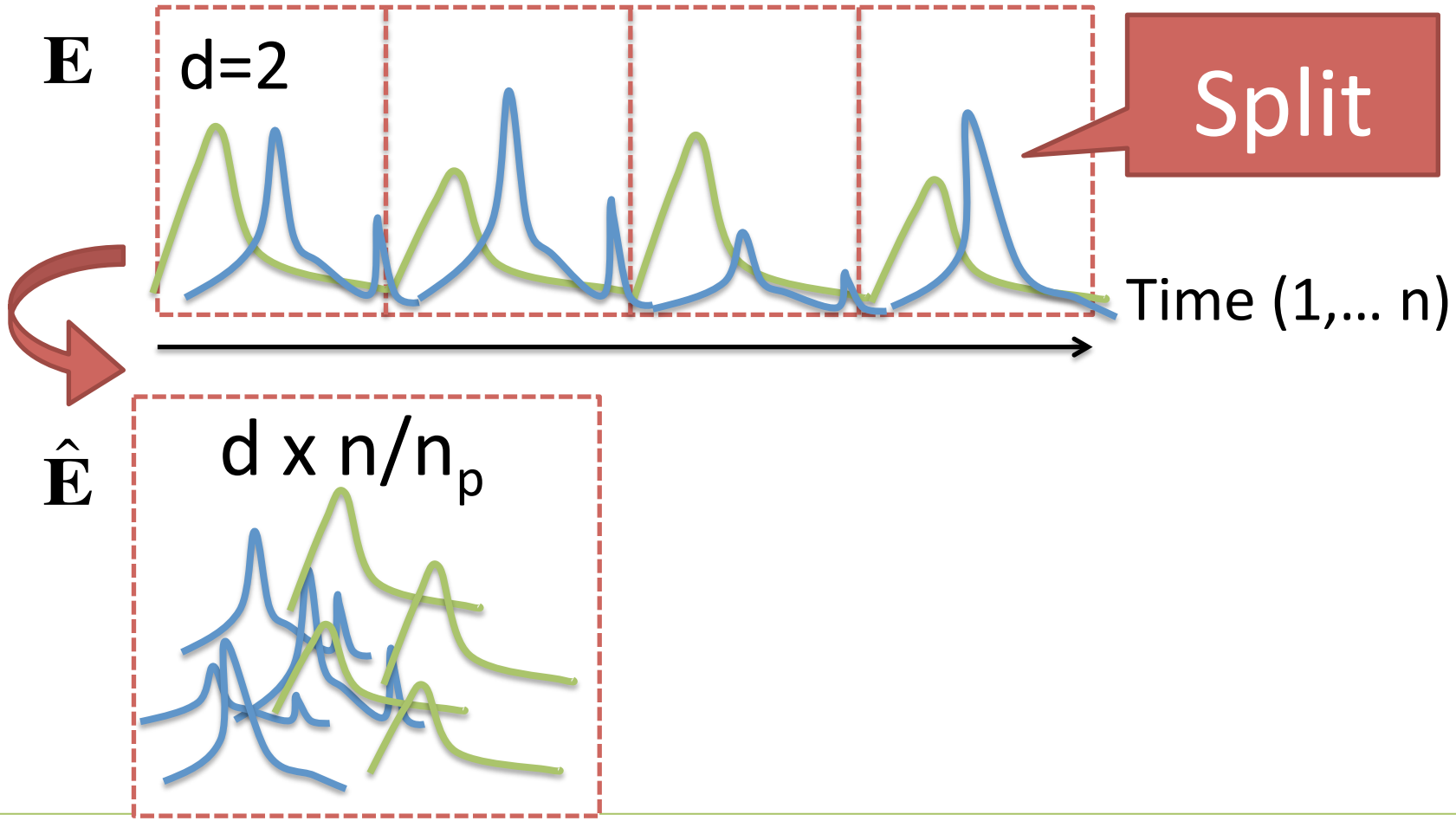
Idea(1-a) Seasonal component detection



Idea (1): Seasonal component analysis

Details

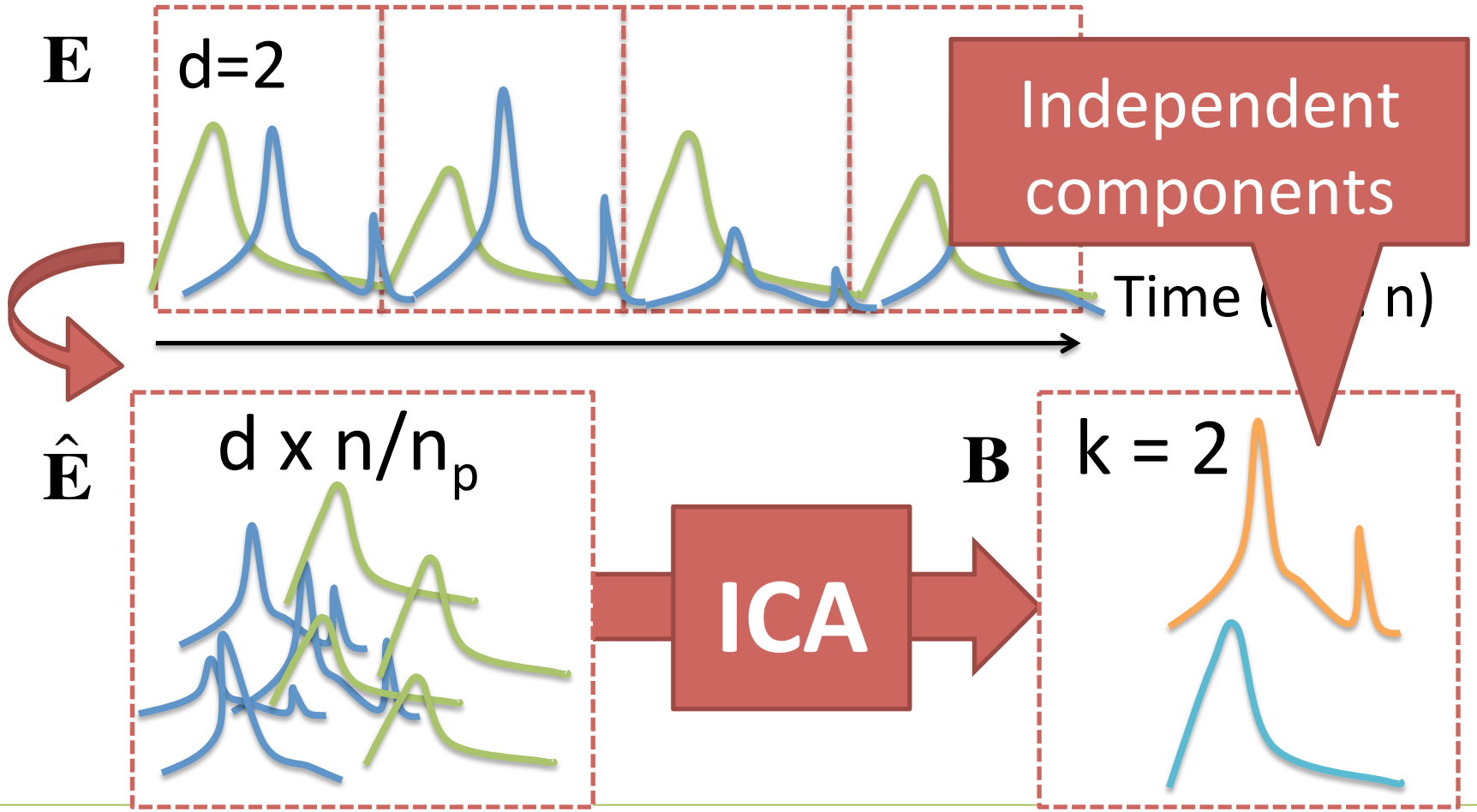
Idea(1-a) Seasonal component detection



Idea (1): Seasonal component analysis

Details

Idea(1-a) Seasonal component detection



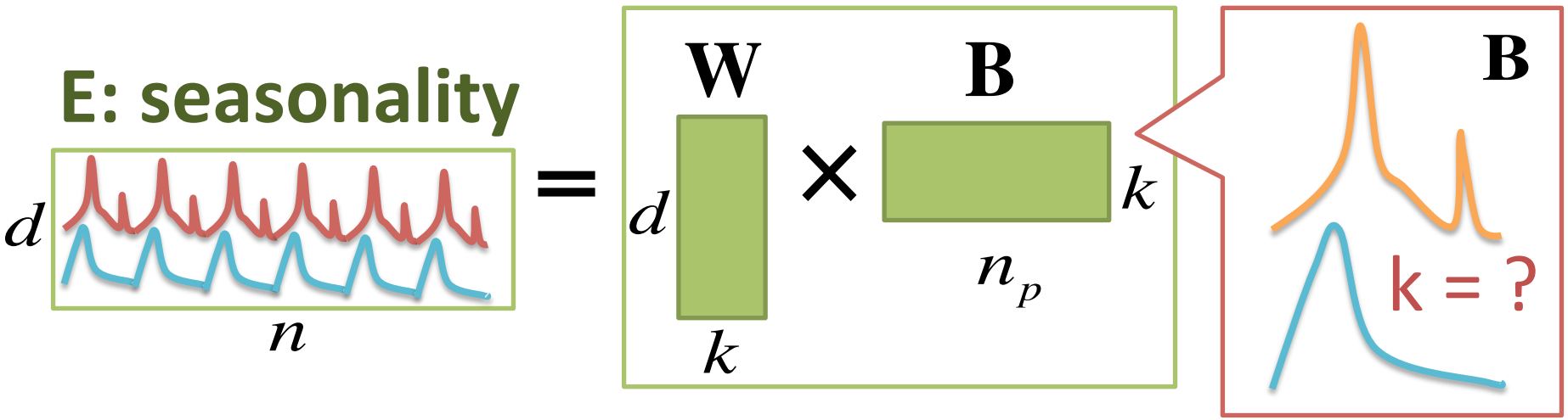
Idea (1): Seasonal component analysis

Details

Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

d: dimension



opt k=?

Idea (1): Seasonal component analysis

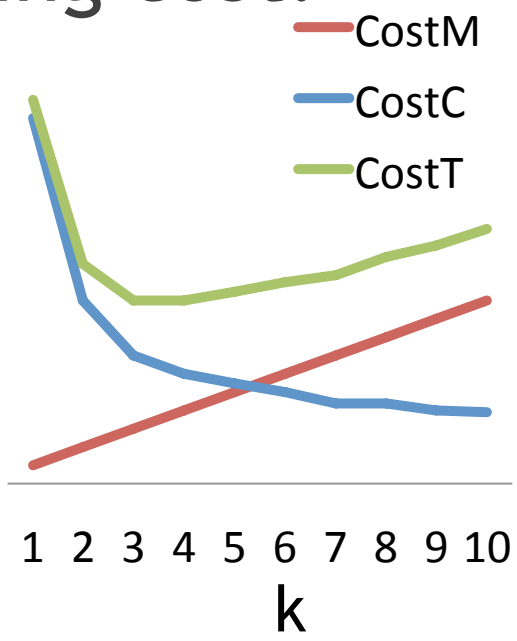
Details

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

$$\min \left(\boxed{\text{Cost}_M(S)} + \boxed{\text{Cost}_c(\mathcal{X} \mid S)} \right)$$

Model cost

Coding cost



Good compression

Good description

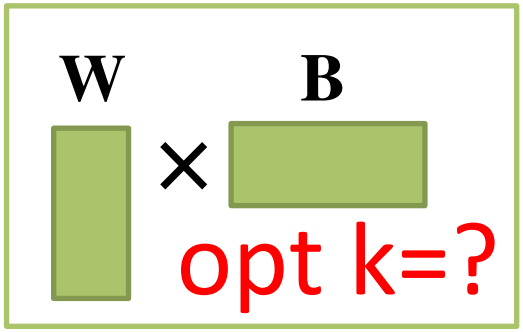
Idea (1): Seasonal component analysis

Details

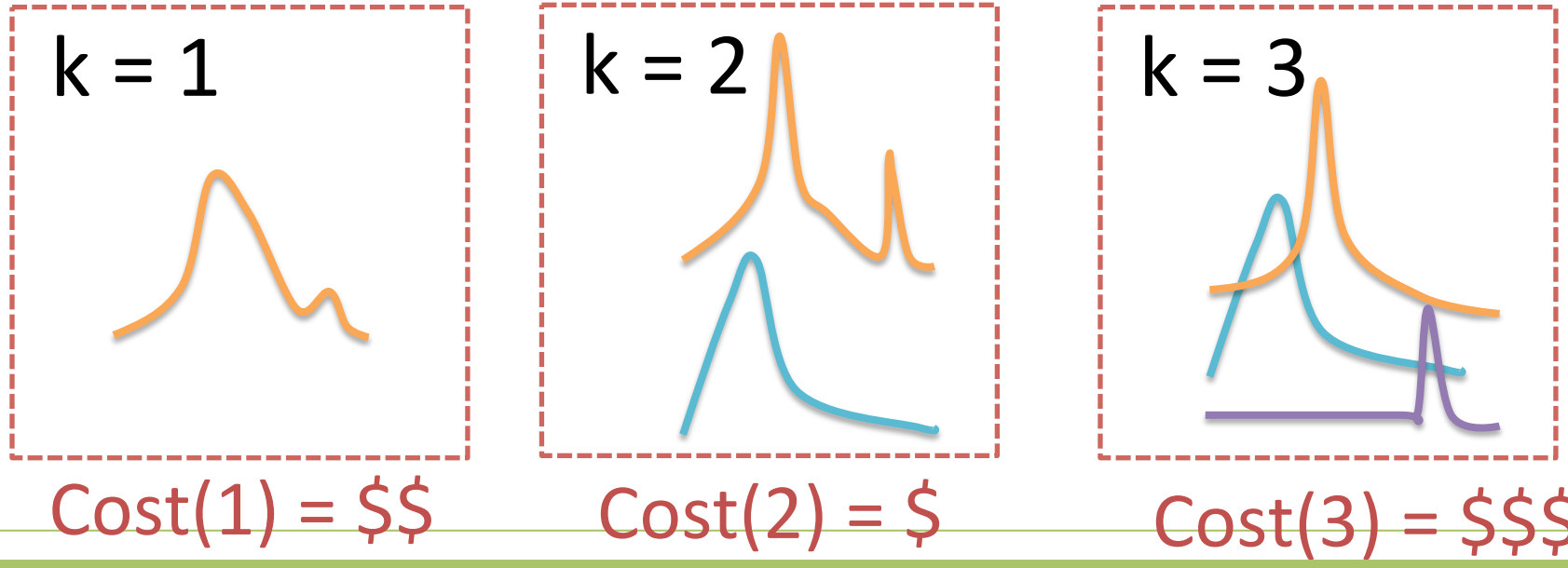
Idea(1-b) Automatic component analysis

Find optimal number k (1≤k≤d)

d: dimension



B

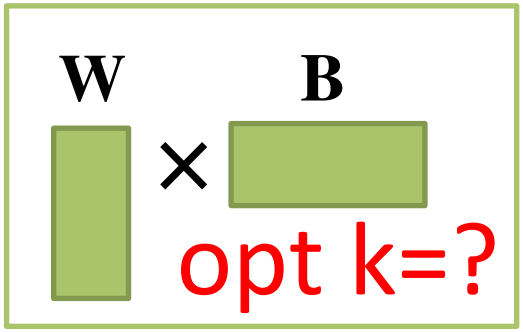


Idea (1): Seasonal component analysis

Details

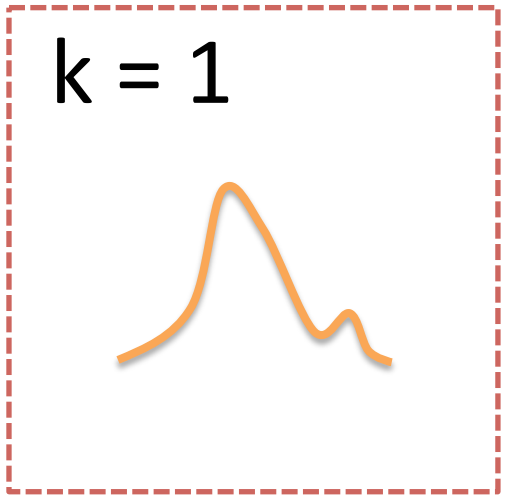
Idea(1-b) Automatic component analysis

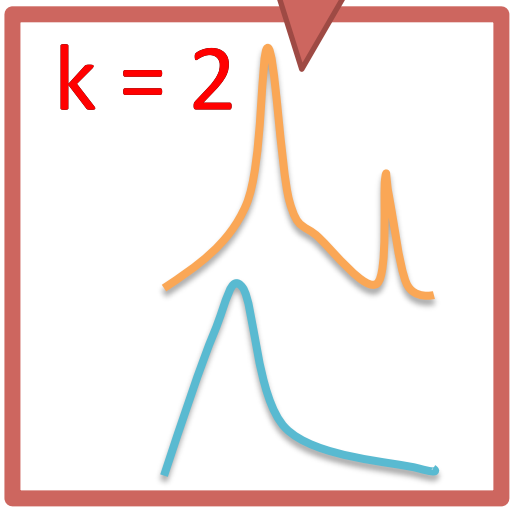
Find optimal number k (1≤k≤d)

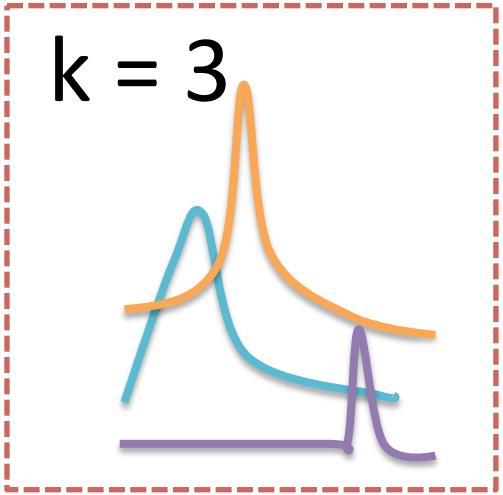


Optimal k

B







Cost(1) = $$

Cost(2) = $

Cost(3) = $$$

Idea (1): Seasonal component analysis

Details

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

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

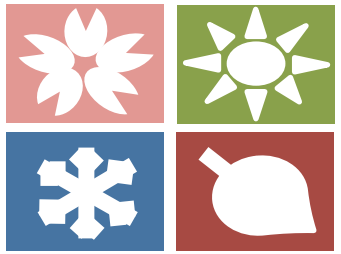
$$k_{opt} = \arg\min_k Cost_T(X;\mathcal{S})$$

Good compression

Good description

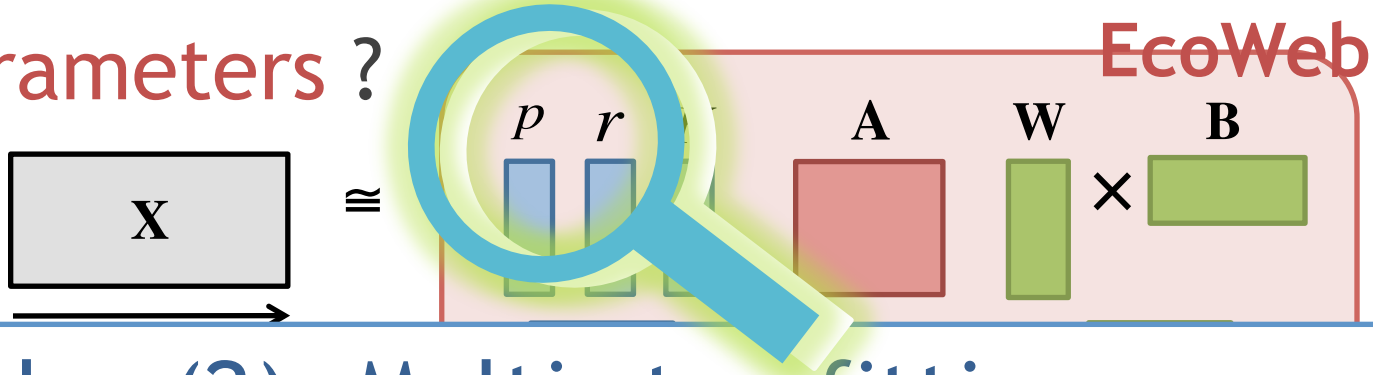
Challenges

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



Idea (1) : Seasonal component analysis

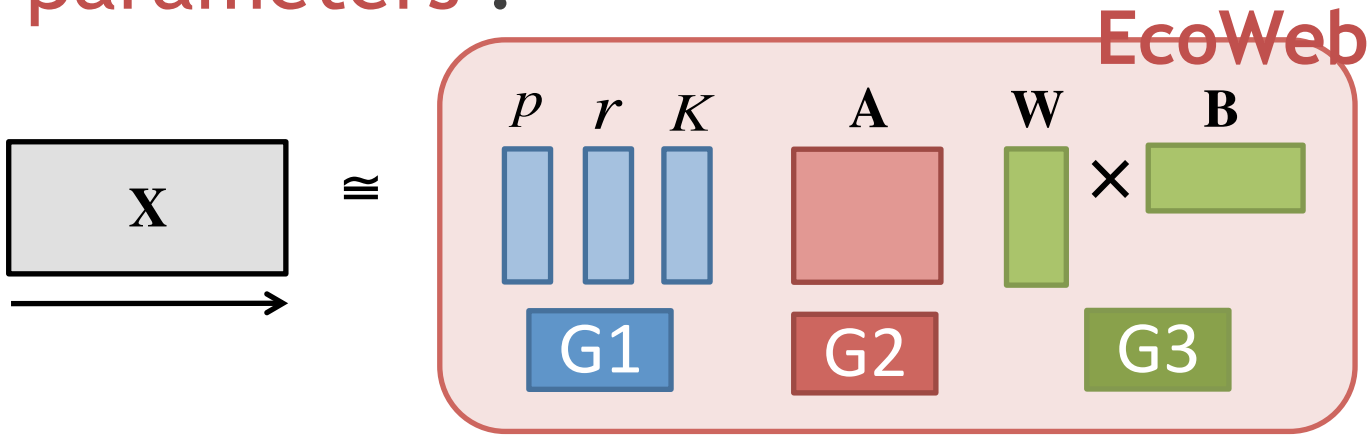
Q2. How can we efficiently estimate full-parameters ?



Idea (2): Multi-step fitting

Idea (2): Multi-step fitting

Q2. How can we efficiently estimate model parameters ?



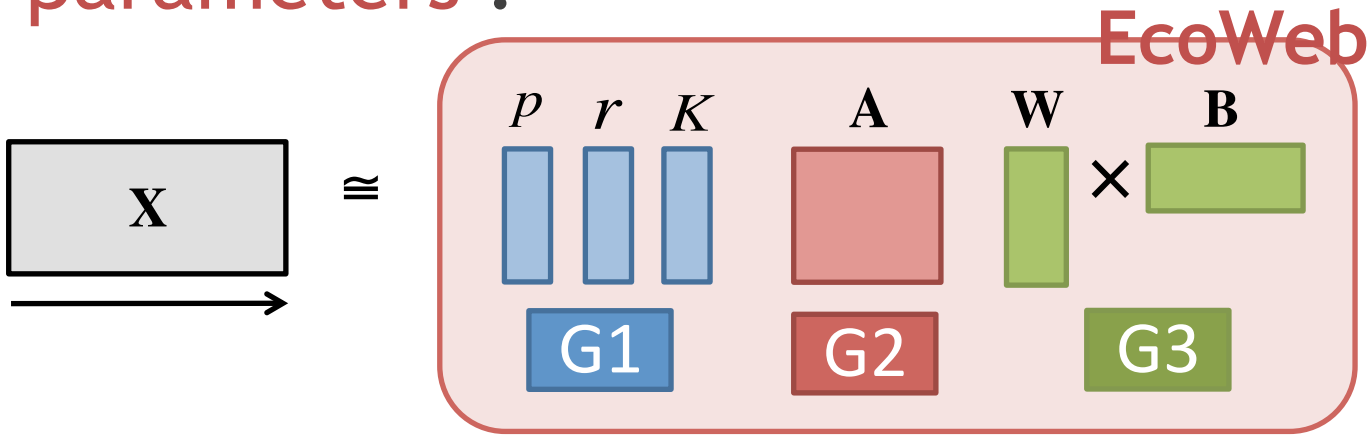
Idea (2): Multi-step fitting

a. **StepFit** (sub)

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

Idea (2): Multi-step fitting

Q2. How can we efficiently estimate model parameters ?

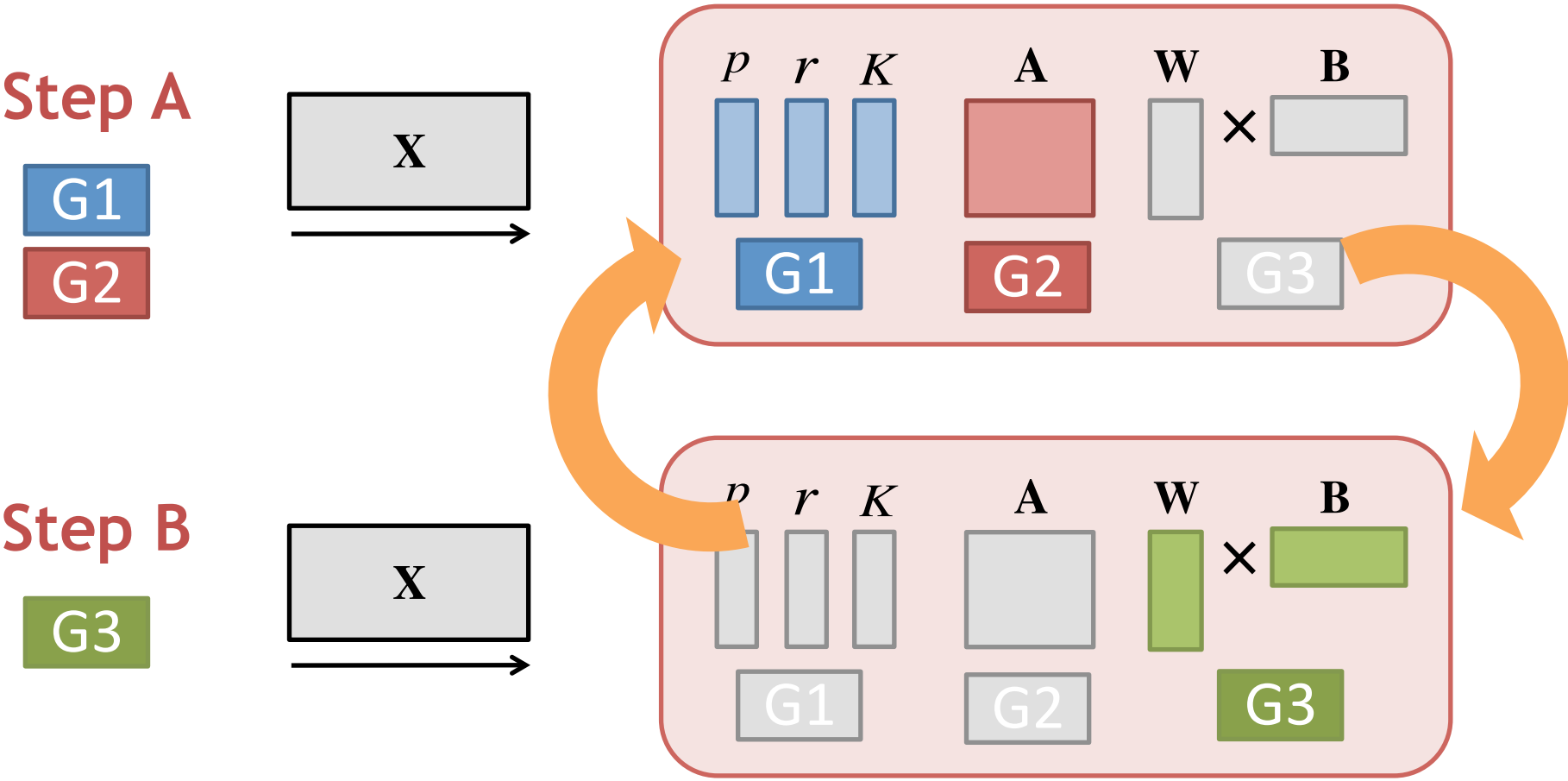


Idea (2): Multi-step fitting

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

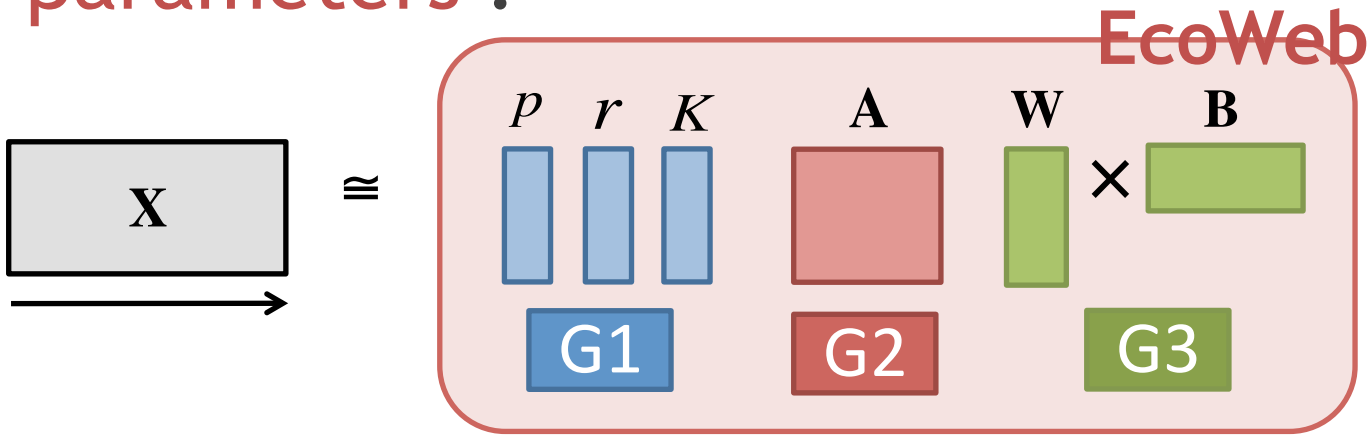
Idea (2): Multi-step fitting

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



Idea (2): Multi-step fitting

Q2. How can we efficiently estimate model parameters ?



Idea (2): Multi-step fitting

a. **StepFit** (sub)

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

Idea (2): Multi-step fitting

EcoWeb-Fit: full algorithm

e.g., 4 keywords: 

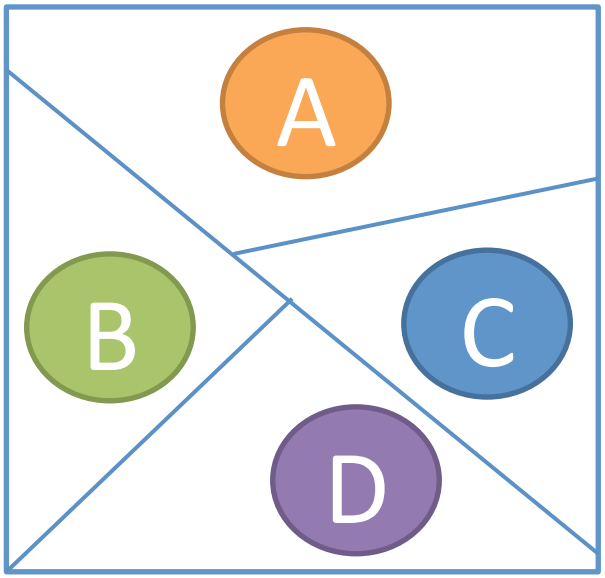
EcoWeb-Fit updates parameters, separately

Idea (2): Multi-step fitting

EcoWeb-Fit: full algorithm

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

1. Individual-Fit



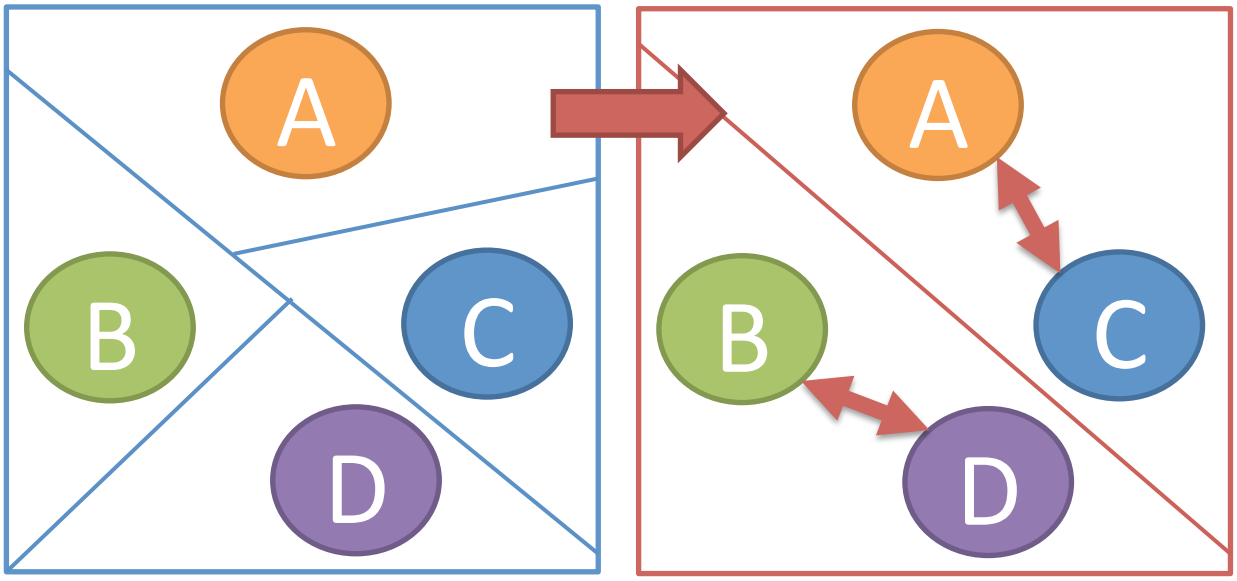
EcoWeb-Fit updates parameters, separately

Idea (2): Multi-step fitting

EcoWeb-Fit: full algorithm

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

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



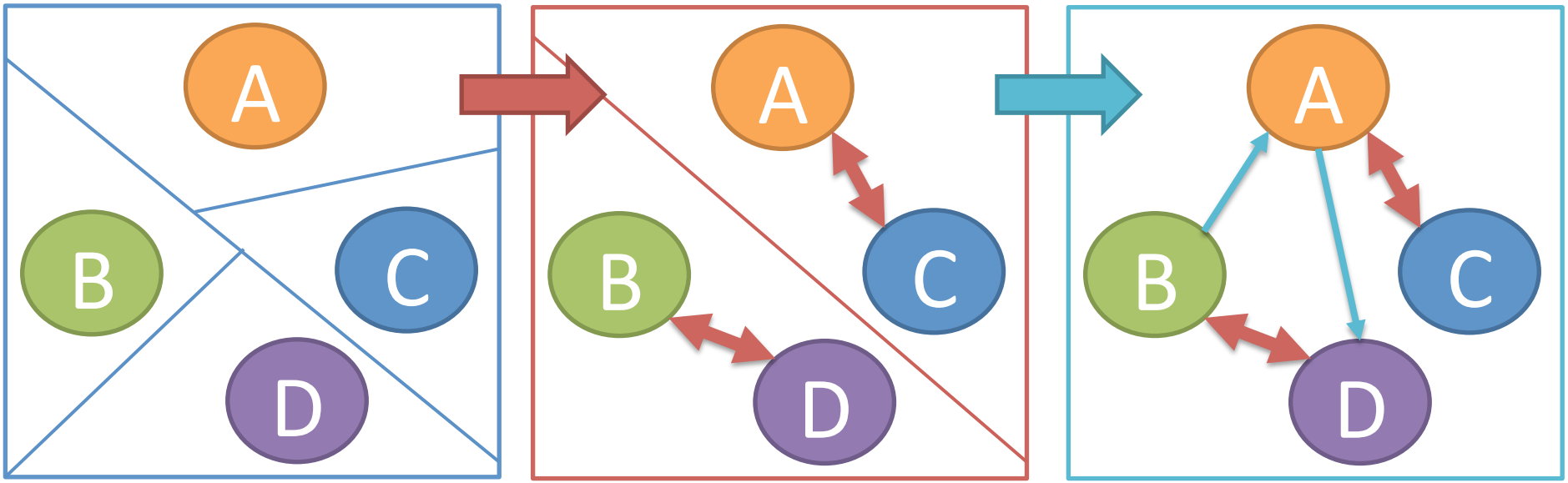
EcoWeb-Fit updates parameters, separately

Idea (2): Multi-step fitting

EcoWeb-Fit: full algorithm

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

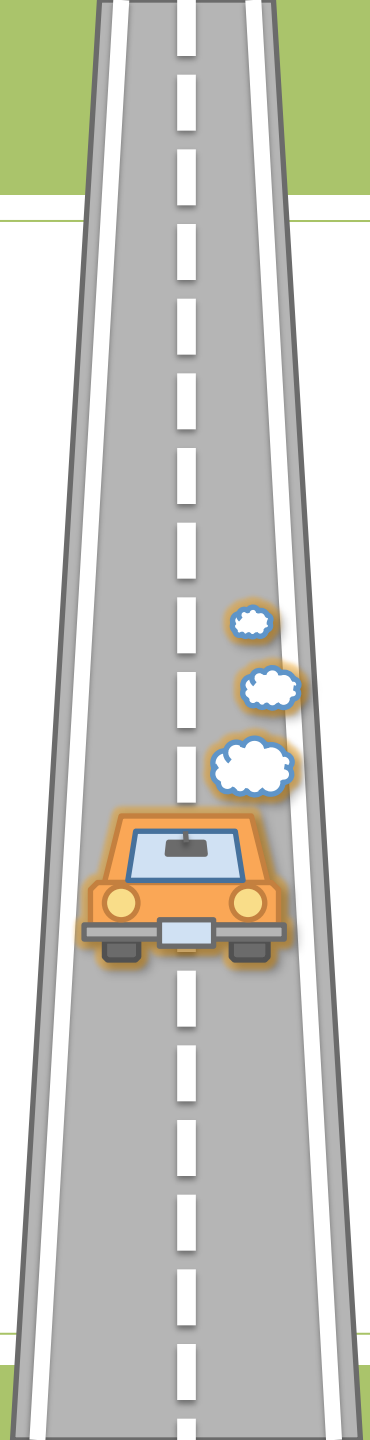
1. Individual-Fit **2. Pair-Fit** **3. Full-Fit**



EcoWeb-Fit updates parameters, separately

Roadmap

- ✔ Motivation
- ✔ Modeling power of EcoWeb
- ✔ Overview
- ✔ Proposed model
- ✔ Algorithm
- Experiments
- EcoWeb – at work
- Conclusions



Experiments

We answer the following questions…

Q1. Effectiveness

How successful is it in spotting patterns?

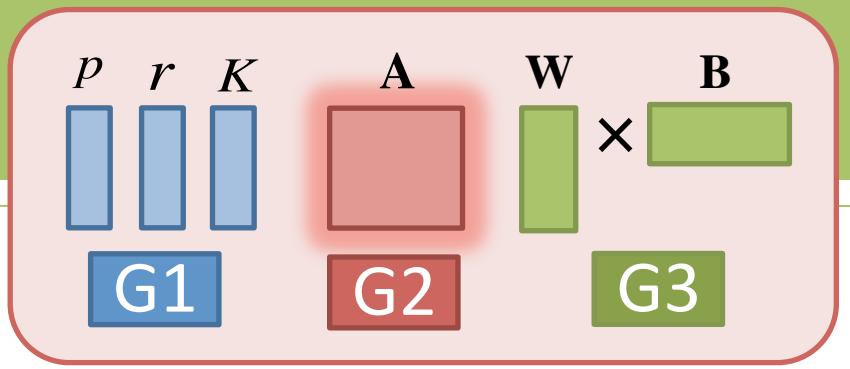
Q2. Accuracy

How well does it match the data?

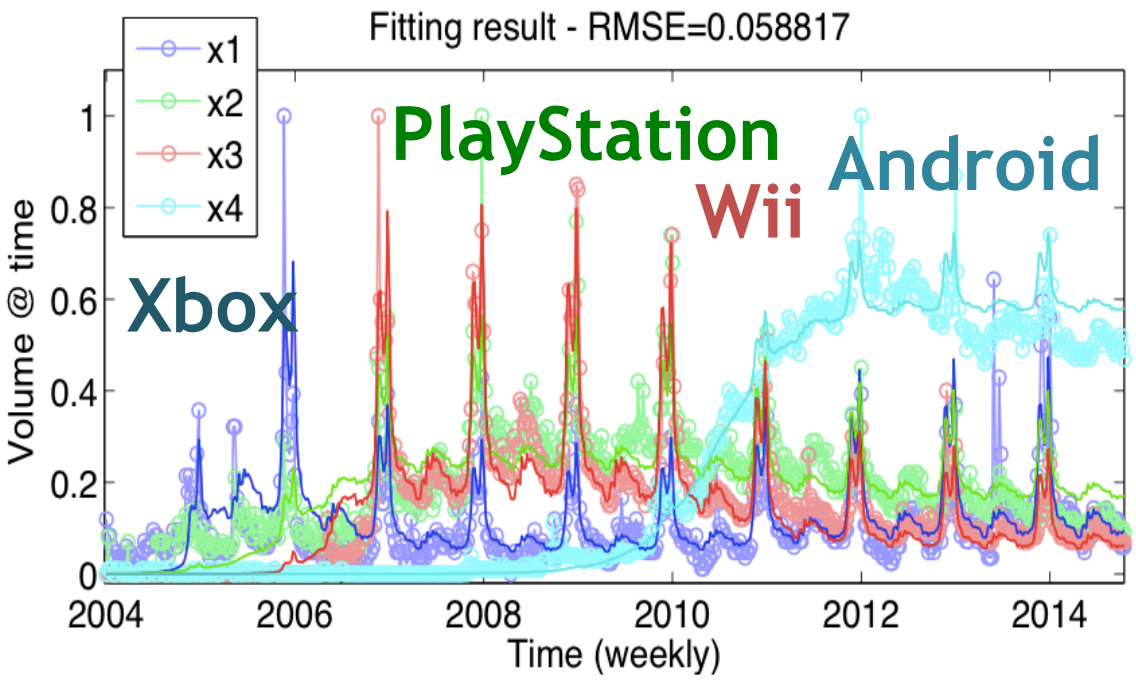
Q3. Scalability

How does it scale in terms of computational time?

Q1. Effectiveness

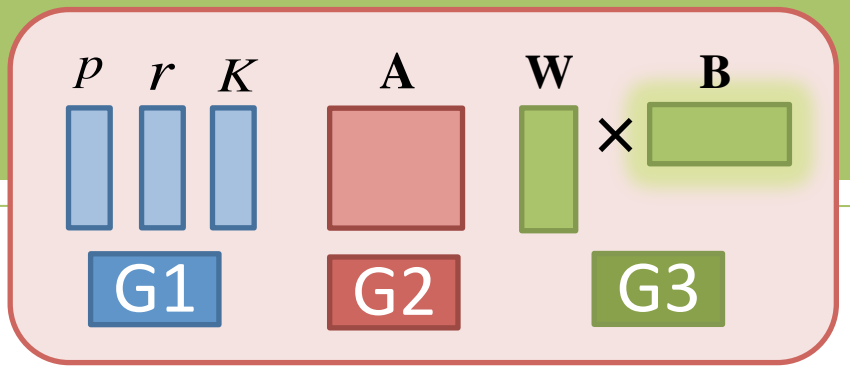


(#1) Video games

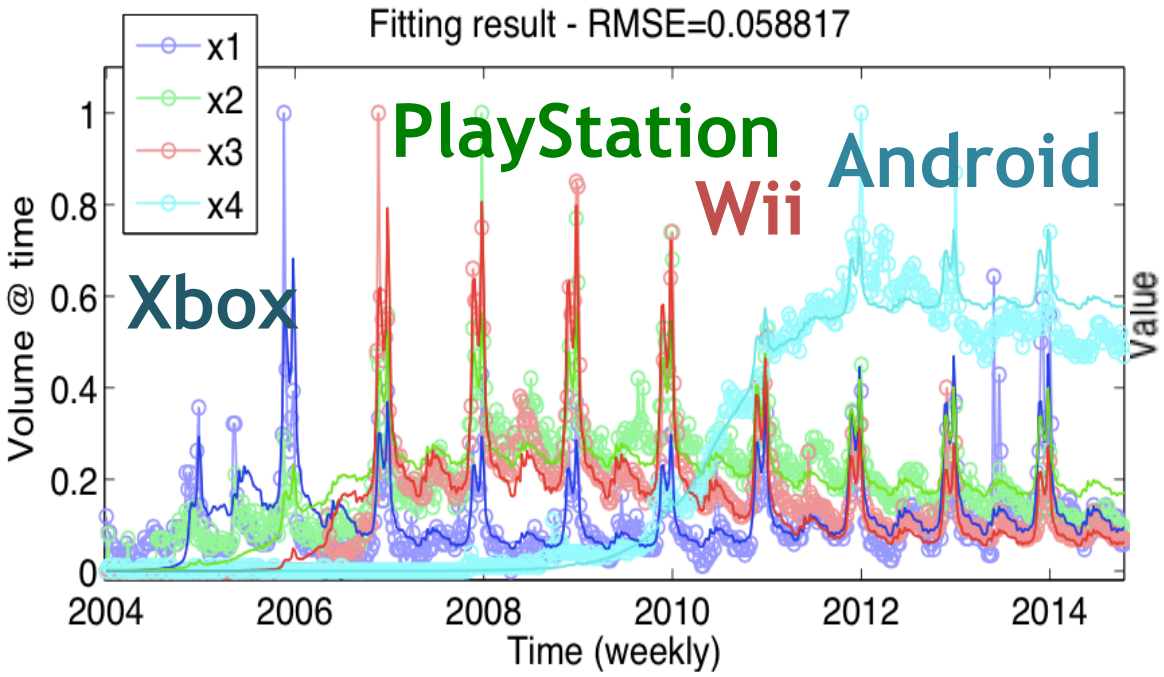


Interactions between keywords

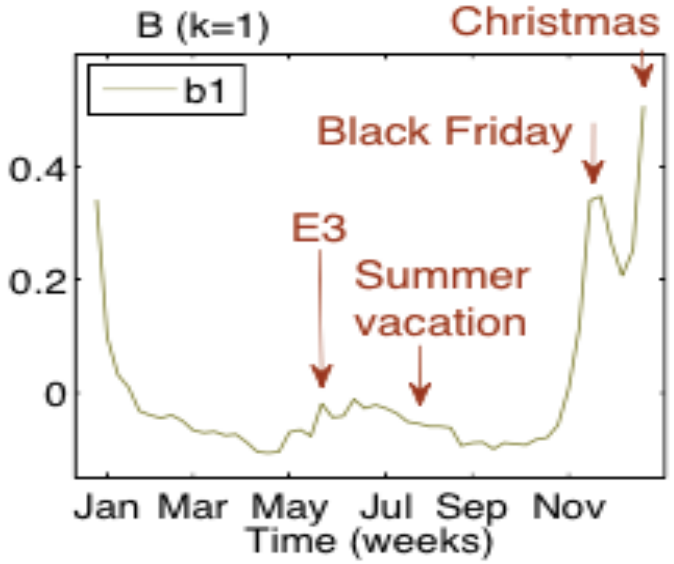
Q1. Effectiveness



(#1) Video games



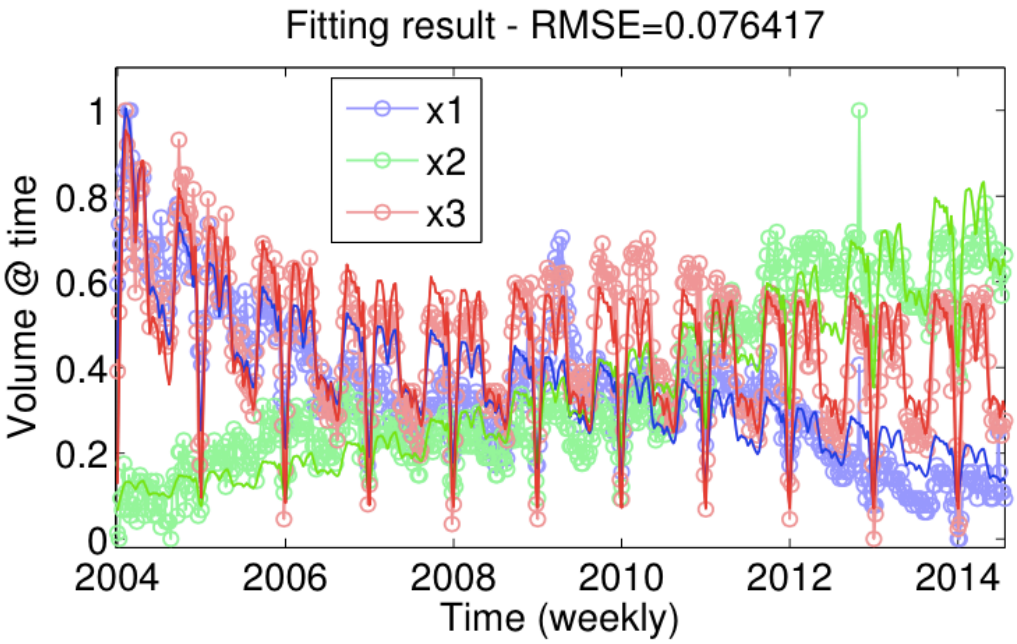
Seasonality



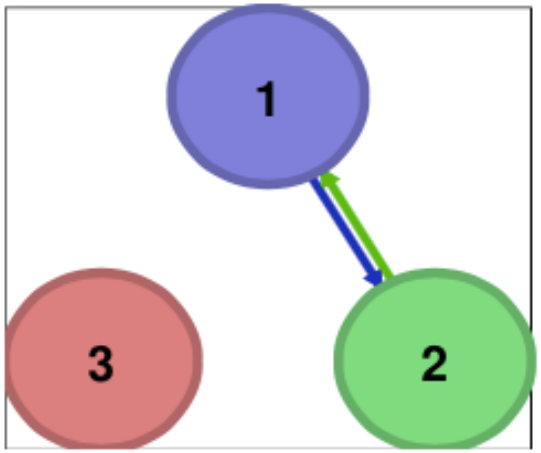
Q1. Effectiveness

(#2) Programming language

C , R , MATLAB



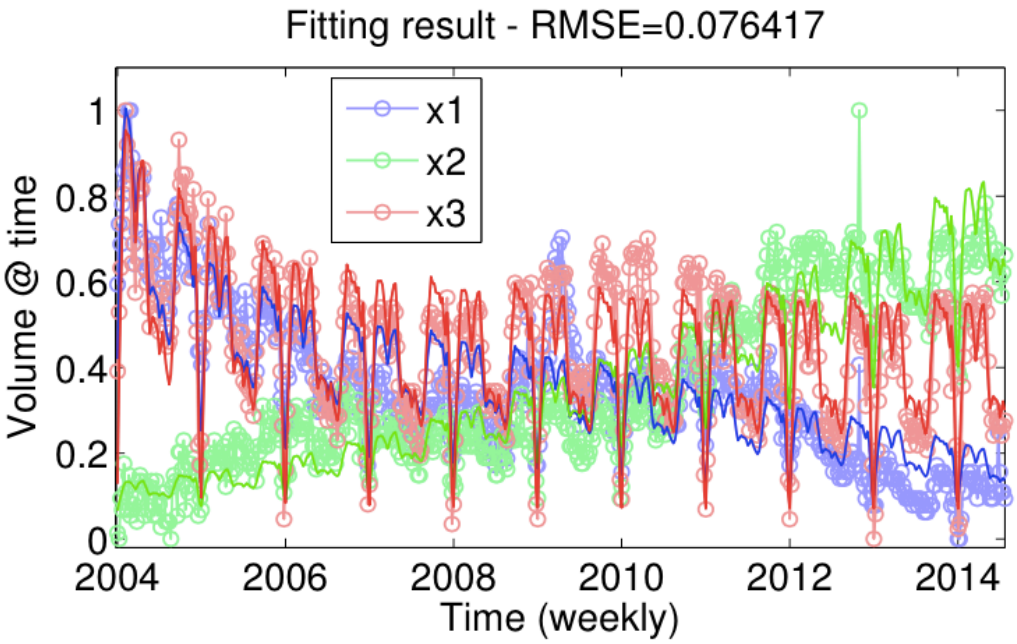
Interactions



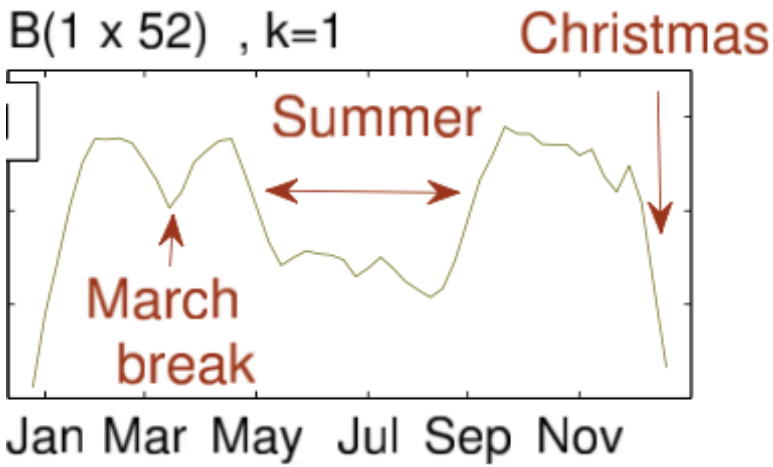
Q1. Effectiveness

(#2) Programming language

C , R , MATLAB



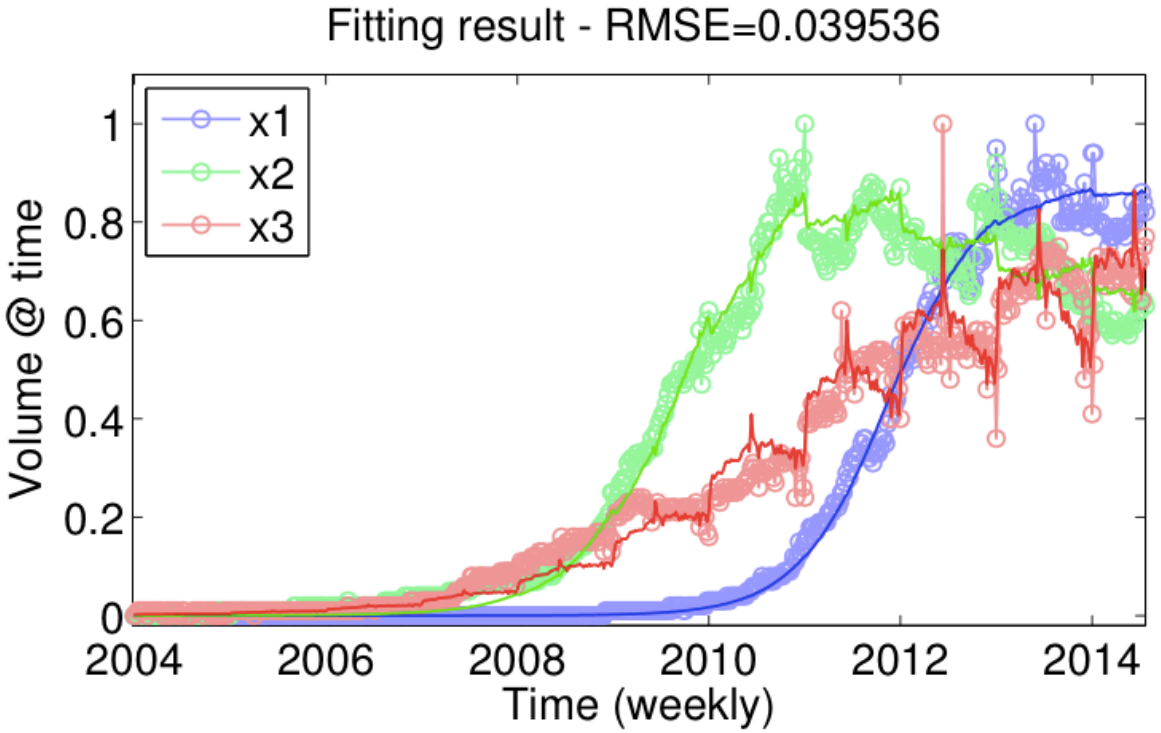
Seasonality



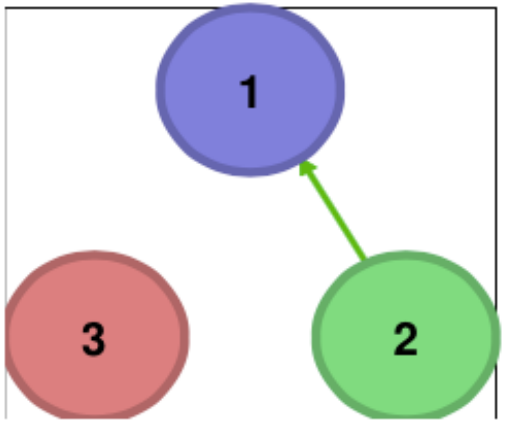
Q1. Effectiveness

(#3) Social media

Tumblr , Facebook , LinkedIn



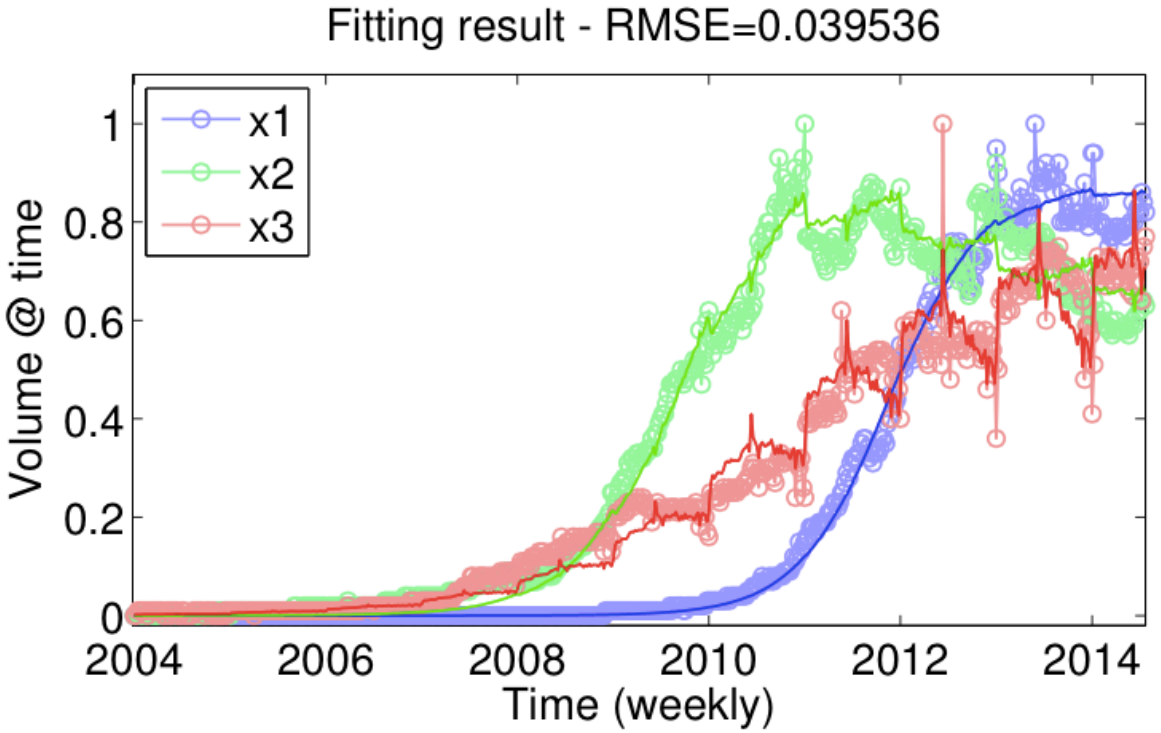
Interactions



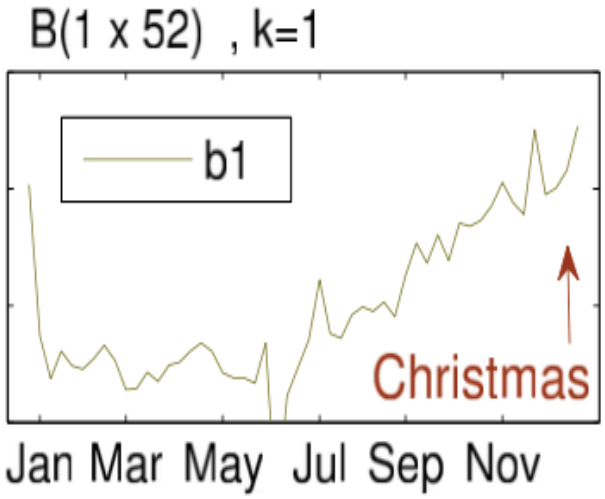
Q1. Effectiveness

(#3) Social media

Tumblr , Facebook , LinkedIn



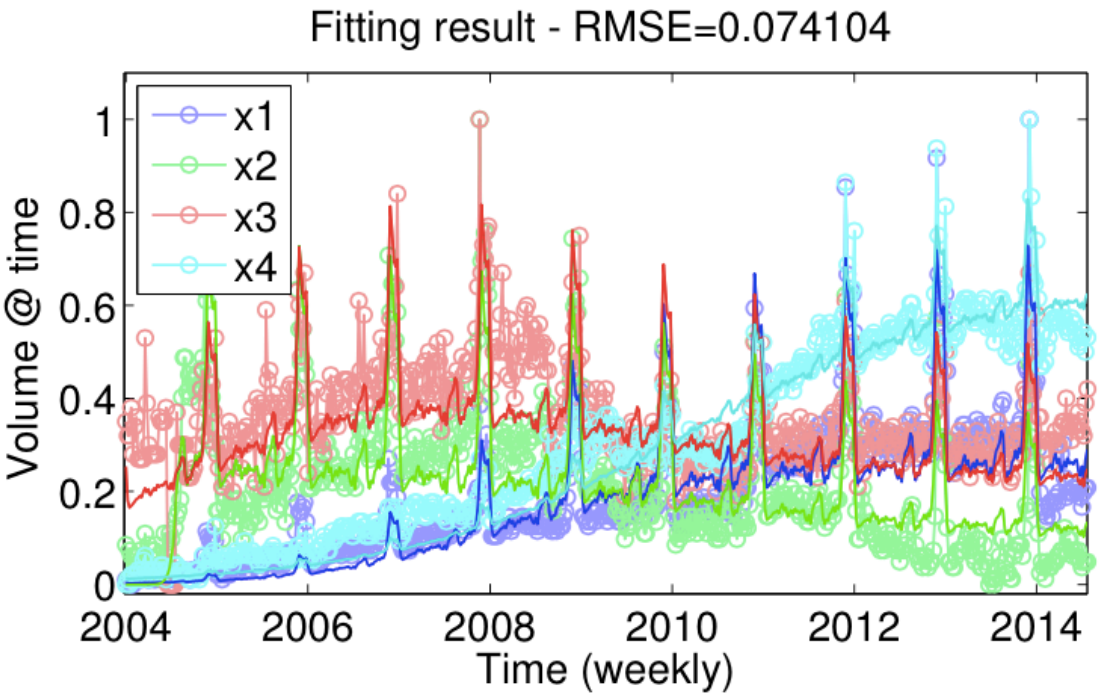
Seasonality



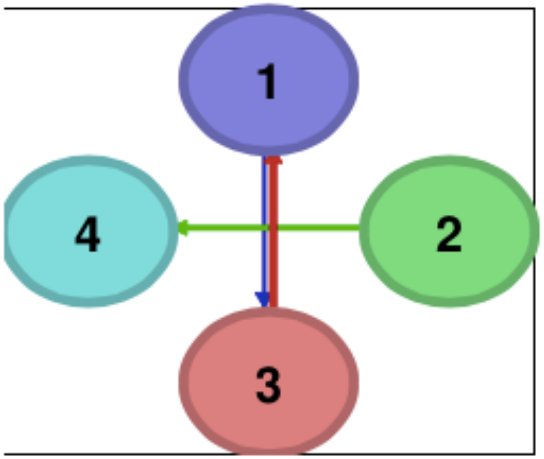
Q1. Effectiveness

(#4) Apparel companies

Kohls , JCPenny , Nordstrom , Forever21



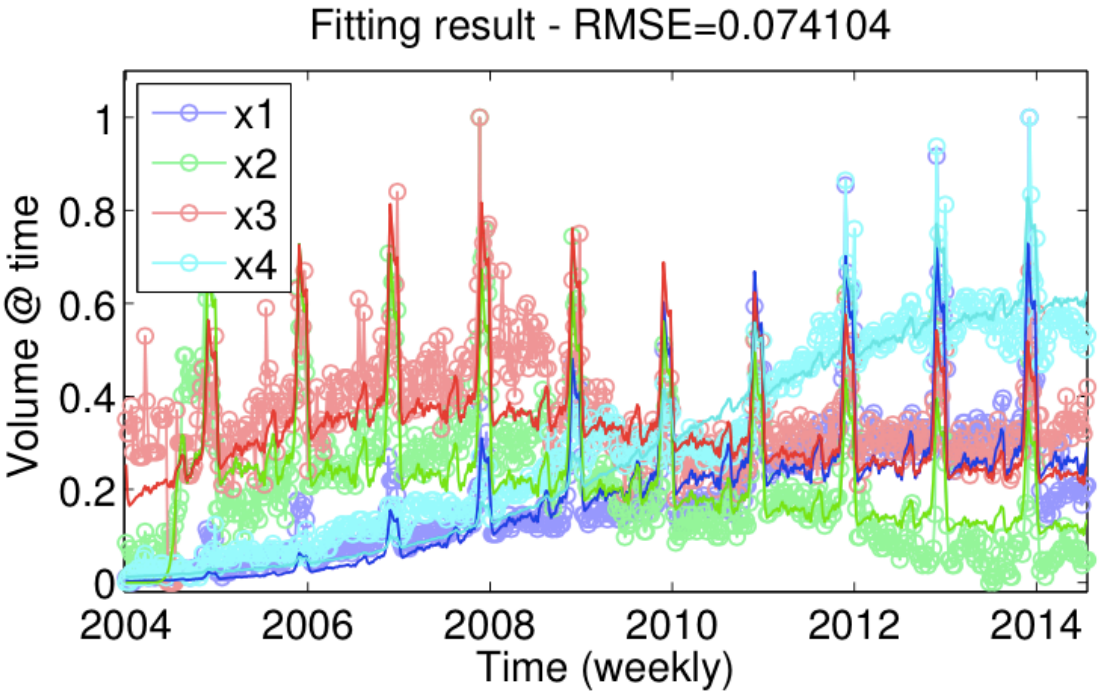
Interactions



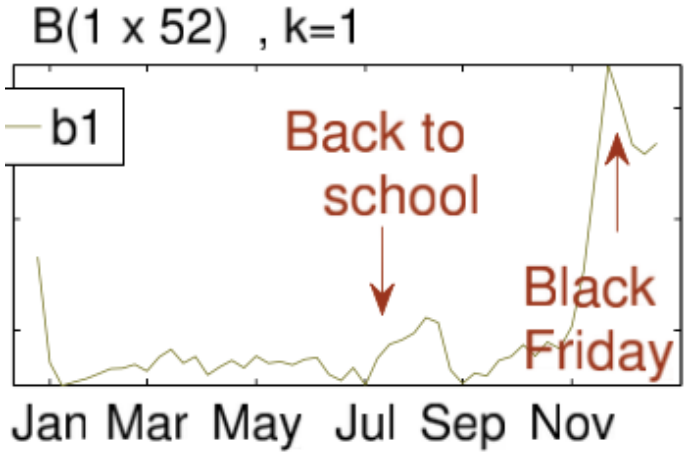
Q1. Effectiveness

(#4) Apparel companies

Kohls , JCPenny , Nordstrom , Forever21



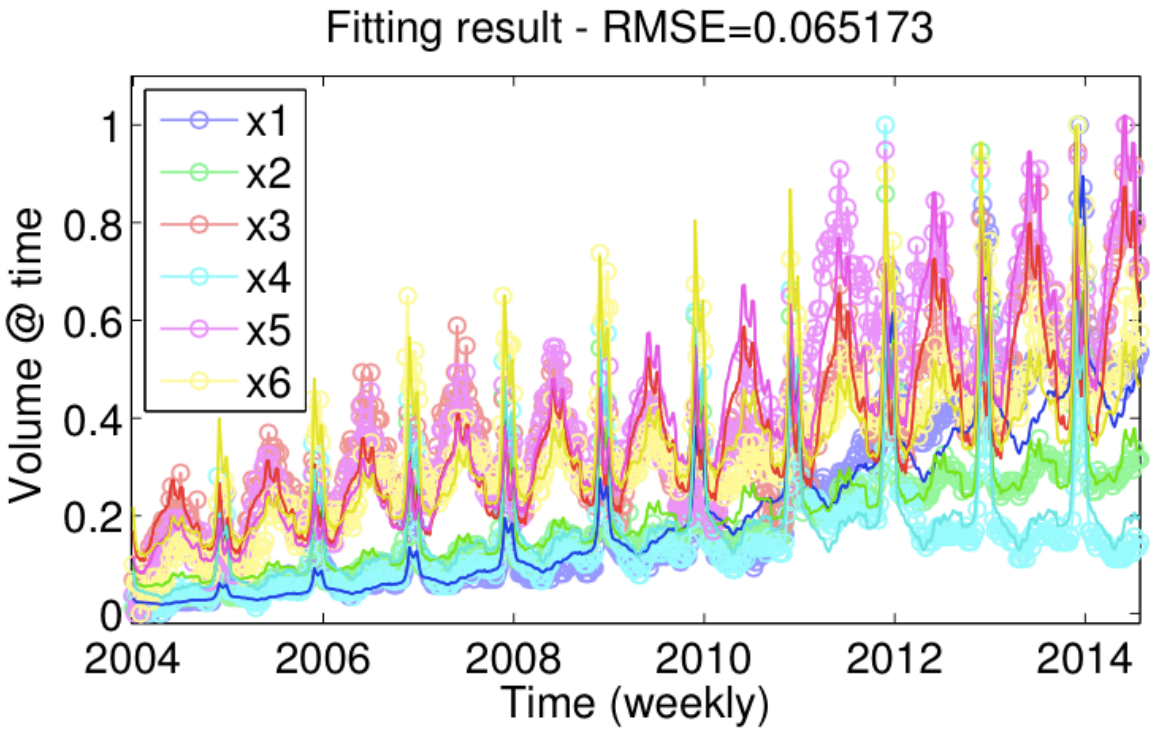
Seasonality



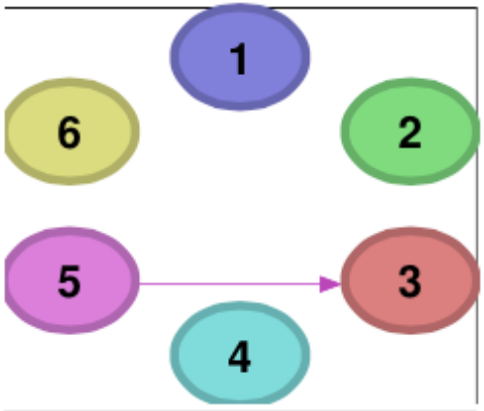
Q1. Effectiveness

(#5) Retail companies

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



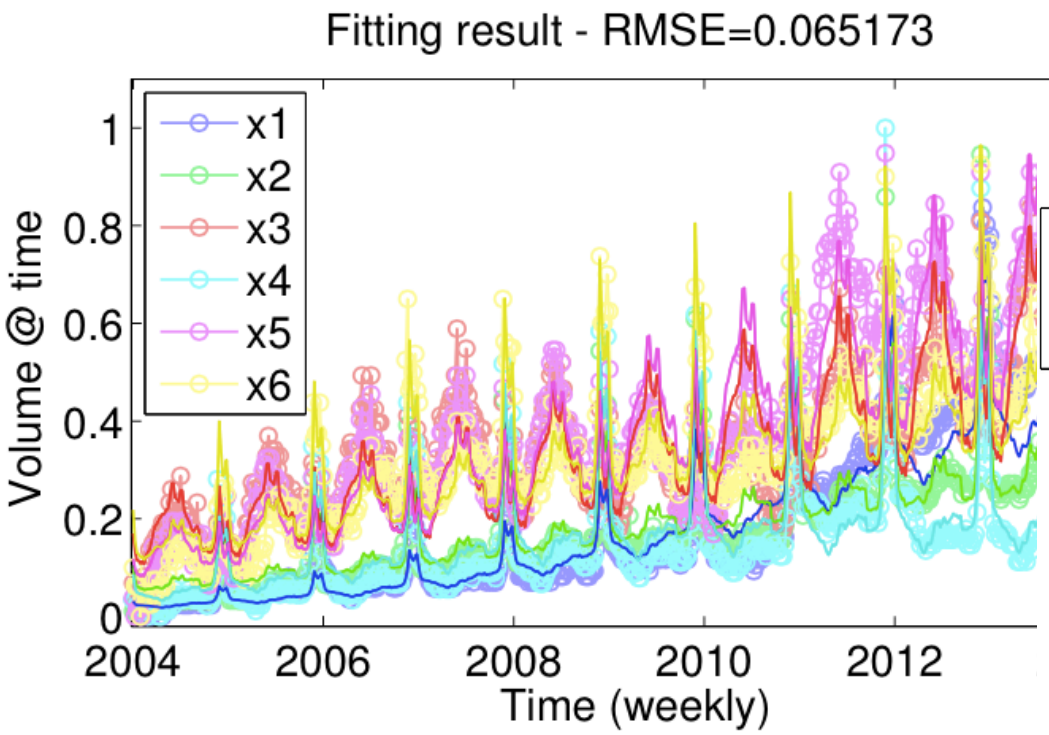
Interaction



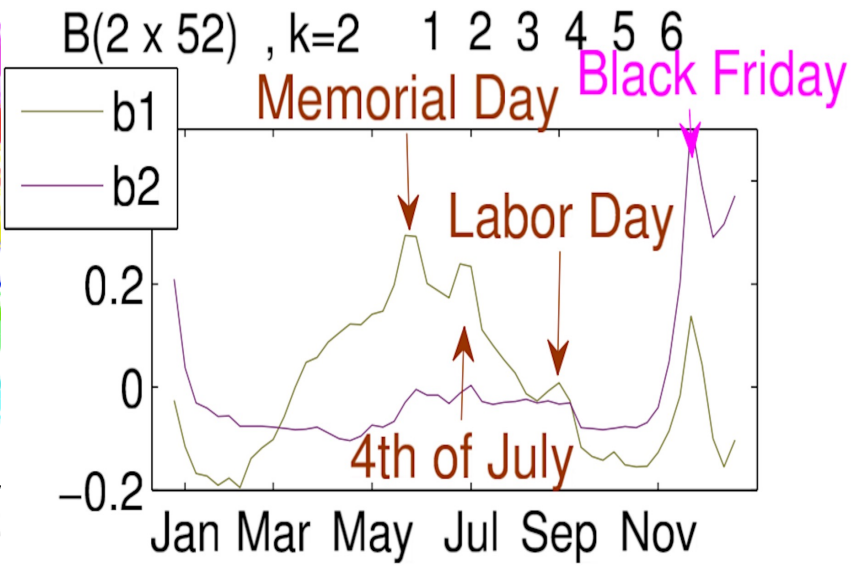
Q1. Effectiveness

(#5) Retail companies

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



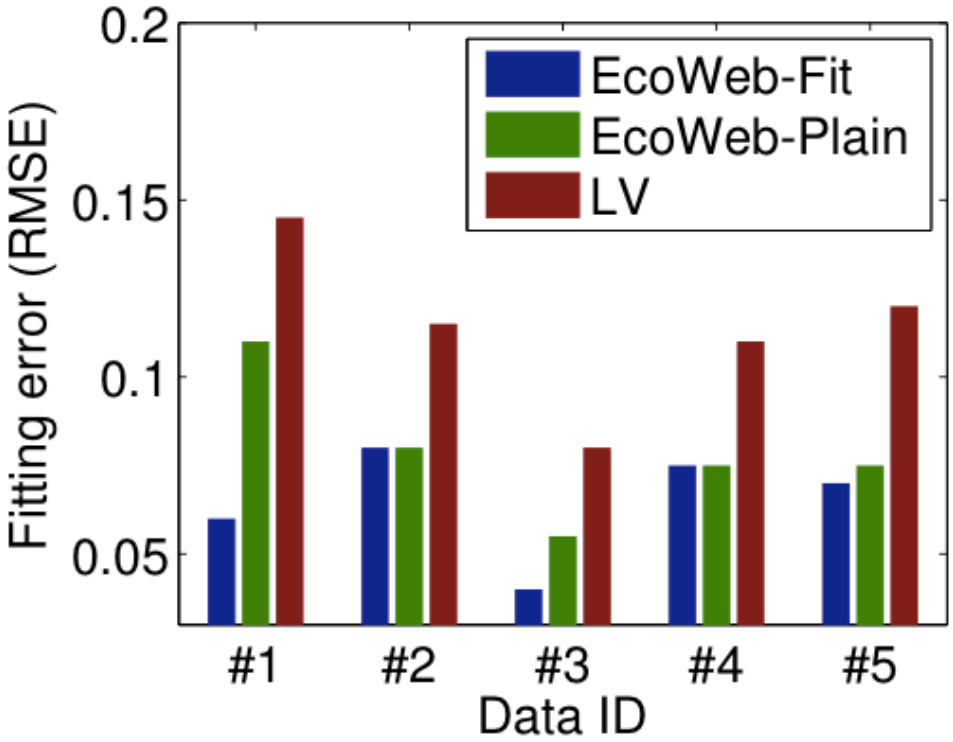
Seasonality



Q2. Accuracy

RMSE between original and fitted volume

(Lower is better)

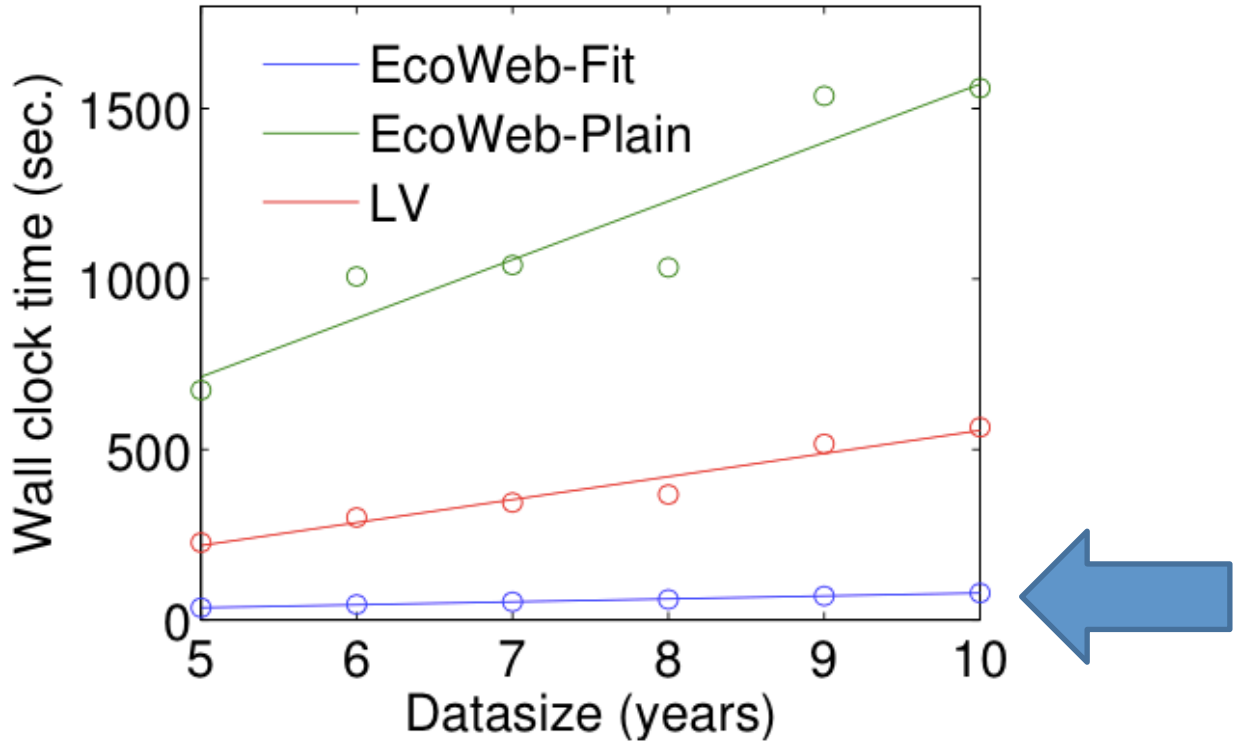


EcoWeb consistently wins!

Q3. Scalability

Wall clock time vs. dataset size (years)

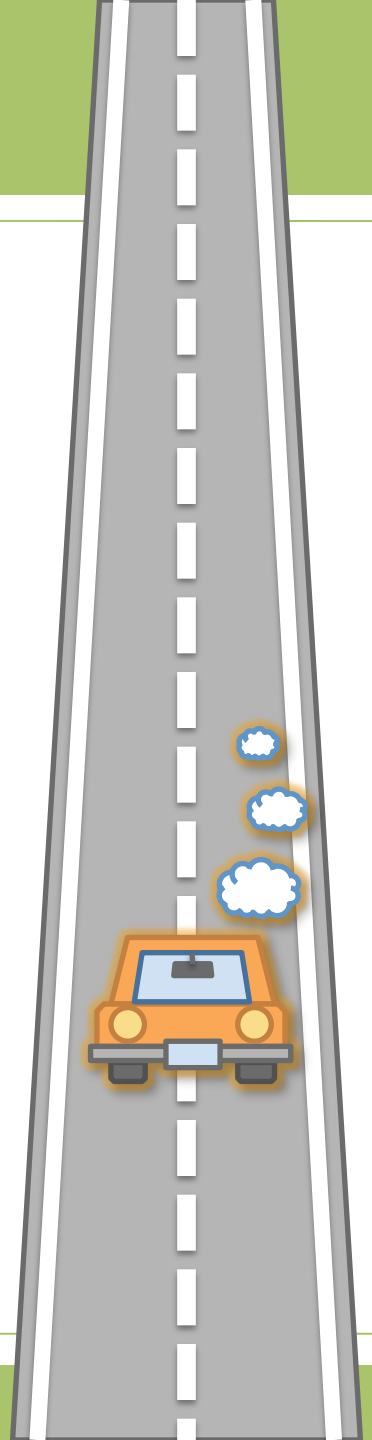
EcoWeb-Fit scales linearly, i.e., O(n)



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

Roadmap

- ✔ Motivation
- ✔ Modeling power of EcoWeb
- ✔ Overview
- ✔ Proposed model
- ✔ Algorithm
- ✔ Experiments
- – EcoWeb – at work
- – Conclusions

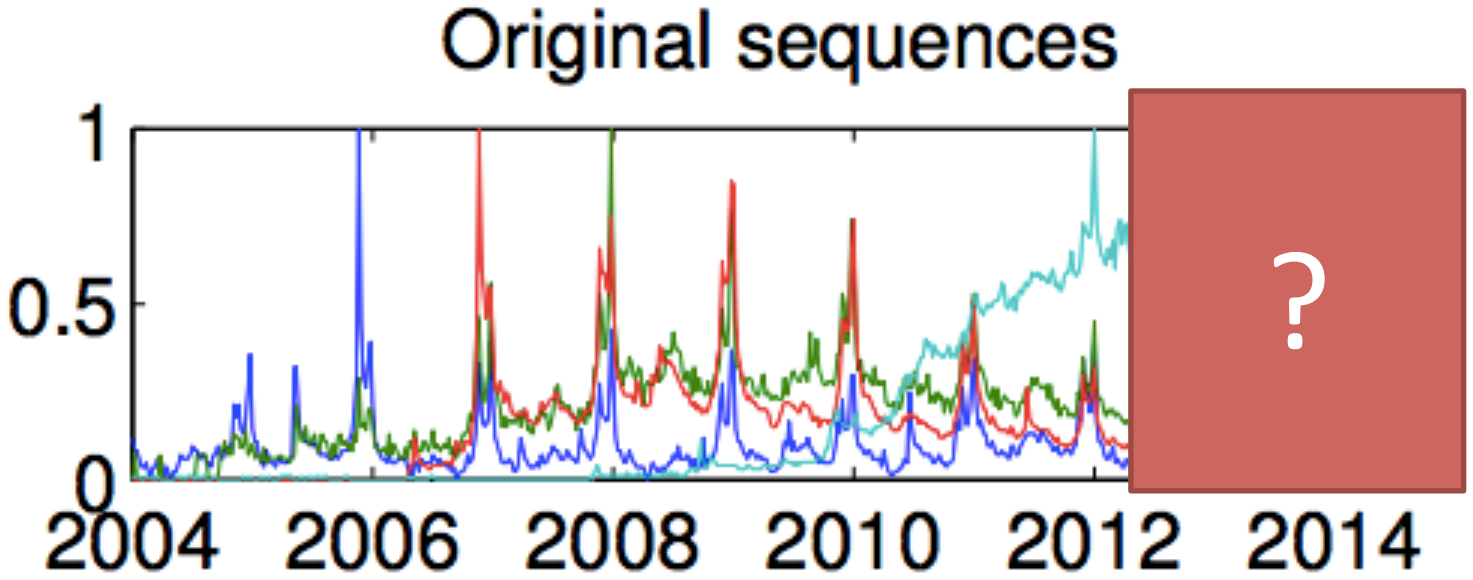


EcoWeb at work - forecasting

Forecasting future activities

Train:
2/3 sequences

Forecast:
1/3 following years

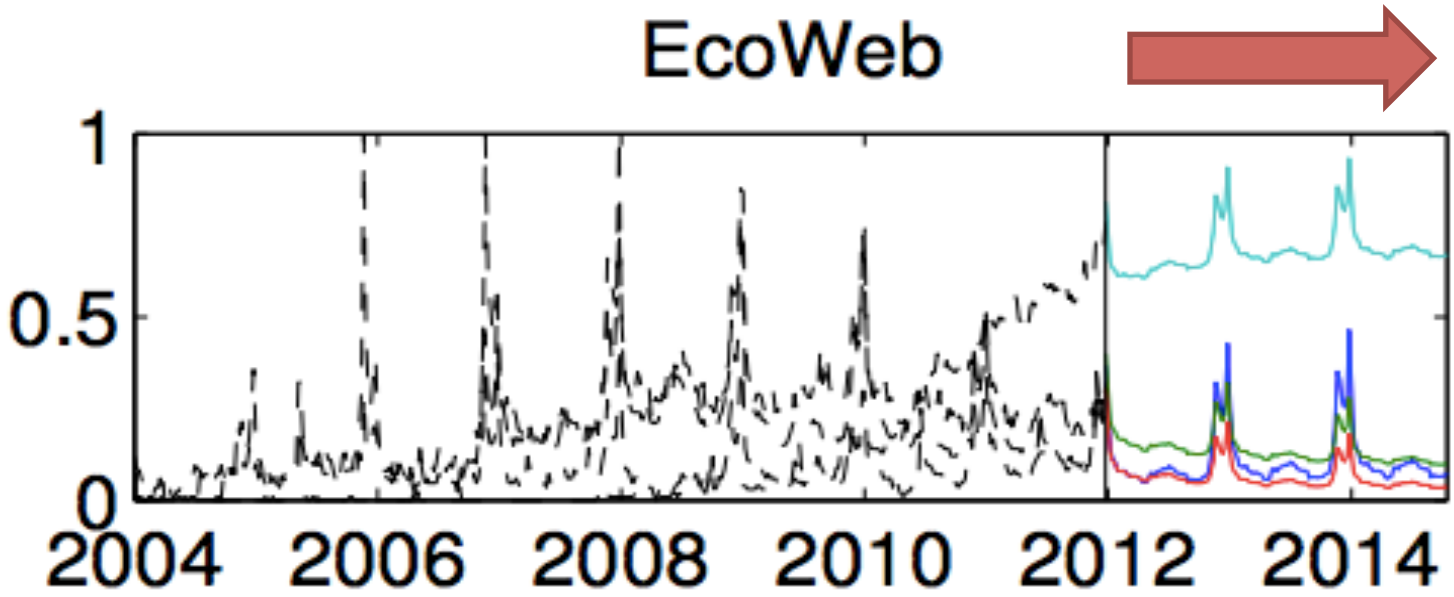


EcoWeb at work - forecasting

Forecasting future activities

Train:
2/3 sequences

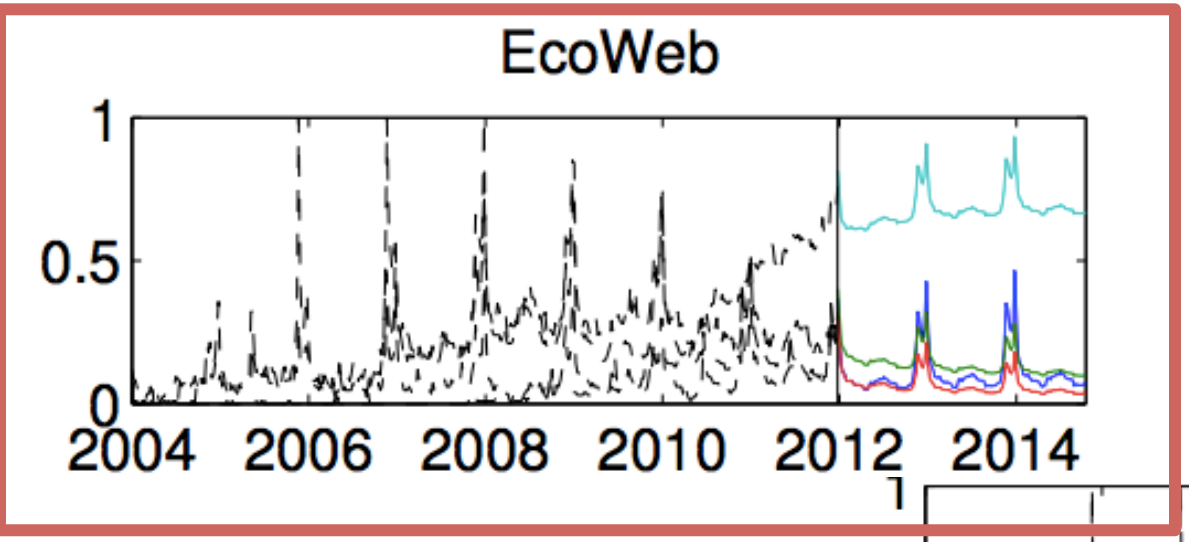
Forecast:
1/3 following years

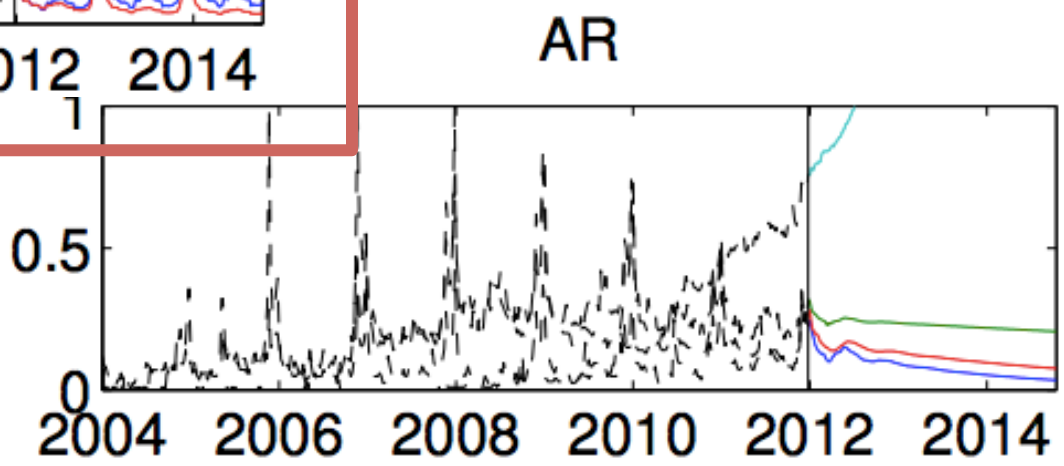


EcoWeb can capture future patterns

EcoWeb at work - forecasting

Forecasting future activities

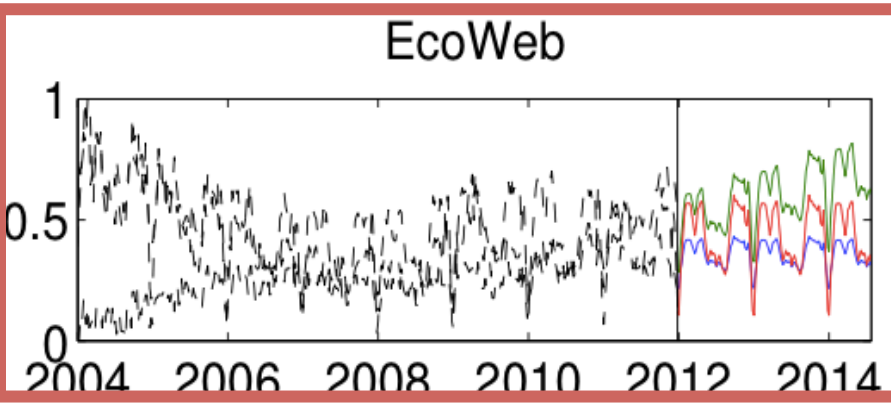


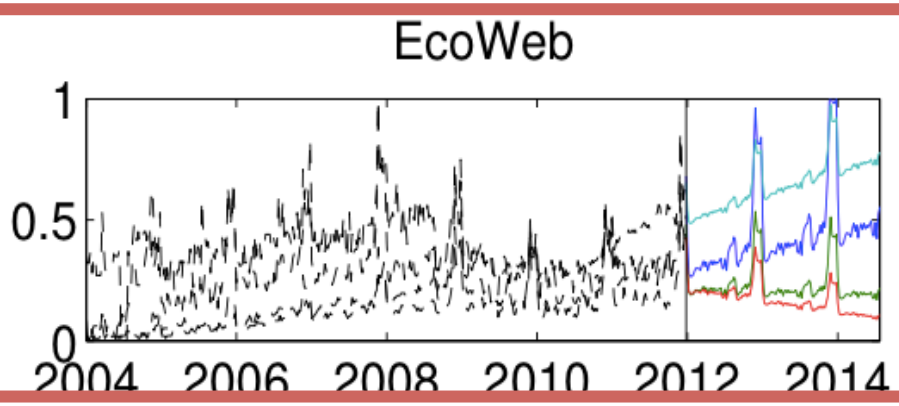


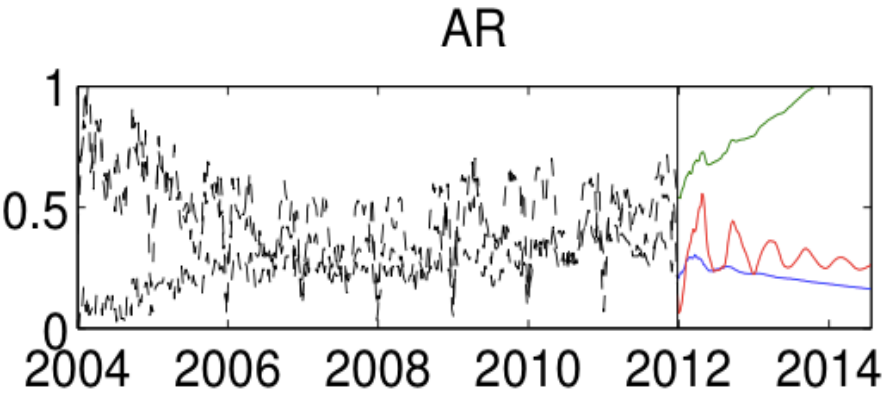
EcoWeb can capture future patterns!

EcoWeb at work - forecasting

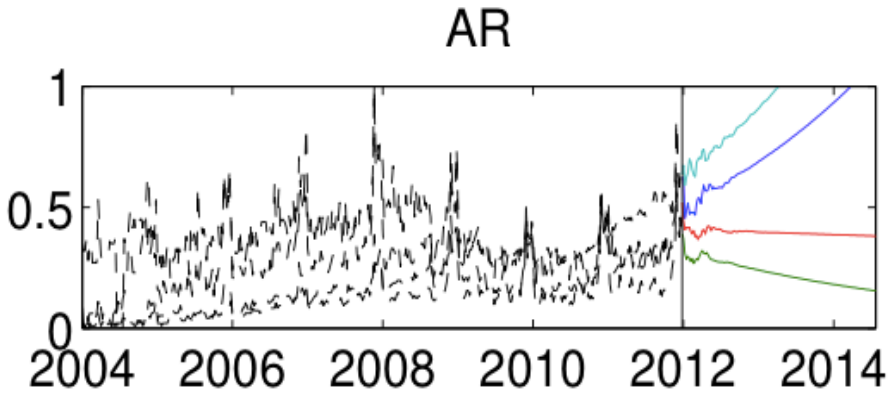
Forecasting future activities







(b) Programming languages (#2)

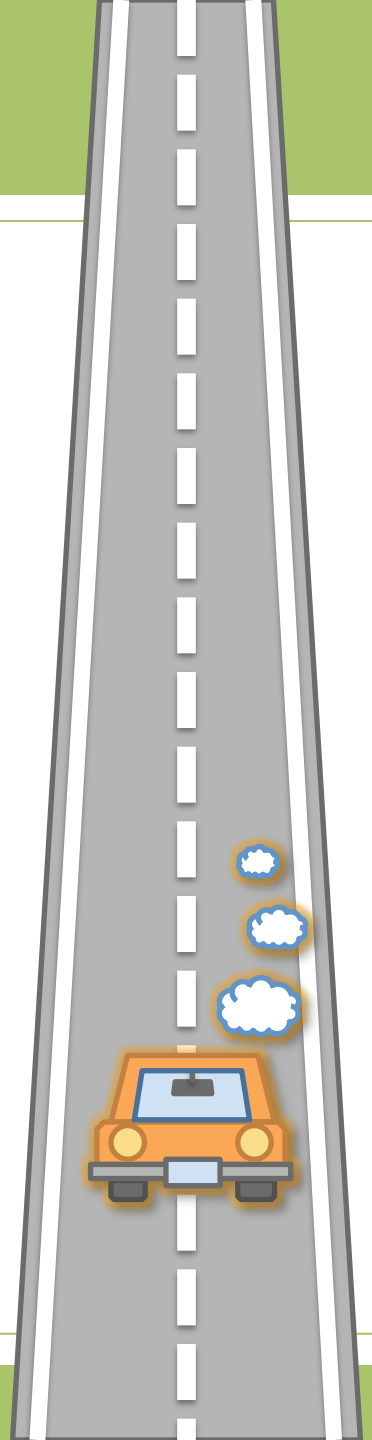


(c) Apparel companies (#4)

EcoWeb can capture future patterns!

Roadmap

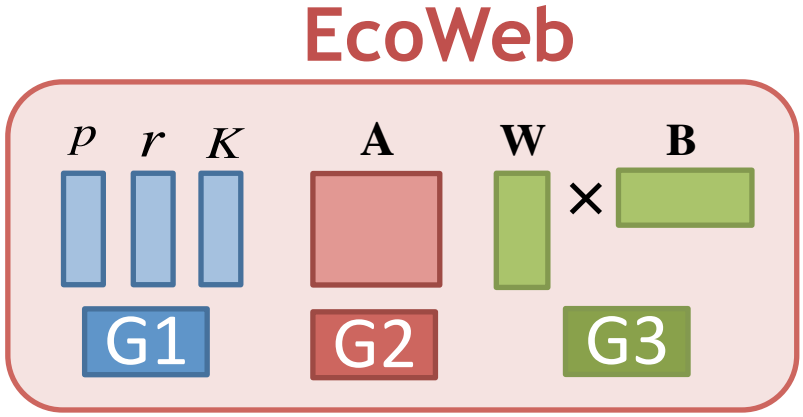
- ✔ Motivation
- ✔ Modeling power of EcoWeb
- ✔ Overview
- ✔ Proposed model
- ✔ Algorithm
- ✔ Experiments
- ✔ EcoWeb – at work
- Conclusions

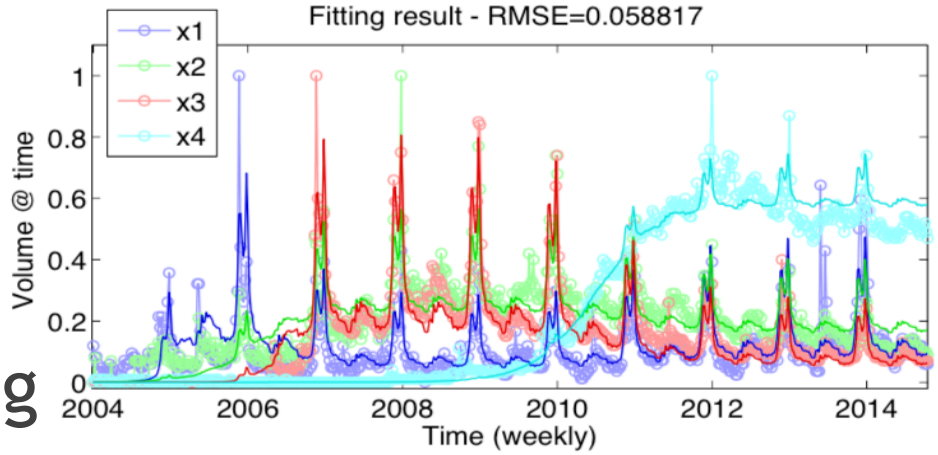


Conclusions

EcoWeb has the following advantages

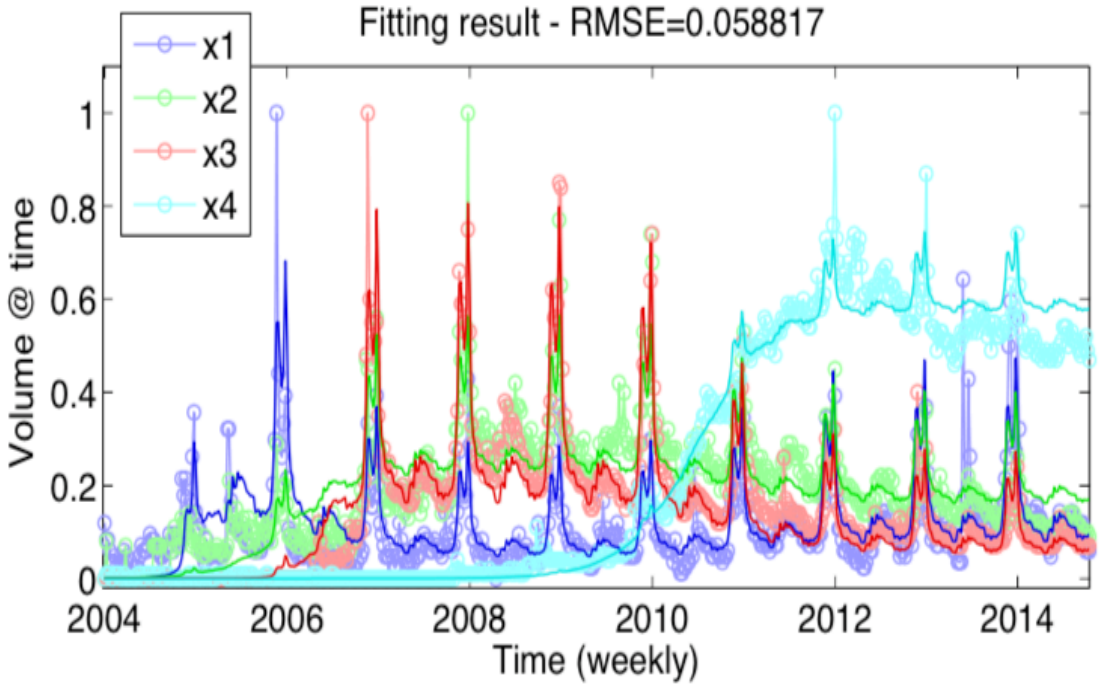
- ✔ **Effective**
  Finds important patterns
- ✔ **Fully-automatic**
  No parameter tuning
- ✔ **Scalable**
  It is linear
- ✔ **Practical**
  Long-range forecasting

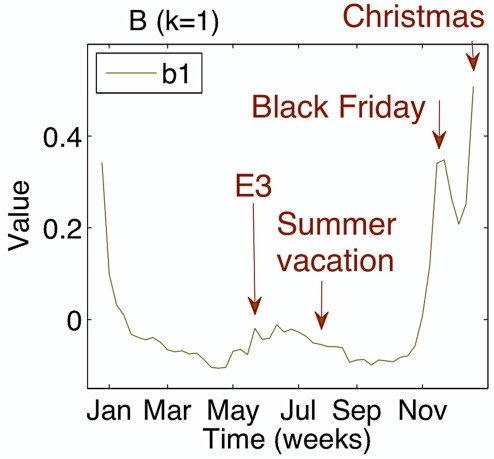




Thank you!







Data & Code:

http://www.cs.kumamoto-u.ac.jp/~yasuko