Canonical Trends

Detecting Trends in Web Data

Felix Bießmann,
Jens-Michalis Papaioannou,
Mikio Braun, Andreas Harth

Berlin Institute of Technology
Department Machine Learning

Karlsruhe Institute of Technology

Temporal Dynamics of Web Data

Web content is copied, repeated or rephrased (Trends/Memes)

This temporal structure contains important information

Growing interest in **temporal dynamics of graphs**

Understanding dynamic graphs [Leskovec et al, KDD, 2005]
Causal Inference [Lozano and Sindhwani, NIPS 2010]
Diffusion of information [Gomez Rodriguez et al, ICML 2011/2012]

Canonical Trend Analysis

- Idea: Predict future features from past features

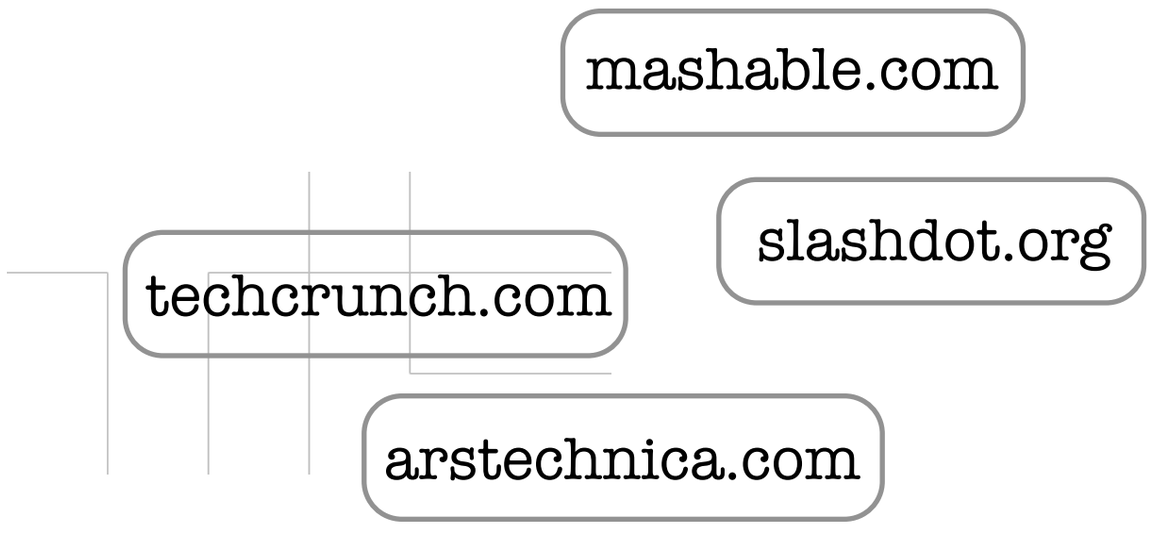
- Finds trends with **highest impact on network**
- Finds web sources that precede/follow trends
- Based on simple algebra, no density estimation involved

Examples:

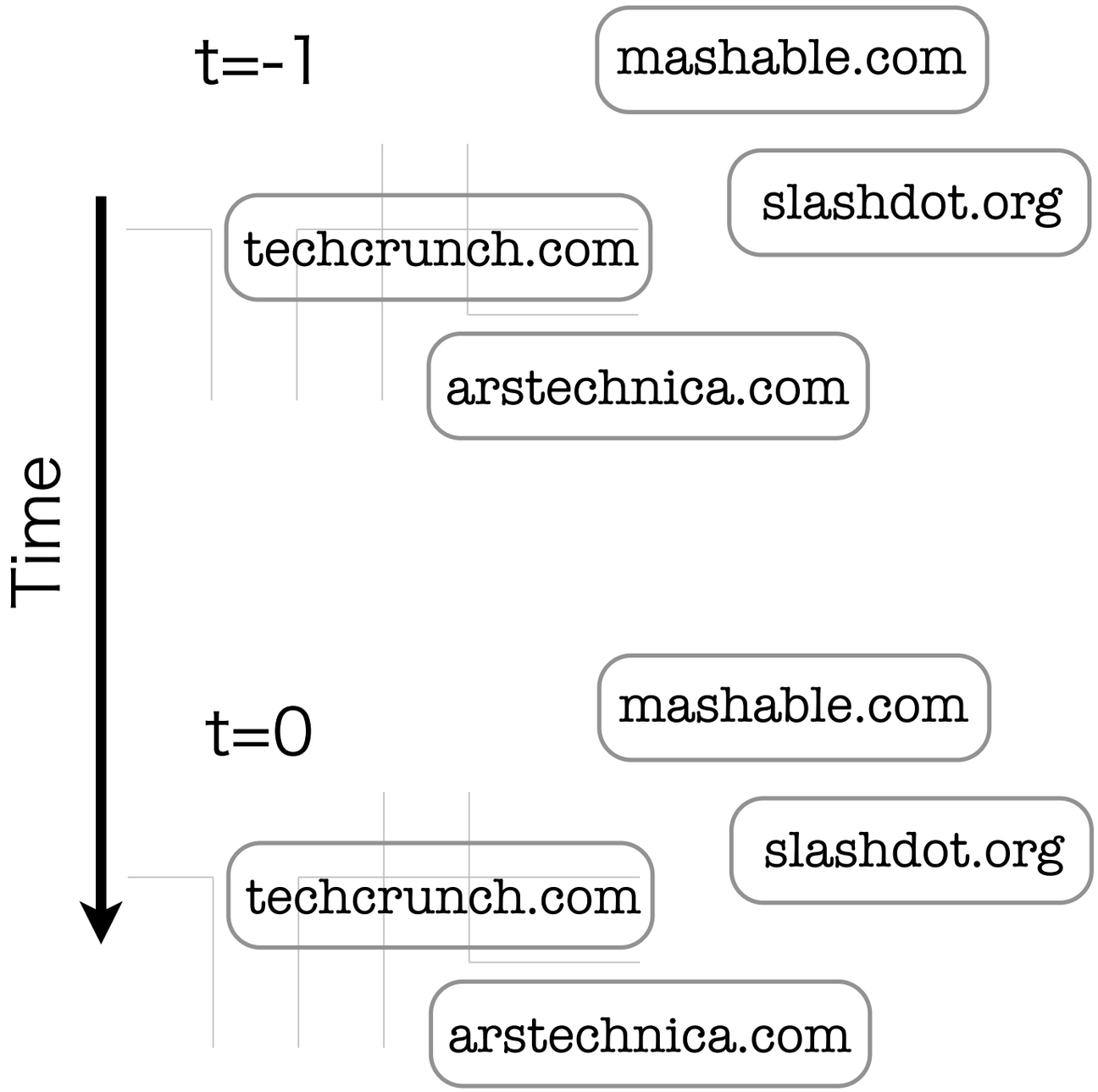
- Music trends on Last.fm [Biessmann et al, AAAI Workshop, 2010]
- Trends in News Articles [Biessmann et al, ICML 2012]
- Spatio-temporal retweet responses to news [Biessmann et al, MLSP 2012]
- Twitter retweet dynamics [Biessmann et al, Nips Workshop 2012]

Canonical Trend Analysis For News Articles

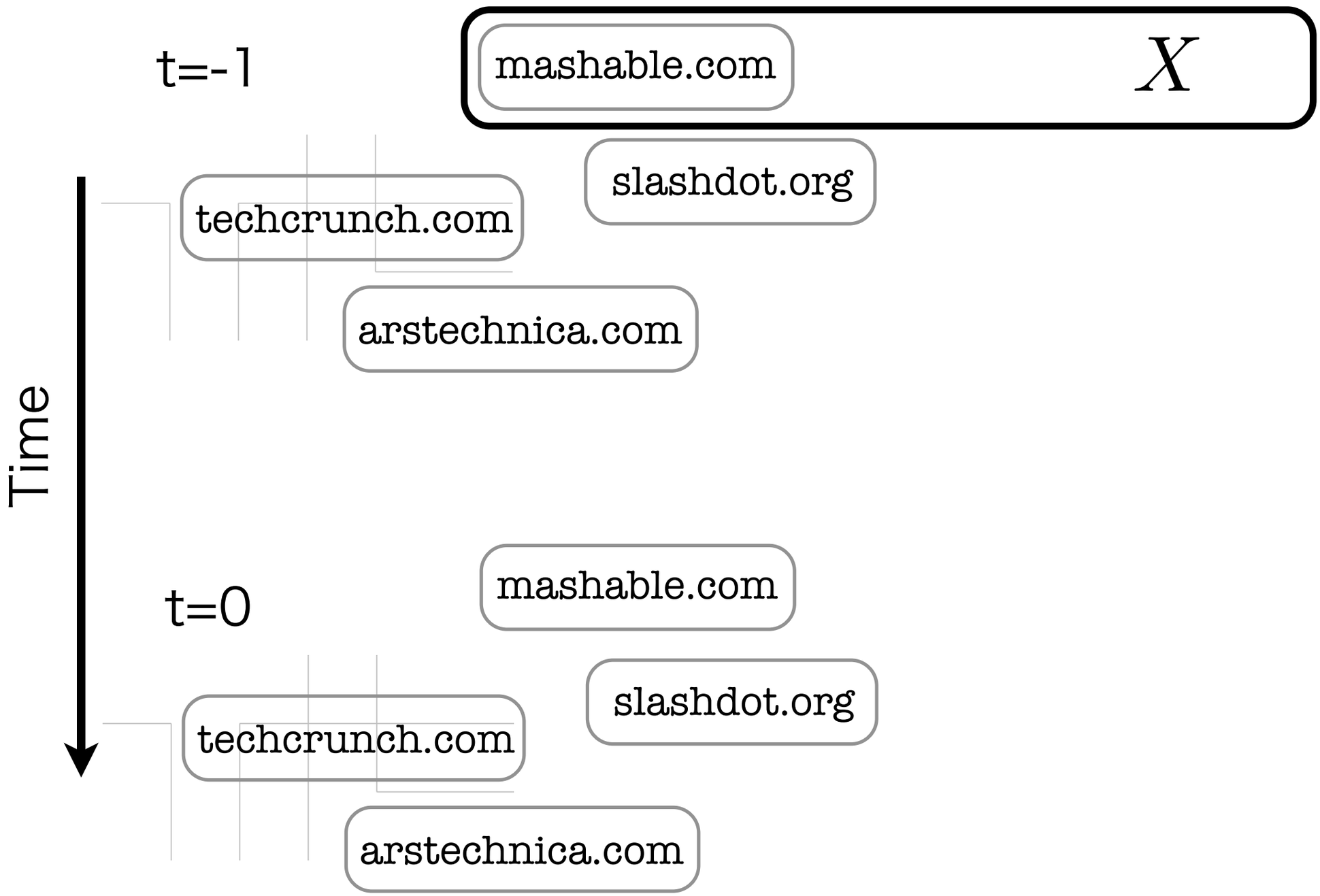
Canonical Trend Analysis For News Articles



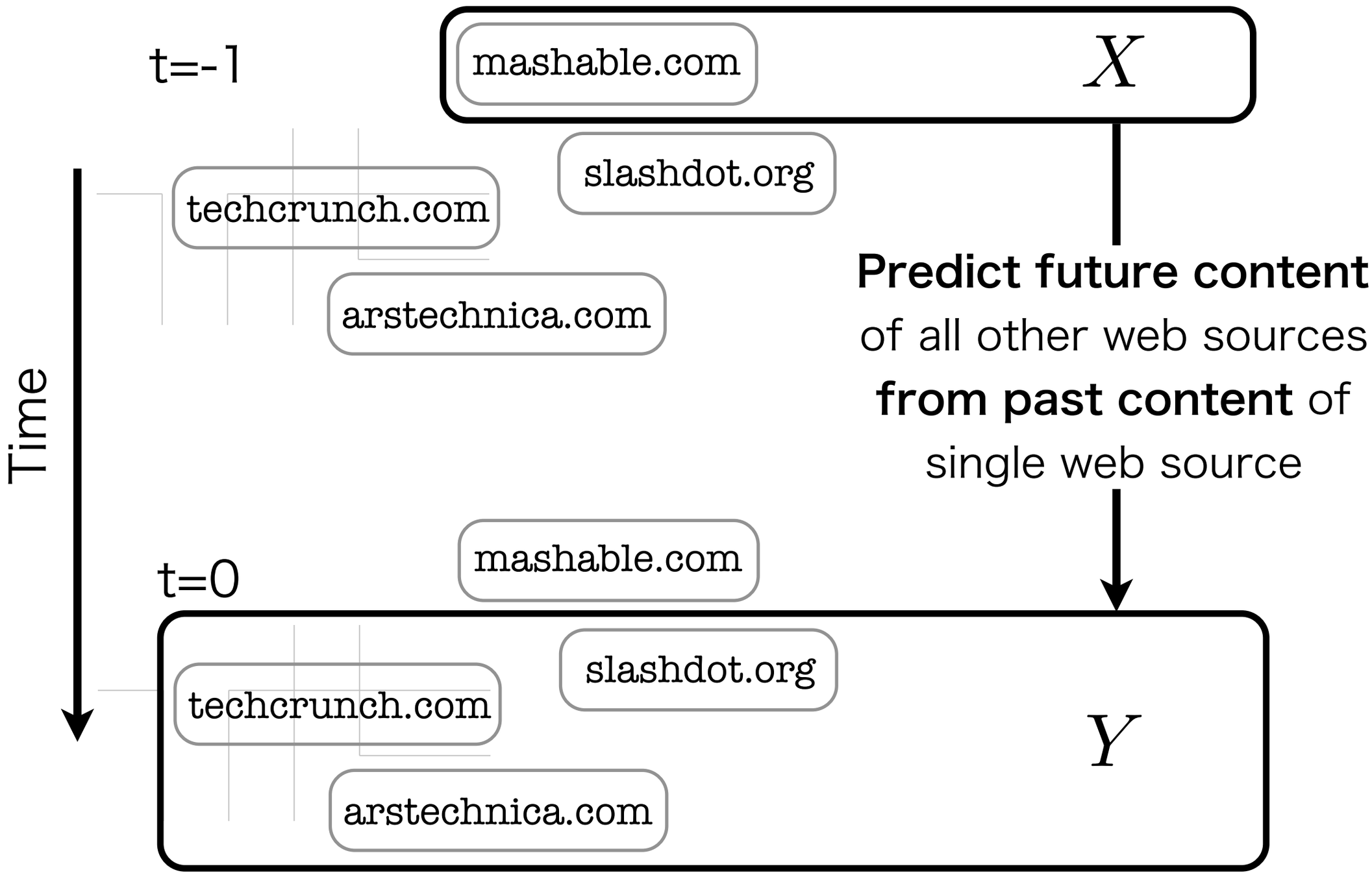
Canonical Trend Analysis For News Articles



Canonical Trend Analysis For News Articles



Canonical Trend Analysis For News Articles



Canonical Trend Analysis For News Articles

Canonical Trend Analysis For News Articles

Canonical Trend Analysis For News Articles

- Which are the main news trends?

Canonical Trend Analysis For News Articles

- Which are the main news trends?
- Who is publishing them first?

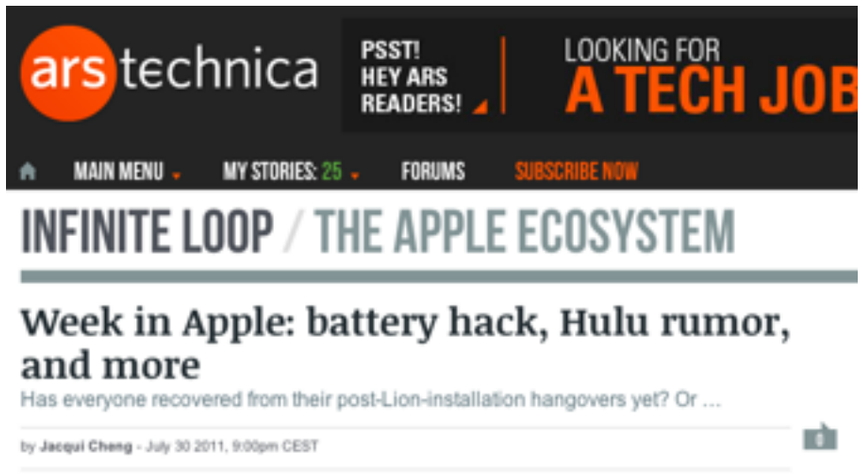
Canonical Trend Analysis For News Articles

- Which are the main news trends?
- Who is publishing them first?
- Who is publishing them later?

Canonical Trend Analysis For News Articles

- Which are the main news trends?
- Who is publishing them first?
- Who is publishing them later?

Canonical Trend Analysis For Social Networks



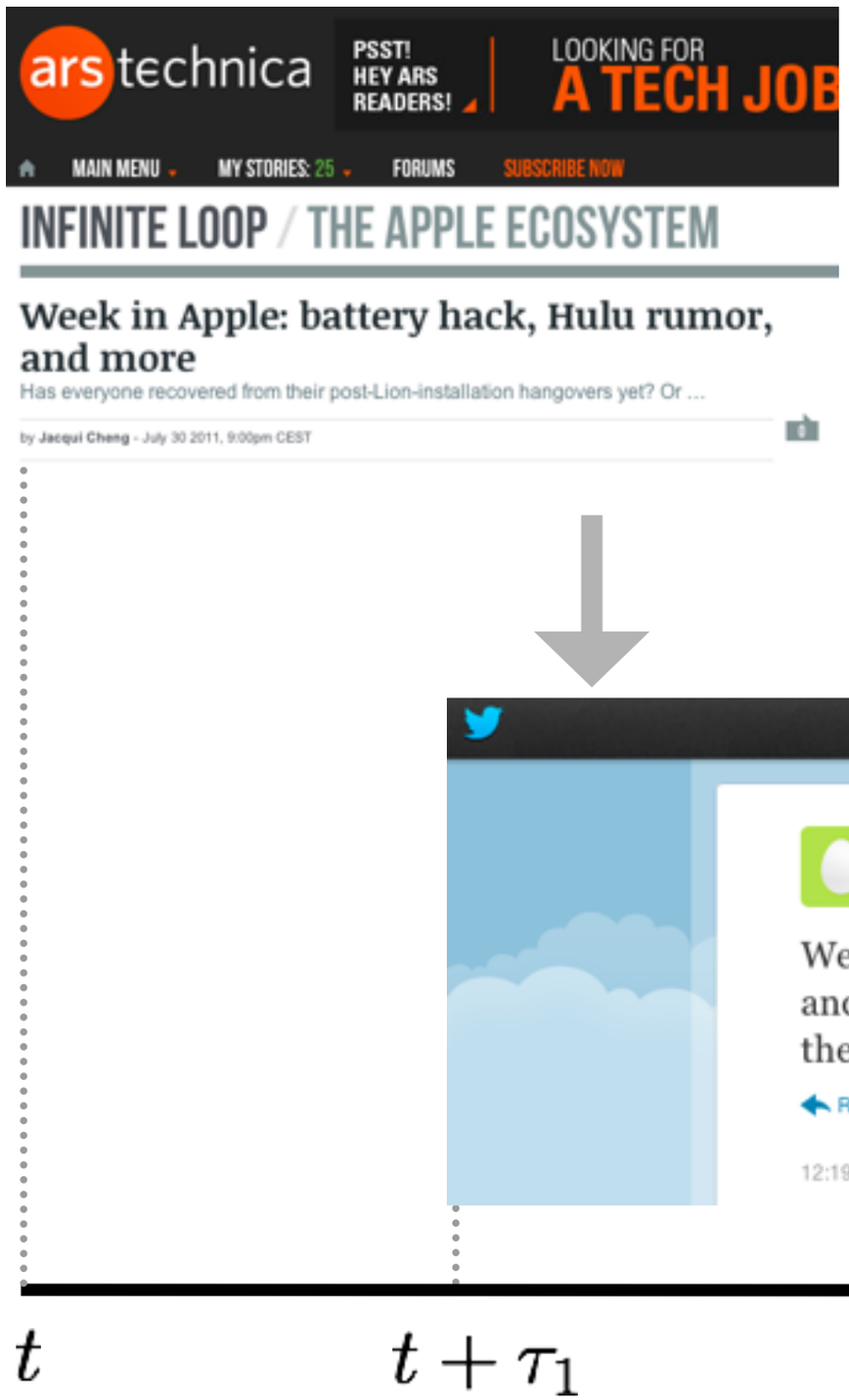
Some **news web site X** publishes some content ...

Canonical Trend Analysis For Social Networks

Some **news web site X** publishes some content ...

... which is **retweeted**

Canonical Trend Analysis For Social Networks



Some **news web site X** publishes some content ...

... which is **retweeted**

Canonical Trend Analysis For Social Networks

Some **news web site X** publishes some content ...

... which is **retweeted**

... at different locations Y

Canonical Trend Analysis For Social Networks

Canonical Trend Analysis For Social Networks

Canonical Trend Analysis For Social Networks

- Can we predict spatio-temporal response of retweets?

Canonical Trend Analysis For Social Networks

- Can we predict spatio-temporal response of retweets?
- Which are the main trends amongst article retweets?

Canonical Trend Analysis For Social Networks

- Can we predict spatio-temporal response of retweets?
- Which are the main trends amongst article retweets?
- How do their spatio-temporal dynamics look like?

Canonical Trend Analysis For Social Networks

- Can we predict spatio-temporal response of retweets?
- Which are the main trends amongst article retweets?
- How do their spatio-temporal dynamics look like?
- Which are the news papers with maximal impact?

Canonical Trend Analysis For Social Networks

- Can we predict spatio-temporal response of retweets?
- Which are the main trends amongst article retweets?
- How do their spatio-temporal dynamics look like?
- Which are the news papers with maximal impact?
- What are the topics with maximal impact?

Canonical Trend Analysis For Social Networks

- Can we predict spatio-temporal response of retweets?
- Which are the main trends amongst article retweets?
- How do their spatio-temporal dynamics look like?
- Which are the news papers with maximal impact?
- What are the topics with maximal impact?

Canonical Trend Analysis

For each web source $f \in \{1, 2, \ldots, F\}$ extract

Feature time series

$$x_f(t) \in \mathbb{R}^W \quad t = \{0, 1, \ldots, T\}$$

Canonical Trend Analysis

For each web source $f \in \{1, 2, \ldots, F\}$ extract

Feature time series

$$x_f(t) \in \mathbb{R}^W \quad t = \{0, 1, \ldots, T\}$$

Single node (user, news site) feature time series

$$X_f = [x_f(t = 1), \ldots, x_f(t = T)] \in \mathbb{R}^{W \times T}$$

Canonical Trend Analysis

For each web source $f \in \{1, 2, \ldots, F\}$ extract

Feature time series

$$x_f(t) \in \mathbb{R}^W \quad t = \{0, 1, \ldots, T\}$$

Single node (user, news site) feature time series

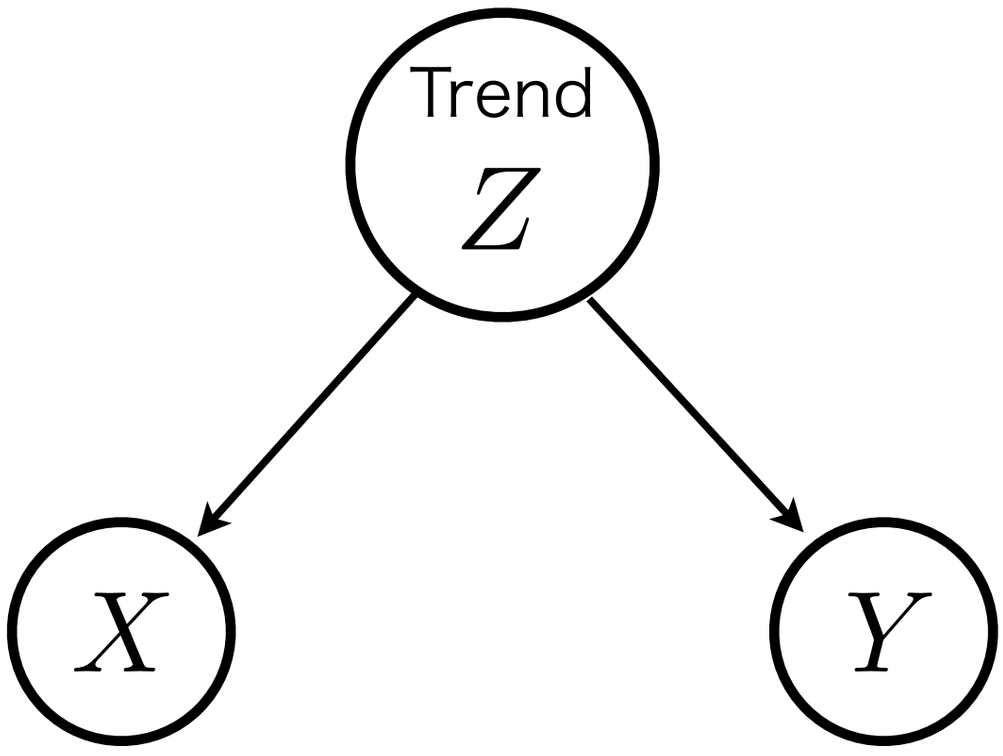
$$X_f = [x_f(t = 1), \ldots, x_f(t = T)] \in \mathbb{R}^{W \times T}$$

Feature time series of **all other nodes**

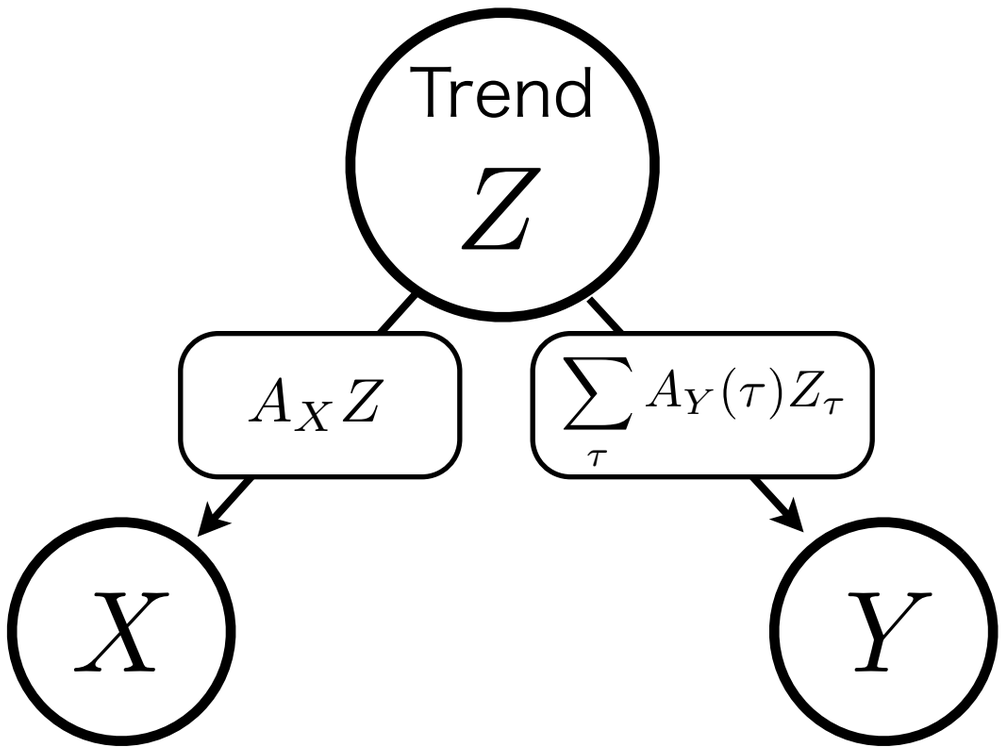
$$Y_f = 1/(F - 1) \sum_{f' \neq f} X_{f'} \in \mathbb{R}^{W \times T}$$


Trend
Z

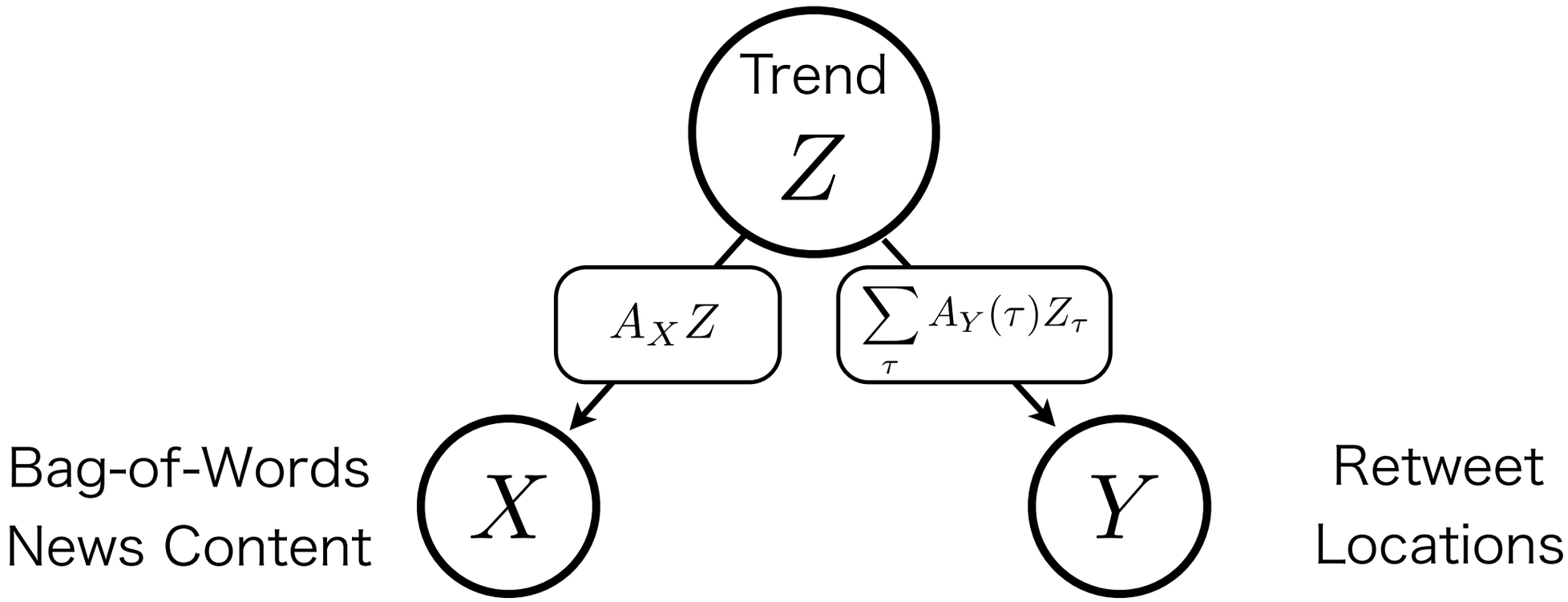
Canonical Trends



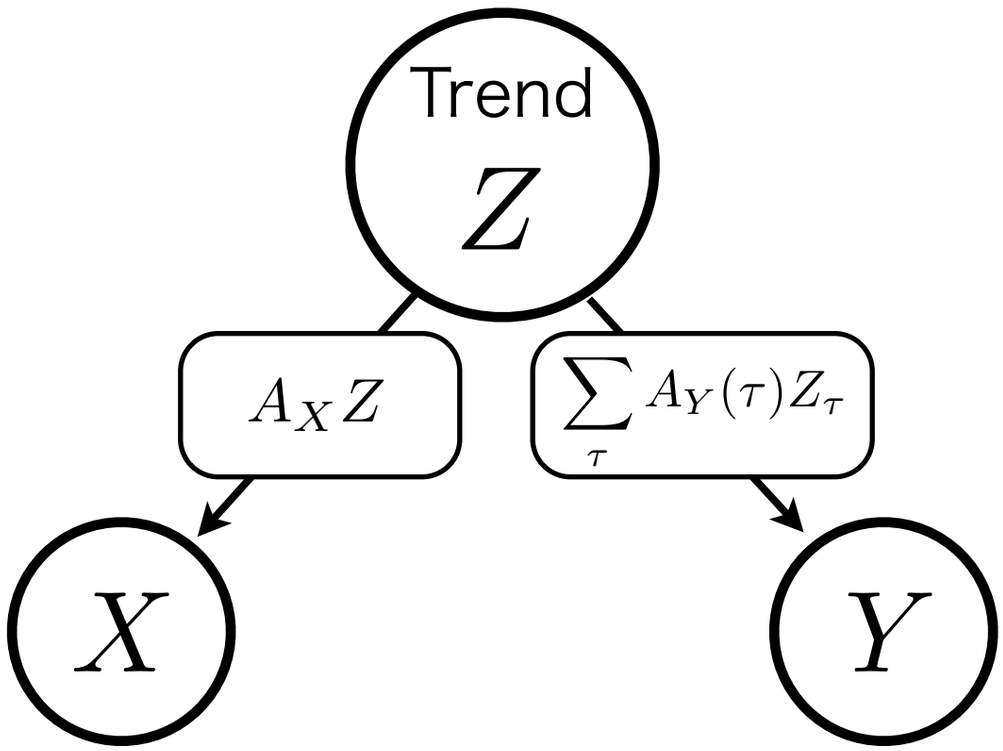
Canonical Trends



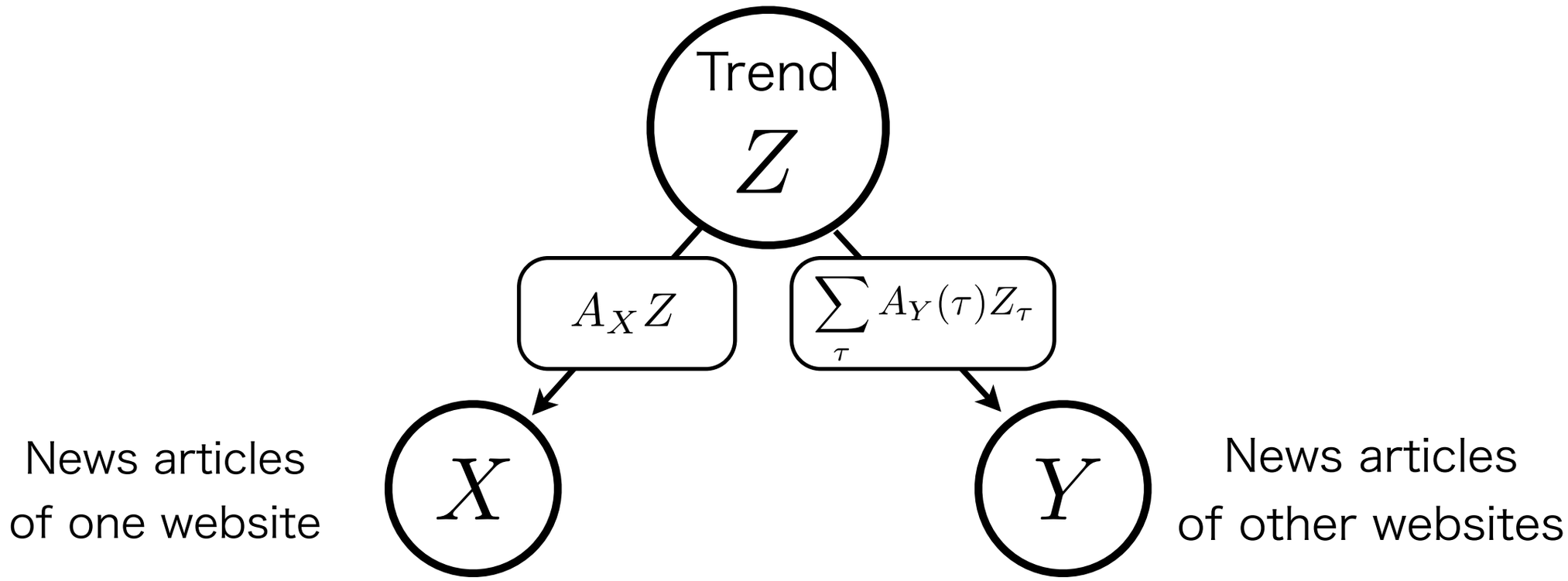
Canonical Trends



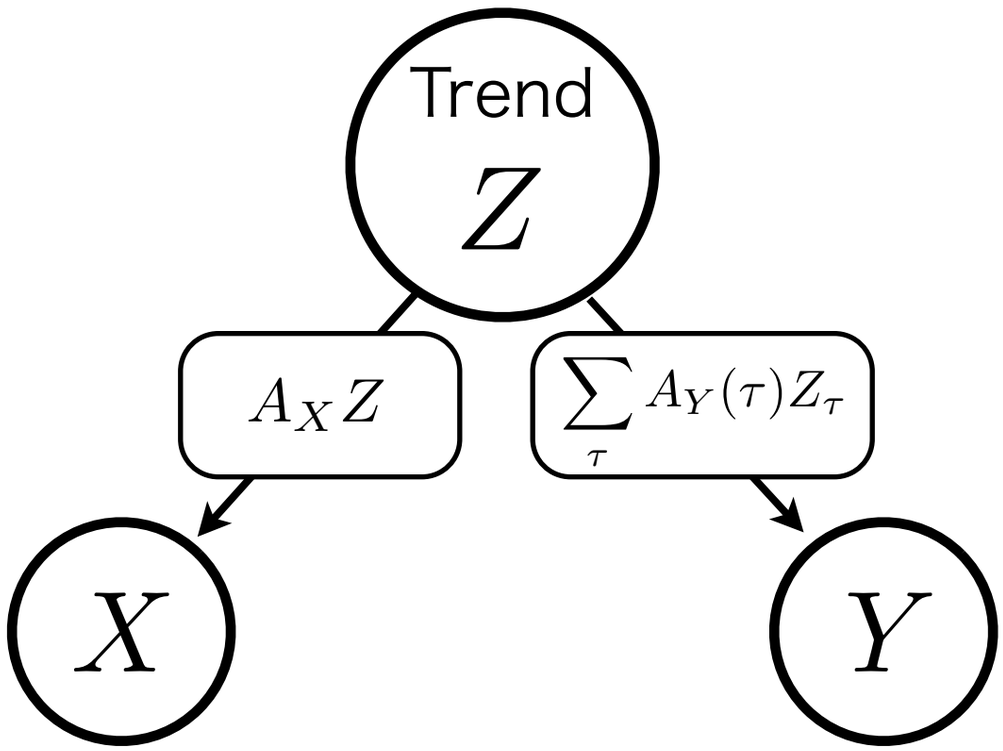
Canonical Trends



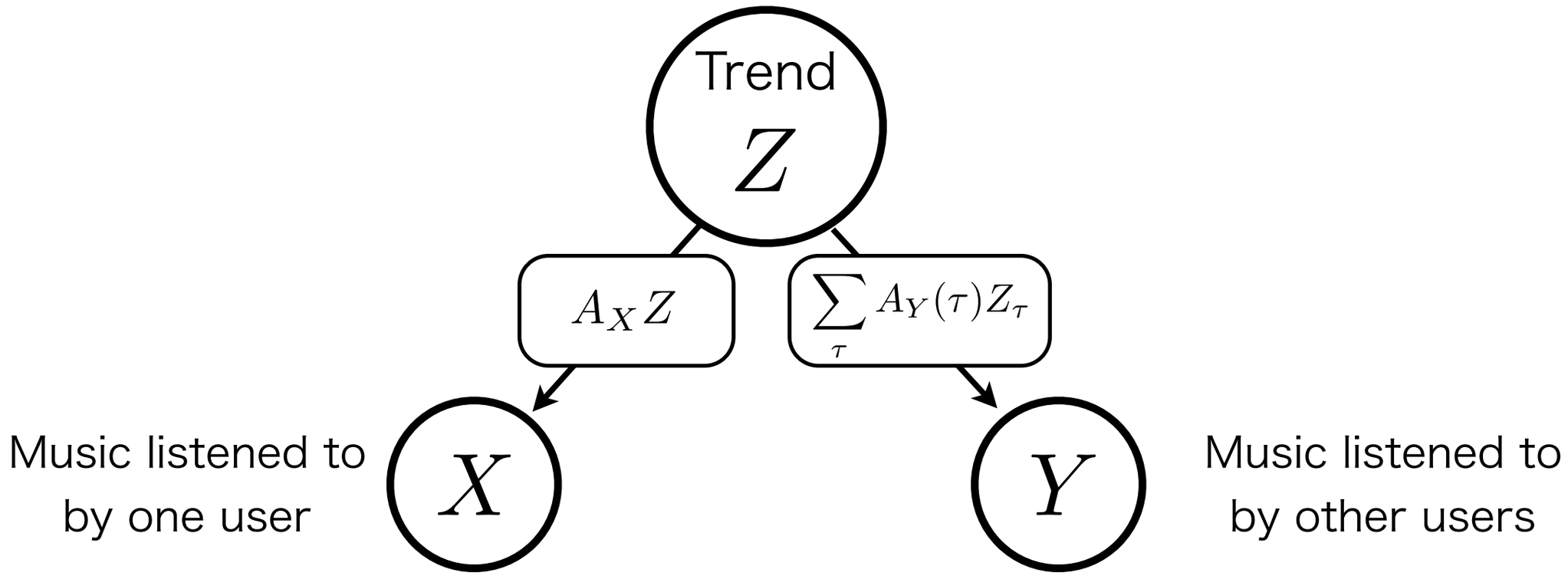
Canonical Trends



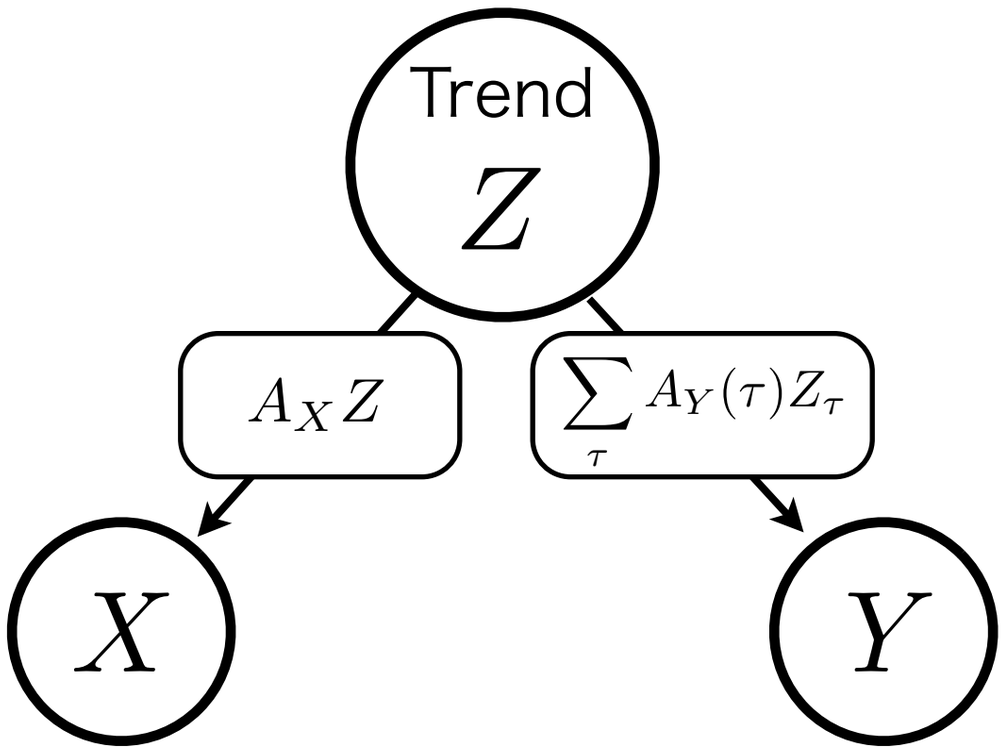
Canonical Trends



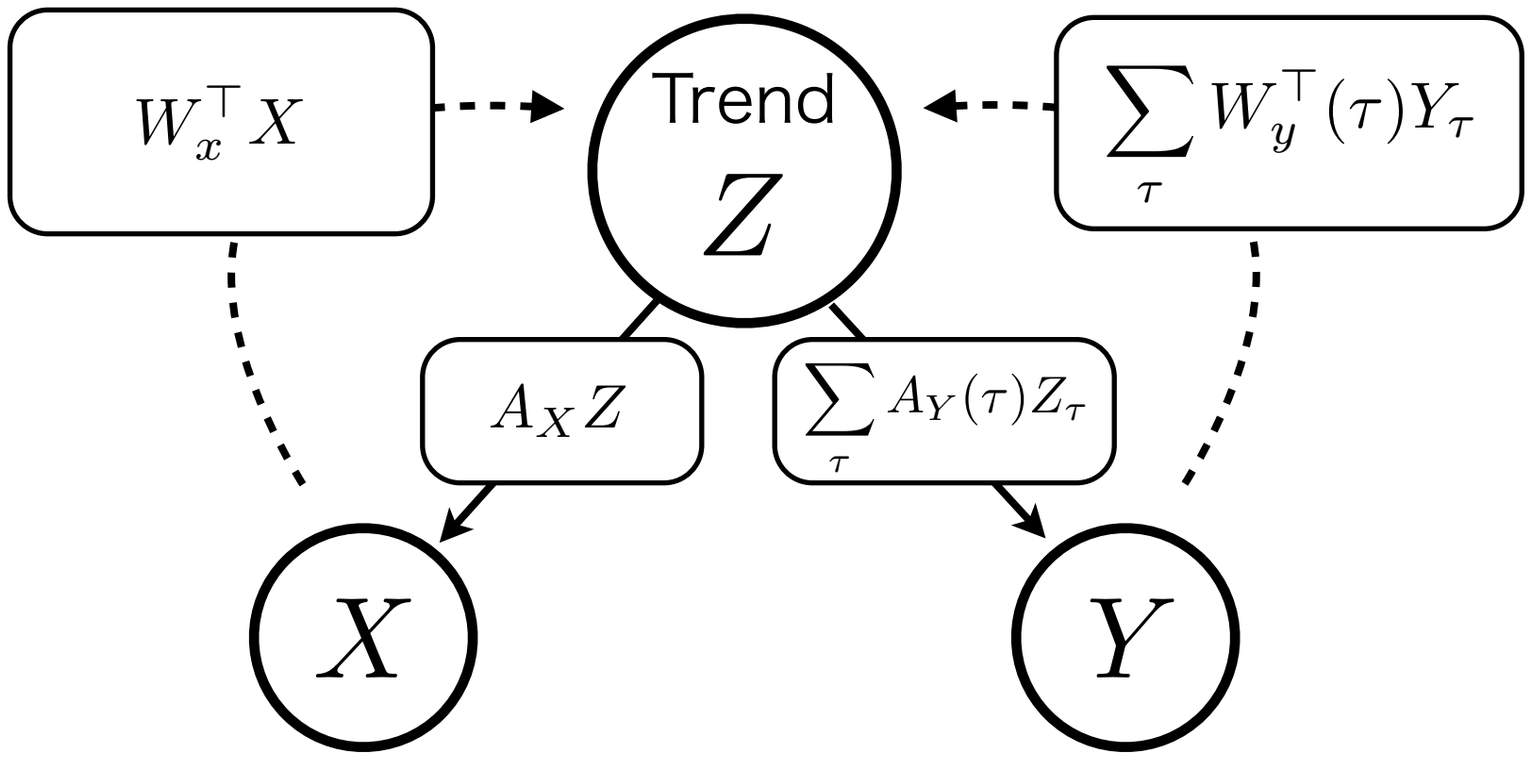
Canonical Trends



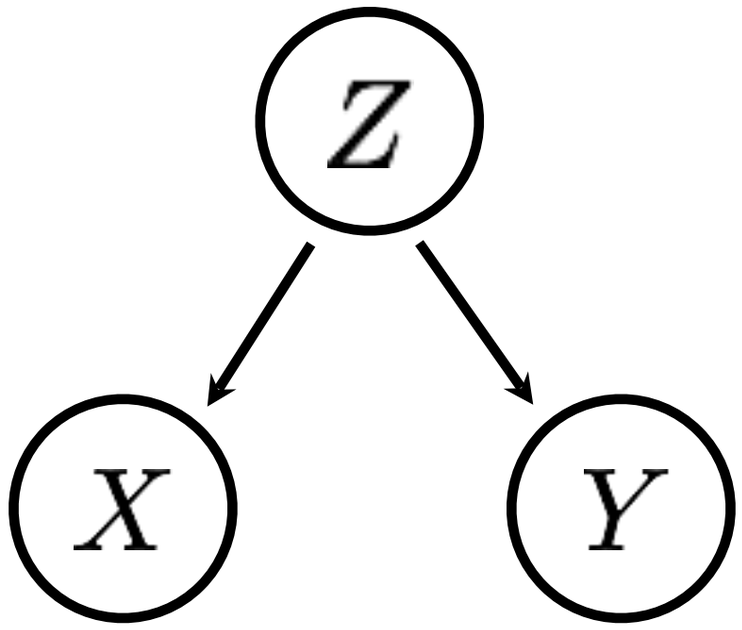
Canonical Trends



Canonical Trends



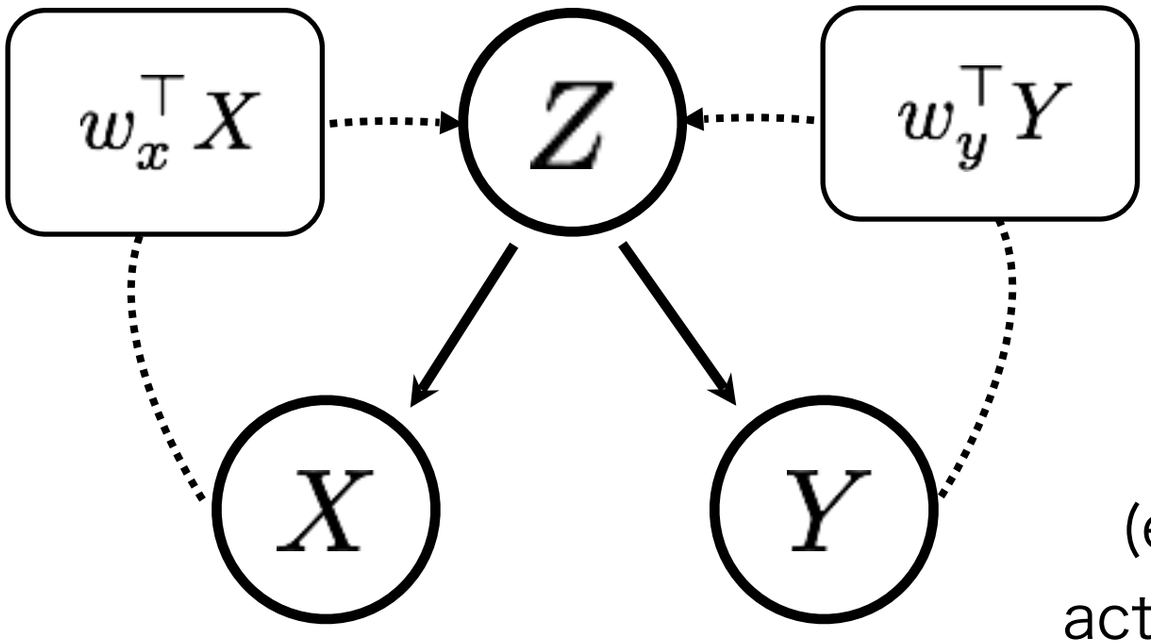
Canonical Correlation Analysis



Latent Variable
(Trend)

Features
(e.g. Bag of Words, User actions, edge histograms, ...)

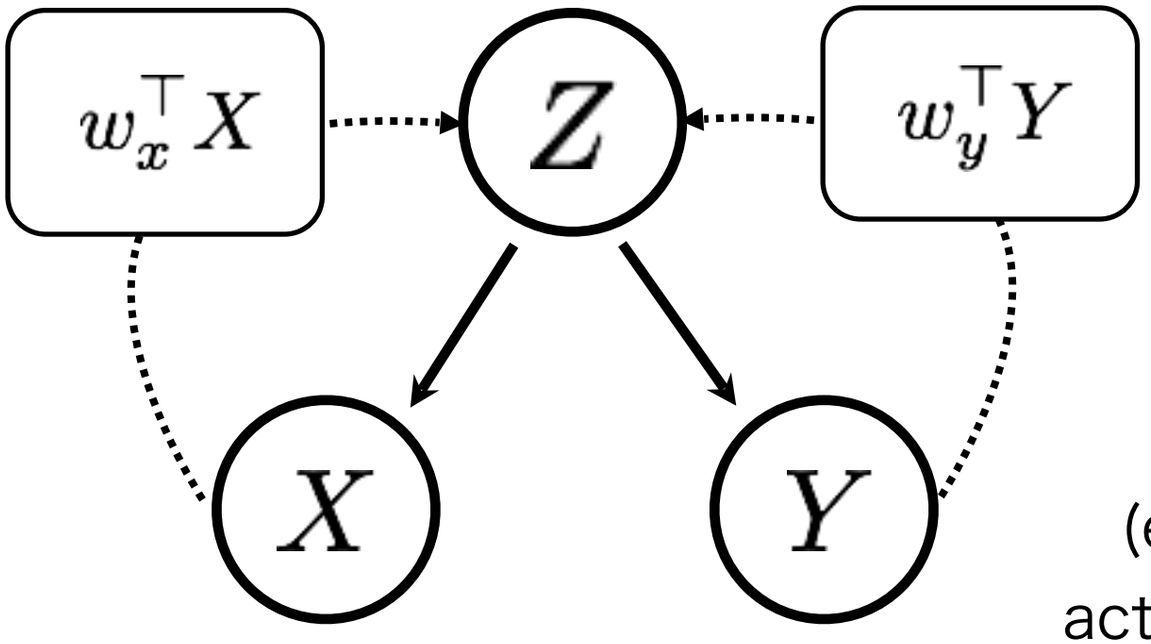
Canonical Correlation Analysis



Latent Variable
(Trend)

Features
(e.g. Bag of Words, User actions, edge histograms, ...)

Canonical Correlation Analysis



Latent Variable
(Trend)

Features
(e.g. Bag of Words, User actions, edge histograms, ...)

$$\underset{w_x,\ w_y}{\operatorname{argmax}} \frac{w_x^\top X Y^\top w_y}{\sqrt{w_x^\top X X^\top w_x w_y^\top Y Y^\top w_y}}$$

[Jordan 1875], [Hotelling 1936], [Bach and Jordan 2006]

Canonical Trend Analysis

Canonical Trend Analysis

Trend

Canonical Trend Analysis

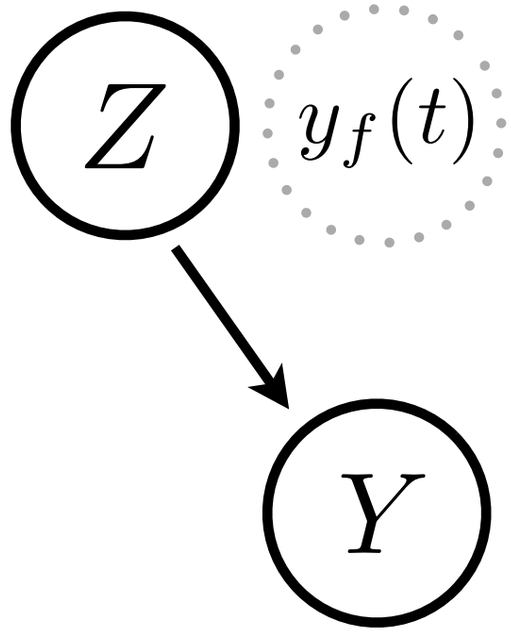
Overall Trend in features of all nodes

$$y_f(t) = w_y^\top Y_f(:, t)$$

Trend

Features

(e.g. Bag of Words)



Canonical Trend Analysis

Overall Trend in features of all nodes

$$y_f(t) = w_y^\top Y_f(:, t)$$

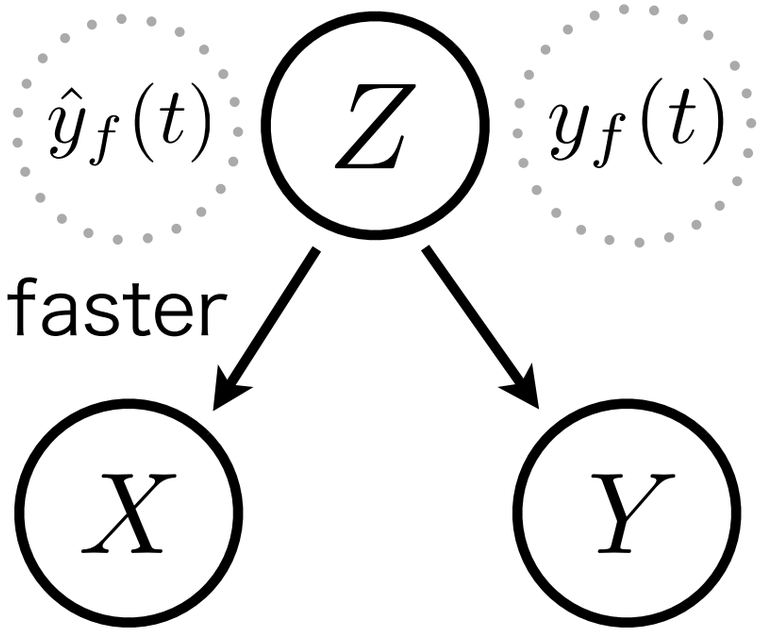
Prediction based on single node features

$$\hat{y}_f(t) = \sum_\tau w_x(\tau)^\top X_f(:, t - \tau)$$

Trend

Features

(e.g. Bag of Words)



Single Topic Example

Overall Trend in features of all nodes

$$y_f(t) = w_y^\top Y_f(:, t)$$

Prediction based on single node features

$$\hat{y}_f(t) = \sum_\tau w_x(\tau)^\top X_f(:, t - \tau)$$

Optimal $w_y \in \mathbb{R}^W$ and $w_x(\tau) \in \mathbb{R}^{W \times N_\tau}$

$$\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}(y_f(t), \hat{y}_f(t))$$

Efficient Computation of Canonical Trends

Efficient Computation of Canonical Trends

Temporal Embedding

$$\tilde{X}_f = \begin{bmatrix} X_{f,\tau=-N_\tau} \\ \vdots \\ X_{f,\tau=-1} \end{bmatrix} \in \mathbb{R}^{W N_\tau \times T}$$

- Converts problem into standard CCA problem

  Jordan 1875, Hotelling 1936, Anderson 1999

Efficient Computation of Canonical Trends

Temporal Embedding

$$\tilde{X}_f = \begin{bmatrix} X_{f,\tau=-N_\tau} \\ \vdots \\ X_{f,\tau=-1} \end{bmatrix} \in \mathbb{R}^{W N_\tau \times T}$$

- Converts problem into standard CCA problem

Jordan 1875, Hotelling 1936, Anderson 1999

(linear) 'Kernel Trick'

$$w_x(\tau) = X_{f,\tau}\alpha$$
$$w_y = Y_f\beta$$

- $2T$ parameters instead of $WN_\tau + W$ parameters
- Very efficient for high-dimensional feature spaces

Fyfe 2000, Fukumizu 2007

Efficient Computation of Canonical Trends

Objective function is maximized in the dual

$$\begin{aligned}\mathrm{Corr}(y(t), \hat{y}(t)) =& \frac{\sum_\tau (w_x(\tau)^\top X_\tau)^\top Y w_y}{\sqrt{\sum_\tau (w_x(\tau)^\top X_\tau X_\tau^\top w_x(\tau)) w_y^\top Y Y^\top w_y}} \\ =& \frac{\alpha^\top K_{\tilde{X}} K_Y \beta}{\sqrt{\alpha^\top K_{\tilde{X}}^2 \alpha \beta^\top K_Y^2 \beta}}\end{aligned}$$

where $K_{\tilde{X}} = \tilde{X}^\top \tilde{X}$, $K_Y = Y^\top Y$ are linear kernels

Efficient Computation of Canonical Trends

Objective function is maximized in the dual

$$\begin{aligned}\mathrm{Corr}(y(t), \hat{y}(t)) &= \frac{\sum_\tau (w_x(\tau)^\top X_\tau)^\top Y w_y}{\sqrt{\sum_\tau (w_x(\tau)^\top X_\tau X_\tau^\top w_x(\tau)) w_y^\top Y Y^\top w_y}} \\ &= \frac{\alpha^\top K_{\tilde{X}} K_Y \beta}{\sqrt{\alpha^\top K_{\tilde{X}}^2 \alpha \beta^\top K_Y^2 \beta}}\end{aligned}$$

Dual coefficients are solution to generalized eigenvalue equation

$$\begin{bmatrix} 0 & K_{\tilde{X}} K_Y \\ K_Y K_{\tilde{X}} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} K_{\tilde{X}}^2 + I\kappa & 0 \\ 0 & K_Y^2 + I\kappa \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Efficient Computation of Canonical Trends

Objective function is maximized in the dual

$$\begin{aligned}\mathrm{Corr}(y(t), \hat{y}(t)) &= \frac{\sum_\tau (w_x(\tau)^\top X_\tau)^\top Y w_y}{\sqrt{\sum_\tau (w_x(\tau)^\top X_\tau X_\tau^\top w_x(\tau)) w_y^\top Y Y^\top w_y}} \\ &= \frac{\alpha^\top K_{\tilde{X}} K_Y \beta}{\sqrt{\alpha^\top K_{\tilde{X}}^2 \alpha \beta^\top K_Y^2 \beta}}\end{aligned}$$

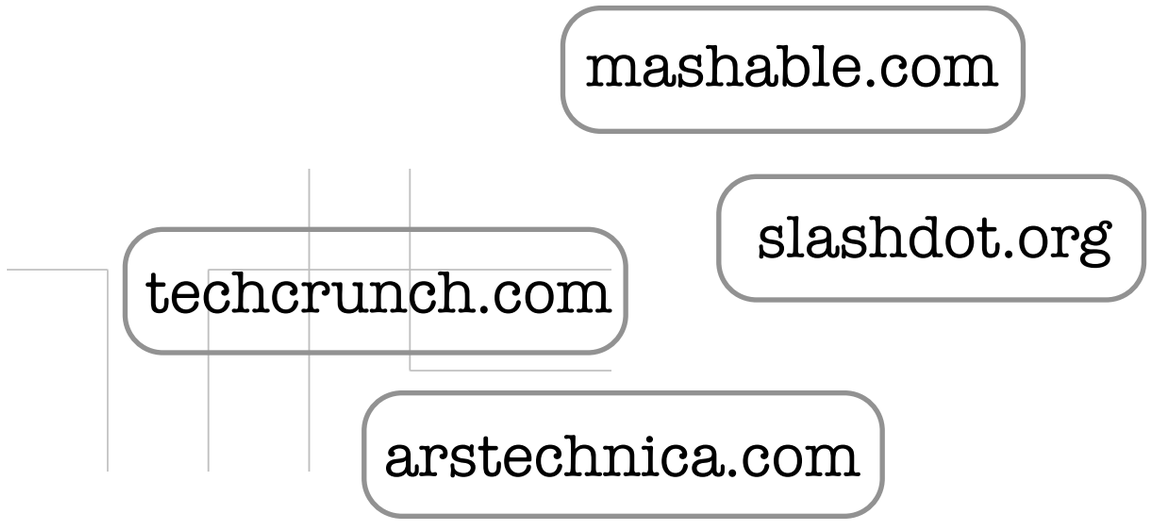
Dual coefficients are solution to generalized eigenvalue equation

$$\begin{bmatrix} 0 & K_{\tilde{X}} K_Y \\ K_Y K_{\tilde{X}} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} K_{\tilde{X}}^2 + I\kappa & 0 \\ 0 & K_Y^2 + I\kappa \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

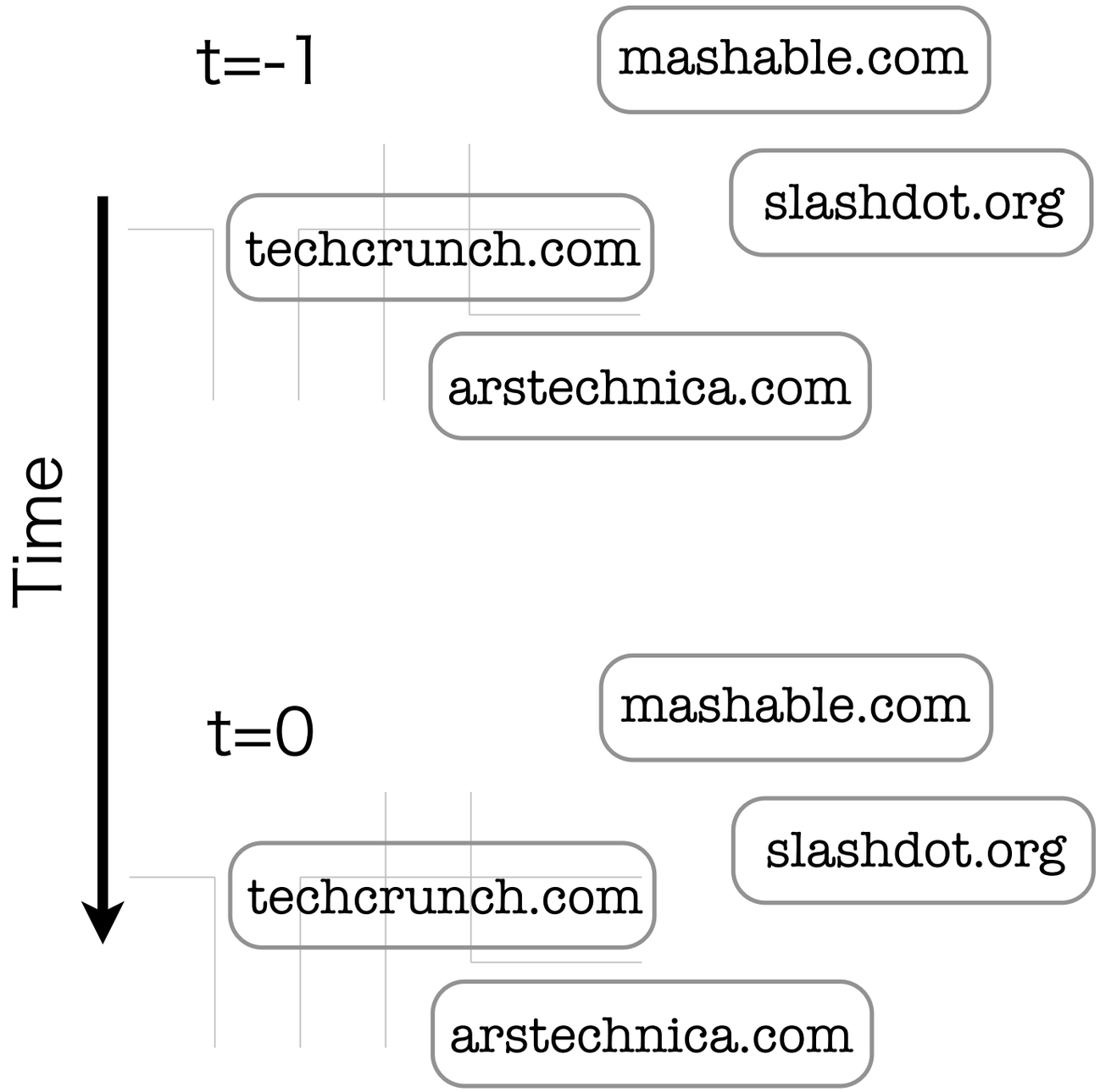
Bießmann et al, Machine Learning, 2010

Canonical Trend Analysis For News Articles

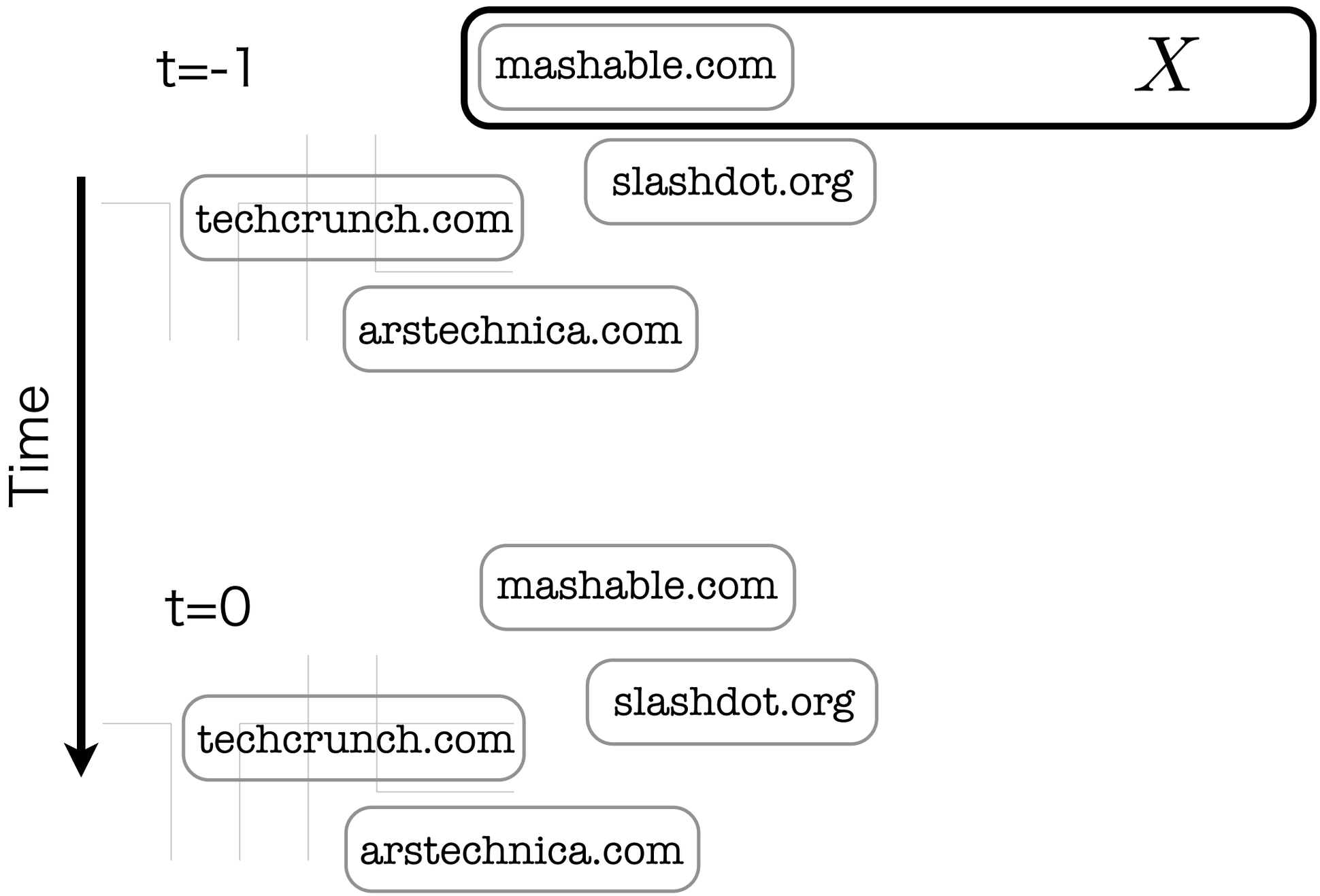
Canonical Trend Analysis For News Articles



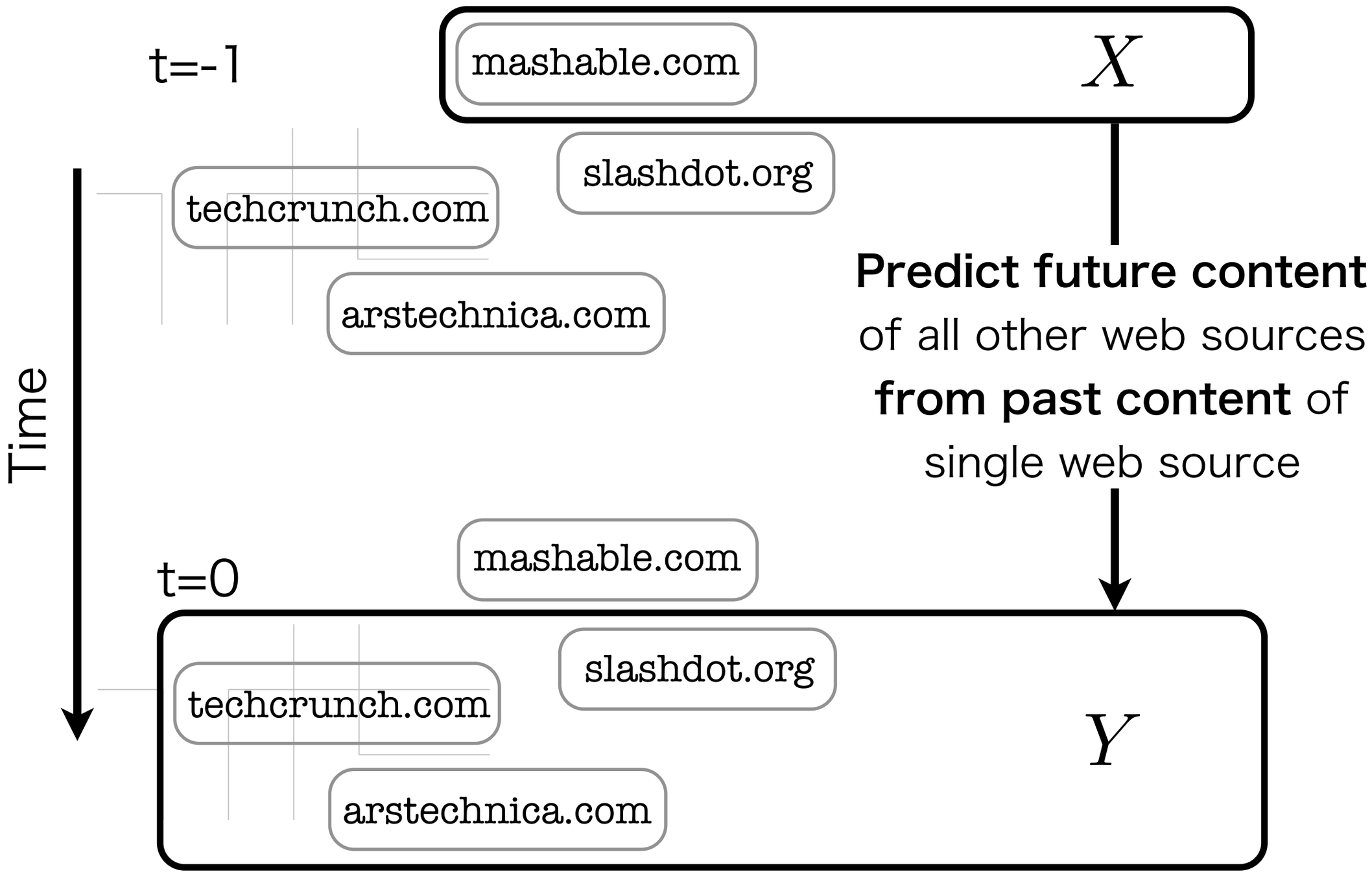
Canonical Trend Analysis For News Articles



Canonical Trend Analysis For News Articles



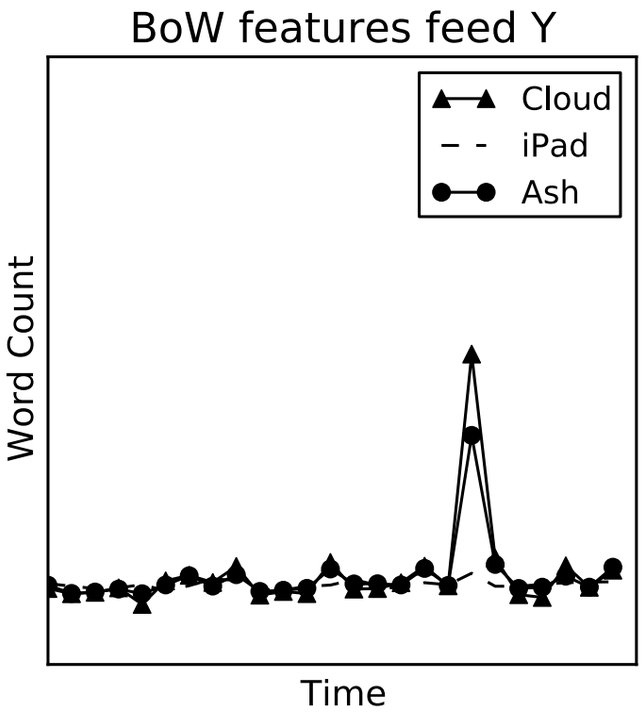
Canonical Trend Analysis For News Articles

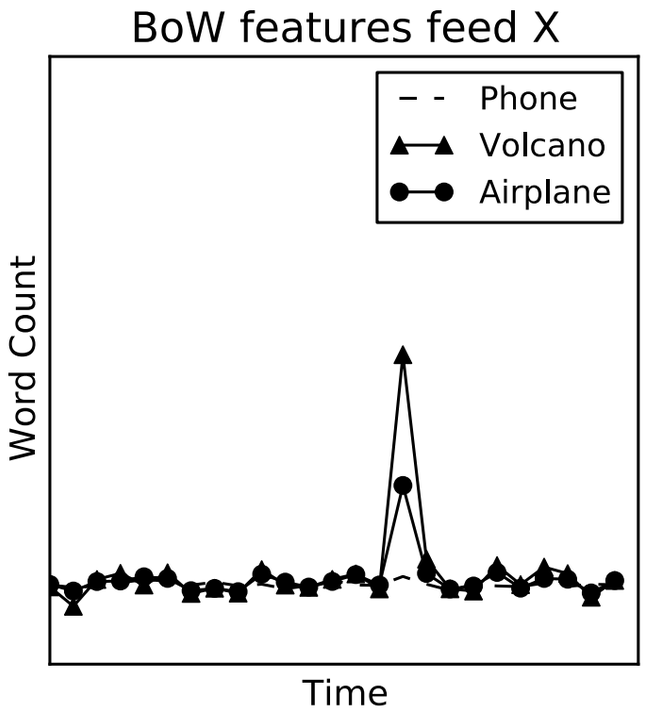


Detecting ‘Trendsetting’ News Websites

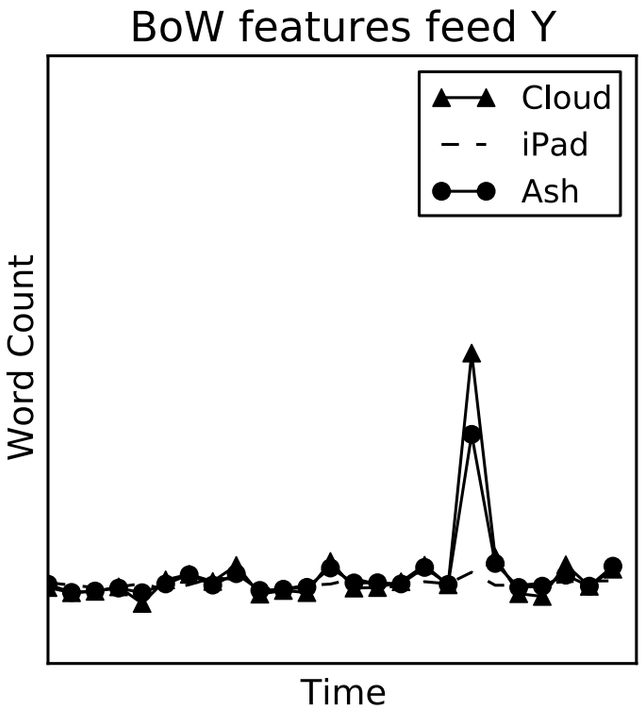
A Toy Data Example:
News on eruption of Eyjafjallajoekull

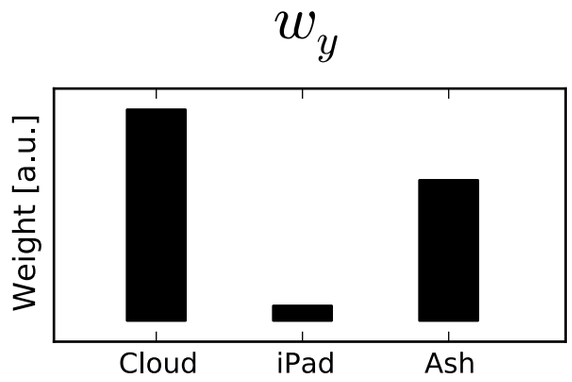
Detecting ‘Trendsetting’ News Websites





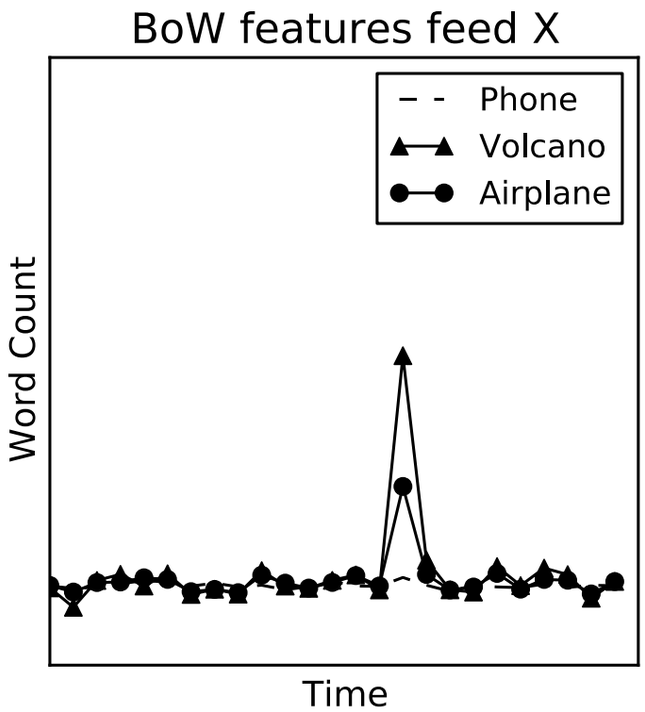
Detecting ‘Trendsetting’ News Websites

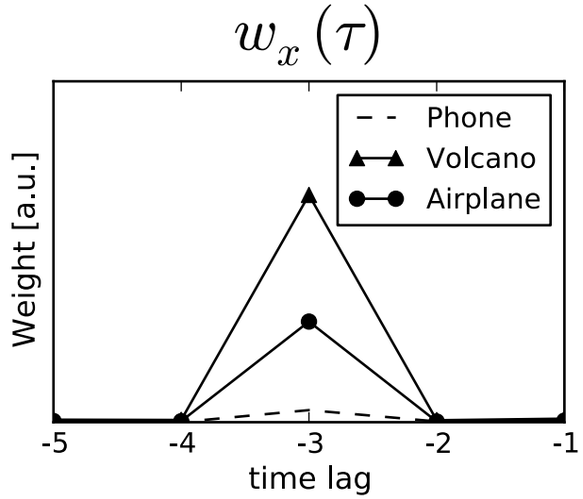




Canonical Trend

Analysis



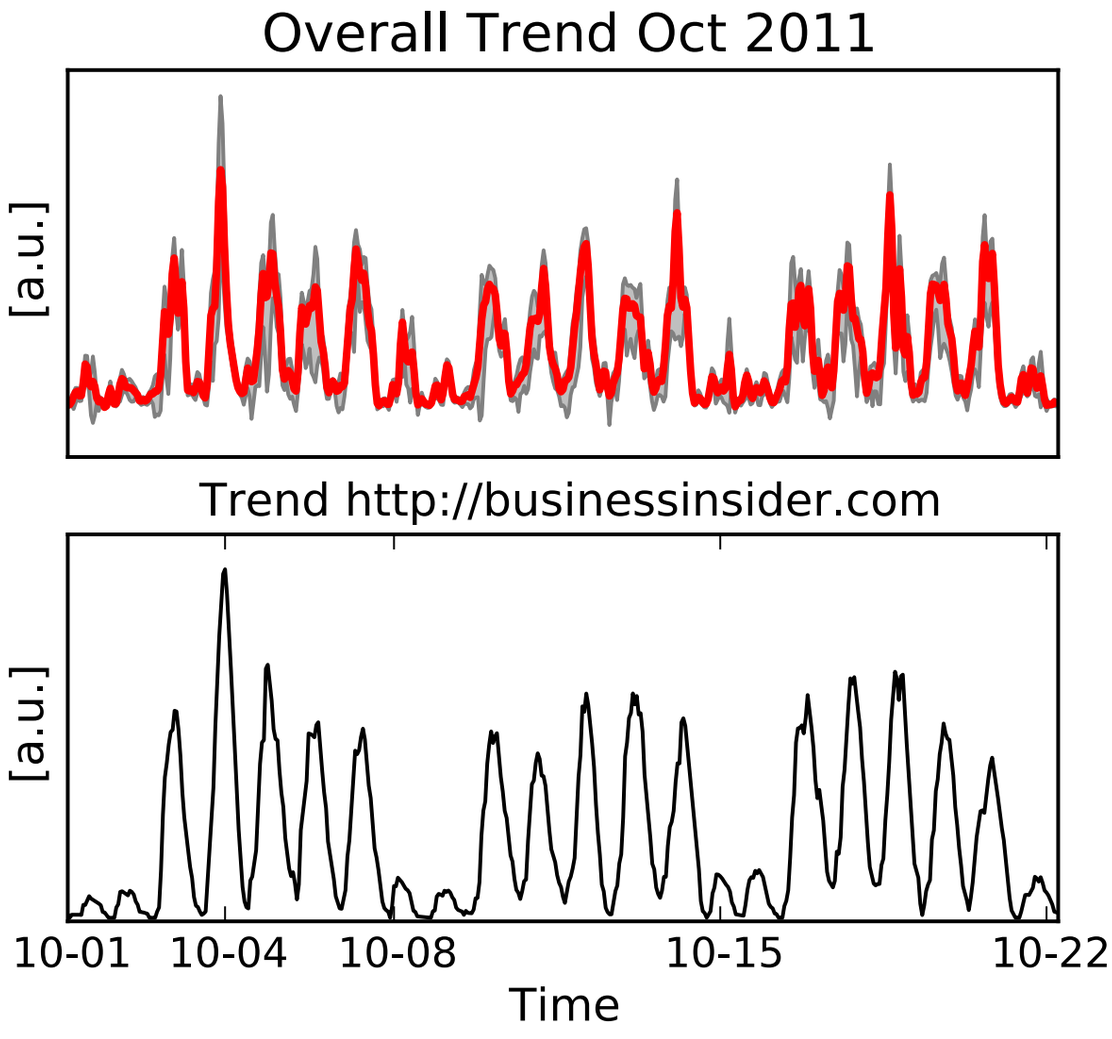


Detecting ‘Trendsetting’ News Websites

Real Data Example:

BoW Features from 96 Technology News Feeds in October 2011

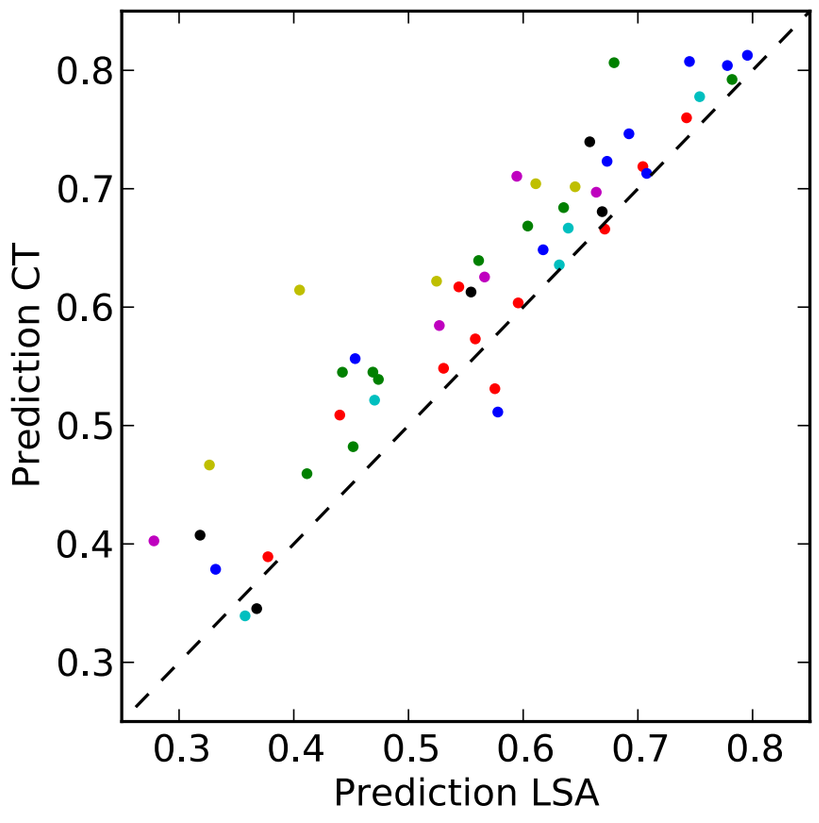
$y_f(t)$



$\hat{y}_f(t)$

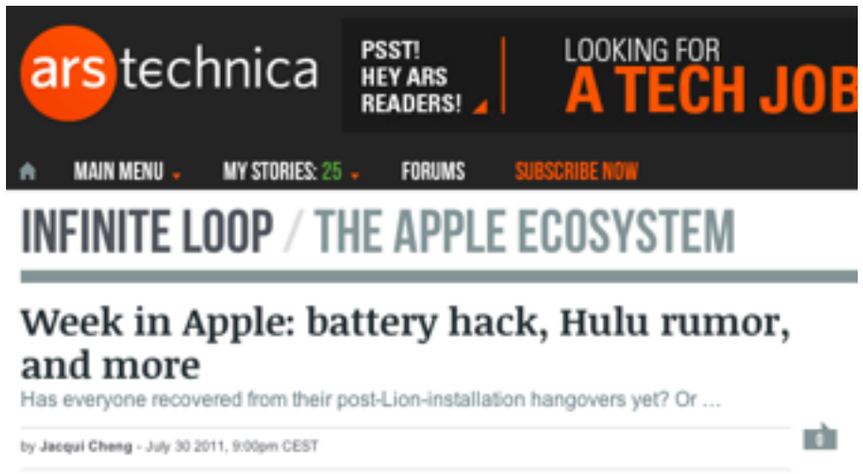
Comparison Canonical Trend Analysis and LSA

Canonical trend analysis **between** X_f and Y_f
vs. LSA on X_f and Y_f **separately**



Canonical Topics predict overall topics
better than Latent Semantic Indexing

Canonical Trend Analysis For Social Networks



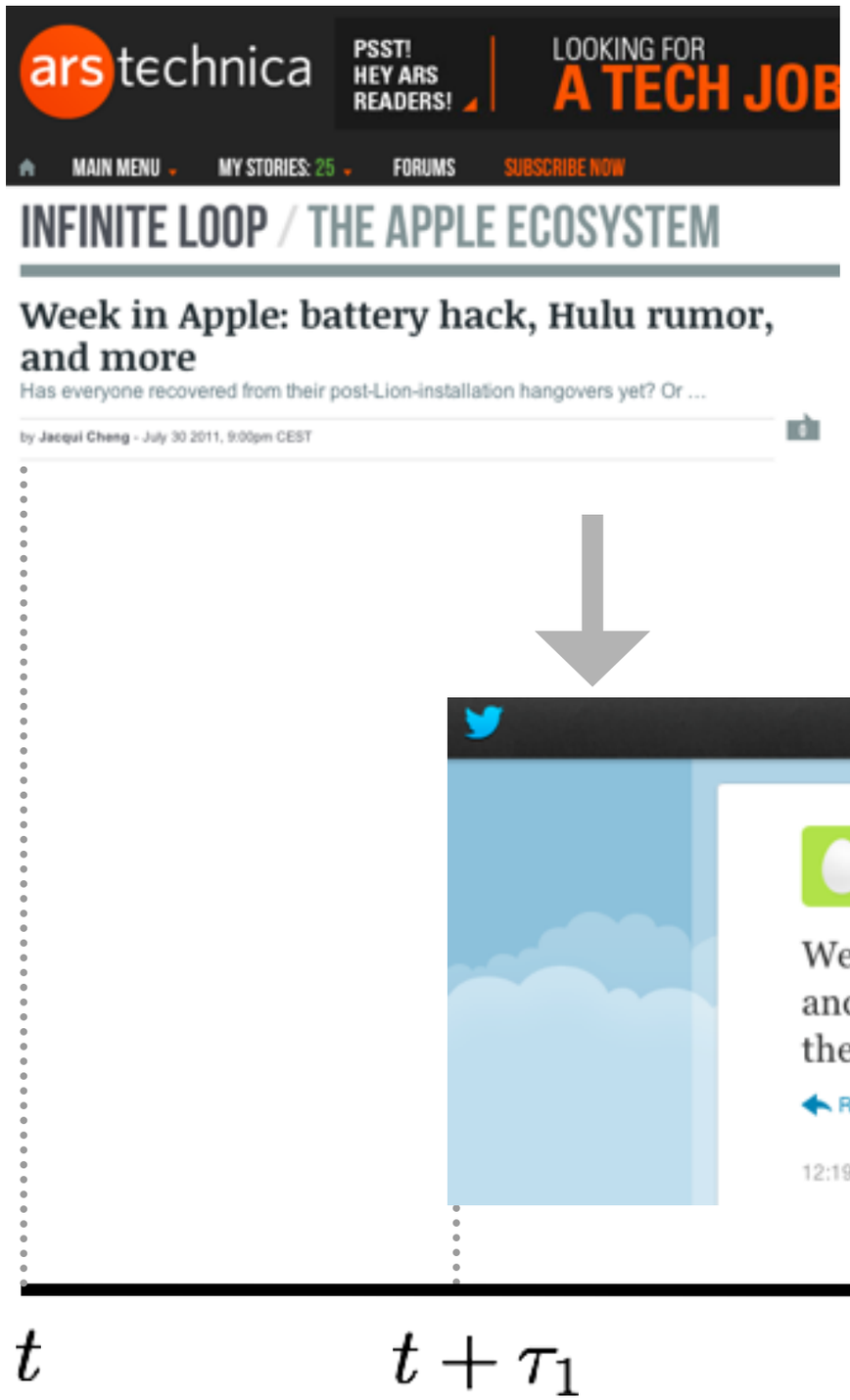
Some **news web site X** publishes some content ...

Canonical Trend Analysis For Social Networks

Some **news web site X** publishes some content ...

... which is **retweeted**

Canonical Trend Analysis For Social Networks



Some **news web site X** publishes some content ...

... which is **retweeted**

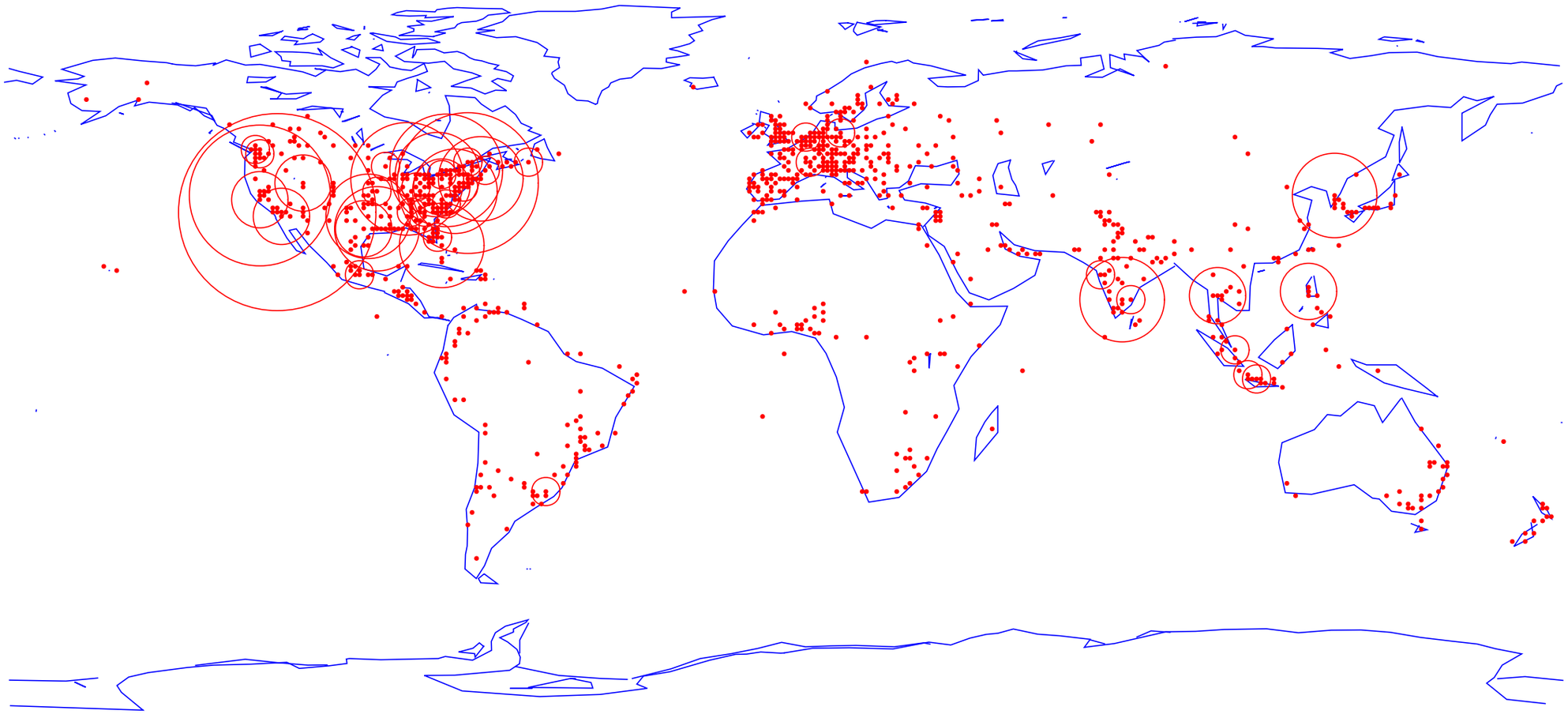
Canonical Trend Analysis For Social Networks

Some **news web site X** publishes some content ...

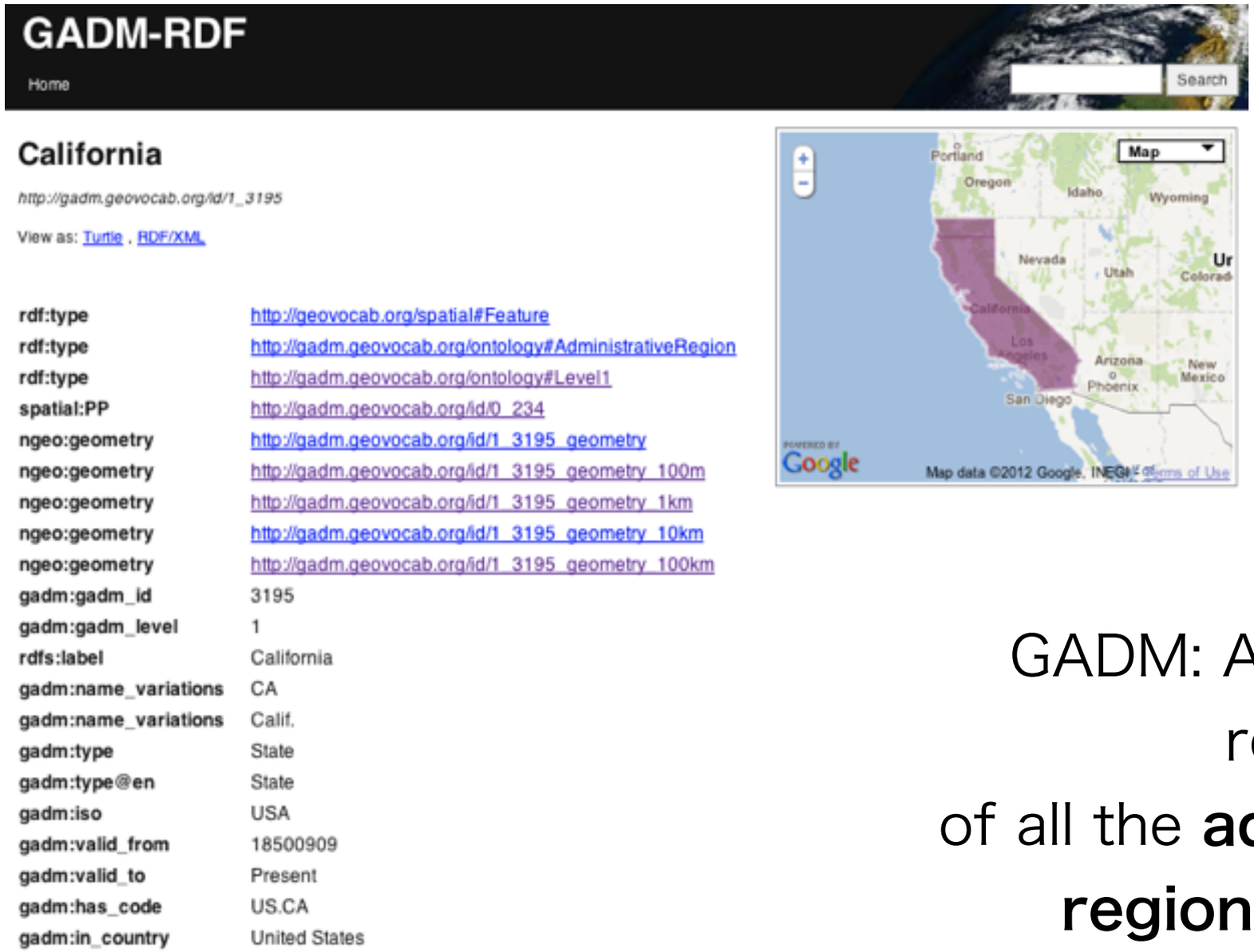
... which is **retweeted**

... at different locations Y

Spatio-temporal Impact of Retweeted News Articles



Downsampling of Geographic Information



GADM: An RDF spatial representation of all the **administrative regions** in the world

Comparisons: Mean, PCA and Canonical Trends

Comparisons: Mean, PCA and Canonical Trends

Canonical Trends

Mean

PCA

Comparisons: Mean, PCA and Canonical Trends

Canonical Trends

$$\underset{w_y(\tau), w_x}{\operatorname{argmax}} \operatorname{Corr}(\hat{x}_f(t), \hat{y}_f(t))$$

Mean

PCA

Comparisons: Mean, PCA and Canonical Trends

Canonical Trends

$$\underset{w_y(\tau), w_x}{\operatorname{argmax}} \operatorname{Corr}(\hat{x}_f(t), \hat{y}_f(t))$$

Mean

$$w_x^\top = \mathbf{1}_x/N, w_y(\tau) = \mathbf{1}_y/N$$

PCA

Comparisons: Mean, PCA and Canonical Trends

Canonical Trends

$$\underset{w_y(\tau), w_x}{\operatorname{argmax}} \operatorname{Corr}(\hat{x}_f(t), \hat{y}_f(t))$$

Mean

$$w_x^\top = \mathbf{1}_x/N, w_y(\tau) = \mathbf{1}_y/N$$

PCA

$$\underset{w_y(\tau)}{\operatorname{argmax}}(w_y(\tau)^\top \tilde{Y}_f \tilde{Y}_f^\top w_y(\tau)),$$

$$\underset{w_x}{\operatorname{argmax}}(w_x^\top X X^\top w_x),$$

$$\text{s.t. } w_y(\tau)^\top w_y(\tau) = w_x^\top w_x = 1$$

Comparisons: Mean, PCA and Canonical Trends

	Hypothesis
Canonical Trends	News Content helps predicting retweet frequency
Mean	Mean Wordcount predicts mean tweet frequency best
PCA	Wordcount variance predicts tweet variance

Comparisons: Mean, PCA and Canonical Trends

Canonical Trends

Mean

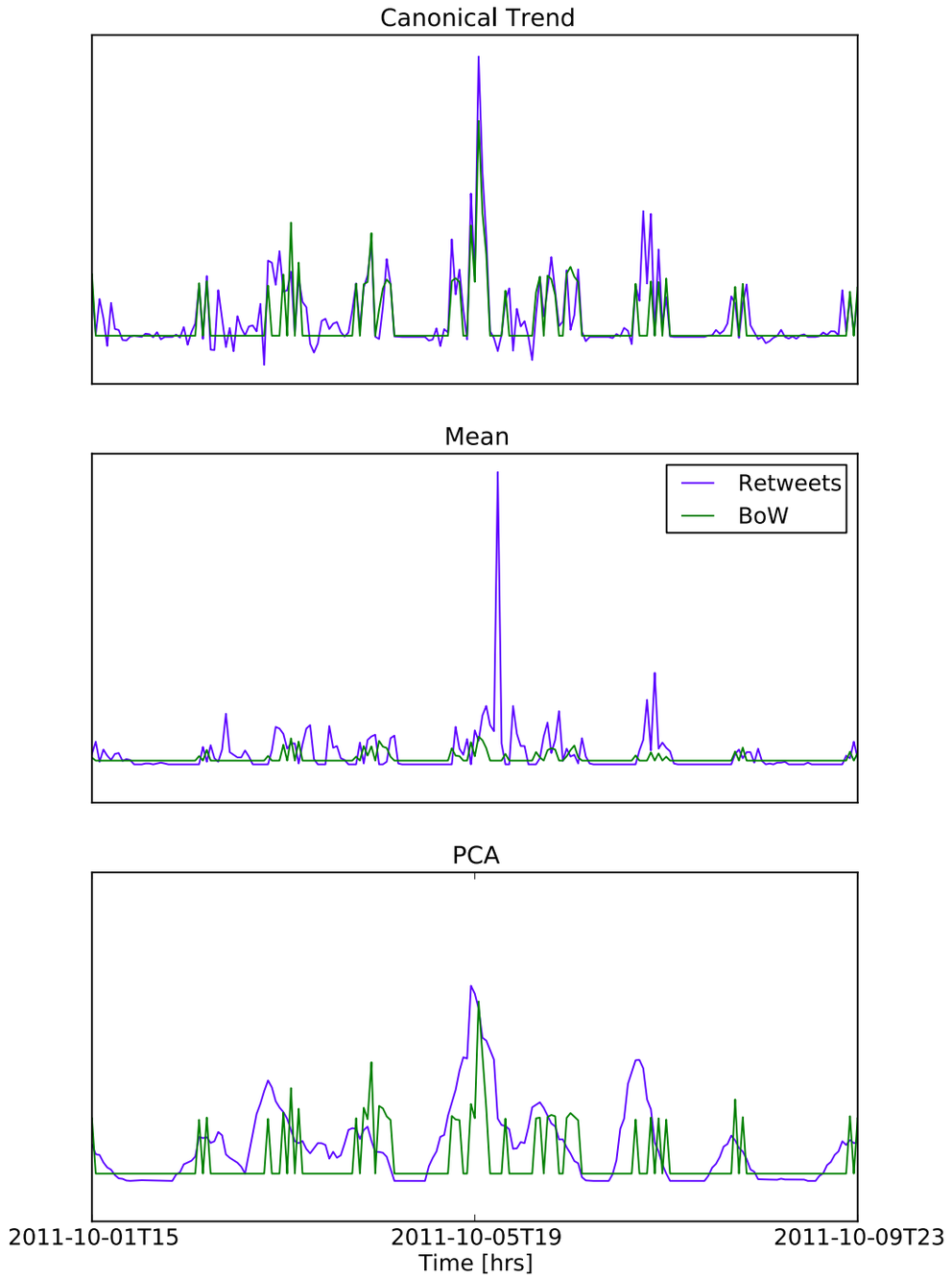
PCA

Comparisons: Mean, PCA and Canonical Trends

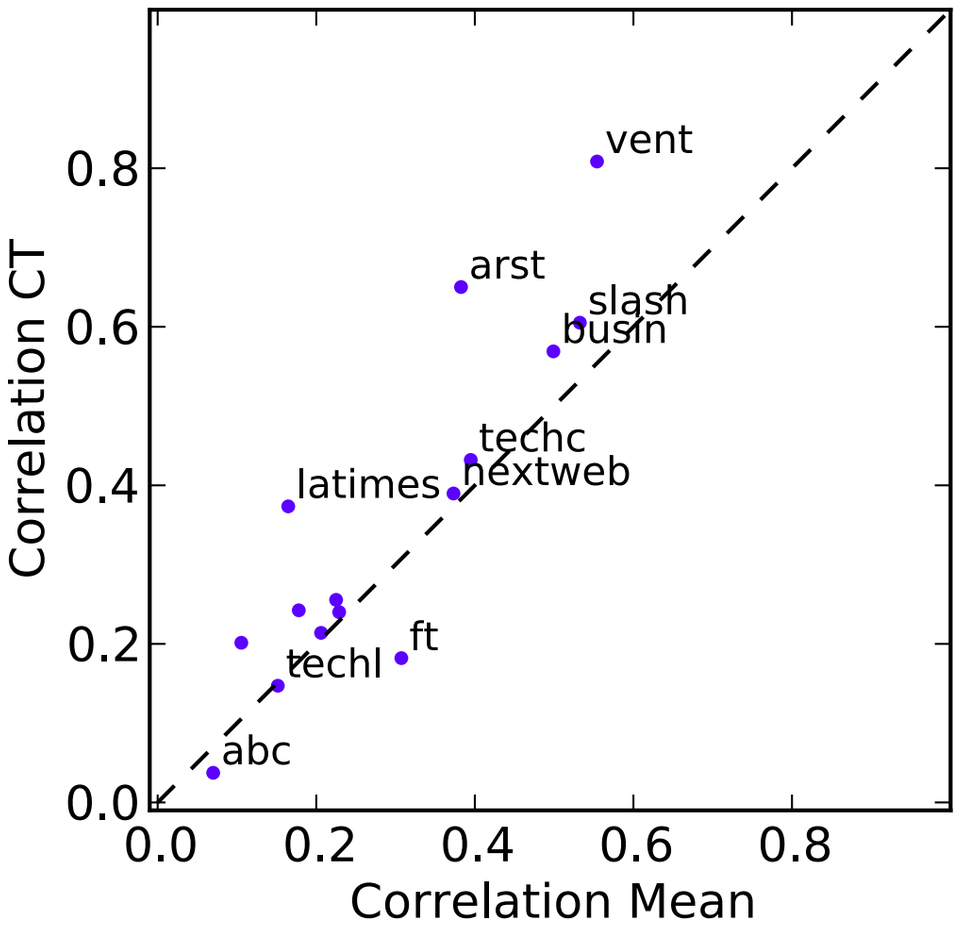
Canonical Trends

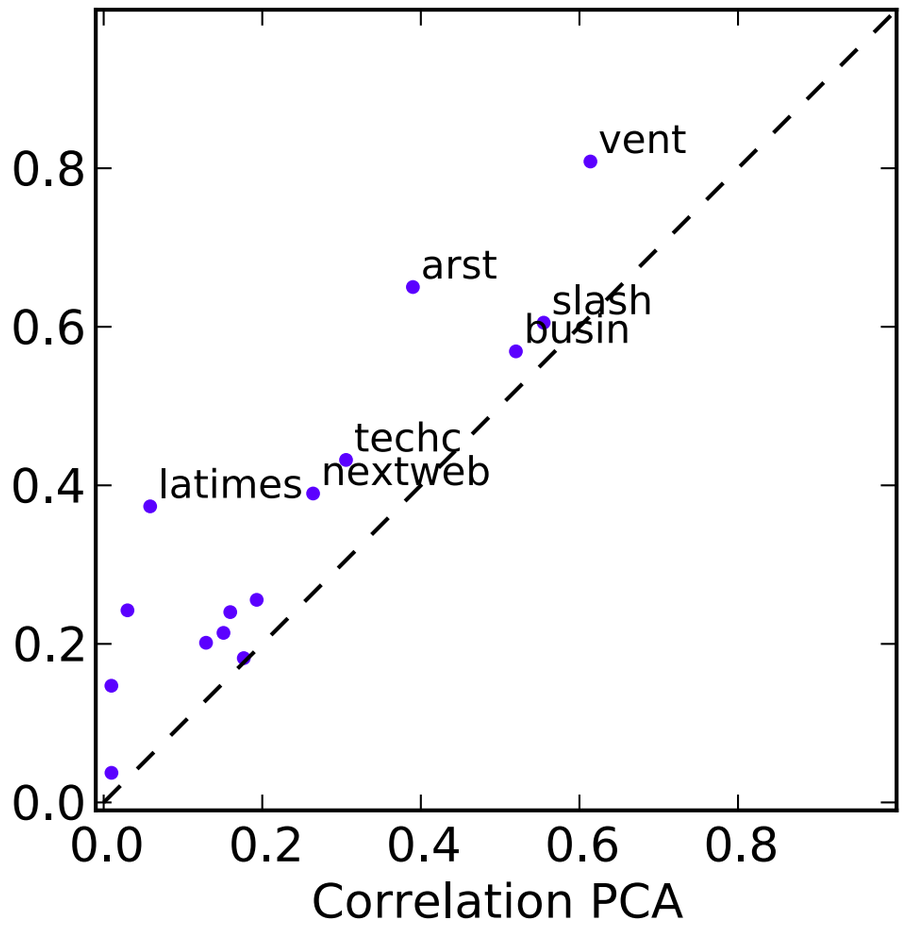
Mean

PCA



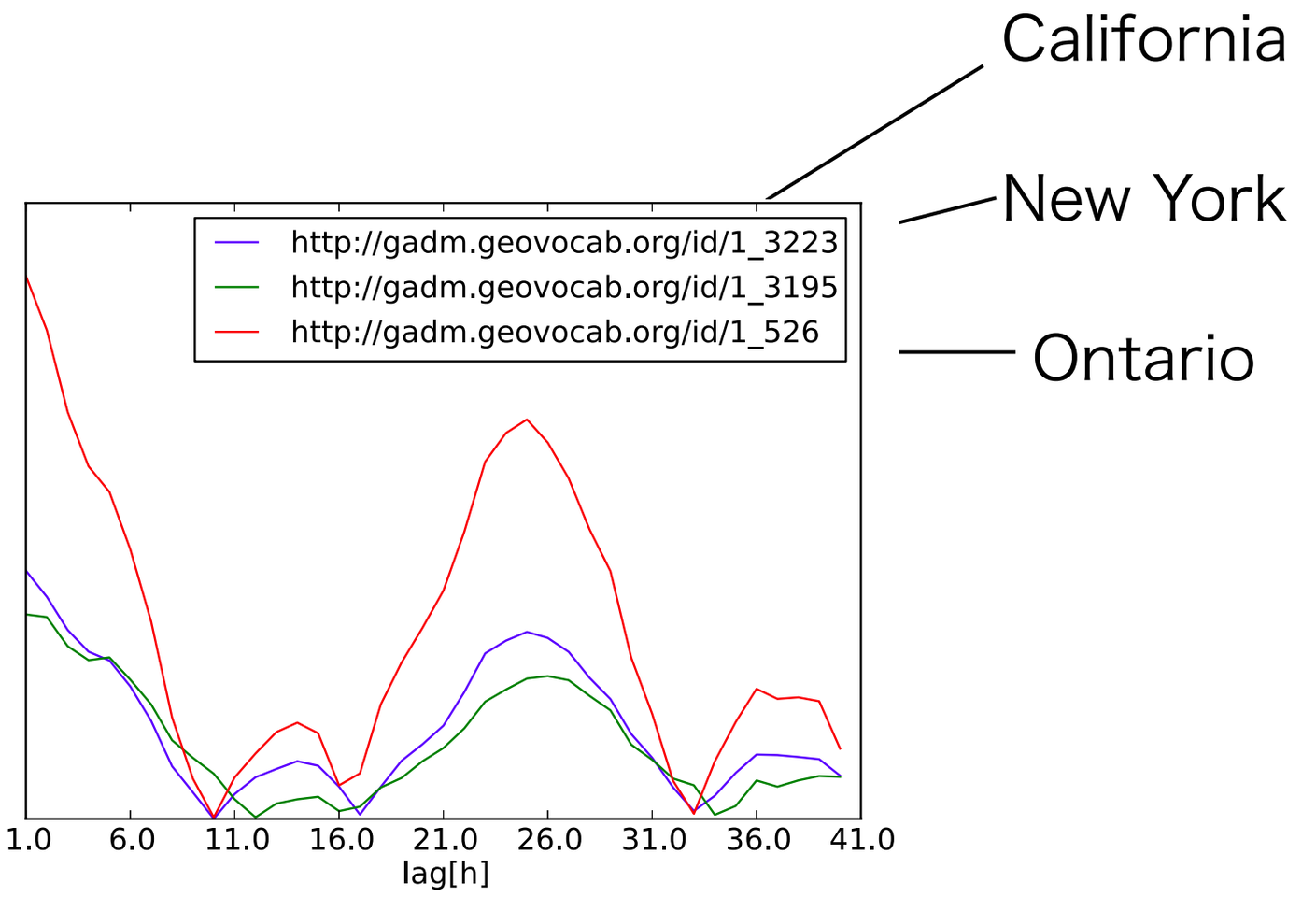
Comparisons: Mean, PCA and Canonical Trends





Canonical Trends

Canonical Convolution



Excerpts from LA Times
Spatiotemporal Response

Canonical Convolution

$$\tau = 1\text{hrs}$$

Canonical Convolution

$$\tau = 12\text{hrs}$$

Spatiotemporal Analysis of Retweets of News

We used canonical correlation analysis to compute

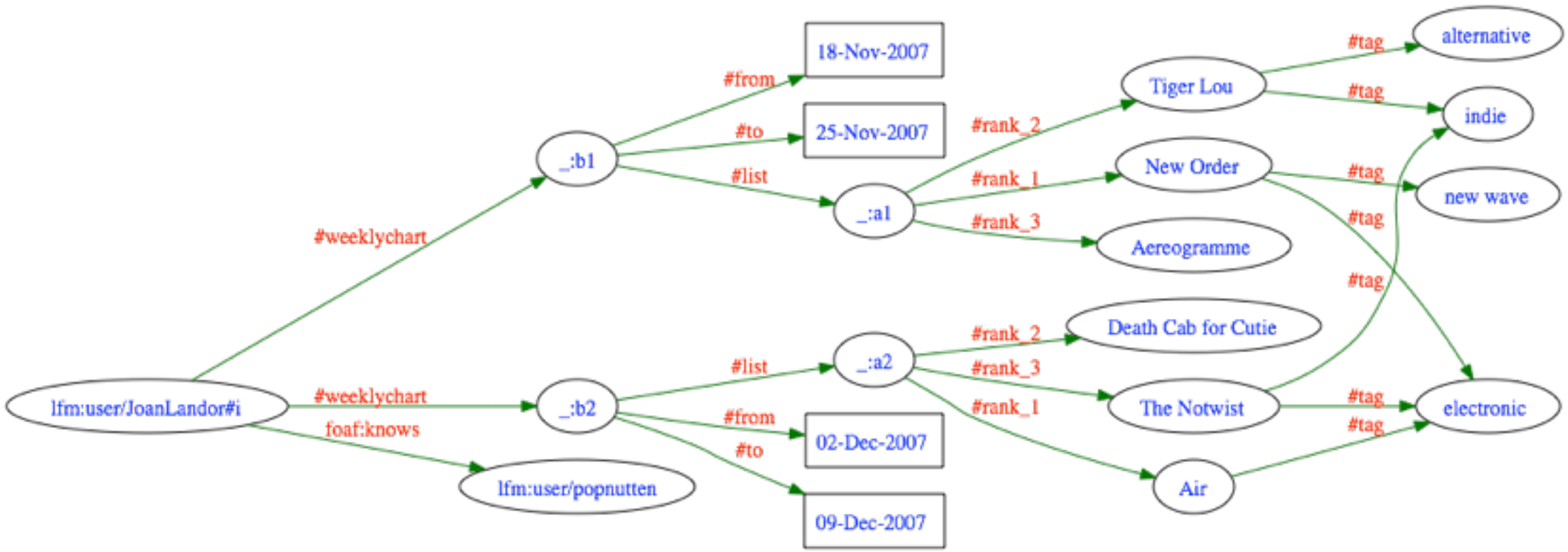
- Bag-of-Word subspace (topic)
- spatiotemporal twitter response patterns

such that news content and retweets are maximally correlated

Results can be interpreted w.r.t

- Impact of web source on social network
- Spatio-temporal evolution of impact

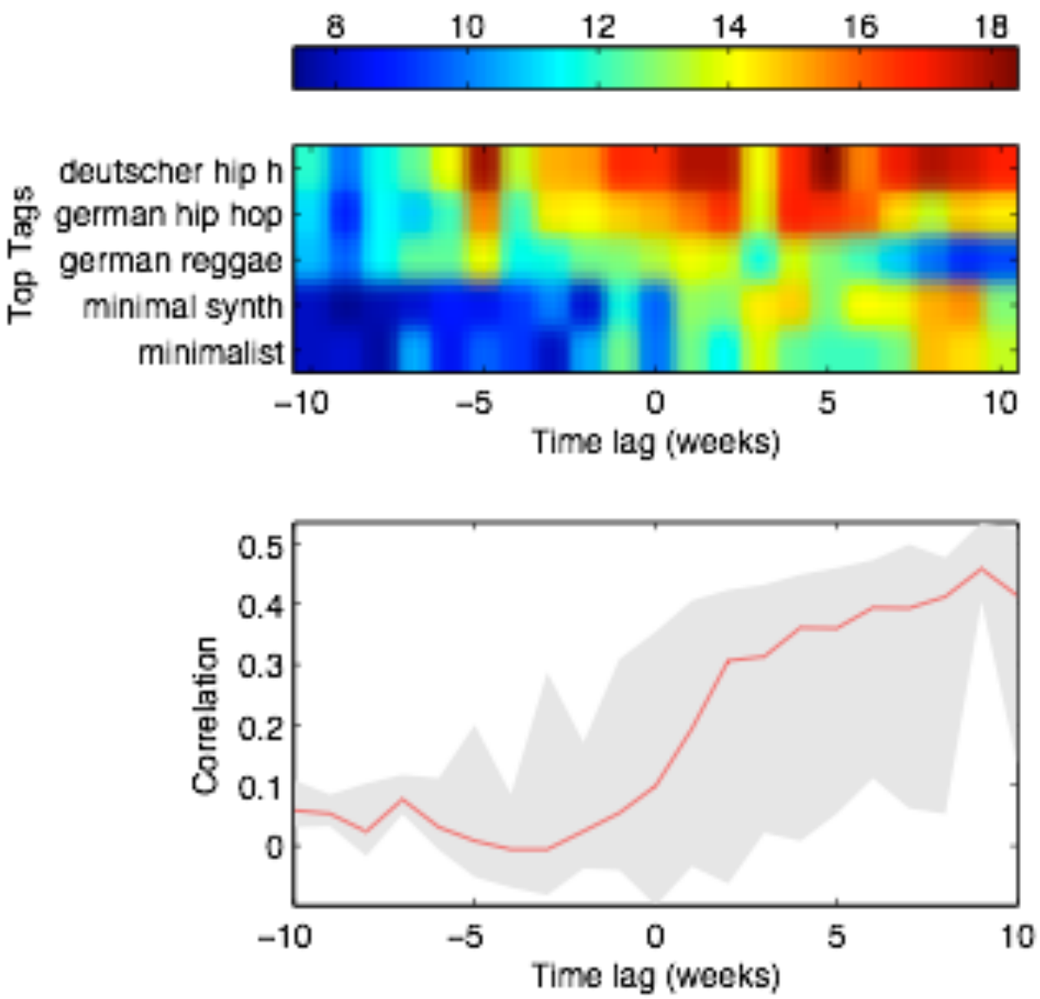
Music Trends on Last.fm



A last.fm user subgraph

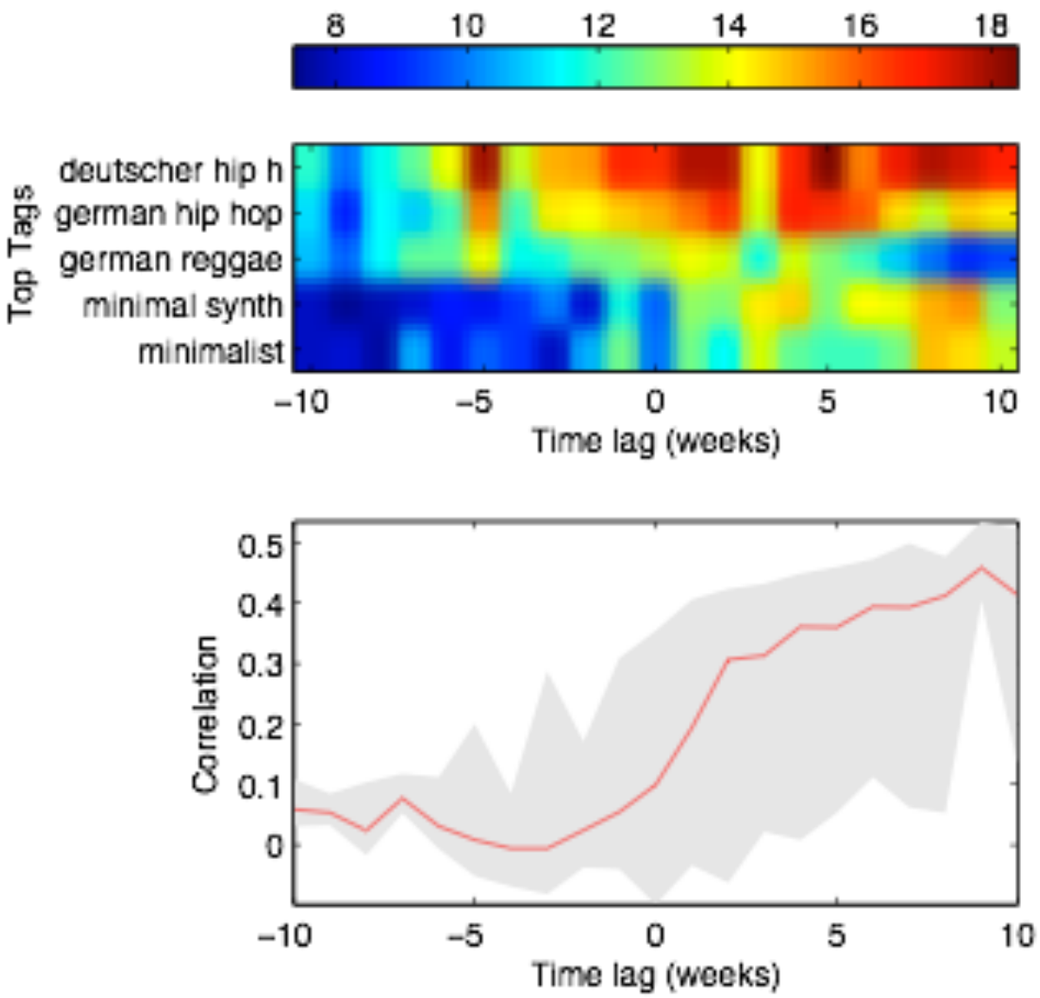
Users and Trends on Last.fm

Behind the Trend

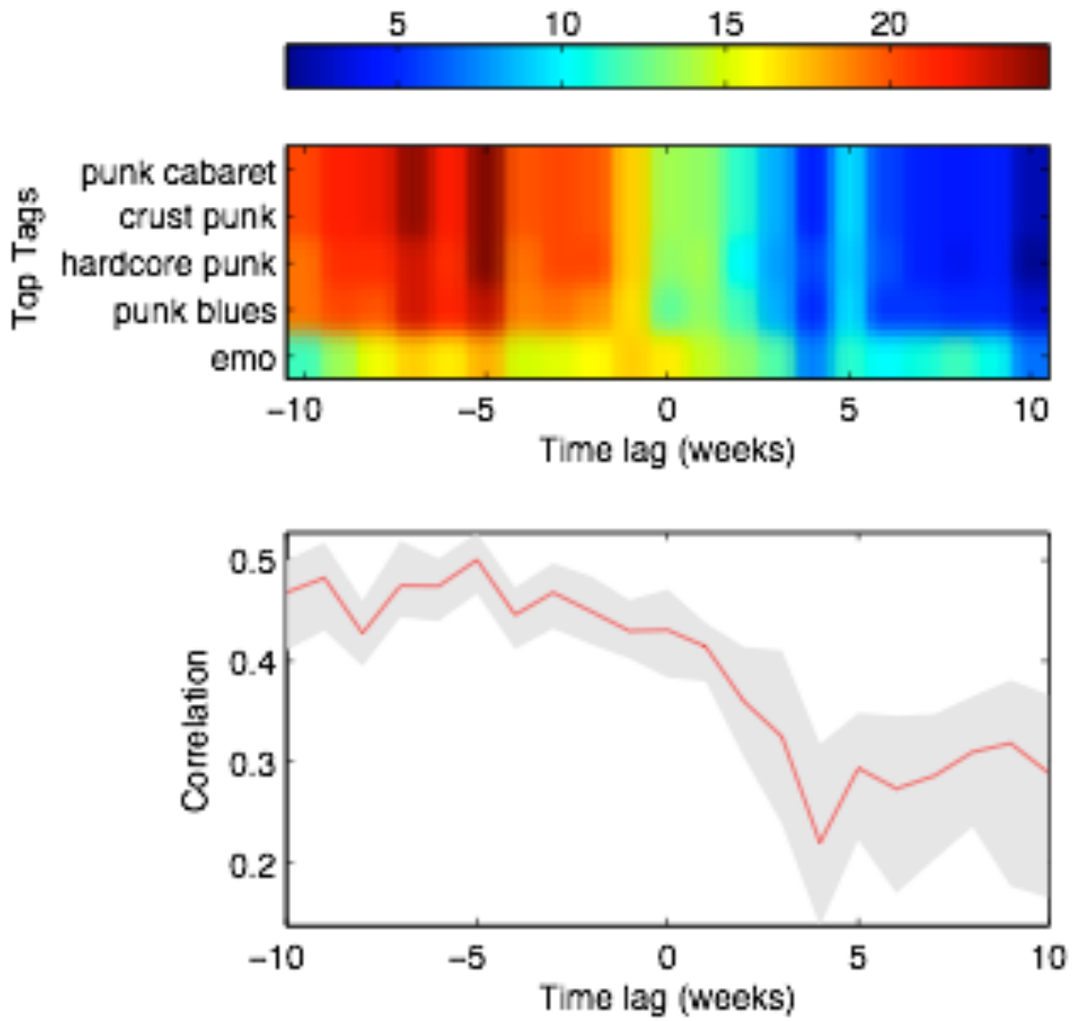


Users and Trends on Last.fm

Behind the Trend



Ahead of Trend



Online Canonical Trend Analysis

For large data sets, kernels become too big (NxN)

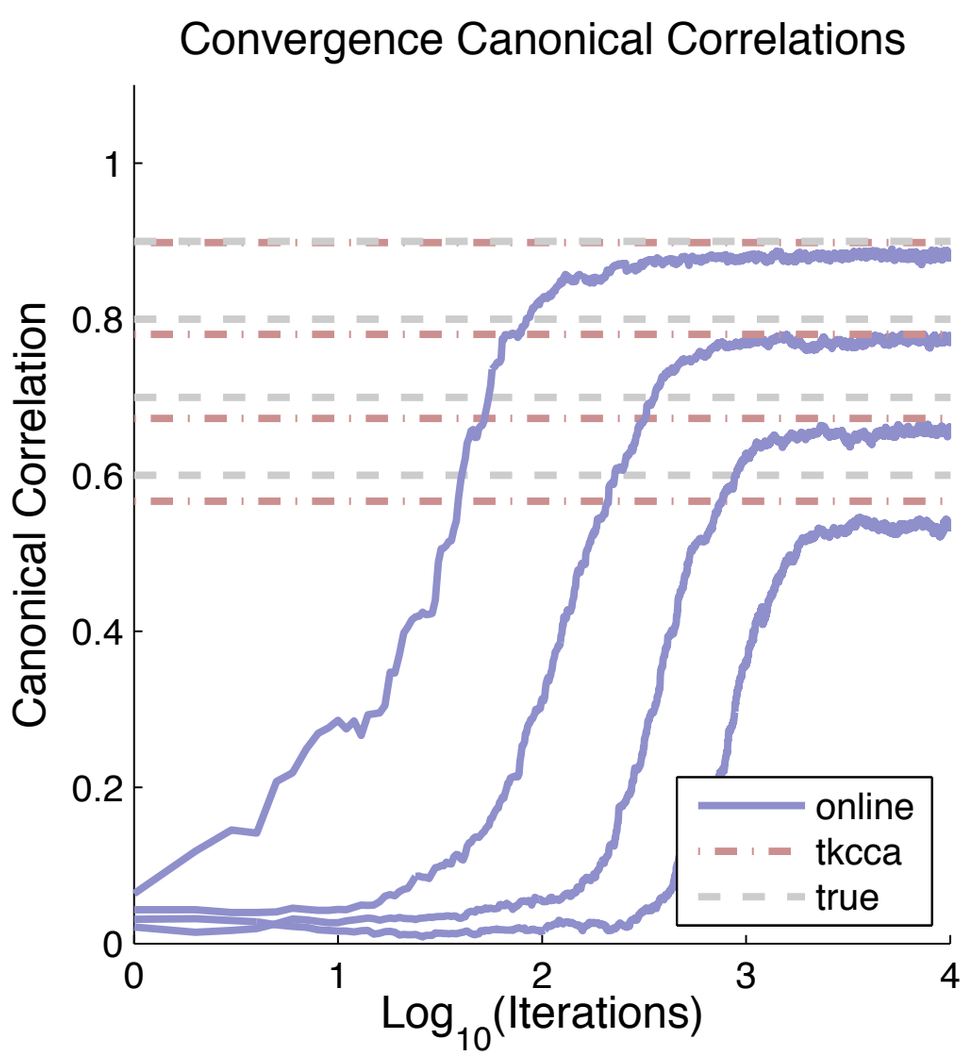
- Subsample
- Change temporal resolution
- Use Stochastic gradient descent (SGD)

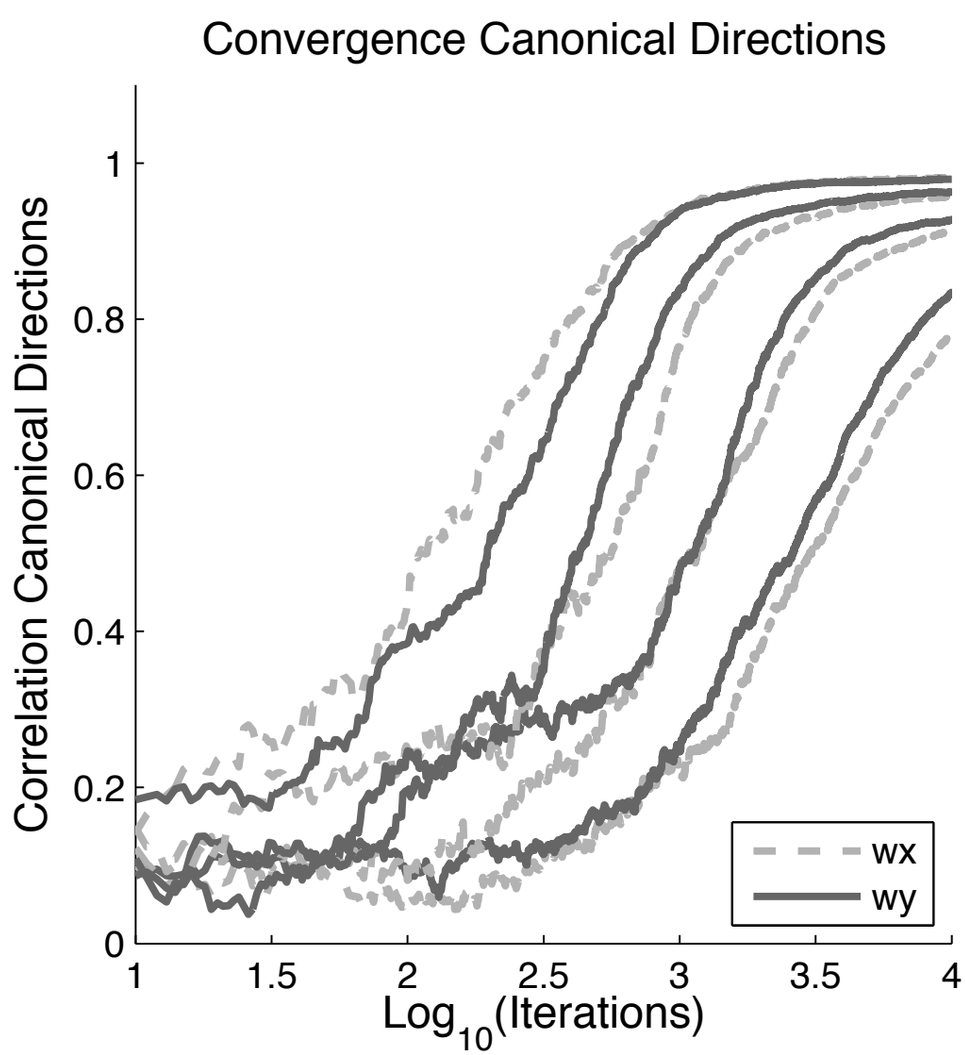
Biessmann et al, Online CCA for Realtime Impact Analysis of Social Media Data, NIPS Workshop, 2012

Algorithm 1: Online Canonical Trend Analysis

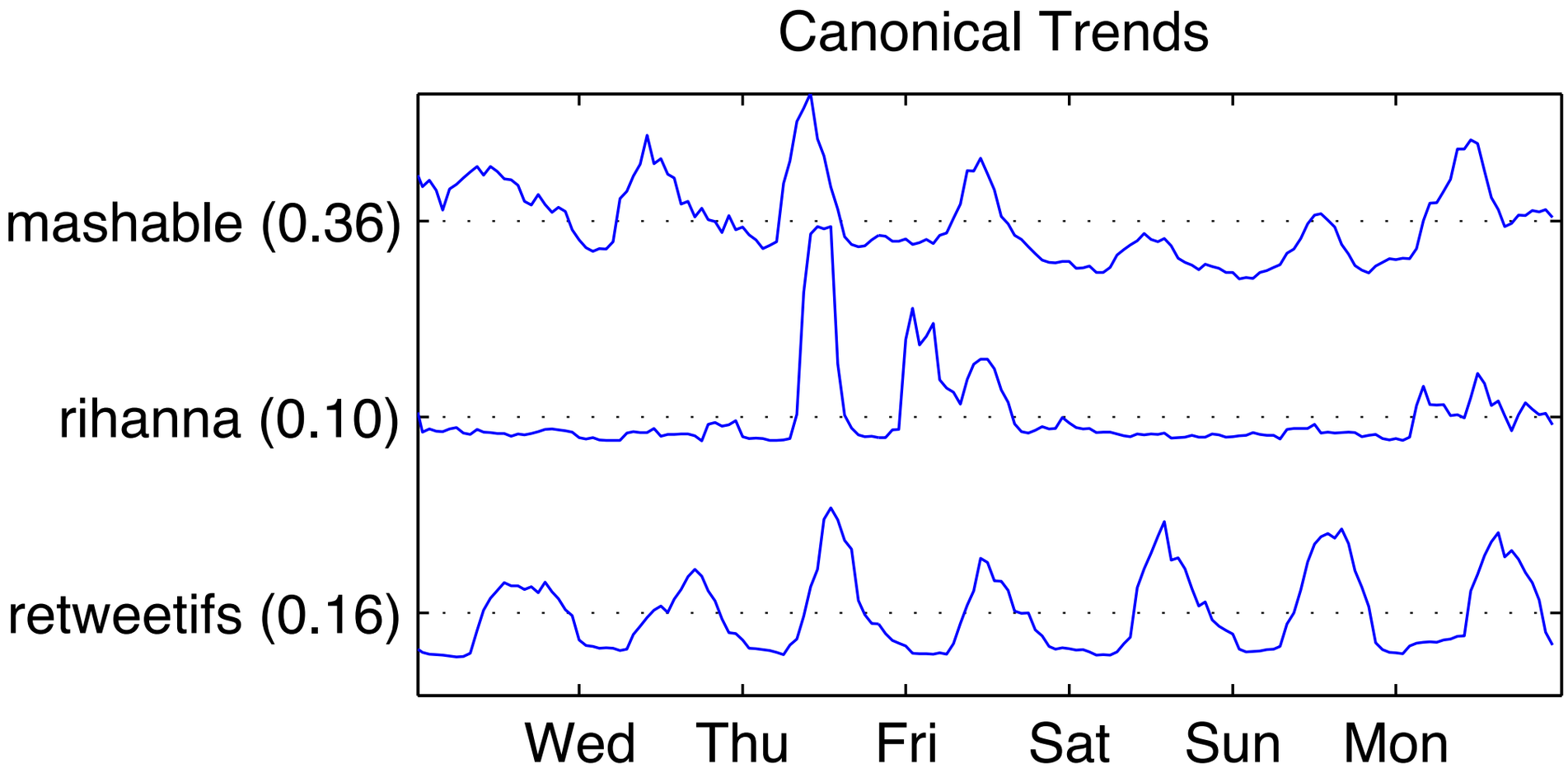
Input: Data streams of web sources $\mathcal{X}_f,\ f = 1, \ldots, F$, learning rate η_0
# For each new sample $x_f(t)$
for $t = 1$ **to** T **do**
  # Loop over all news feeds
  for $f = 1$ **to** F **do**
    $y_f(t) = 1/(1-F) \sum_{f' \neq f} x_{f'}(t)$
    # Temporal Embedding (eq. 2)
    $\tilde{x}_f(t) = [x_f(t - N_\tau)^\top, \ldots, x_f(t-1)^\top]^\top$
    # Update $w_{\tilde{X}},\ w_y$ (eq. 7)
    $w_{\tilde{X}} \leftarrow w_{\tilde{X}} + \eta_0/t\ \ \tilde{x}_f(t) y(t)_f^\top w_y$
    $w_y \leftarrow w_y + \eta_0/t\ \ w_{\tilde{X}}^\top \tilde{x}_f(t) y(t)_f^\top$
    # Project canonical directions on unit ball
    $w_{\tilde{X}} \leftarrow w_{\tilde{X}} \|w_{\tilde{X}}\|_2^{-1}$
    $w_y \leftarrow w_y \|w_y\|_2^{-1}$
  end for
  Rank Feeds according to Corr$(w_{\tilde{X}}^\top x_f, w_y^\top y_f)$
end for

Online Canonical Trend Analysis





Online Canonical Trend Analysis



Summary

Summary

Canonical Trend Analysis

Summary

Canonical Trend Analysis

Simple trend detection algorithm

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do
- Allows to visualize temporal dynamics

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do
- Allows to visualize temporal dynamics
- Non-linear extensions trivial (Kernel Trick)

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do
- Allows to visualize temporal dynamics
- Non-linear extensions trivial (Kernel Trick)
- Easily parallelizable

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do
- Allows to visualize temporal dynamics
- Non-linear extensions trivial (Kernel Trick)
- Easily parallelizable
- Efficient to solve in the dual*

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do
- Allows to visualize temporal dynamics
- Non-linear extensions trivial (Kernel Trick)
- Easily parallelizable
- Efficient to solve in the dual*
- SGD version for scalable inference

Summary

Canonical Trend Analysis

- Simple trend detection algorithm
- Detects web sources that publish trends before others do
- Allows to visualize temporal dynamics
- Non-linear extensions trivial (Kernel Trick)
- Easily parallelizable
- Efficient to solve in the dual*
- SGD version for scalable inference

*(for reasonably chosen temporal resolutions)

Summary

Summary

Thank you