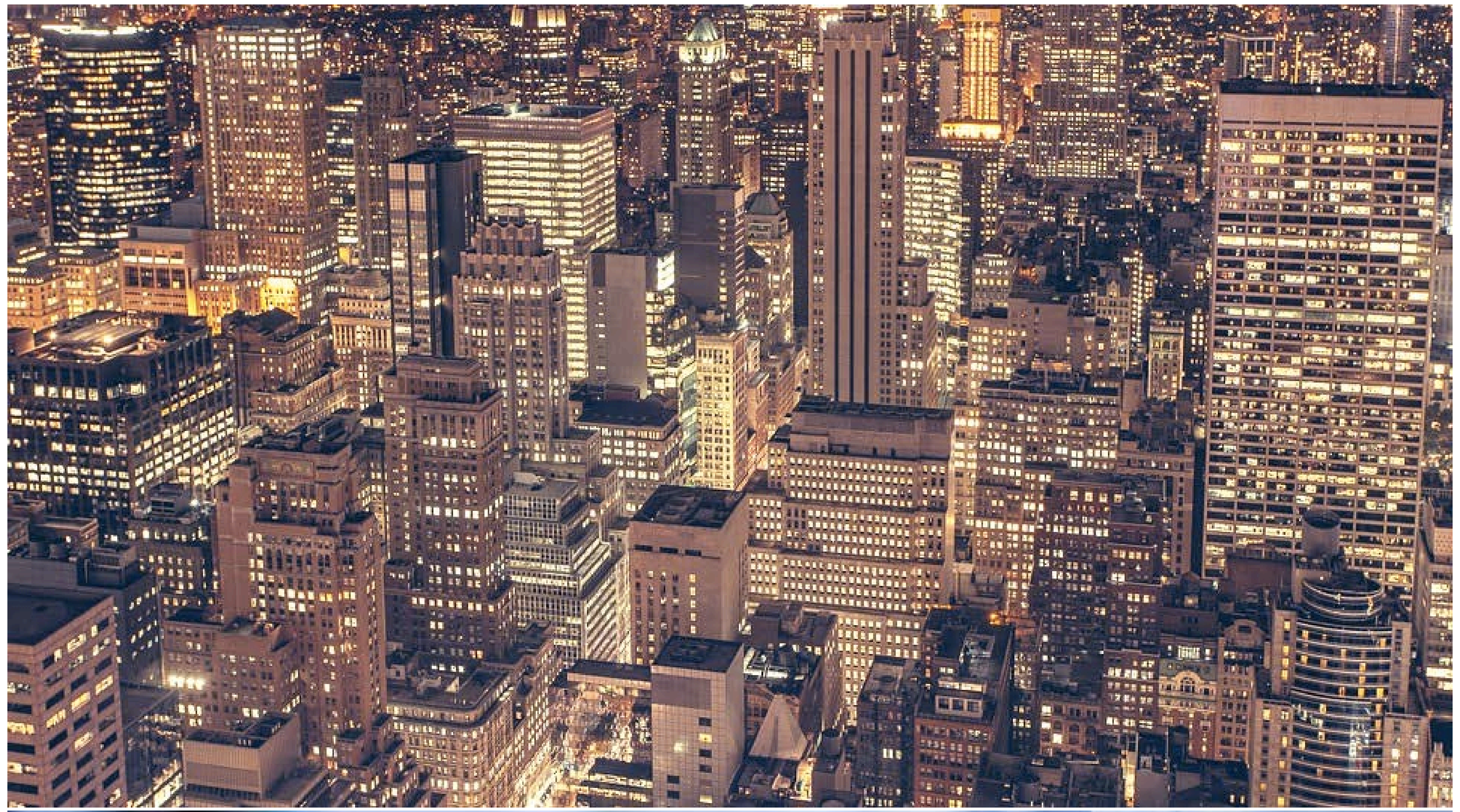
Analyzing and Visualizing Word Semantics over Centuries

Stephan Mandt
Computer Science, UCI
www.stephanmandt.com

In many situations, we want to make sense of complex data sets.

Motivation Probabilistic Modeling

- Allows to build models for customized data analysis
- Allows users to include domain knowledge into a ML framework
- Assumes that the data are generated by a generative process that the modeler specifies

“Prerequisites”

Some probability

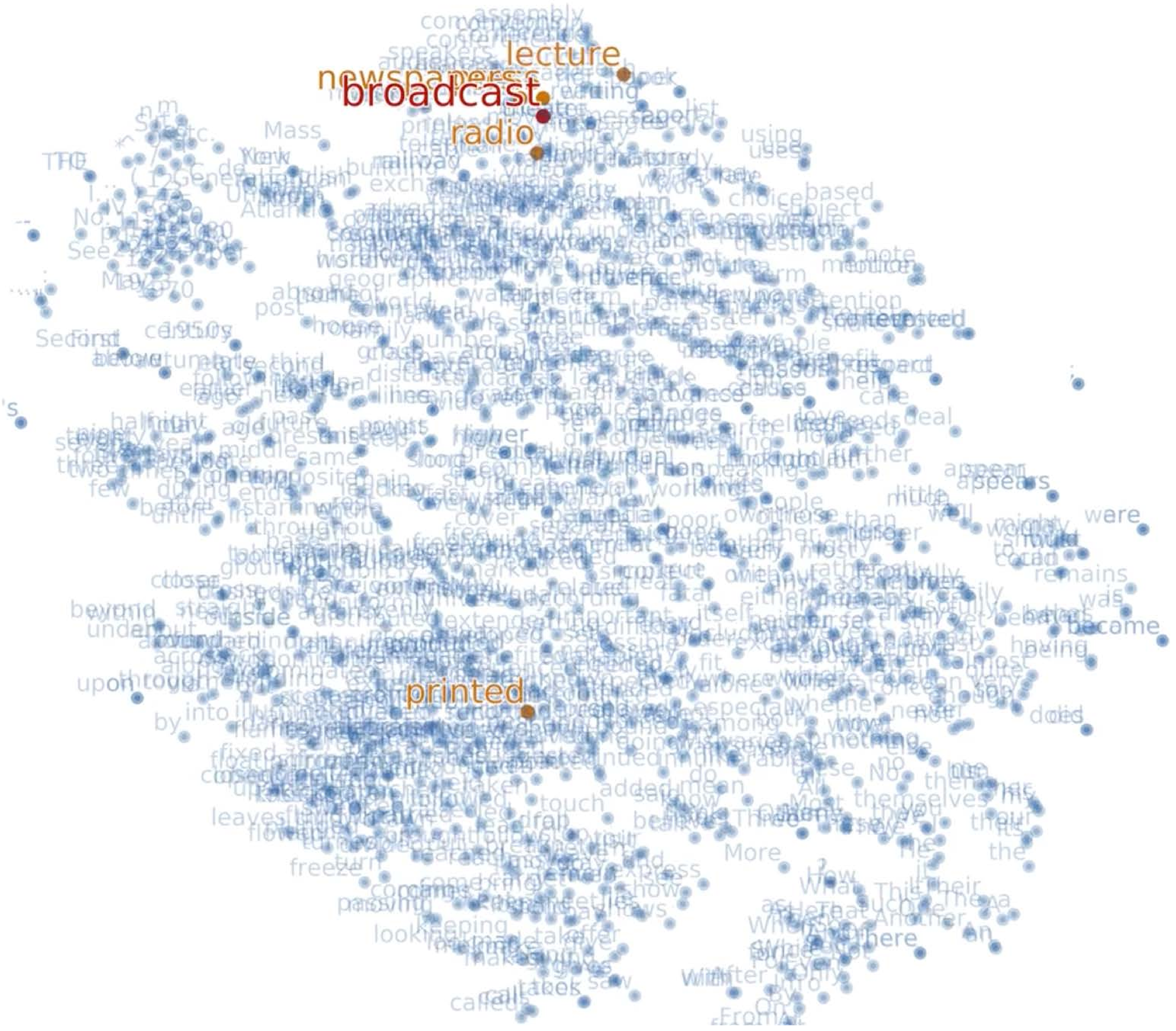
- joint distribution, conditional distribution: $p(\text{snow}, \text{cold}) = p(\text{snow}|\text{cold})p(\text{cold})$
- expectation

Some optimization theory

- gradient-based optimization

Example Dynamic Word Embeddings

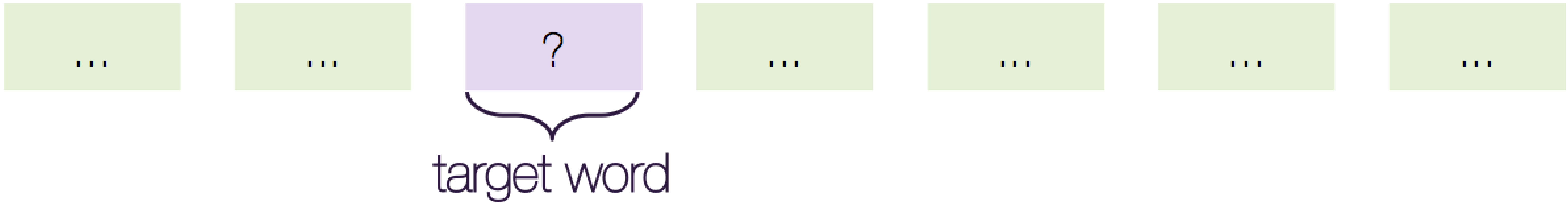
Word Embeddings Introduction



- Goal: for each word, learn a vector that captures its semantic meaning
- Input: massive amounts of unstructured text
- Output: Vector representations of words

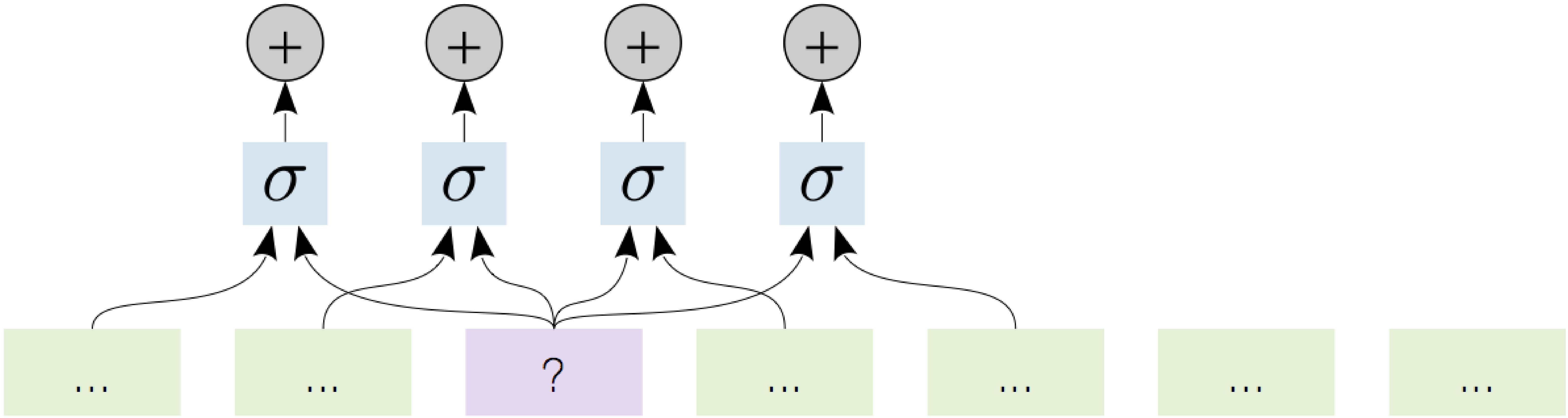
Background Word Embeddings

[Mikolov et al., ICLR 2013 & NIPS 2013]



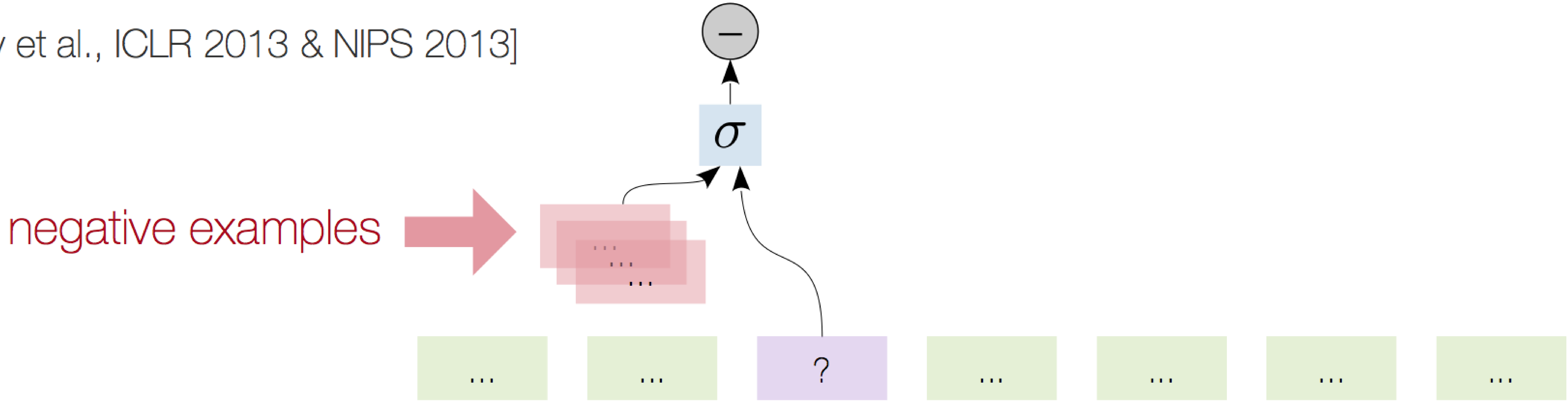
Background Word Embeddings

[Mikolov et al., ICLR 2013 & NIPS 2013]



Background Word Embeddings

[Mikolov et al., ICLR 2013 & NIPS 2013]

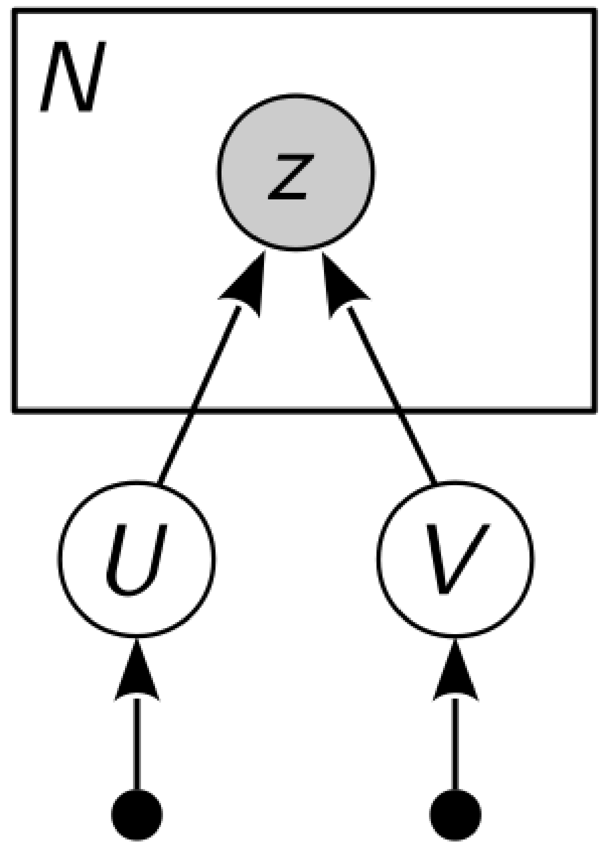


Not a generative model for text!

Word Embeddings as Probabilistic Models (Barkan 2017)

First, two important definitions:

$$\mathbf{n}^{+}_{ij} = \#\{\text{counts of words } i \text{ and } j \text{ co-occurring withing range } L\}$$
$$\mathbf{n}^{-}_{ij} = \text{same quantity for randomly shuffled corpus}$$



Idea:

- assign every word in the vocabulary *V* to two vectors (*u* and *v*)
- train these vectors to predict whether a given pairing of words is more likely to occur in the true or shuffled corpus

$$\ell = -\sum_{i,j=1}^{V} \mathbf{n}^{+}_{\mathbf{ij}} \log \sigma(u_i^\top v_j) - \sum_{i,j=1}^{V} \mathbf{n}^{-}_{\mathbf{ij}} \log \sigma(-u_i^\top v_j)$$

== loss function of word2vec

positive samples **negative samples**

Application Word Embeddings Over Time


The Scranton Times.
GREAT TITANIC SINKS WITH 1500 ON BOARD
CARPATHIA STEAMING TO NEW YORK WITH HANDFUL OF SURVIVORS OF THE GREATEST TRAGEDY IN THE HISTORY OF NAVIGATION
NEW YORK IS STUNNED BY SEA HORROR
ALL HOPE NOW ABANDONED; NOTABLES WHO ARE VICTIMS
Titanic Now Rests On the Bottom of the Sea


Visit WSJ.com to See Our New Look and Features
THE WALL STREET JOURNAL.
TUESDAY, SEPTEMBER 16, 2008
AIG, Lehman Shock Hits World Markets
Focus Moves to Fate of Giant Insurer After U.S. Allows Investment Bank to Fail; Barclays in Talks to Buy Core Lehman Unit
AIG Faces Cash Crisis As Stock Dives 61%

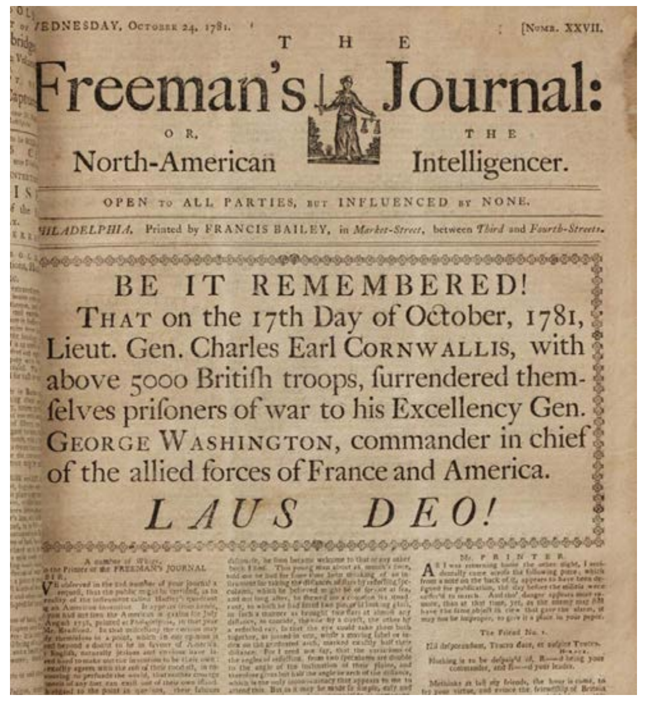
1781

1912

1963

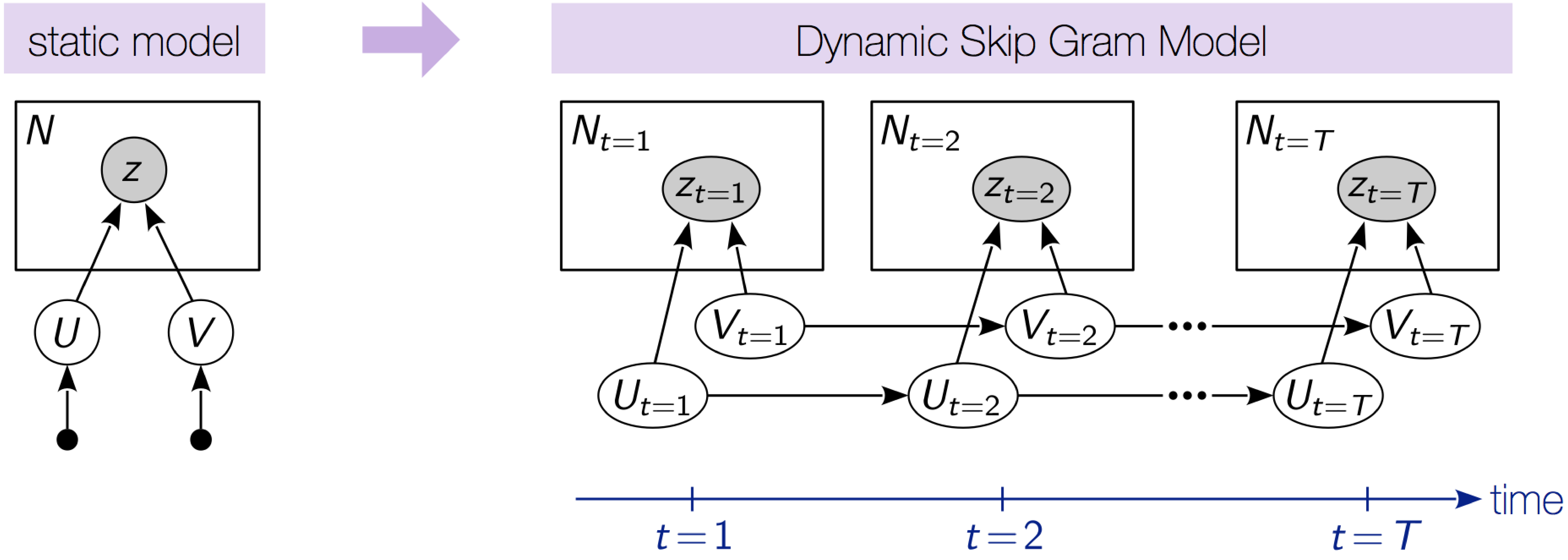
2008

time stamp


THE Freeman's Journal: OR, North-American Intelligencer.
OPEN to ALL PARTIES, but INFLUENCED by NONE.
BE IT REMEMBERED!
THAT on the 17th Day of October, 1781, Lieut. Gen. Charles Earl CORNWALLIS, with above 5000 British troops, surrendered themselves prisoners of war to his Excellency Gen. GEORGE WASHINGTON, commander in chief of the allied forces of France and America.
LAUS DEO!


The Evening Bulletin
Sniper Kills Kennedy
Presidency Taken Over By Johnson

Dynamic Word Embeddings Skip-Gram as a Probabilistic Time Series Model



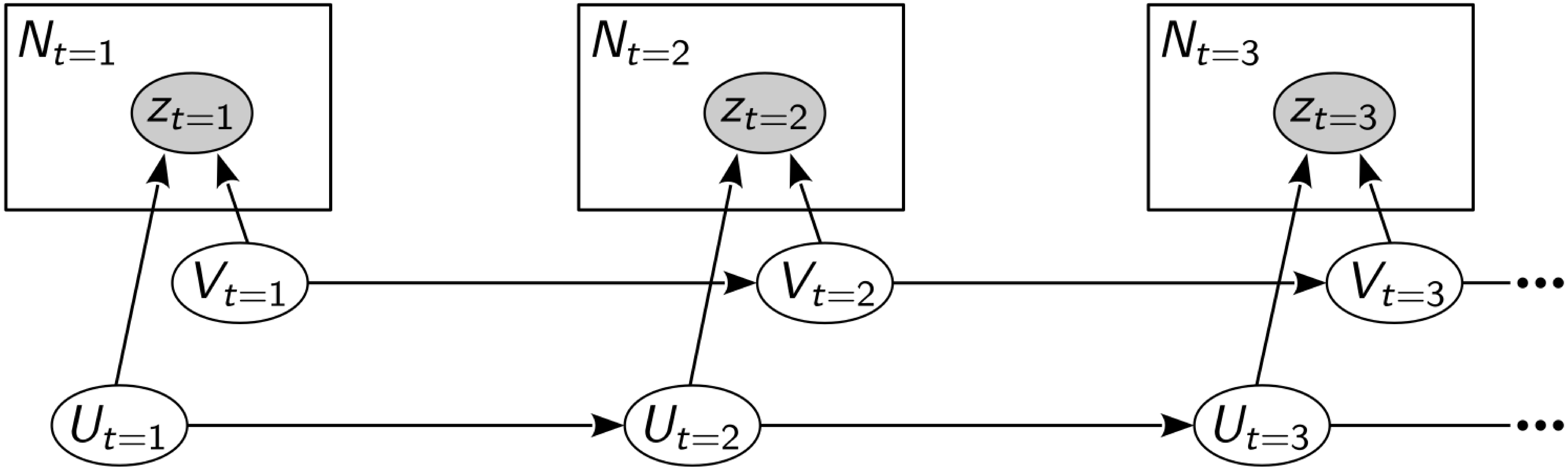
Imposed prior dynamics: probabilistic Kalman filter.

$$U_{t+1}|U_t \sim \mathcal{N}(U_t, \sigma_t^2 I)$$
$$V_{t+1}|V_t \sim \mathcal{N}(V_t, \sigma_t^2 I)$$

$$\sigma_t^2 = D(\tau_{t+1} - \tau_t)$$

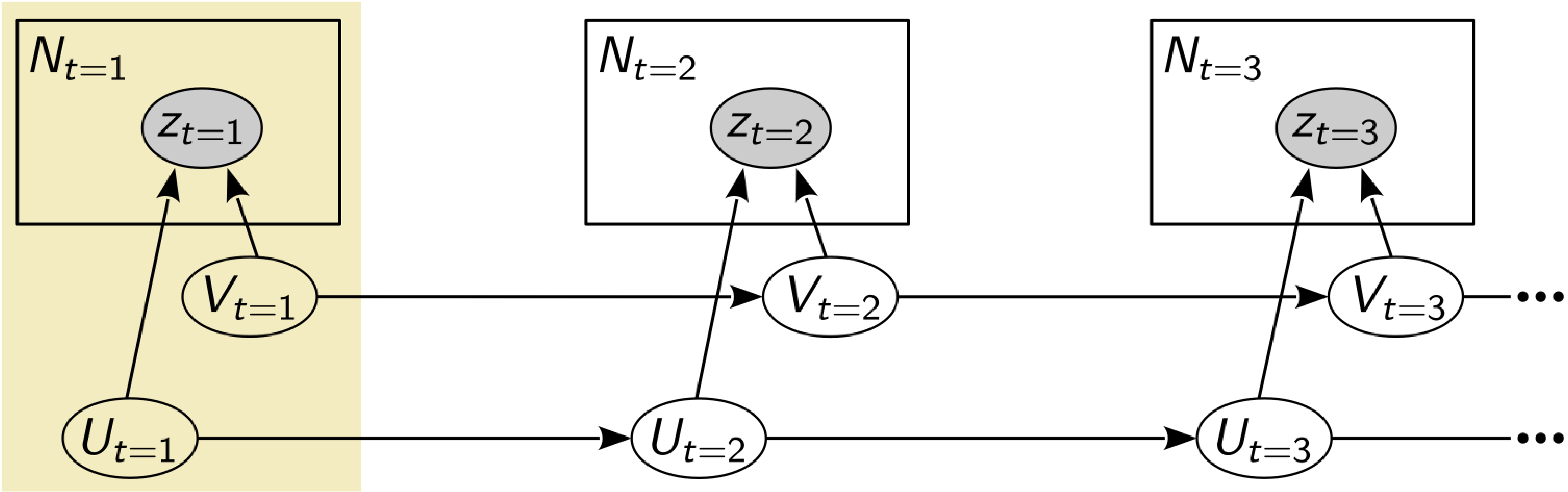
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



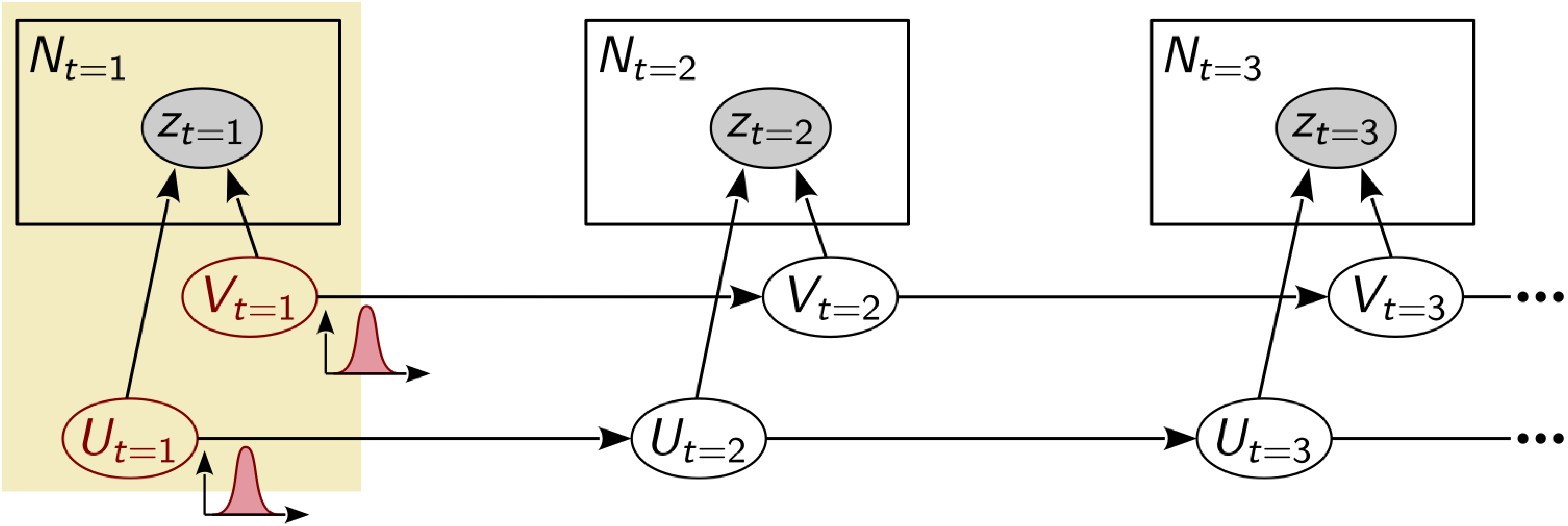
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



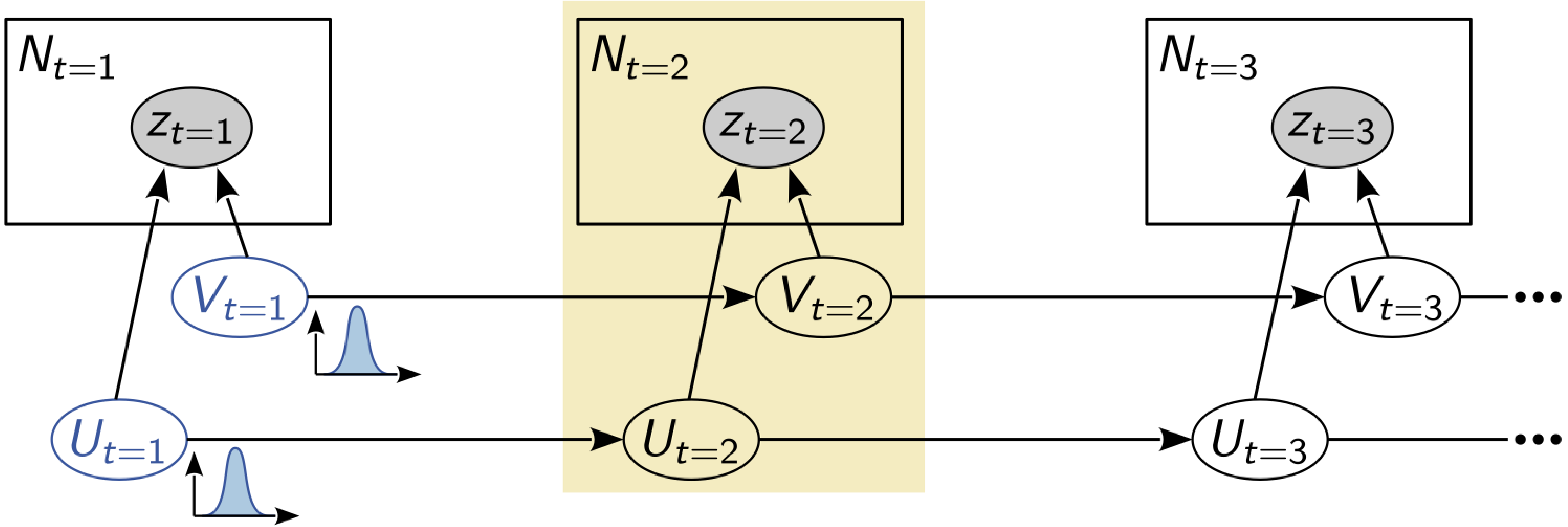
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



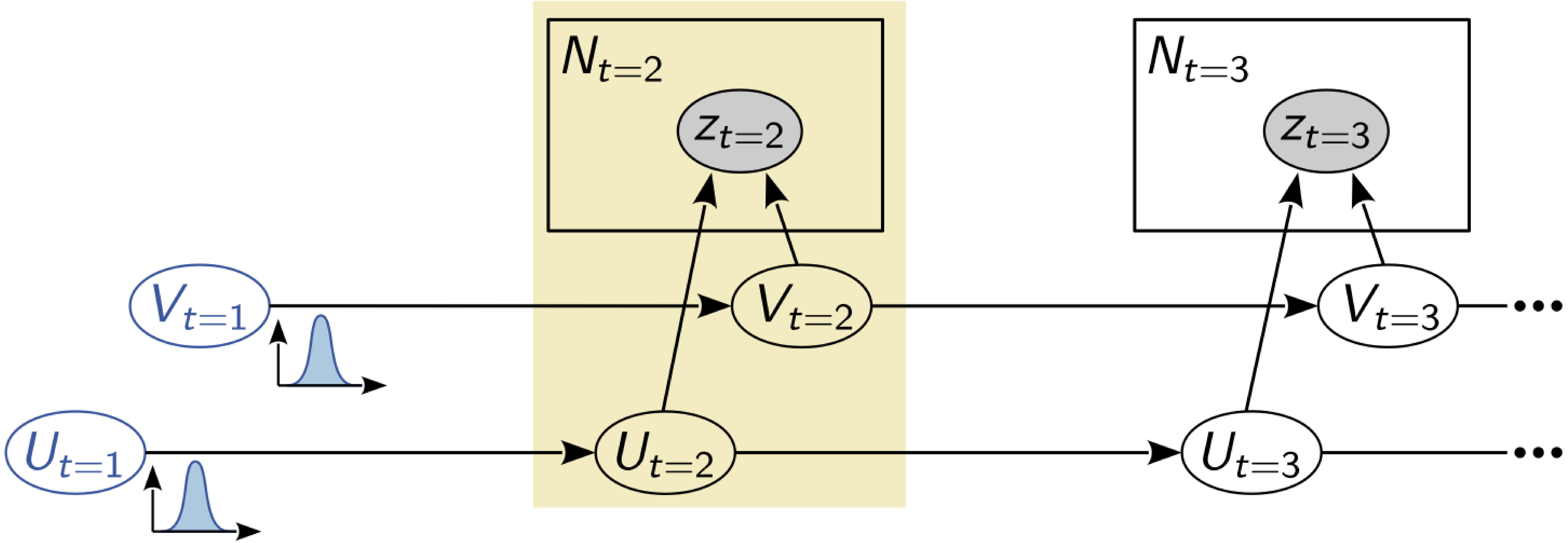
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



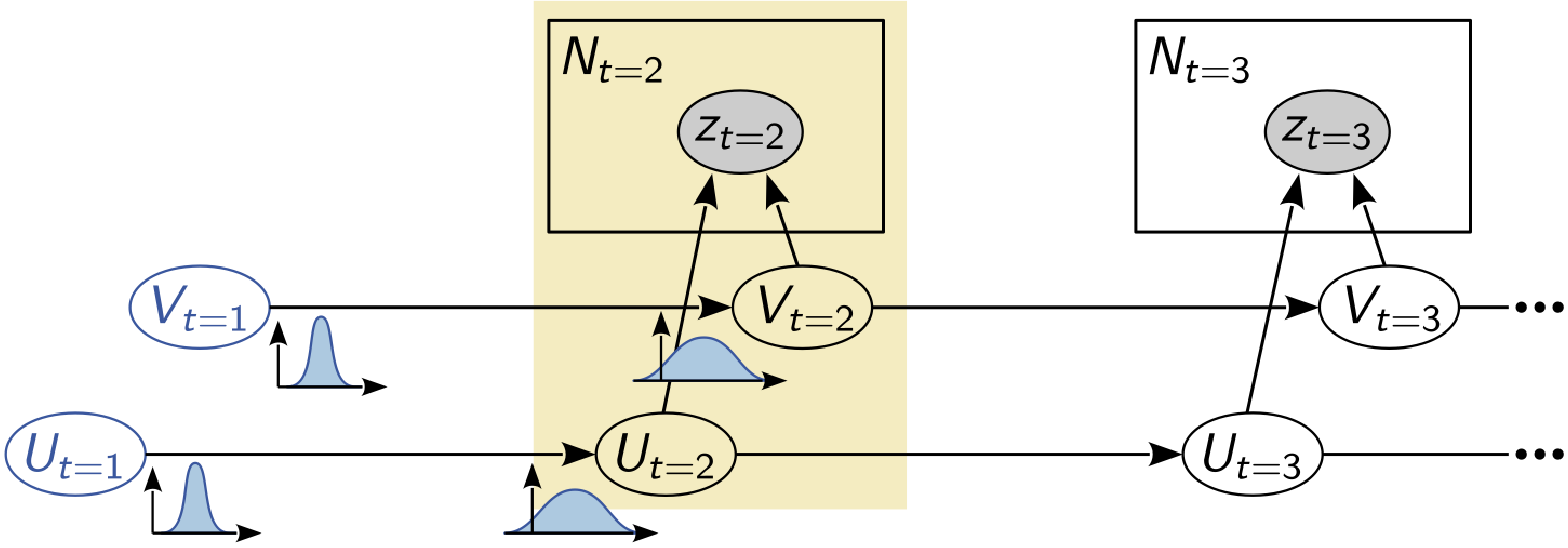
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



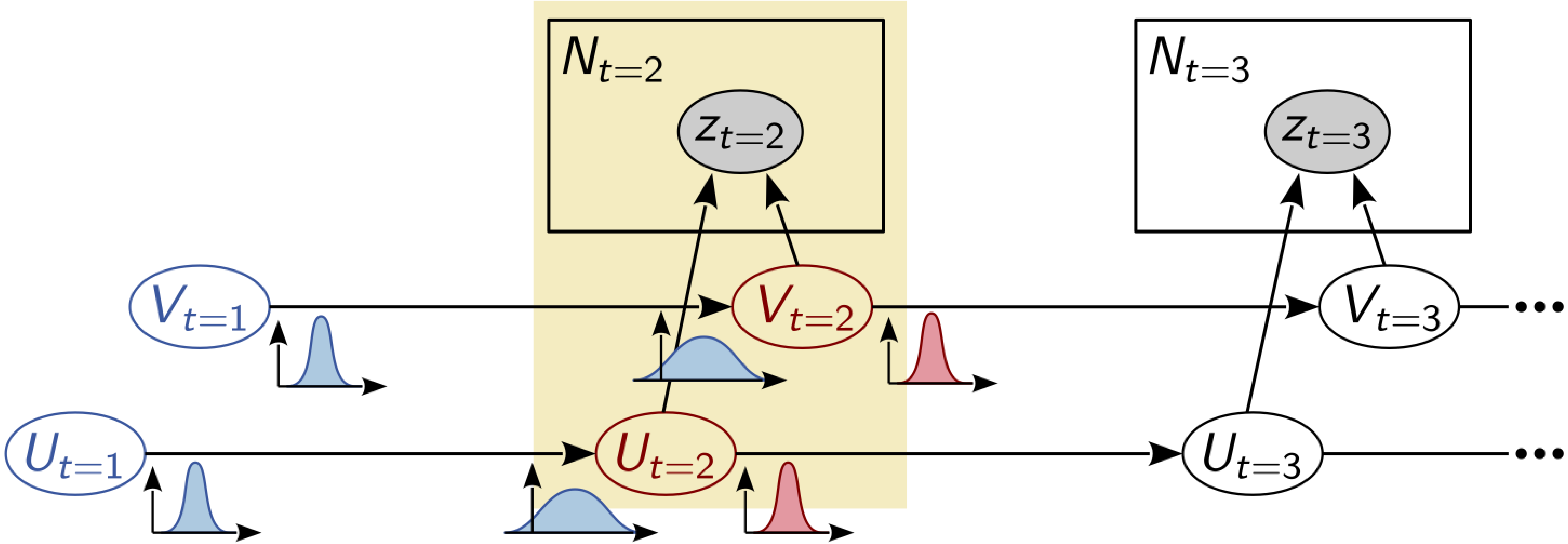
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



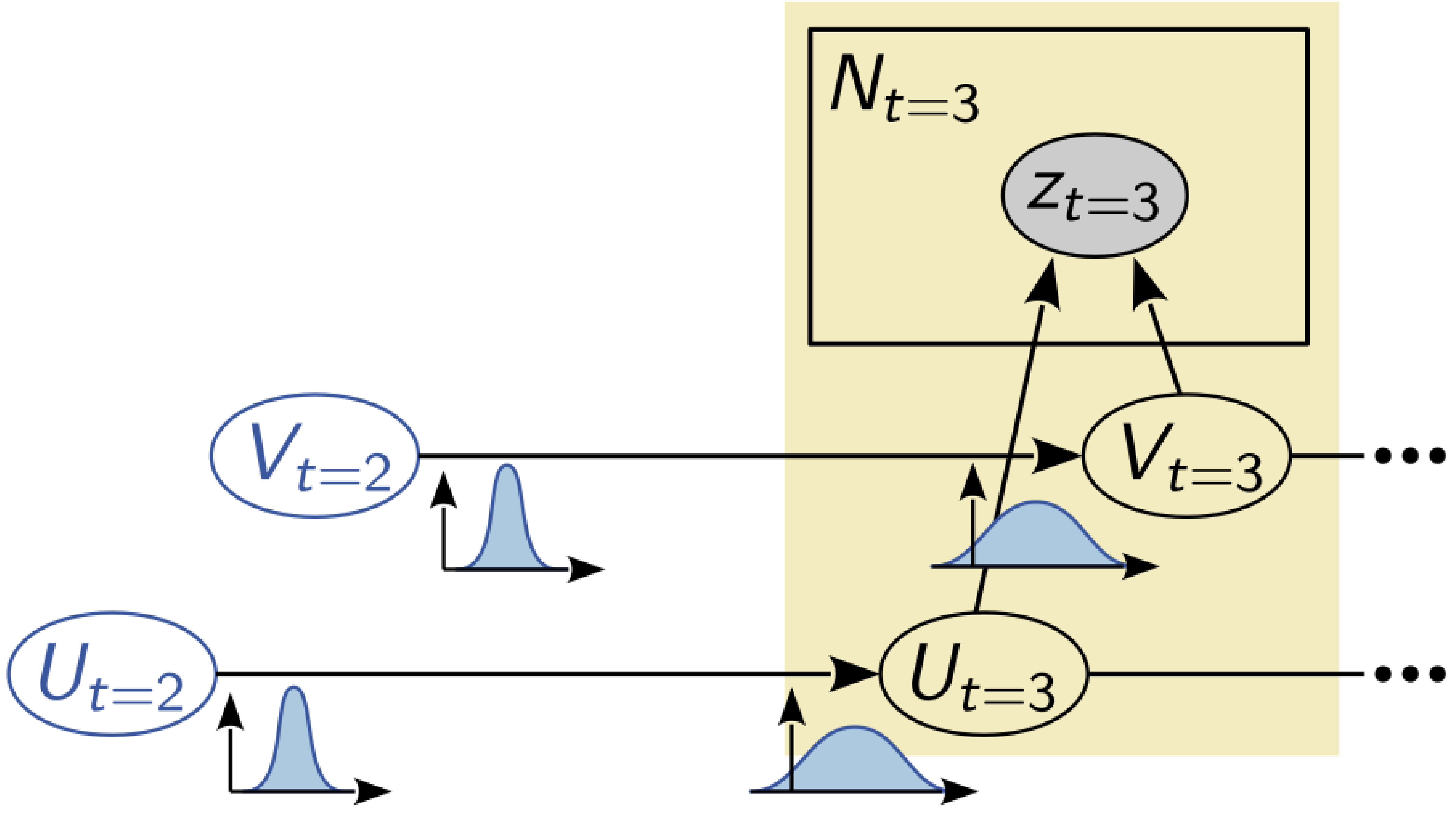
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



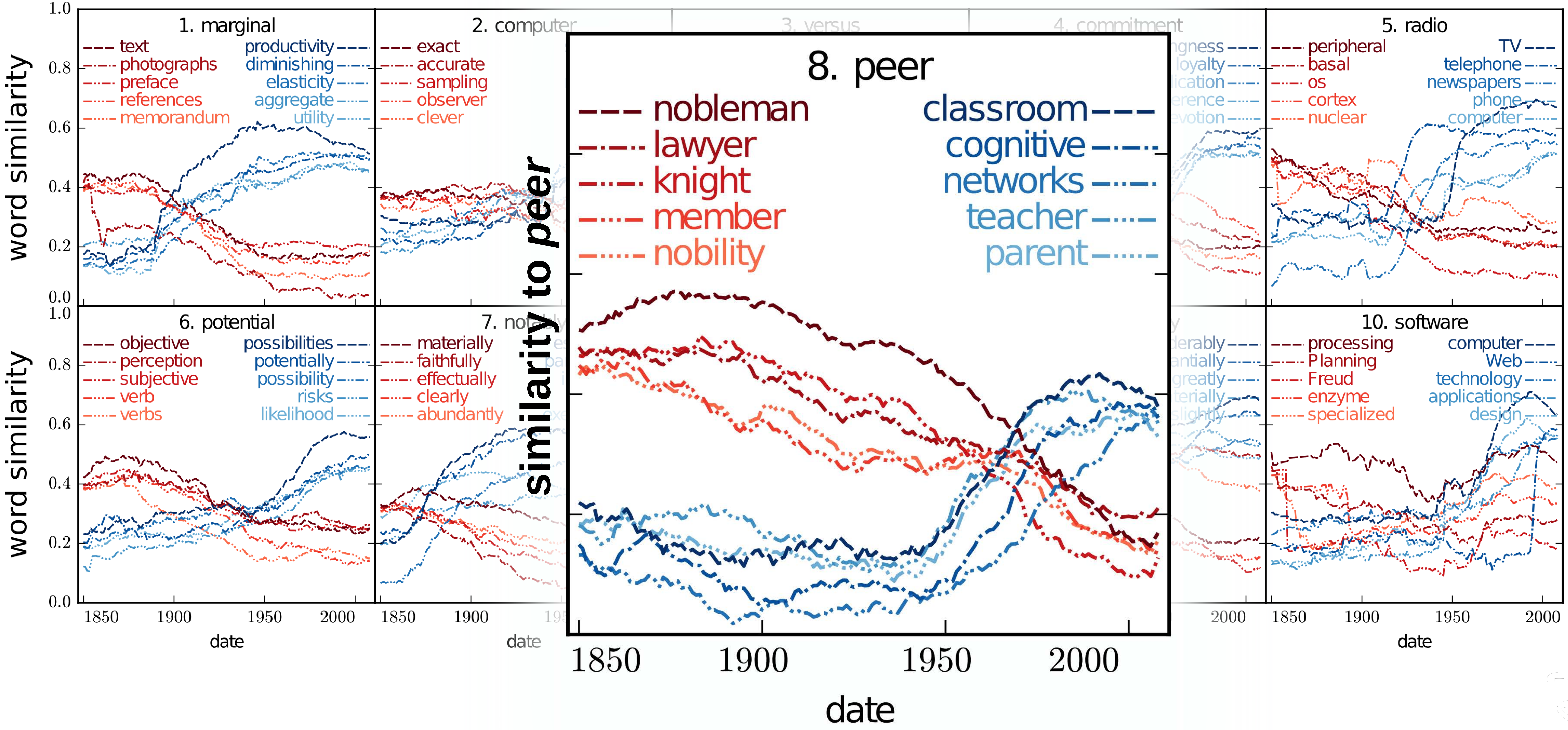
Dynamic Skip Gram Model

Skip Gram Filtering → online learning on real time data



Dynamic Sk

Results from Historical Books Word Similarities over Time



Project Description

Part I

- Re-implement and publish a user-friendly version of dynamic word embeddings code (skip-gram filtering)
- Required: sufficient familiarity with probabilistic ML / maths

Part II

- Create a platform for visualizing dynamic word embeddings
- Required: creativity and data visualization skills

Bamler & Mandt. Dynamic Word Embeddings. ICML 2017.
www.stephanmandt.com