# Dynamic Topic Modeling for Monitoring Market Competition from Online Text and Image Data

Hao Zhang[1] Gunhee Kim[2] Eric P. Xing[1]

[1]: School of Computer Science, Carnegie Mellon University

[2]: Department of Computer Science and Engineering, Seoul National University

August 9, 2015

# Outline

- Introduction
- Model
- Learning and Inference
- Evaluation
- Visualization -- Dynamics and Competitions
- Conclusion

# Background

**The increasing pervasiveness of the Internet has lead to a wealth of consumer-created data over a multitude of online platforms**

## What can we learn?

General public's experience towards different companies' products and service

Performance evaluations in different market conditions (time, location etc.)

# Background

**The increasing pervasiveness of the Internet has lead to a wealth of consumer-created data over a multitude of online platforms**

## What does marketers want to see?

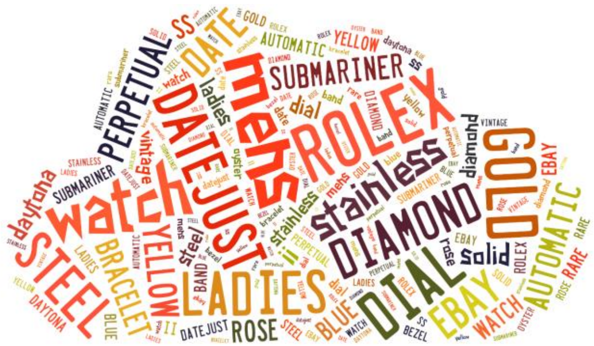
- **Detection:** Listen in consumers' opinions towards their products and their competitors
- **Summarization:** Summarize/visualize how a shared market is occupied by different brands
- **Dynamics:** Monitoring the changes of market competition over time

# Problem Statement

**SuperBowl + beer**


compete

**Watch + luxury**




compete

corona


budlight


guiness

rolex

burberry

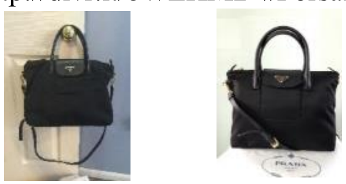
omega

# Problem Statement

## (a) Input: Tweets and associated images of competing brands

*Prada*

#Style #**Prada** Black Leather & Nylon Tessuto Saffiano Shoulder #**Bag** http://dlvr.it/8WZKM2 #Forsale #Auction



Coat from @ASOS , top from @FreePeople, jeans from Rag & Bone, boots from #ChristianLouboutin & **bag** from @**Prada** .

....

*Gucci*

What is the most beautifully-designed perfume bottle? Tell us on the blog here: http://smarturl.it/ie2fka and win **Gucci**

Designer Kate Spade, Invicta, **Gucci** & More Watches from $22 & Extra 20% Off http://www.dealsplus.com/t/1zr85Y

....

*Chanel*

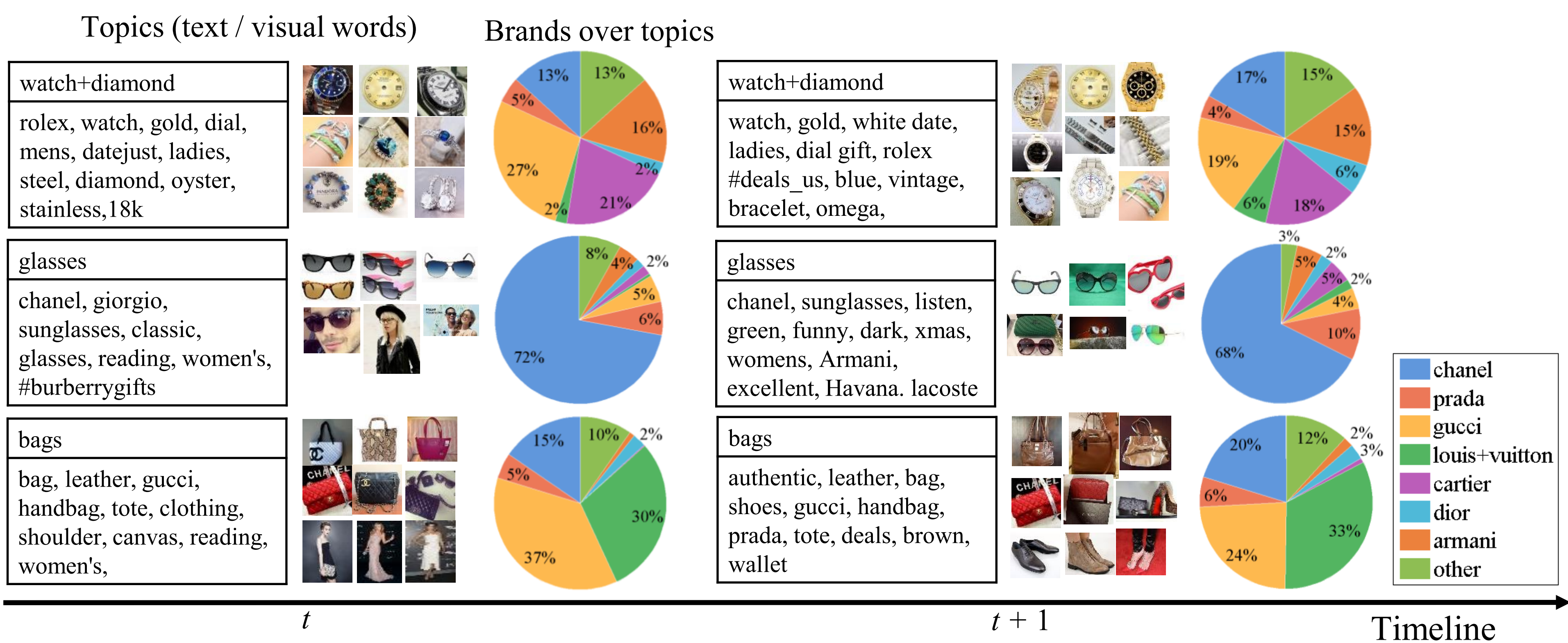
The latest crop of #**Chanel** Pre-Spring bags have arrived! See the full collection now: http://bit.ly/1z3PnKG

Pretty In Pink: From @**Chanel** to @nailsinc, the best petal-hued make-up launches this spring http://vogue.uk/8p6UOi

....

# Problem Statement

**(b) Output: Temporal evolution of topics and brands' proportion over the topics**



# Our Approach: Joint Analysis of Text and Images

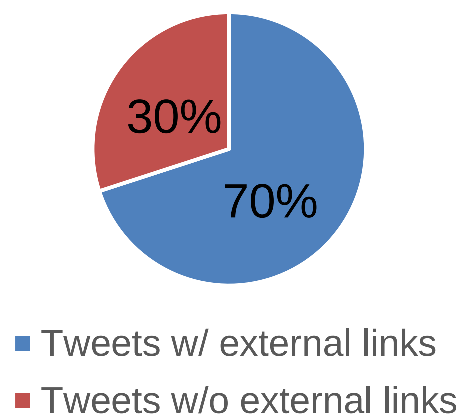
**Take advantage of the pervasiveness of images on the social media**

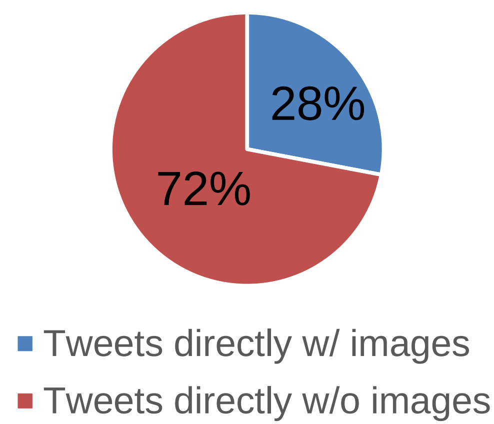
- No previous attempts so far to jointly leverage text and pictures for online market intelligence

## Why are *joint interpretation of text and images* helpful for online market intelligence?

A large portion of tweets simply show images&links without any meaningful text in them. Images play an important role for representing topics in this type of documents

Oh, it's really the most beautifully-designed perfume bottle I have ever seen!!!

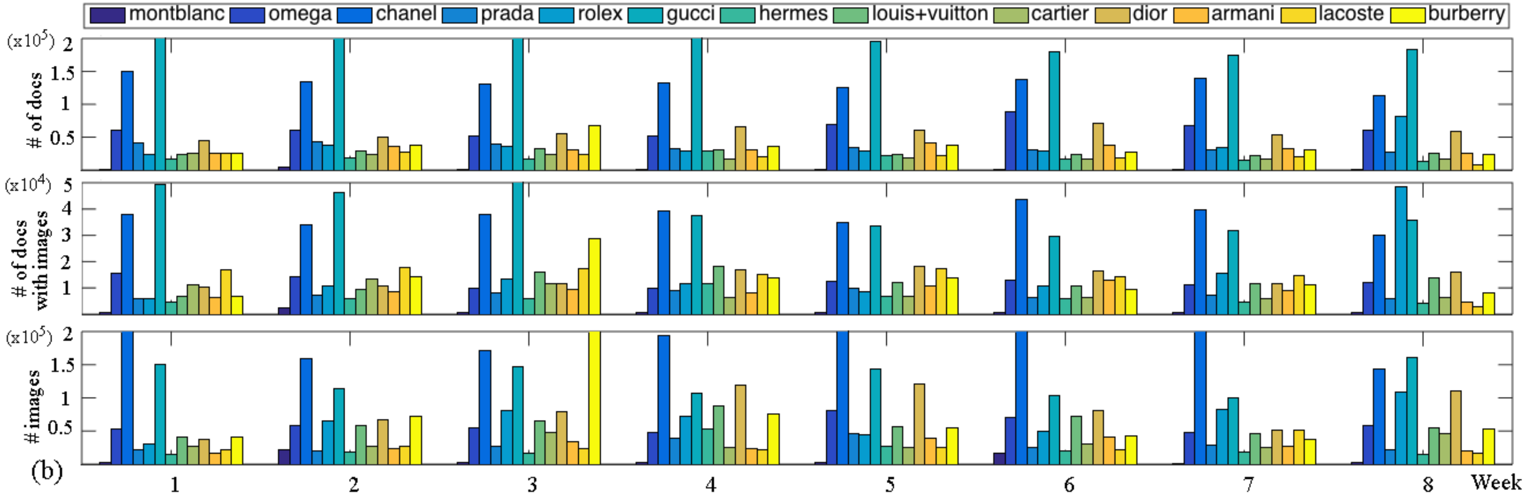
# Our Approach: Joint Analysis of Text and Images

**Take advantage of the pervasiveness of images on the social media**

- No previous attempts so far to jointly leverage text and pictures for online market intelligence

## Why are *joint interpretation of text and images* helpful for online market intelligence?

Many users prefer to use images to deliver their idea more clearly and broadly, and thus the topic detection with images reflects users' intents better.




5.5 million tweets

6.6 million images

# Our Approach: Joint Analysis of Text and Images

**Take advantage of the pervasiveness of images on the social media**

- No previous attempts so far to jointly leverage text and pictures for online market intelligence

## Why are *joint interpretation of text and images* helpful for online market intelligence?

The joint use of images with text also helps marketers interpret the discovered topics.

140 characters limit

What a wonderfullllllll night!!!!!


marketers may need to see the associated images to understand key ideas of tweets easier and quicker

winter
dior
nude
nutrition
Hydrations

# Related Work

## Online Market Intelligence

**BrandPluse[KDD05]**

**Market-Structure[2012]**

**Competitive Intelligence[2011]**

**Brand Monitoring[2011]**

**Show me the money! [KDD 2007]**

- Competitive brands on latent topics
- Jointly leverage text and images

## Topic Model for Econometrics

**Financial TM [2009]**

**Purchase Behavior [2009]**

**Geo TM [2013]**

**Topic Sentiment Mixture [2007]**

**Online Reviews TM [2008]**

- Modeling brands and competitions
- Jointly leverage text and images

# Related Work

## Dynamic and Multi-view Topic Models

**Dynamic TM[2006]**

**Latent Subspace Learning [2012]**

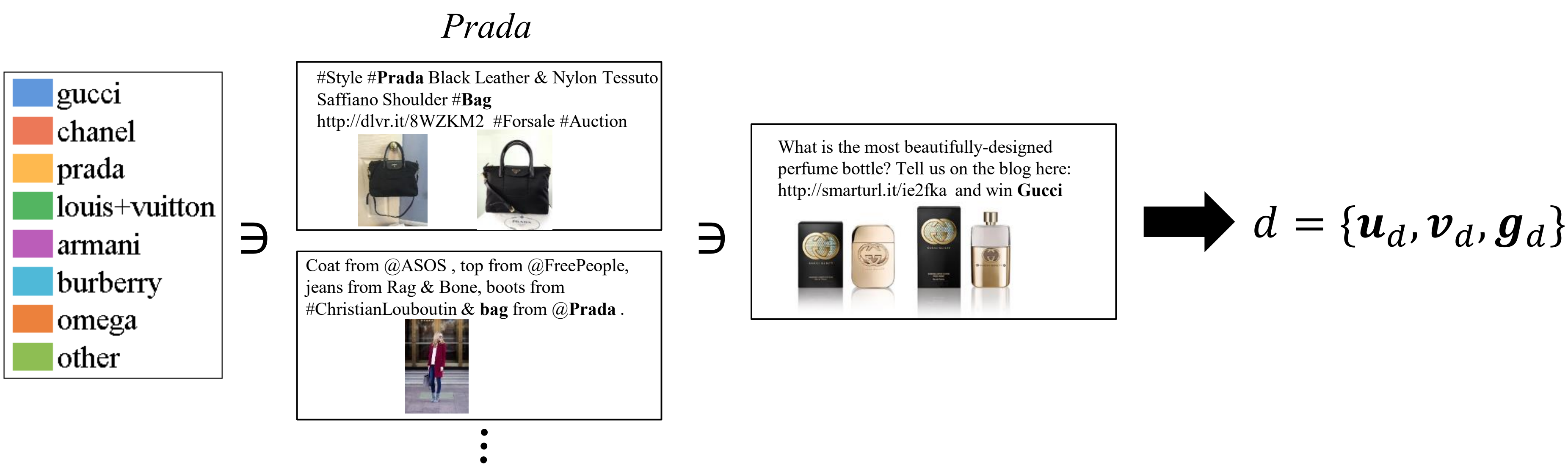
**Topic Models for Image Annotation and Text illustration[2010]**

**Bilateral Correspondence Model [2014]**

- Directly modeling the competition of multiple entities (e.g. brands) over shared topic spaces
- Modeling the interaction between multiple brands and entities

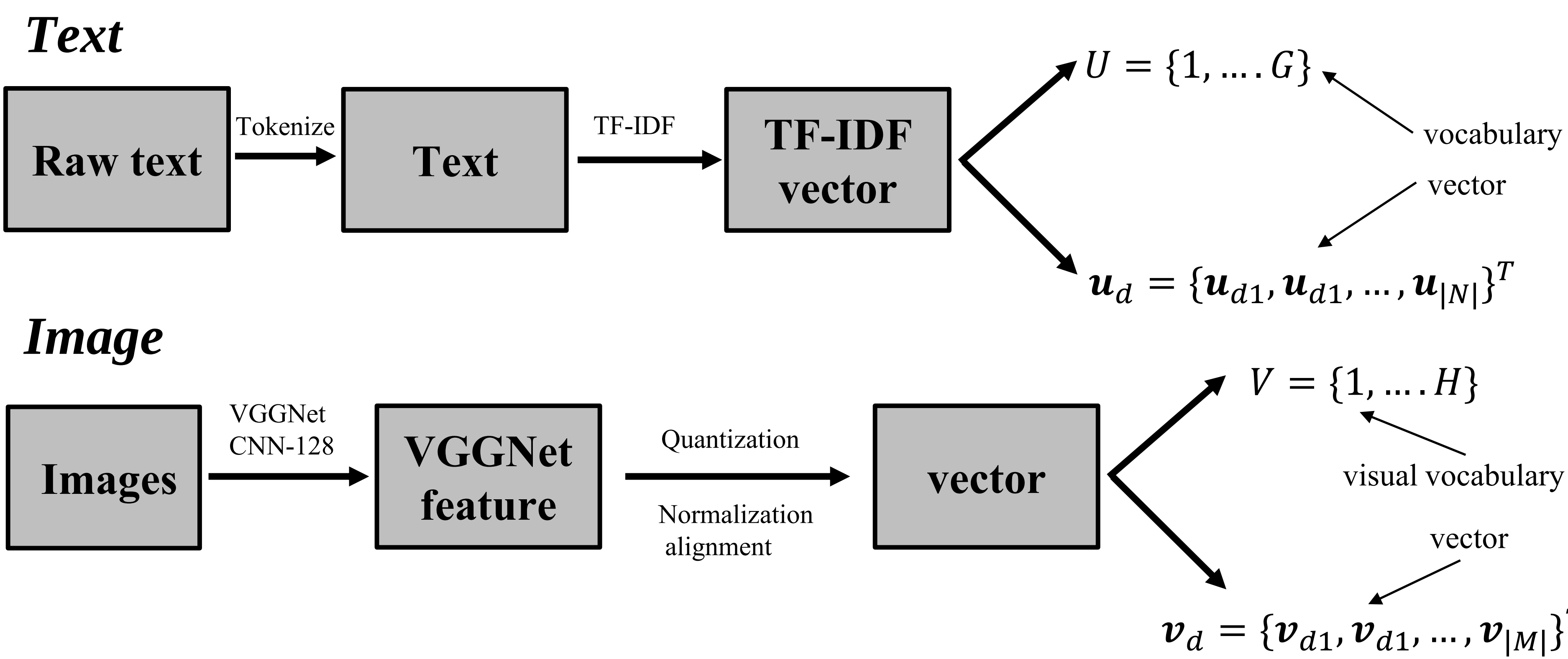
# Model

- **Input**:
  - $\boldsymbol{\mathcal{B}} = \{1, \dots, \boldsymbol{\mathcal{B}}^L\}$ a set of competition brands of interest
  - $\boldsymbol{\mathcal{B}}^L$ is a set of documents related with brand $l$
  - $d = \{\boldsymbol{u}_d, \boldsymbol{v}_d, \boldsymbol{g}_d\} \in \boldsymbol{\mathcal{B}}^L$ is a document consisting of text and images
  - $\boldsymbol{u}_d$ vector representation of the text document
  - $\boldsymbol{v}_d$ vector representation of the images
  - $\boldsymbol{g}_d \in R^L$ vector notation which brands are associated with document $d$



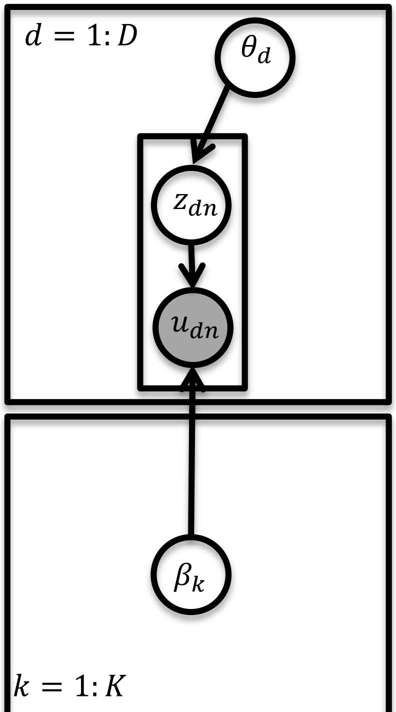
# Dataset

- We collect raw tweets and associated images using Twitter REST API
- Two groups of bands: ***Luxury*** (13 brands) and ***Beer*** (12 brands)
- Total **6.6 million** of tweets and **7.5 million** of images, ranging from ***10/20/2014*** to ***02/01/2015***
- Get the vector representations



# Model

- **Base Model: Sparse Topical Coding**

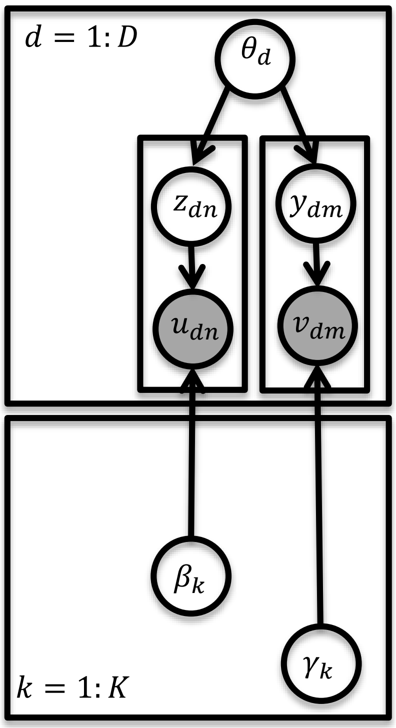



Advantages:

- We encourage each document to be associated with only a small number of strong topics for better analysis of the interaction between multiple brands
- Sparsity leads to a more robust text/image representation in topic space, especially for short documents like tweets (140 characters' limt)

# Model

- **Multi-view Extension**




- Both text and image words share a same document code $\boldsymbol{\theta}$
- $\gamma$: visual topic-word matrix
- Define the distributions as follows:

sample the prior

$$p(\boldsymbol{\theta}) \propto \exp(-\lambda \|\boldsymbol{\theta}\|_1)$$

sparsity on document code

sample the word code

$$p(z_{dn}|\boldsymbol{\theta}_d) \propto \exp(-\delta_u \|z_{dn} - \theta_d\|_2^2 - \rho_u ||z_{dn}||_1)$$

$$p(y_{dm}|\boldsymbol{\theta}_d) \propto \exp(-\delta_v \|y_{dm} - \theta_d\|_2^2 - \rho_v ||y_{dm}||_1)$$

sparsity on word code

sample the word count

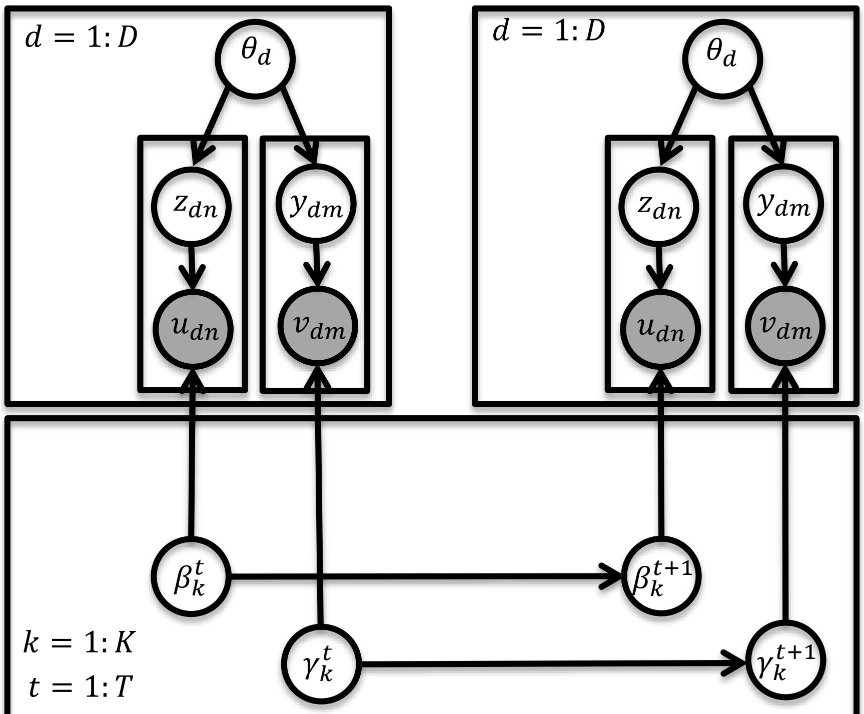
$$p(u_{dn}|\boldsymbol{z}_{dn}, \boldsymbol{\beta}) \propto N\left(u_{dn}; z_{dn}^T \boldsymbol{\beta}_{.n}, \sigma_u^2 \boldsymbol{I}\right)$$

$$p(v_{dm}|\boldsymbol{y}_{dm}, \boldsymbol{\gamma}) \propto N\left(v_{dm}; y_{dm}^T \boldsymbol{\gamma}_{.m}, \sigma_v^2 \boldsymbol{I}\right)$$

exponential family

# Model

- **Dynamic extension**




- Based on the discrete dTM [Blei06]
- Divide a corpus of documents into sequential groups, so that $\beta$ and $\gamma$ change over time
- State space model with a Gaussian noise:

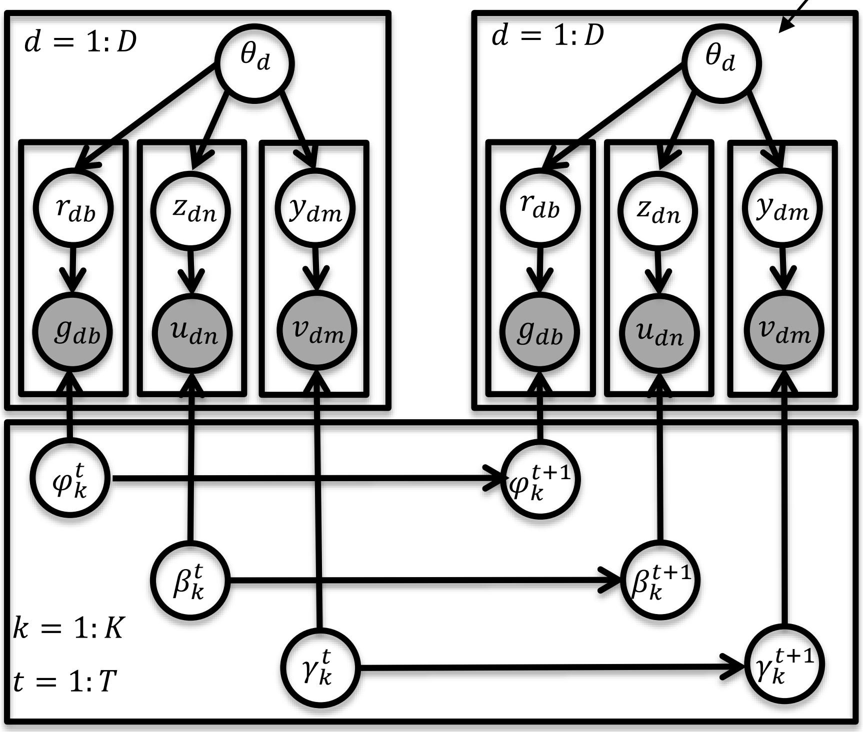
$$p\left(\boldsymbol{\beta}_{k.}^{t} \middle| \boldsymbol{\beta}_{k.}^{t-1}\right) = N(\boldsymbol{\beta}_{k.}^{t-1}, \sigma_\beta^2 I)$$

$$p\left(\boldsymbol{\gamma}_{k.}^{t} \middle| \boldsymbol{\gamma}_{k.}^{t-1}\right) = N(\boldsymbol{\gamma}_{k.}^{t-1}, \sigma_\gamma^2 I)$$

# Model

- **Competition Extension**

bridge




- **Competition:**

$\boldsymbol{\phi} : \boldsymbol{R}^{\boldsymbol{K}\times\boldsymbol{L}}$, proportions of brands on latent topics, $g_d \in R^L$ brand vector for document d, $r_{db} \in R^K$ brand code in topic space

- **Dynamics:**

$\boldsymbol{\phi}$ is evolved over time using Gaussian state space model

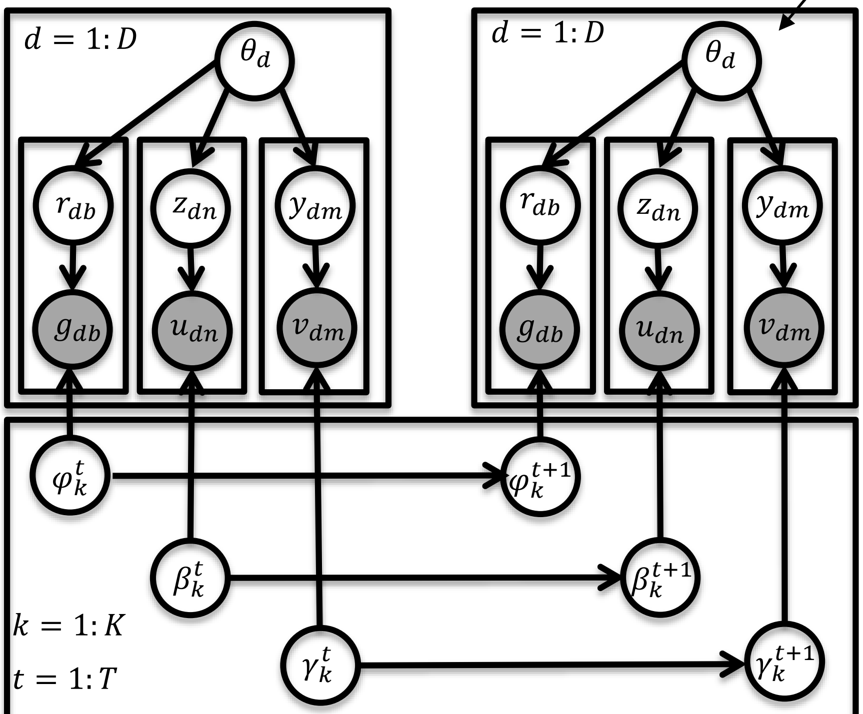
- Distributions:

$$p(r_{db}|\boldsymbol{\theta}_d) \propto \exp(-\delta_b\|r_{db} - \theta_d\|_2^2 - \rho_b||r_{db}||_1)$$
$$p(g_{db}|\boldsymbol{r}_{db},\boldsymbol{\phi}) \propto N\left(g_{db}; r_{db}^T\boldsymbol{\phi}_{.b}, \sigma_b^2\boldsymbol{I}\right)$$
$$p\left(\boldsymbol{\phi}_{k.}^t\middle|\boldsymbol{\phi}_{k.}^{t-1}\right) = N(\boldsymbol{\phi}_{k.}^{t-1}, \sigma_\phi^2 I)$$

# Model

- **Competition Extension**

bridge




For each time slice $t$:

1. Draw a text topic matrix $\boldsymbol{\beta}^t|\boldsymbol{\beta}^{t-1} \sim \mathcal{N}(\boldsymbol{\beta}^{t-1}, \sigma_\beta^2 I)$.
2. Draw an image topic matrix $\boldsymbol{\gamma}^t|\boldsymbol{\gamma}^{t-1} \sim \mathcal{N}(\boldsymbol{\gamma}^{t-1}, \sigma_\gamma^2 I)$.
3. Draw a brand topic matrix with two options: (i) dynamic $\boldsymbol{\phi}^t|\boldsymbol{\phi}^{t-1} \sim \mathcal{N}(\boldsymbol{\phi}^{t-1}, \sigma_\phi^2 I)$, or (ii) independent $\boldsymbol{\phi}^t \sim Unif(0,1)$.
4. For each document $d = (\boldsymbol{u}, \boldsymbol{v})$ in $D^t$,

(a) Sample a document code $\boldsymbol{\theta}_d \sim$ prior $p(\boldsymbol{\theta})$.

(b) For each observed text word $n \in N$,
- i. Sample a word code $\boldsymbol{z}_{dn} \sim p(\boldsymbol{z}_{dn}|\boldsymbol{\theta}_d)$.
- ii. Sample a word count $u_{dn} \sim p(u|\boldsymbol{z}_{dn}, \boldsymbol{\beta})$.

(c) If $M$ is not an empty set:
- i. For each observed visual word $m \in M$,
  - A. Sample a visual word code $\boldsymbol{y}_{dm} \sim p(\boldsymbol{y}_{dm}|\boldsymbol{\theta}_d)$.
  - B. Sample a visual word count $v_{dm} \sim p(v|\boldsymbol{y}_{dm}, \boldsymbol{\gamma})$.

(d) For each observed brand $b \in B$,
- i. Sample a latent brand code $\boldsymbol{r}_{db} \sim p(\boldsymbol{r}_{db}|\boldsymbol{\theta}_d)$
- ii. Sample a brand association $g_{db} \sim p(g|\boldsymbol{r}_{db}, \boldsymbol{\phi})$

# Learning and Inference

- **Map Formulation**
  - Joint Probability

$$\boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{g} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi})$$

$$= \boldsymbol{p}(\boldsymbol{\theta}) \prod_{\boldsymbol{n} \in \boldsymbol{N}} \boldsymbol{p}(\boldsymbol{z_n} | \boldsymbol{\theta}) \boldsymbol{p}(\boldsymbol{u_n} | \boldsymbol{z_n}, \boldsymbol{\beta}) \prod_{\boldsymbol{m} \in \boldsymbol{M}} \boldsymbol{p}(\boldsymbol{y_m} | \boldsymbol{\gamma}) \boldsymbol{p}(\boldsymbol{v_m} | \boldsymbol{y_m}, \boldsymbol{\gamma})$$

$$\prod_{\boldsymbol{b} \in \boldsymbol{B}} \boldsymbol{p}(\boldsymbol{r_b} | \boldsymbol{\phi}) \boldsymbol{p}(\boldsymbol{g_b} | \boldsymbol{r_b}, \boldsymbol{\phi})$$

  - Denote $\Theta^t = \{\theta_d^t, z_d^t, y_d^t, r_d^t\}_{d=1}^{D^t}$(i.e., add the superscript $t$)
  - Negative log posterior

$$-\log \boldsymbol{p}(\Theta^t, \boldsymbol{\beta^t}, \boldsymbol{\gamma^t}, \boldsymbol{\phi^t} | \{\boldsymbol{u_d^t}, \boldsymbol{v_d^t}, \boldsymbol{g_d^t}\}_{\boldsymbol{d=1}}^{\boldsymbol{D^t}})$$

$$\propto -\log \boldsymbol{p}(\Theta^t, \{\boldsymbol{u_d^t}, \boldsymbol{v_d^t}, \boldsymbol{g_d^t}\}_{\boldsymbol{d=1}}^{\boldsymbol{D^t}} | \boldsymbol{\beta^t}, \boldsymbol{\gamma^t}, \boldsymbol{\phi^t})$$

# Learning and Inference

- **Minimize the negative log posterior:**

$$\min_{\{\Theta^t,\boldsymbol{\beta}^t,\boldsymbol{\gamma}^t,\phi^t\}_{t=1}^T} \sum_{t=1}^{T}\sum_{d=1}^{D} \lambda||\boldsymbol{\theta}_d^t||_1$$

sparse term for document code

$$+\sum_{t=1}^{T}(\pi_1||\boldsymbol{\beta}^t-\boldsymbol{\beta}^{t-1}||_2^2+\pi_2||\boldsymbol{\gamma}^t-\boldsymbol{\gamma}^{t-1}||_2^2+\pi_3||\boldsymbol{\phi}^t-\boldsymbol{\phi}^{t-1}||_2^2)$$

evolving chain

$$+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{n\in N_d^t}(\nu_1||\boldsymbol{z}_{dn}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_1||\boldsymbol{z}_{dn}^t||_1+L(\boldsymbol{z}_{dn}^t,\boldsymbol{\beta}^t))$$

text

$$+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{m\in N_d^t}(\nu_2||\boldsymbol{y}_{dm}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_2||\boldsymbol{y}_{dm}^t||_1+L(\boldsymbol{y}_{dm}^t,\boldsymbol{\gamma}^t))$$

image

$$+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{b\in B_d^t}(\nu_3||\boldsymbol{r}_{db}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_3||\boldsymbol{r}_{db}^t||_1+L(\boldsymbol{r}_{db}^t,\boldsymbol{\phi}^t))$$

brand

$$s.t.\ \boldsymbol{\theta}_d^t>0,\forall d,t.\ \boldsymbol{z}_{dn}^t,\boldsymbol{y}_{dm}^t,\boldsymbol{r}_{db}^t>0,\forall d,n,m,b,t$$

$$\beta_k^t\in P_U,\gamma_k^t\in P_V,\phi_k^t\in P_B,\forall k,t$$

simplex constraint

# Learning and Inference

$$
\min_{\{\Theta^t,\boldsymbol{\beta}^t,\boldsymbol{\gamma}^t,\phi^t\}_{t=1}^T} \sum_{t=1}^{T}\sum_{d=1}^{D} \lambda||\boldsymbol{\theta}_d^t||_1
$$

$$
+\sum_{t=1}^{T}(\pi_1||\boldsymbol{\beta}^t-\boldsymbol{\beta}^{t-1}||_2^2+\pi_2||\boldsymbol{\gamma}^t-\boldsymbol{\gamma}^{t-1}||_2^2+\pi_3||\boldsymbol{\phi}^t-\boldsymbol{\phi}^{t-1}||_2^2)
$$

$$
+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{n\in N_d^t}(\nu_1||\boldsymbol{z}_{dn}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_1||\boldsymbol{z}_{dn}^t||_1+L(\boldsymbol{z}_{dn}^t,\boldsymbol{\beta}^t))
$$

$$
+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{m\in N_d^t}(\nu_2||\boldsymbol{y}_{dm}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_2||\boldsymbol{y}_{dm}^t||_1+L(\boldsymbol{y}_{dm}^t,\boldsymbol{\gamma}^t))
$$

$$
+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{b\in B_d^t}(\nu_3||\boldsymbol{r}_{db}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_3||\boldsymbol{r}_{db}^t||_1+L(\boldsymbol{r}_{db}^t,\boldsymbol{\phi}^t))
$$

$$
s.t.\ \boldsymbol{\theta}_d^t>0,\forall d,t.\ \boldsymbol{z}_{dn}^t,\boldsymbol{y}_{dm}^t,\boldsymbol{r}_{db}^t>0,\forall d,n,m,b,t
$$

$$
\beta_k^t\in P_U,\gamma_k^t\in P_V,\phi_k^t\in P_B,\forall k,t
$$

# Learning and Inference

$$\min_{\{\Theta^t, \boldsymbol{\beta}^t, \boldsymbol{\gamma}^t, \phi^t\}_{t=1}^T} \sum_{t=1}^{T} \sum_{d=1}^{D} \lambda ||\boldsymbol{\theta}_d^t||_1$$

$$+ \sum_{t=1}^{T} (\pi_1 ||\boldsymbol{\beta}^t - \boldsymbol{\beta}^{t-1}||_2^2 + \pi_2 ||\boldsymbol{\gamma}^t - \boldsymbol{\gamma}^{t-1}||_2^2 + \pi_3 ||\boldsymbol{\phi}^t - \boldsymbol{\phi}^{t-1}||_2^2)$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D^t} \sum_{n \in N_d^t} (\nu_1 ||\boldsymbol{z}_{dn}^t - \boldsymbol{\theta}_d^t||_2^2 + \rho_1 ||\boldsymbol{z}_{dn}^t||_1 + L(\boldsymbol{z}_{dn}^t, \boldsymbol{\beta}^t))$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D^t} \sum_{m \in N_d^t} (\nu_2 ||\boldsymbol{y}_{dm}^t - \boldsymbol{\theta}_d^t||_2^2 + \rho_2 ||\boldsymbol{y}_{dm}^t||_1 + L(\boldsymbol{y}_{dm}^t, \boldsymbol{\gamma}^t))$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D^t} \sum_{b \in B_d^t} (\nu_3 ||\boldsymbol{r}_{db}^t - \boldsymbol{\theta}_d^t||_2^2 + \rho_3 ||\boldsymbol{r}_{db}^t||_1 + L(\boldsymbol{r}_{db}^t, \boldsymbol{\phi}^t))$$

$$s.t. \boldsymbol{\theta}_d^t > 0, \forall d, t. \boldsymbol{z}_{dn}^t, \boldsymbol{y}_{dm}^t, \boldsymbol{r}_{db}^t > 0, \forall d, n, m, b, t$$

$$\beta_k^t \in P_U, \gamma_k^t \in P_V, \phi_k^t \in P_B, \forall k, t$$

# Learning and Inference

$$\min_{\{\Theta^t, \boldsymbol{\beta}^t, \boldsymbol{\gamma}^t, \phi^t\}_{t=1}^T} \sum_{t=1}^{T} \sum_{d=1}^{D} \lambda ||\boldsymbol{\theta}_d^t||_1$$

$$+ \sum_{t=1}^{T} (\pi_1 ||\boldsymbol{\beta}^t - \boldsymbol{\beta}^{t-1}||_2^2 + \pi_2 ||\boldsymbol{\gamma}^t - \boldsymbol{\gamma}^{t-1}||_2^2 + \pi_3 ||\boldsymbol{\phi}^t - \boldsymbol{\phi}^{t-1}||_2^2)$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D^t} \sum_{n \in N_d^t} (\nu_1 ||\boldsymbol{z}_{dn}^t - \boldsymbol{\theta}_d^t||_2^2 + \rho_1 ||\boldsymbol{z}_{dn}^t||_1 + L(\boldsymbol{z}_{dn}^t, \boldsymbol{\beta}^t))$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D^t} \sum_{m \in N_d^t} (\nu_2 ||\boldsymbol{y}_{dm}^t - \boldsymbol{\theta}_d^t||_2^2 + \rho_2 ||\boldsymbol{y}_{dm}^t||_1 + L(\boldsymbol{y}_{dm}^t, \boldsymbol{\gamma}^t))$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D^t} \sum_{b \in B_d^t} (\nu_3 ||\boldsymbol{r}_{db}^t - \boldsymbol{\theta}_d^t||_2^2 + \rho_3 ||\boldsymbol{r}_{db}^t||_1 + L(\boldsymbol{r}_{db}^t, \boldsymbol{\phi}^t))$$

$$s.t.\ \boldsymbol{\theta}_d^t > 0, \forall d, t.\ \boldsymbol{z}_{dn}^t, \boldsymbol{y}_{dm}^t, \boldsymbol{r}_{db}^t > 0, \forall d, n, m, b, t$$

$$\beta_k^t \in P_U, \gamma_k^t \in P_V, \phi_k^t \in P_B, \forall k, t$$

# Learning and Inference

$$\min_{\{\Theta^t,\boldsymbol{\beta}^t,\boldsymbol{\gamma}^t,\phi^t\}_{t=1}^T} \sum_{t=1}^{T}\sum_{d=1}^{D} \lambda||\boldsymbol{\theta}_d^t||_1$$

$$+\sum_{t=1}^{T}(\pi_1||\boldsymbol{\beta}^t-\boldsymbol{\beta}^{t-1}||_2^2+\pi_2||\boldsymbol{\gamma}^t-\boldsymbol{\gamma}^{t-1}||_2^2+\pi_3||\boldsymbol{\phi}^t-\boldsymbol{\phi}^{t-1}||_2^2)$$

$$+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{n\in N_d^t}(\nu_1||\boldsymbol{z}_{dn}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_1||\boldsymbol{z}_{dn}^t||_1+L(\boldsymbol{z}_{dn}^t,\boldsymbol{\beta}^t))$$

$$+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{m\in N_d^t}(\nu_2||\boldsymbol{y}_{dm}^t-\boldsymbol{\theta}_d^t||_2^2+\rho_2||\boldsymbol{y}_{dm}^t||_1+L(\boldsymbol{y}_{dm}^t,\boldsymbol{\gamma}^t))$$

$$+\sum_{t=1}^{T}\sum_{d=1}^{D^t}\sum_{b\in B_d^t}(\nu_3||\boldsymbol{r}_d^t-\boldsymbol{\theta}_d^t||_2^2+\rho_3||\boldsymbol{r}_{db}^t||_1+L(\boldsymbol{r}_{db}^t,\boldsymbol{\phi}^t))$$

$$s.t.\ \boldsymbol{\theta}_d^t>0,\forall d,t.\ \boldsymbol{z}_{dn}^t,\boldsymbol{y}_{dm}^t,\boldsymbol{r}_{db}^t>0,\forall d,n,m,b,t$$

$$\beta_k^t\in P_U,\gamma_k^t\in P_V,\phi_k^t\in P_B,\forall k,t$$

# Model Evaluation

## As a Topic Model: Topic Quality Evaluation

– Argument 1: Lower perplexity ≠ higher quality [J. Chang 2009]

– Argument 2: Perplexity is not a fair metric for models with different distributions

- We directly evaluate the ***Coherence*** and ***Validity*** of the learned topics [Xie 2013]

  – Define the **Coherence Measure (CM)**:

$$\boldsymbol{CM = \frac{\#\ of\ relevant\ words}{\#\ of\ words\ in\ valid\ topics}}$$

  – Define the **Validity Measure (VM)**:

$$\boldsymbol{VM = \frac{\#\ of\ valid\ topics}{\#\ of\ topics}}$$

- Both textual and visual topics are evaluated on the Amazon Mechanical Turk

# Model Evaluation

## As a Topic Model: Topic Quality Evaluation

- **Average VM/CM on text topics**

| | VM (Beer / Luxury) | CM (Beer / Luxury) |
|---|---|---|
| *dLDA* | 0.53 / 0.68 | 0.55 / 0.52 |
| *STC + dyn* | 0.44 / 0.66 | 0.57 / 0.57 |
| *cdSTC + multi* | 0.51 / 0.70 | **0.63 / 0.59** |
| *cdSTC + text* | **0.605 / 0.71** | 0.61 / **0.59** |

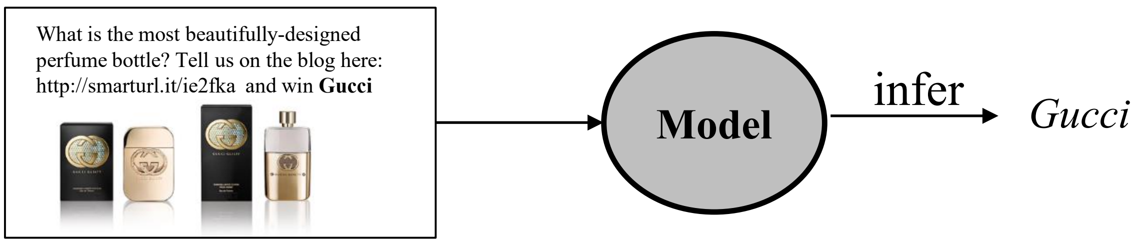
- **Average VM/CM on visual topics**

| | VM (Beer / Luxury) | CM (Beer / Luxury) |
|---|---|---|
| *Kmeans* | 0.39 / 0.56 | 0.59 / 0.64 |
| *LDA + multi* | **0.57** / 0.63 | 0.51 / 0.69 |
| *cdSTC + multi* | **0.57 / 0.65** | **0.66 / 0.71** |

# Model Evaluation

## As a Topic Model: Evaluation on Prediction

- **Task I**: Given a novel tweet, can we predict its most associated brand?




novel tweets

  – Supervised dSTC (sdSTC): infer the most associated brand

$$\max_{\{\Theta^t, \mathcal{M}^t, \boldsymbol{\eta}^t\}_{t=1}^{T}} \sum_{t=1}^{T} f(\Theta^t, \mathcal{M}^t, D^t) + CR(\Theta^t, \boldsymbol{\eta}^t) + \frac{1}{2}\|\boldsymbol{\eta}^t\|_2^2$$

$$s.t.\ \boldsymbol{\theta}_d^t > 0, \forall d, t.\ \boldsymbol{z}_{dn}^t, \boldsymbol{y}_{dm}^t > 0, \forall d, n, m, t$$

$$\beta_k^t \in P_U, \gamma_k^t \in P_V, \forall k, t$$

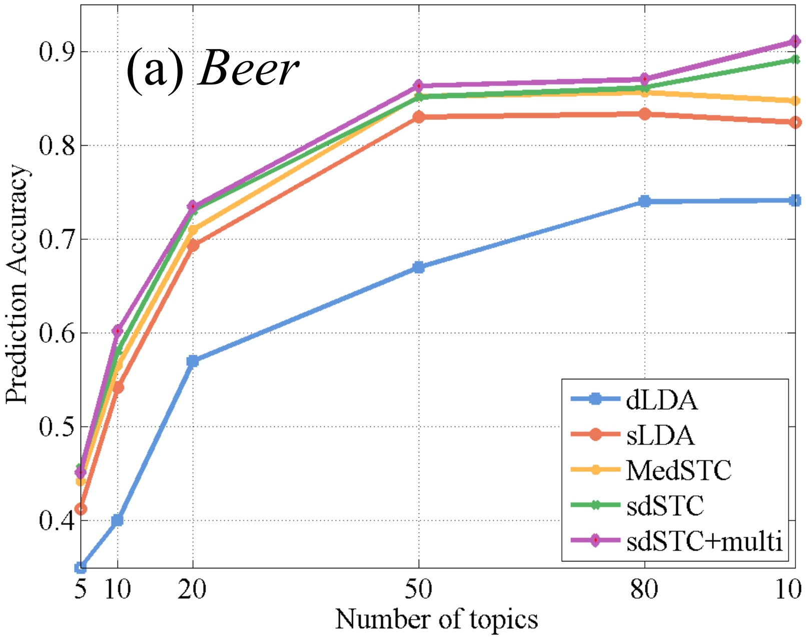
where $R$ is the multi-class hinge loss.

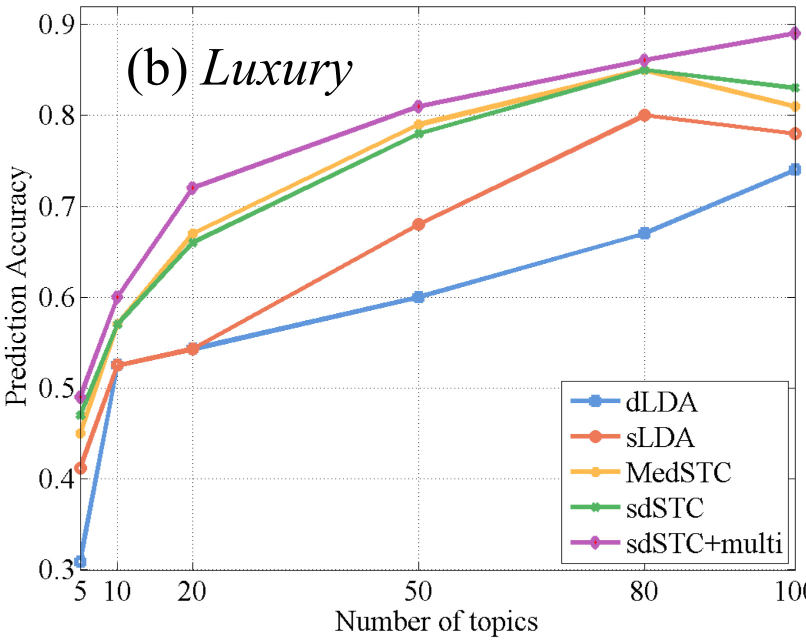
- Solved using coordinated descent

# Model Evaluation

## As a Topic Model: Evaluation on Prediction

- **Task I-I:**
  - Randomly split data in every time slice into 90% for training and 10% for testing
  - Motivation: let the model see data in every time slice
  - Text and images complement each other to detect more representative topics

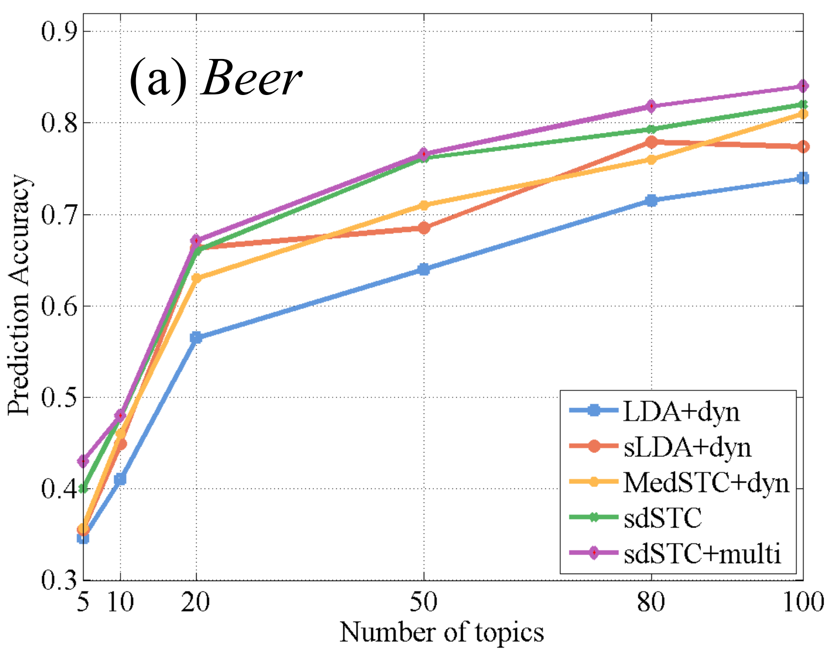


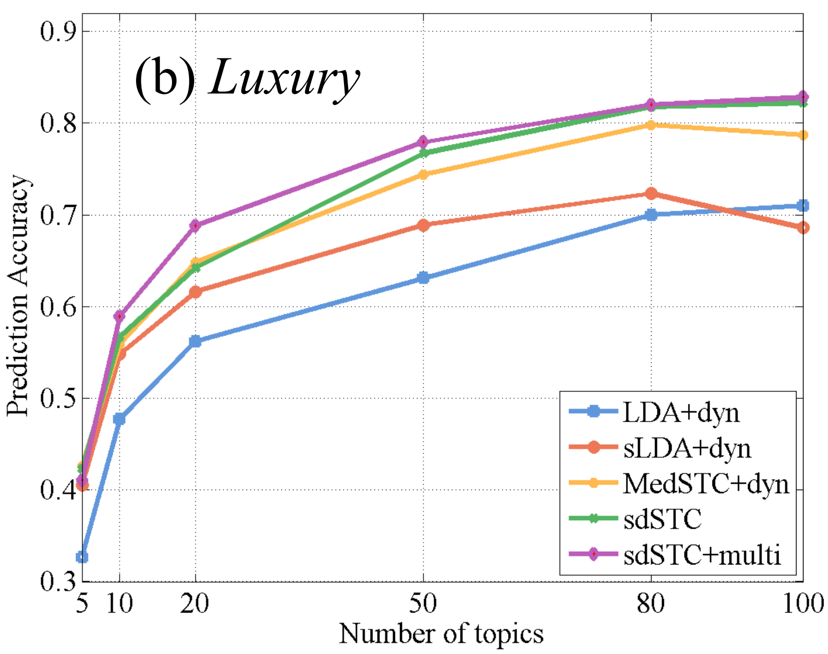



# Model Evaluation

## As a Topic Model: Evaluation on Prediction

- **Task I-II:**
  - Use the data in $[1, t-1]$ for training, $[t-1,\ t]$ for testing
  - Motivation: let the model only see data in past time slices
  - Image data is very helpful to predict the *future*

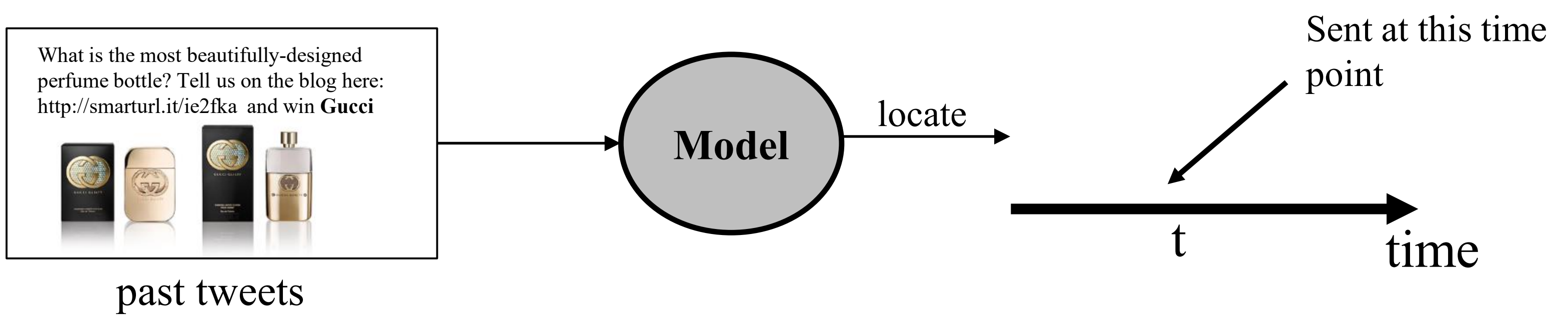


# Model Evaluation

## As a Topic Model: Evaluation on Prediction

- **Task II:** given an unseen past document, can we predict which time slice it is likely to belong?



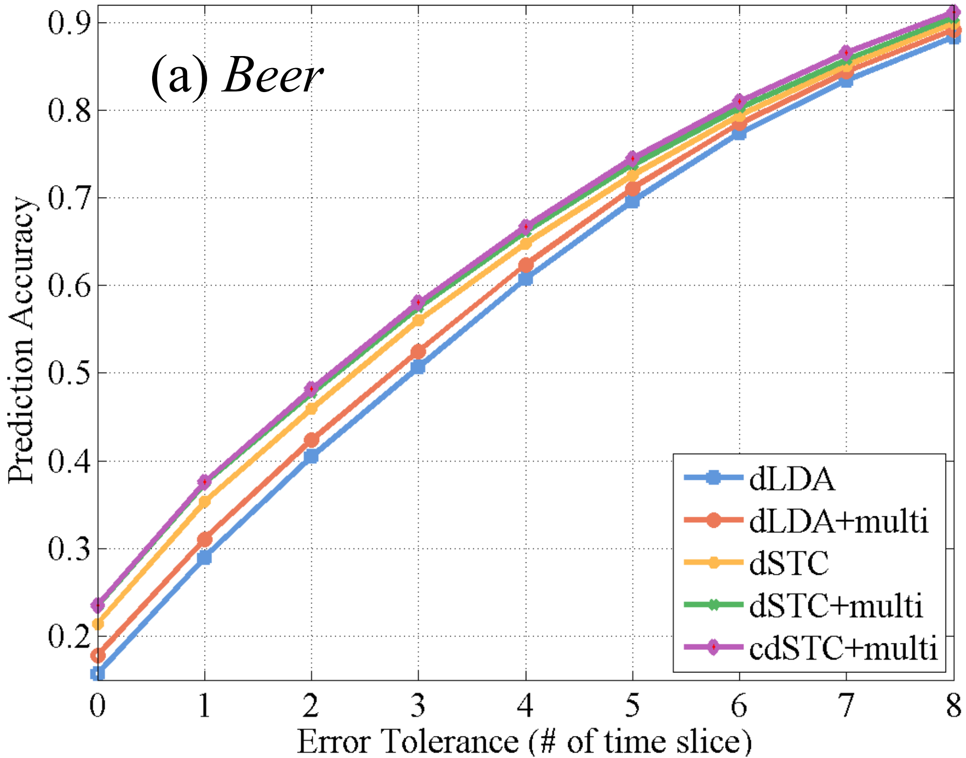

$$\max_{t}\ p(d|\boldsymbol{\mathcal{M}}^t)\ , where$$

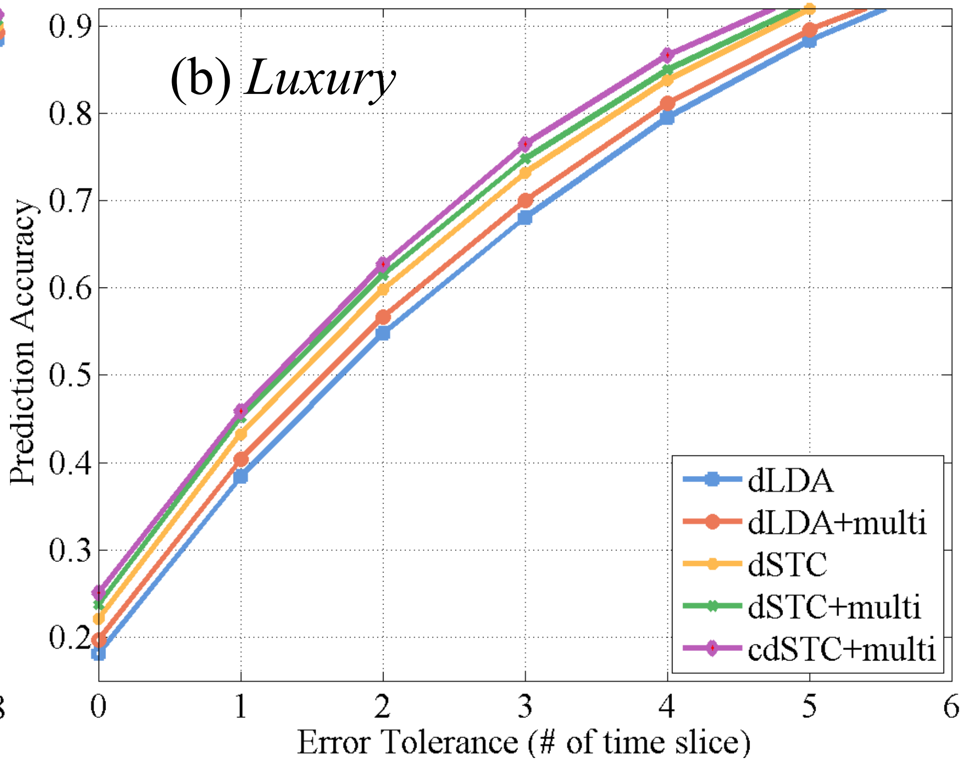
$$p(d|\boldsymbol{\mathcal{M}}^t) = \prod_{n\in N_d} p(u_n|\boldsymbol{\beta}^t) \prod_{m\in M_d} p(v_m|\boldsymbol{\gamma}^t) \prod_{b\in B_d} p(g_b|\boldsymbol{\phi}^t)$$

# Model Evaluation

## As a Topic Model: Evaluation on Prediction

- **Task II**
  - Randomly split the data of every time slice into 90% for training and 10% for localization test.
  - The explicit modeling of brand information does help improve the performance

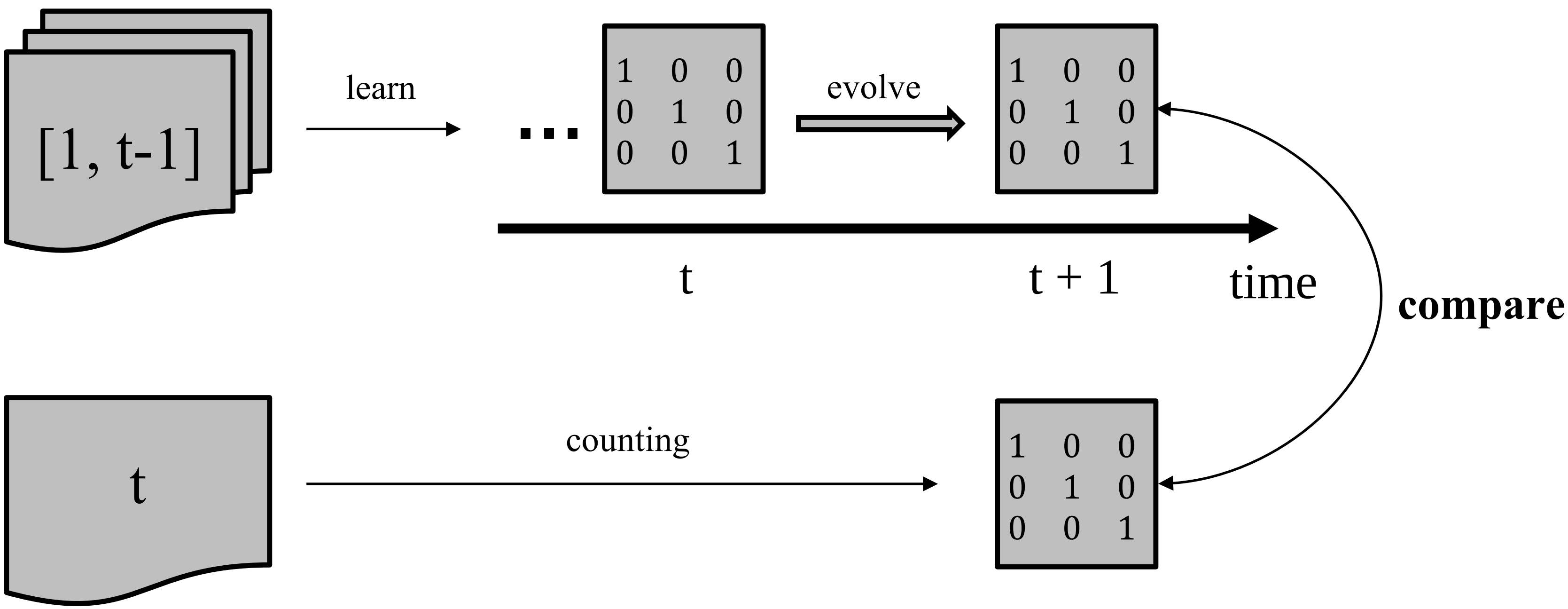


# Model Evaluation

## An Interesting Prediction Task

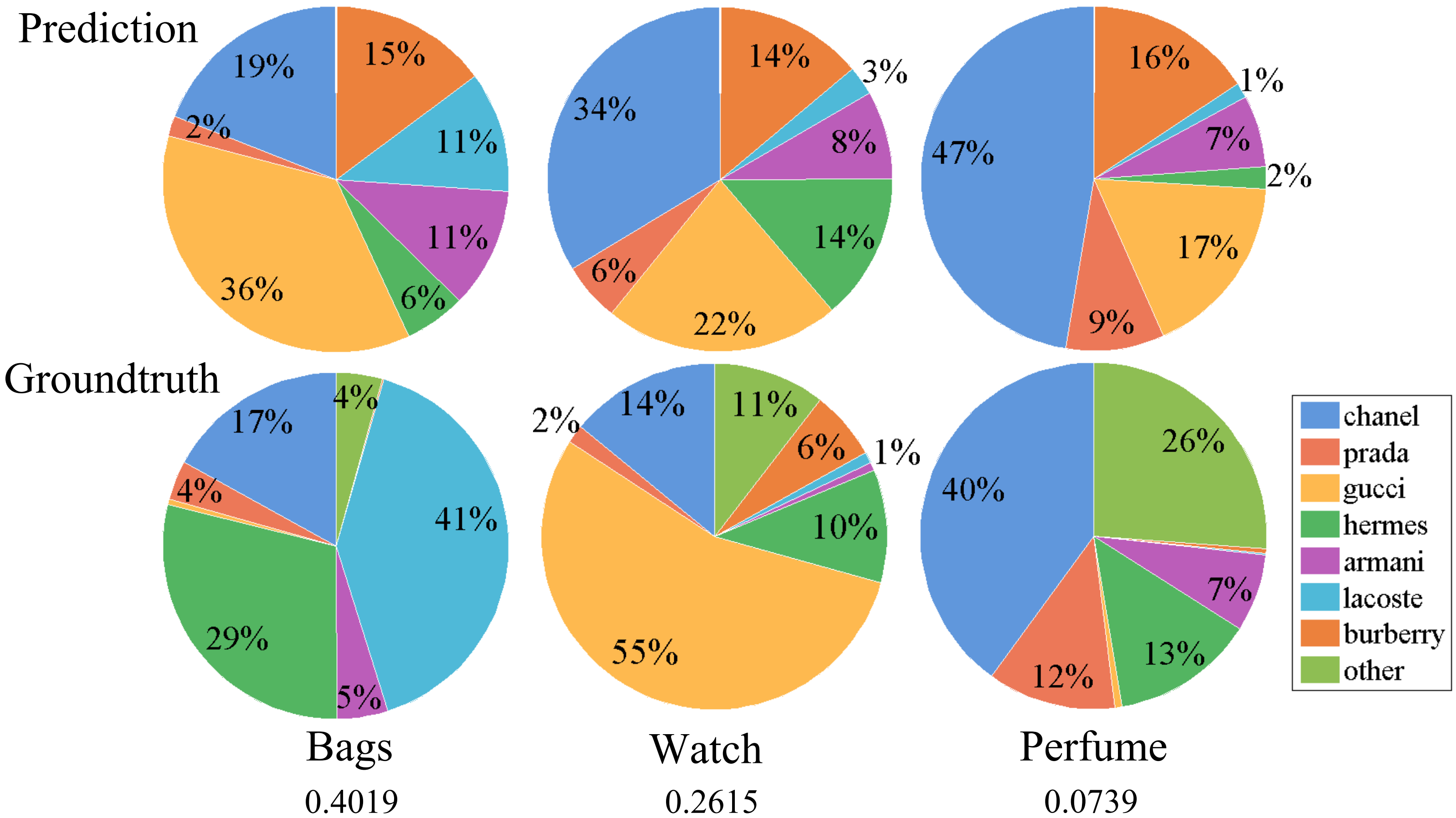
- **Task III:** what if we want to predict the future competition trends according to past data?
- How? Given past data, we evolve the occupation matrix $\phi$ over time



# Model Evaluation

## An Interesting Prediction Task

- **Task III**
  - Evaluated using the KL divergence



# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

### As a monitor, we aim to answer:

- **Static:** how brands occupy the market in one time slice?
- **Dynamic:**
  - how each textual/visual topic evolves over time?
  - how each brand's occupation changes over time? (local)
  - how's the competition trends between multi-brands like over time? (global)

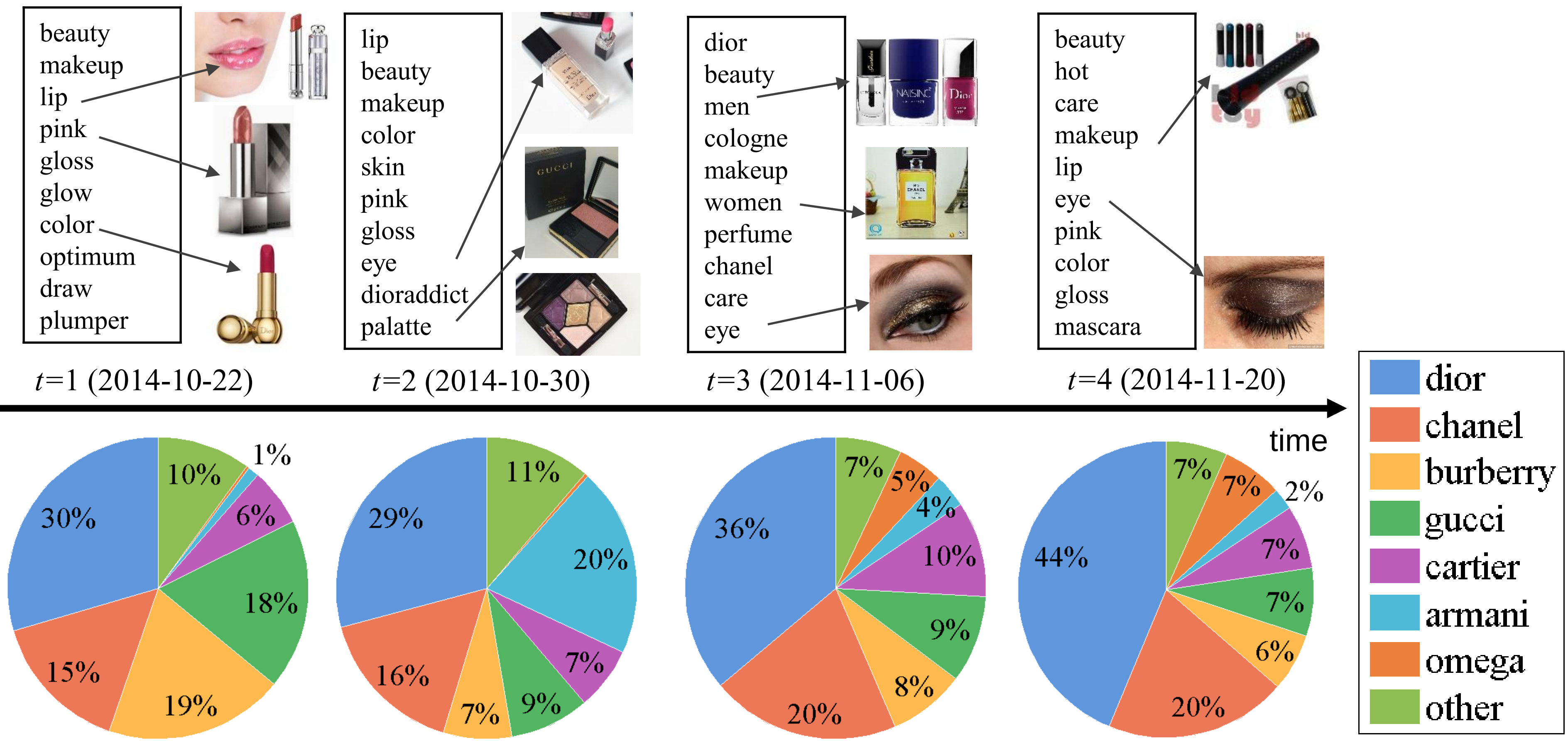
easy

difficult

# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

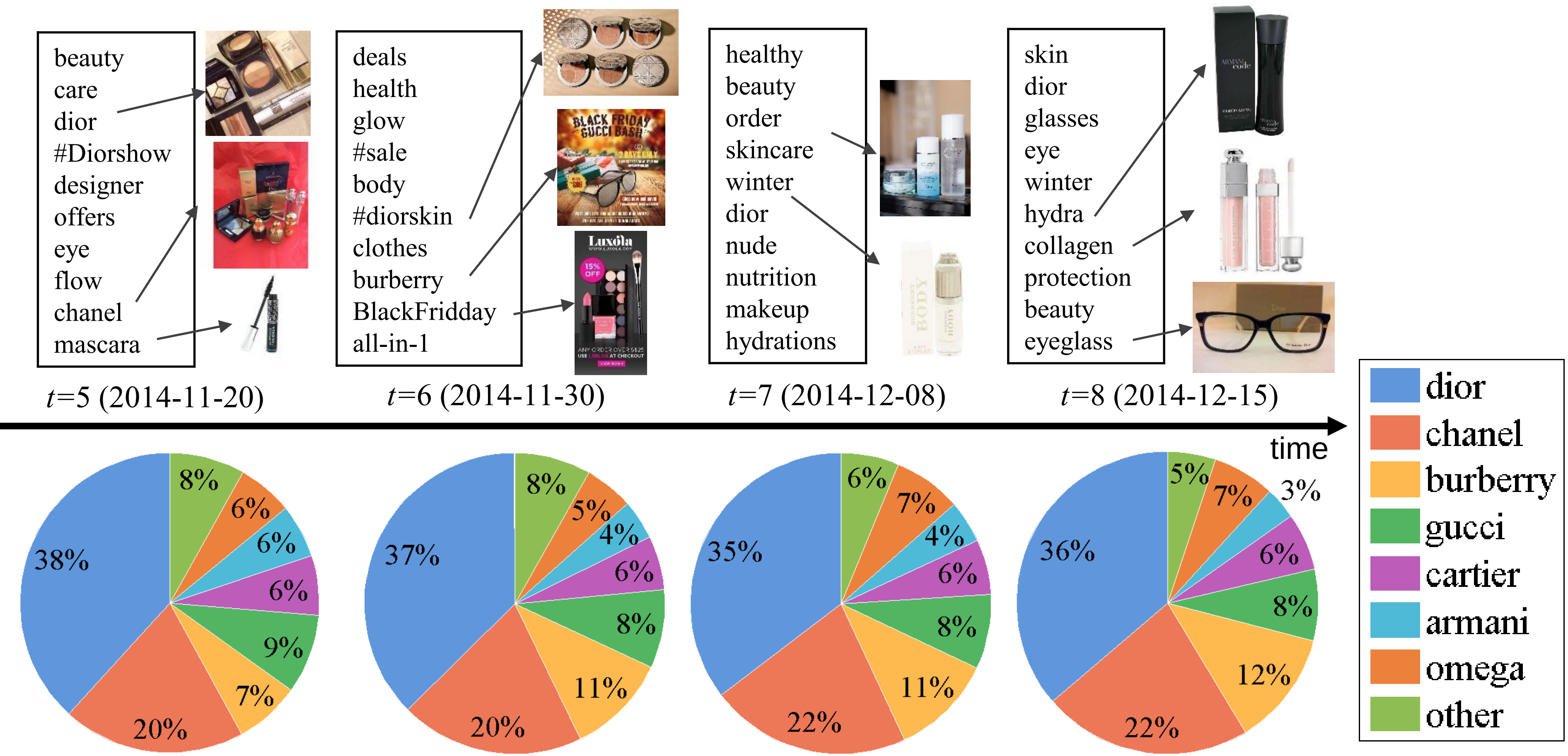
Topic: **beauty**



# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

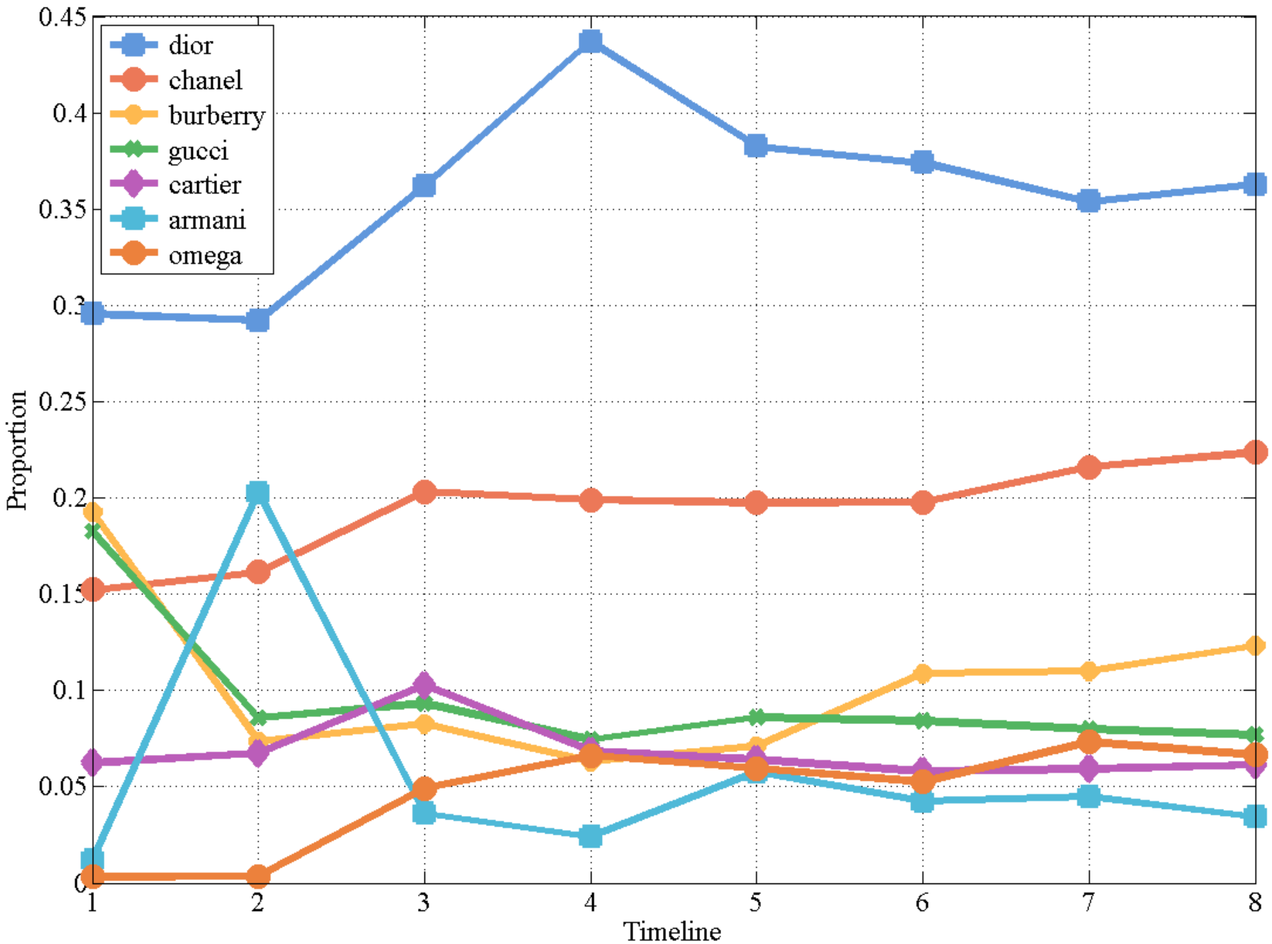
Topic: **beauty**



# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

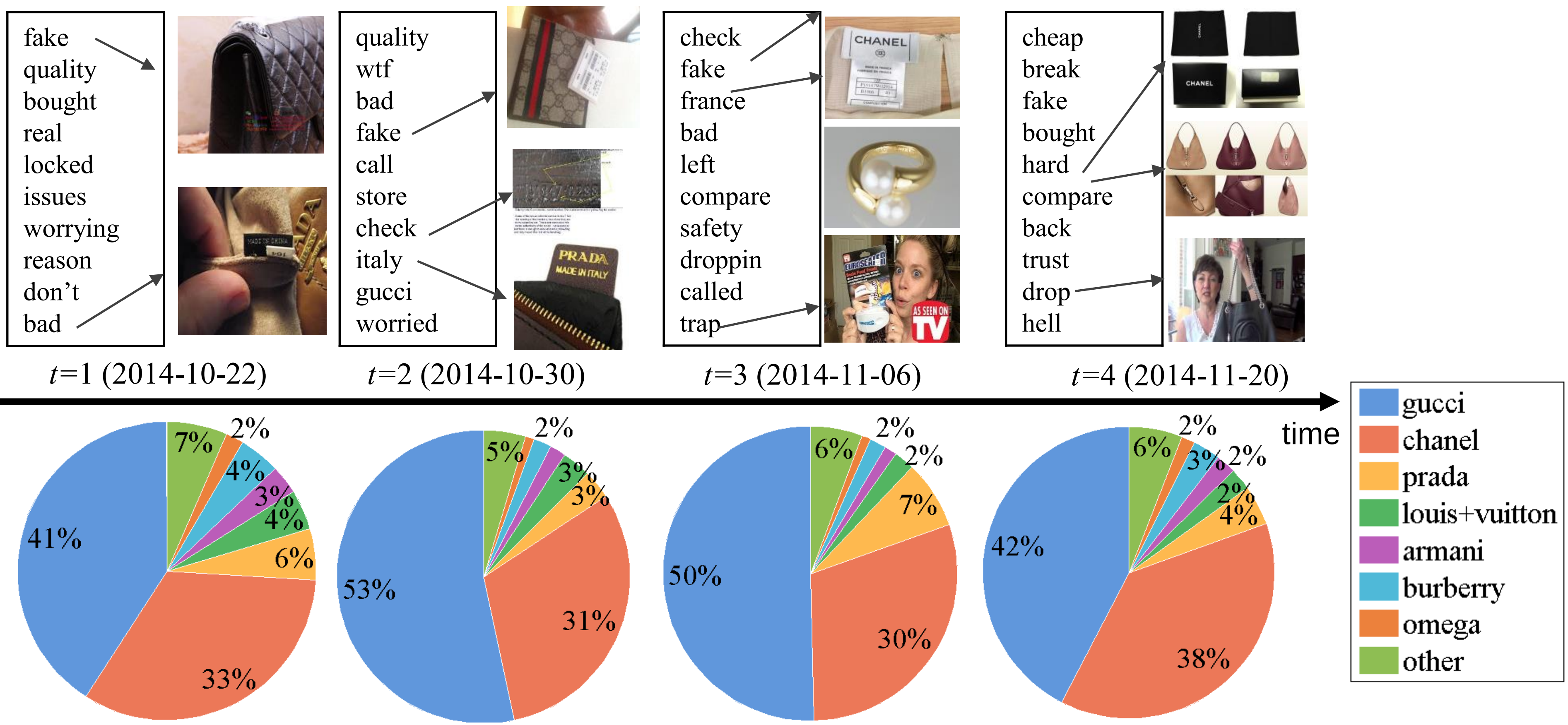
Topic: **beauty**



# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

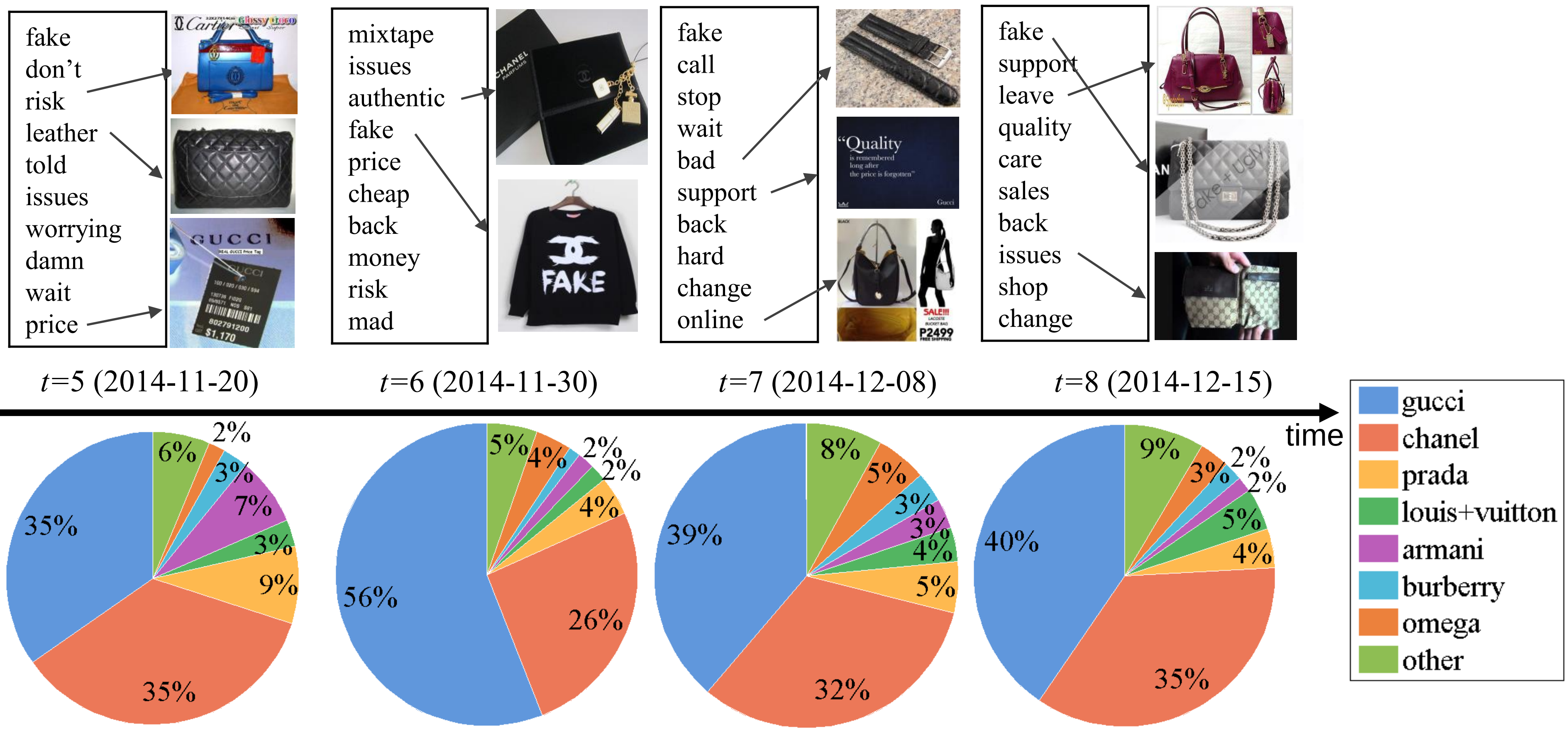
Topic: **fake+bad**



# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

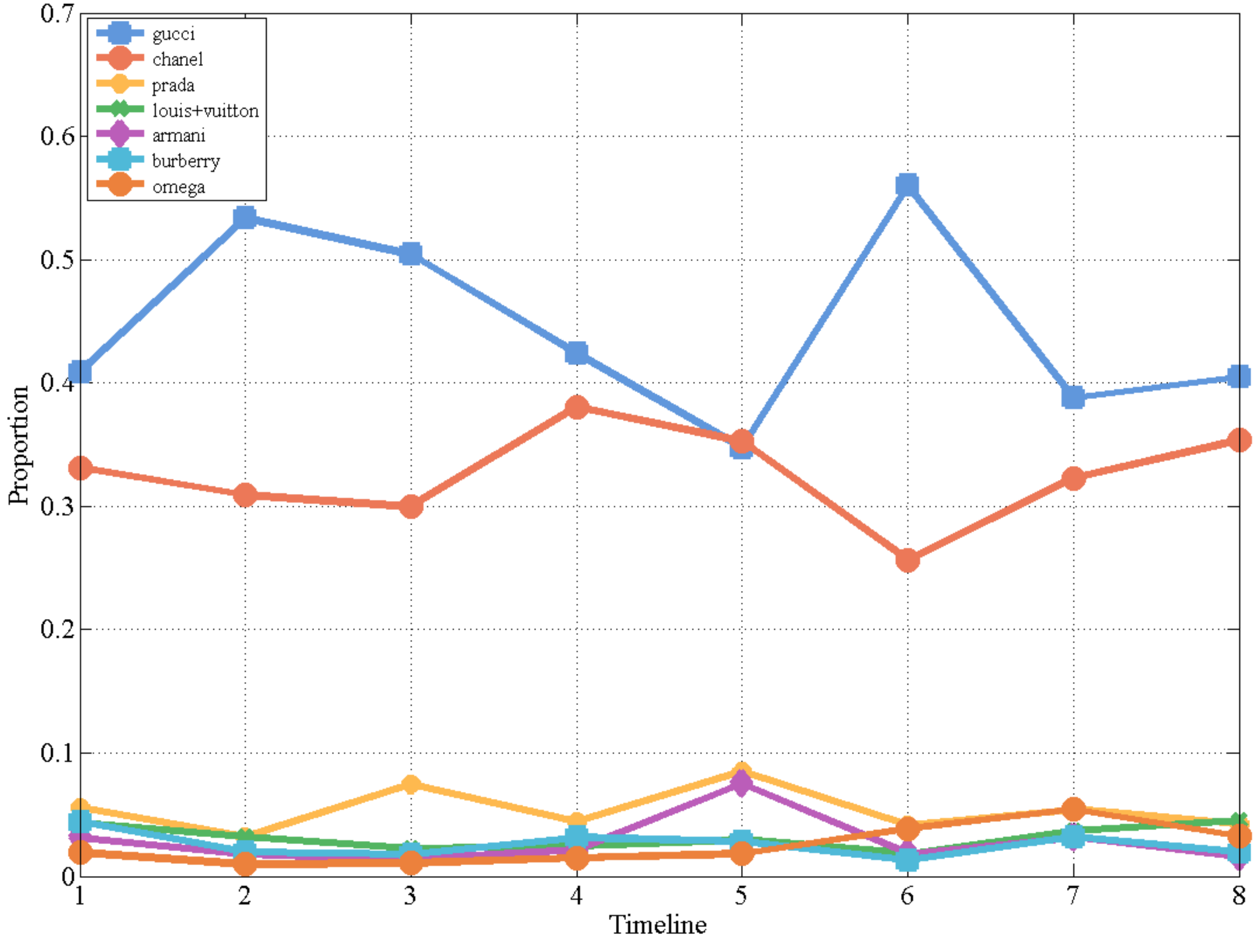
Topic: **fake+bad**



# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

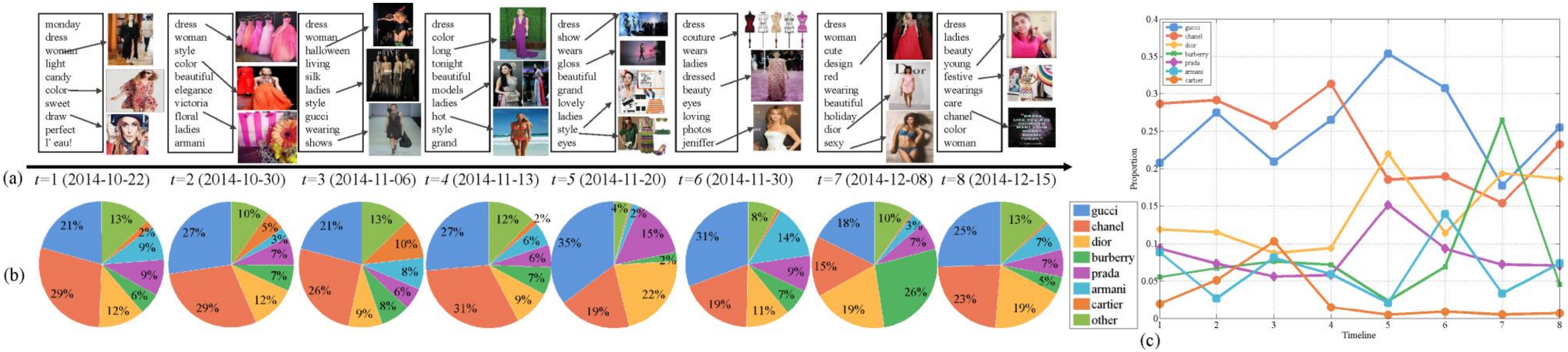
Topic: **fake-bad**



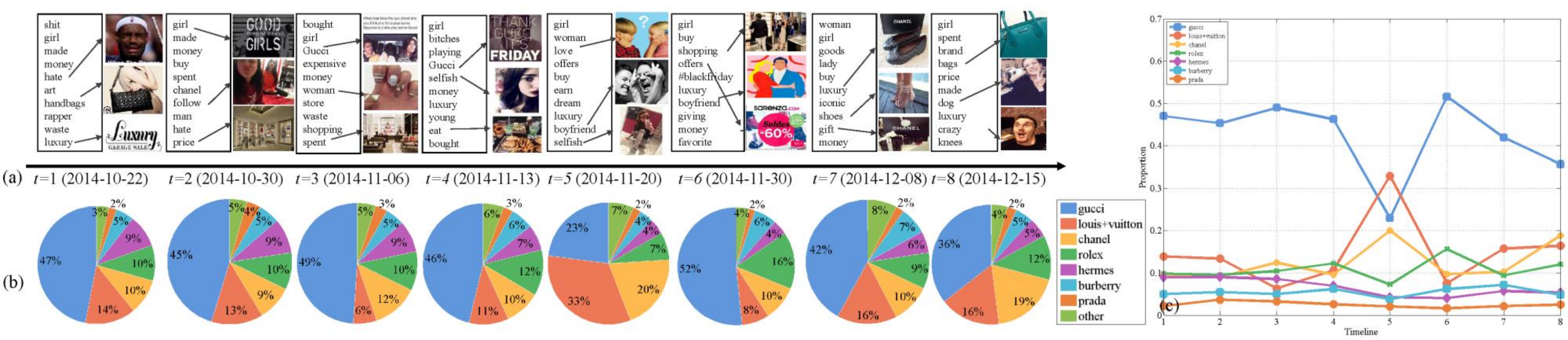
# Visualization: Brand Competition

## Monitoring Competitions and Dynamics

### Topics: woman + dress




### Topics: girl + waste



# Conclusion

- **First attempt so far to propose a principled topic model to**
  - Discover the topics that are competitively shared between multiple brands
  - Track the temporal evolution of dominance of brands over topics by leveraging both text and image data

- **We propose a novel dynamic topic model to correctly address three major challenges:**
  - Multi-view representation of text and images
  - Modeling of latent topics that are competitively shared by multiple brands
  - Tracking temporal evolution of the topics and brand occupations

# Conclusion

- **We evaluate our algorithm using newly collected dataset from Twitter from October 2014 to February 2015:**
  - 10 million tweets with 8 million of associated images
  - Superior performance for dynamic topic modeling and three prediction tasks:
    - Prediction of the most associated brands
    - Most-likely created time
    - Competition trends for unseen tweet
  - Visualizations of competition trends extracted from tons of data

- **Various potential applications**
  - Social media monitoring and visualization
  - Joint analysis of online multi-modal data
  - Online market intelligence

# Thank You!

# Q & A

Project page http://www.cs.cmu.edu/~hzhang2/projects/BrandCompetition/brand competition.html