



## Introduction

### Image Tweets

- Constitute a large traffic in microblogs (e.g., 45% of posts in *Weibo*)
- Attract larger viewership and survive longer than text-only posts

### Research Questions

- Why do people post image tweets?
- What is the relationship between the image and text?
- Can we design a model to interpret how image tweets are generated?

### Our Contributions

- Identify multiple image and text relationships (i.e., visual and emotional)
- Develop VELDA, a novel topic model to capture the two relations and model the generative process

## Multiple Image and Text Correlations

From our previous work [1], we know image and text in microblog can be correlated from visual and emotional aspect.

Back in #London, #tea in #hand-decorated **china** by my mum, **strawberry** from my garden and best read @BritishVogue



Visually Relevant

I have been missing you for such a long time. We taste sweetness in every bitterness. This is life. Have faith. Love life.



Emotionally Relevant

We further confirm this by surveying 109 real *Weibo* users:

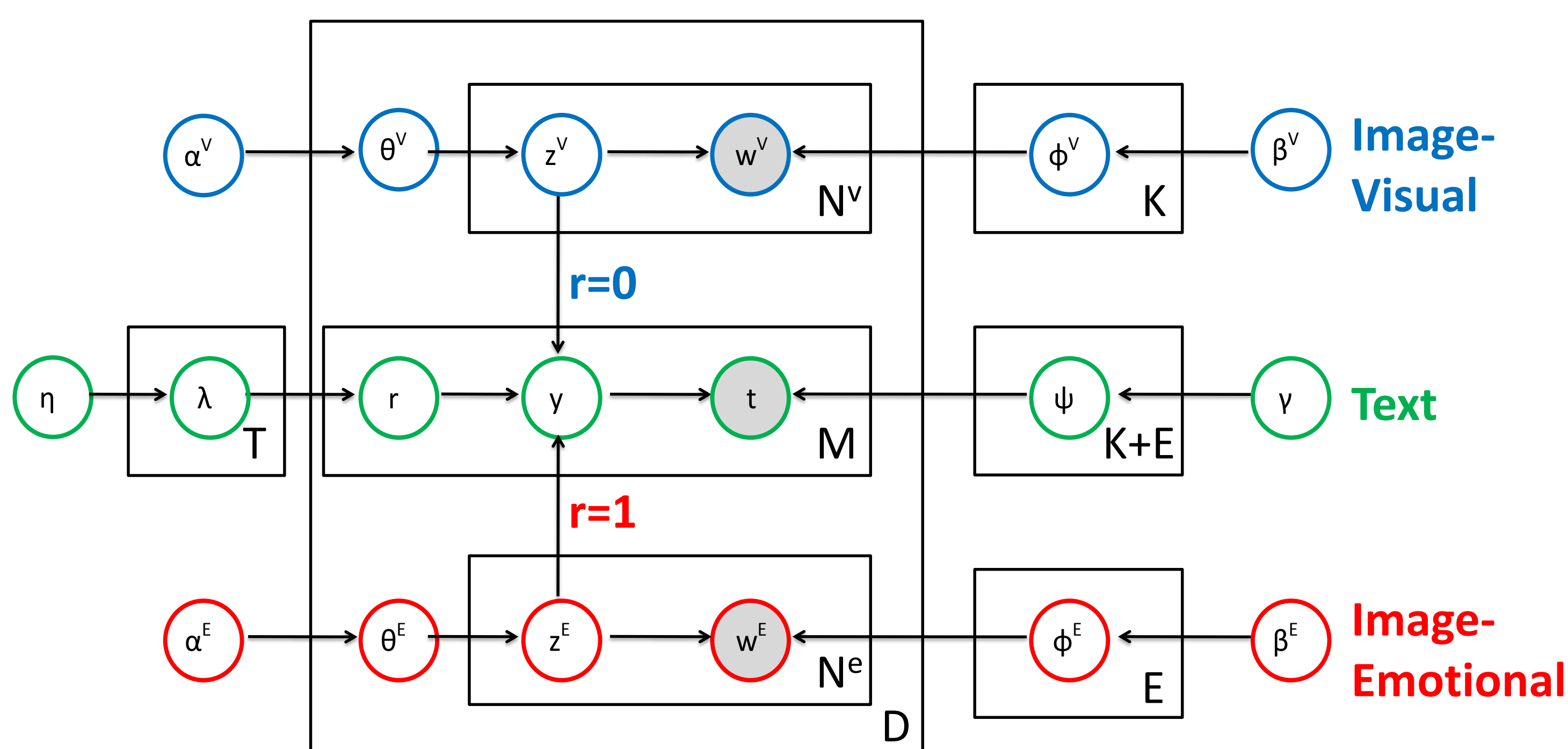
*Why do you embed an image in a tweet?*

- 66.6%: Enhancing the emotion of the text
- 29.4%: Visual correspondence to the text
- 3.7%: Pure visual attractiveness

[1] Tao Chen et al (2013). Understanding and Classifying Image Tweet. ACM MM'13.

## Visual-Emotional LDA (VELDA)

Our VELDA model captures the two correlations at the topic level, while existing methods (e.g., Canonical Correlation Analysis, Correspondence LDA) are only able to capture a single correlation.



- Models two views of an image via two standard LDAs
- Introduces a switch variable  $r$  to indicate the relevance between image and a textual word
- Models per-term relevance distribution  $\lambda$
- Infers the parameters through collapsed Gibbs sampling

### Feature Representation

- Textual** words: Segment Chinese text and tokenize English text
- Visual** words: Quantize SIFT descriptors with K-Means
- Emotional** words: Segment image into patches and quantize the emotional features in each patch with K-Means

## Experiments


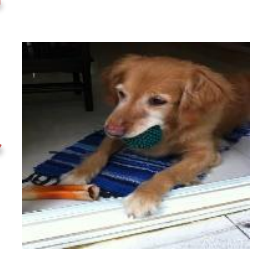
### Five Datasets

	Weibo	Twitter	Google-Zh	Google-En	Wiki POTD
Size	22.7K	16.4K	38.8K	26.9K	2.5K
Language	Chinese	English	Chinese	English	English

### Evaluation with Cross-modality Image Retrieval

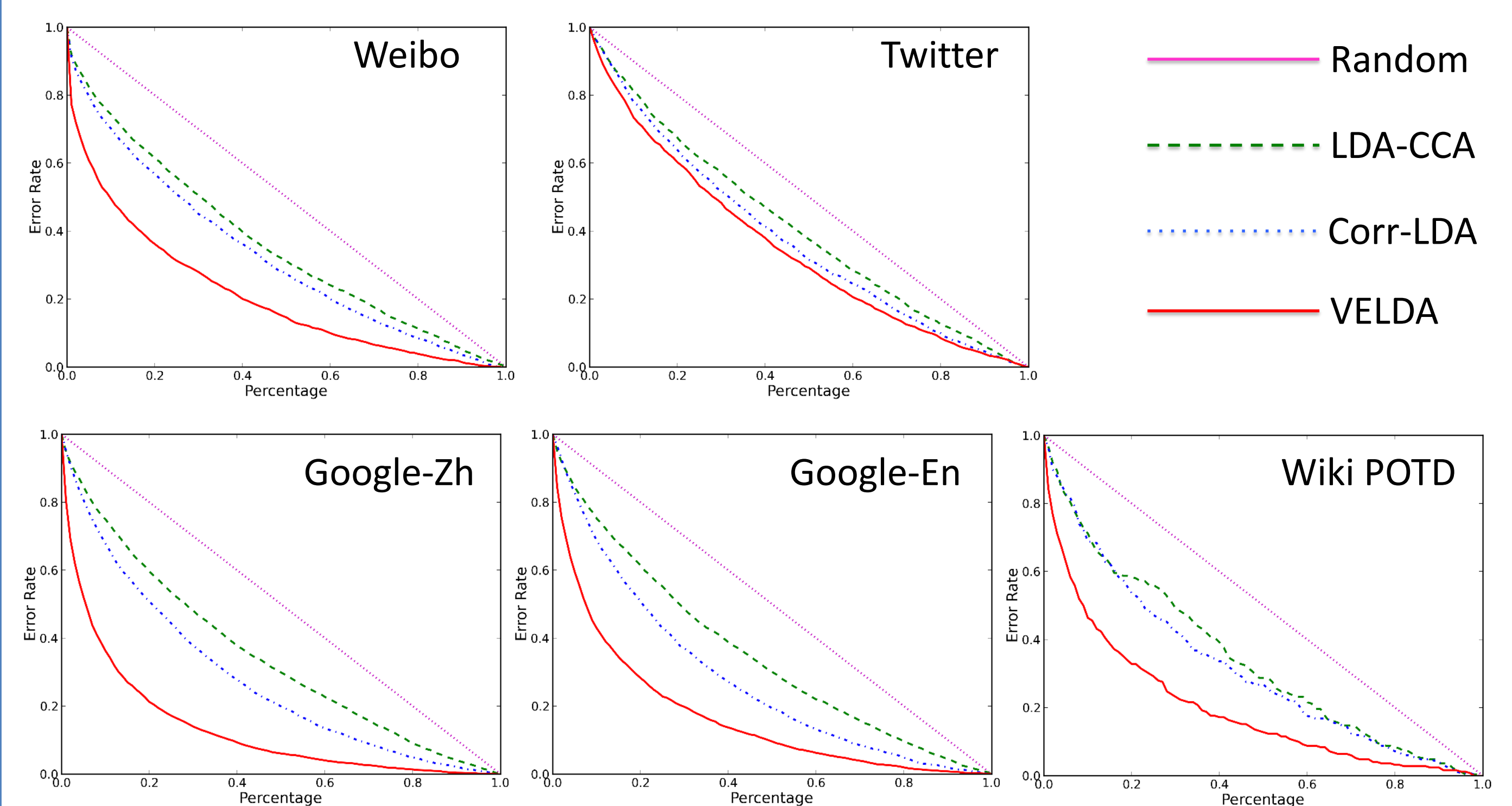
- Given a textual query, rank images by

$$= \prod_{n=1}^N (\underbrace{\lambda_{t_n,0} \sum_{k=1}^K \psi_{k,t_n} \theta_{i,k}^V}_{\text{Visual relevance}} + \underbrace{\lambda_{t_n,1} \sum_{e=1}^E \psi_{e,t_n} \theta_{i,e}^E}_{\text{Emotional relevance}})$$

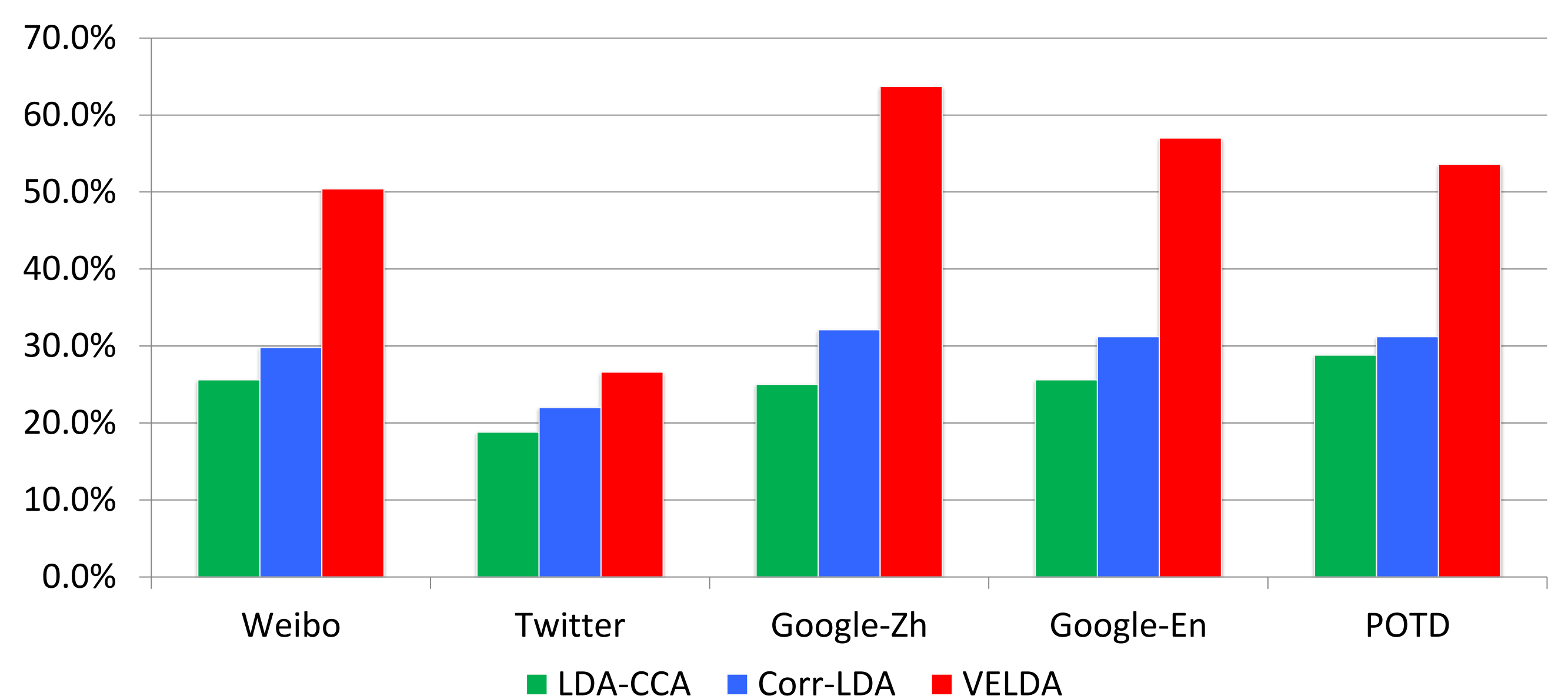
query   } top t%

- Correct: if the ground truth image appears at the top  $t$  % of retrieved results
- 90% as training and 10% as testing

### Error Rate with Varied $t$ %



### Accuracy at the Top 10% of the Ranked List



- VELDA significantly outperforms the other methods
- Robustness and good generalization
- More challenging on Twitter due to the extreme brevity of tweets (average 6.7 textual words)

## Towards Microblog Illustration

[Have some #nuts at noon] Nuts such as **walnuts**, **peanuts**, sunflower seeds, hazelnuts, cedar nuts and chestnuts, should be part of our daily diet. They are rich in Omega-3 and Omega-6 fatty acids and other essential amino acids. These are essential for good health and have anti-aging benefits too.



#Upset I am **hungry** but I cannot eat now as I have to **wait** for someone else. What if there is a blackout now? Let me amuse myself by reading up some **jokes**.

