Item Recommendation with Variational Autoencoders and Heterogenous Priors

Giannis Karamanolakis
Columbia

Kevin Raji Cherian
Columbia

Ananth Ravi Narayan
Columbia

Jie Yuan
Columbia

Da Tang
Columbia

Tony Jebara
Columbia, Netflix

Item Recommendation - Collaborative Filtering (CF)

$U \times I$

Item Recommendation - Collaborative Filtering (CF)

Item Recommendation - Collaborative Filtering (CF)

Item Recommendation - Collaborative Filtering (CF)

Item Recommendation - Collaborative Filtering (CF)

$U \times I$

<u>Latent Factor Models</u>

$(U \times K) \times (K \times I)$

Item Recommendation - Collaborative Filtering (CF)

<u>Latent Factor Models</u>

$$(U \times K) \times (K \times I)$$

(-) linear $\longrightarrow$ limited modeling capacity

Item Recommendation - Collaborative Filtering (CF)

Latent Factor Models

$$(U \times K) \times (K \times I)$$

(-) linear $\longrightarrow$ limited modeling capacity

Non-linear Features

Item Recommendation - Collaborative Filtering (CF)

$U \times I$

Latent Factor Models

$$(U \times K) \times (K \times I)$$

(-) linear $\longrightarrow$ limited modeling capacity

Non-linear Features

Neural Networks

Item Recommendation - Collaborative Filtering (CF)

<u>**Latent Factor Models**</u>

$$(U \times K) \times (K \times I)$$

(-) linear $\longrightarrow$ limited modeling capacity

Non-linear Features

Neural Networks

<u>**Variational Autoencoders (VAEs)**</u>

"Auto-encoding Variational Bayes" D. P. Kingma, M. Welling, ICLR 2014

"Variational Autoencoders for Collaborative Filtering" D. Liang, RG. Krishnan, MD. Hoffman, T. Jebara, WWW 2018

Item Recommendation - Collaborative Filtering (CF)

Latent Factor Models

$$(U \times K) \times (K \times I)$$

(-) linear $\longrightarrow$ limited modeling capacity

Non-linear Features

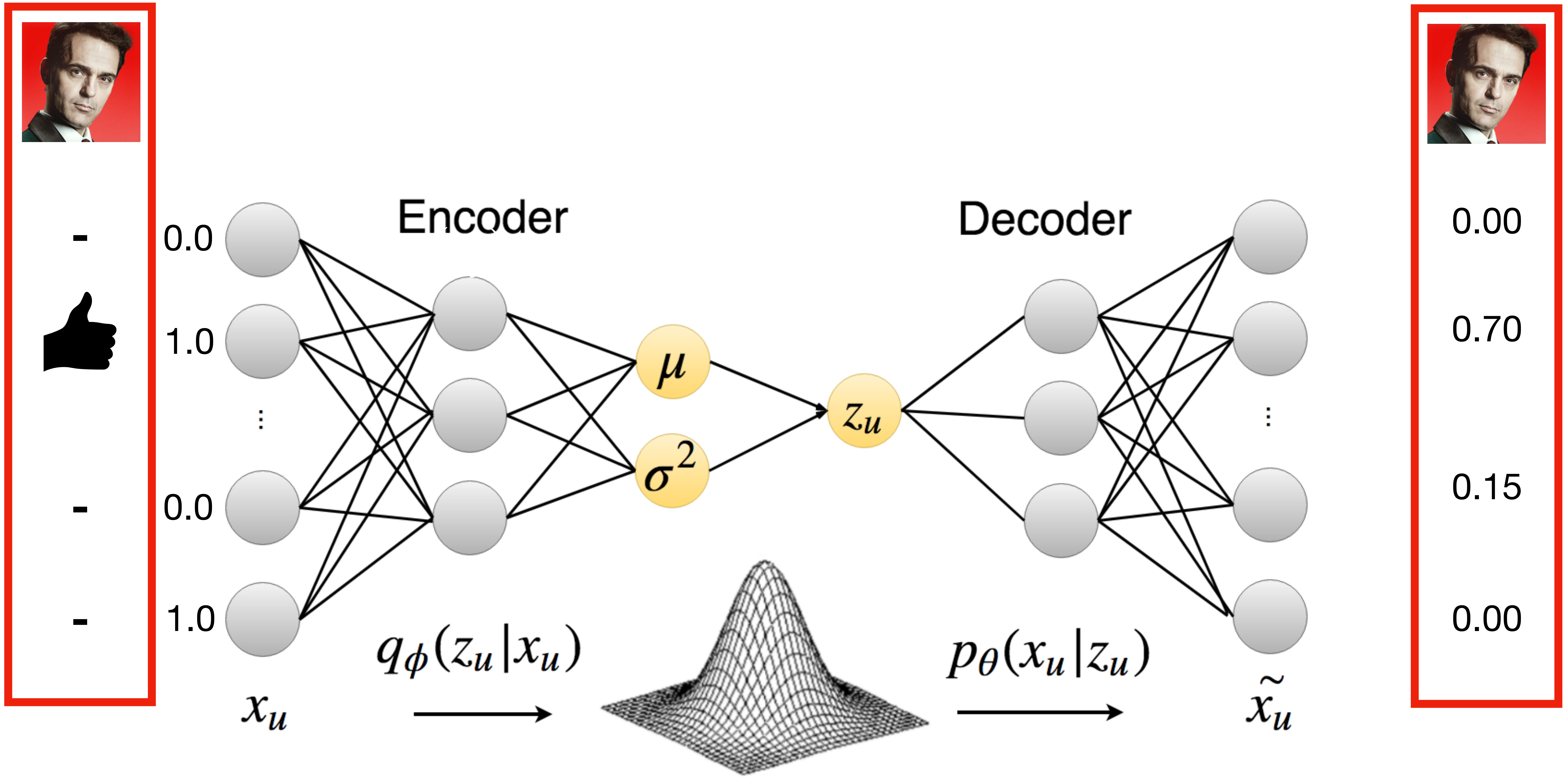
Neural Networks

Variational Autoencoders (VAEs)

(+) have larger modeling capacity
(+) generalize linear latent factor models

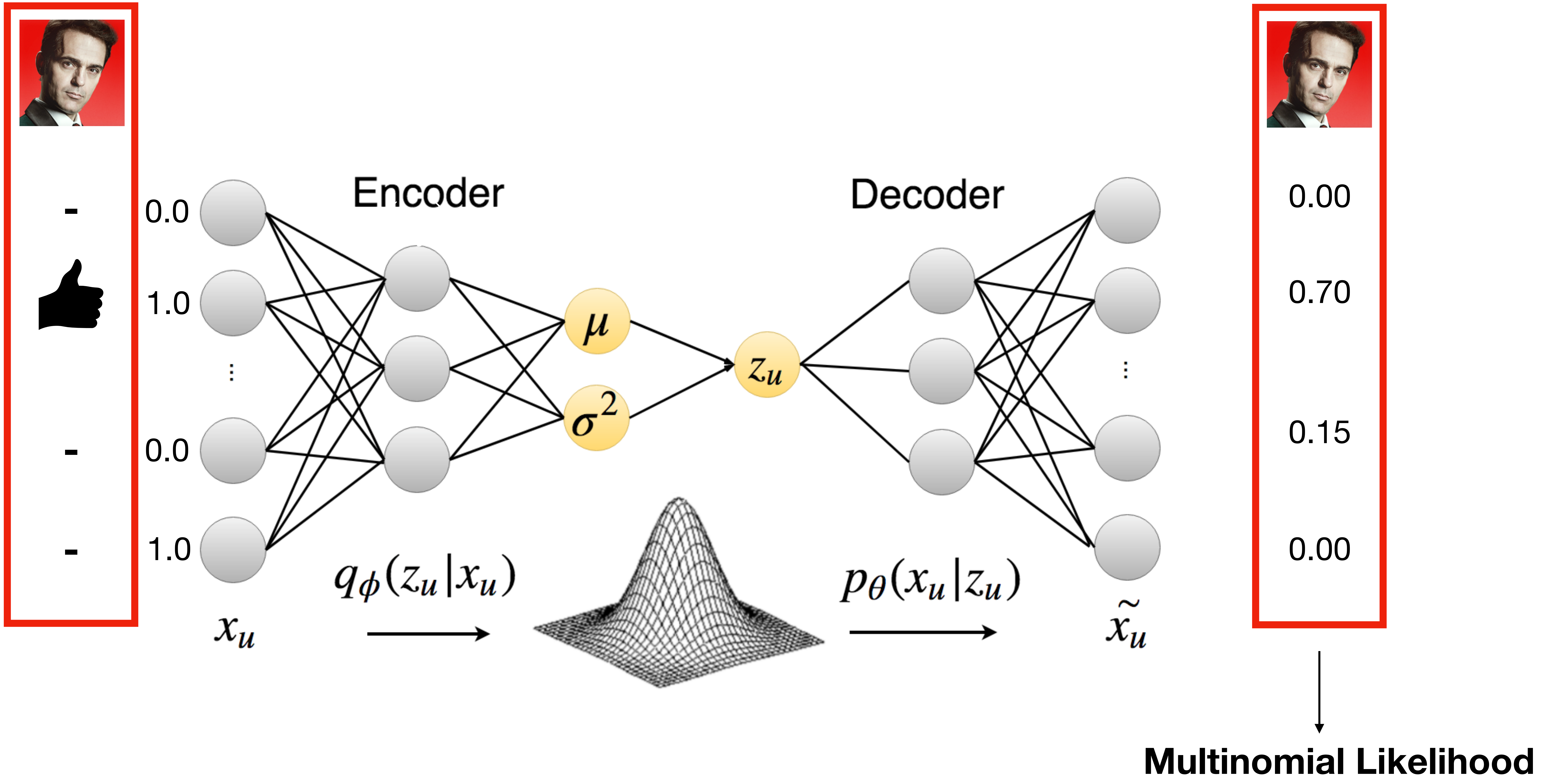
"Auto-encoding Variational Bayes" D. P. Kingma, M. Welling, ICLR 2014
"Variational Autoencoders for Collaborative Filtering" D. Liang, RG. Krishnan, MD. Hoffman, T. Jebara, WWW 2018

VAEs for Collaborative Filtering



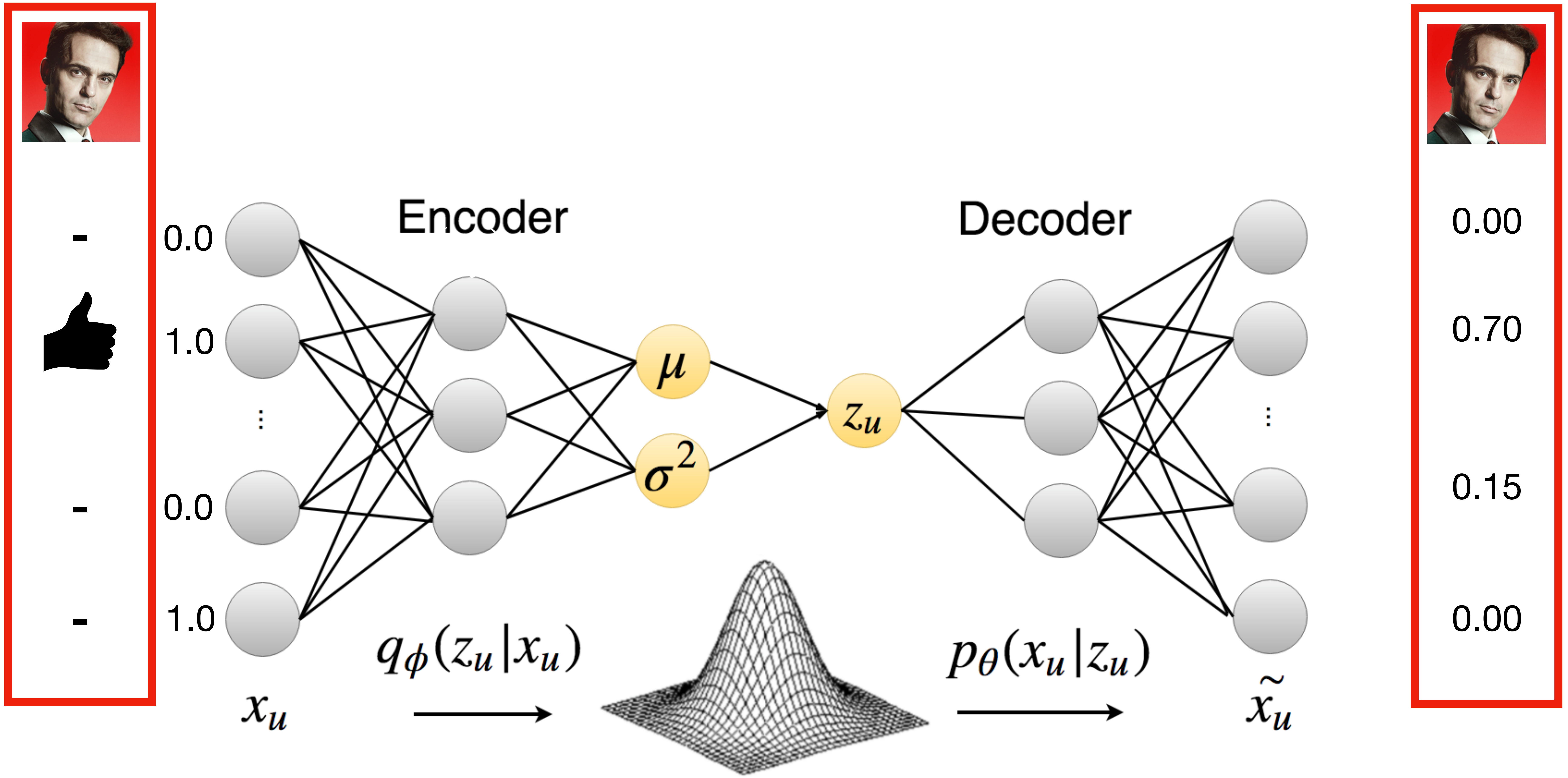
"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

VAEs for Collaborative Filtering



"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

VAEs for Collaborative Filtering



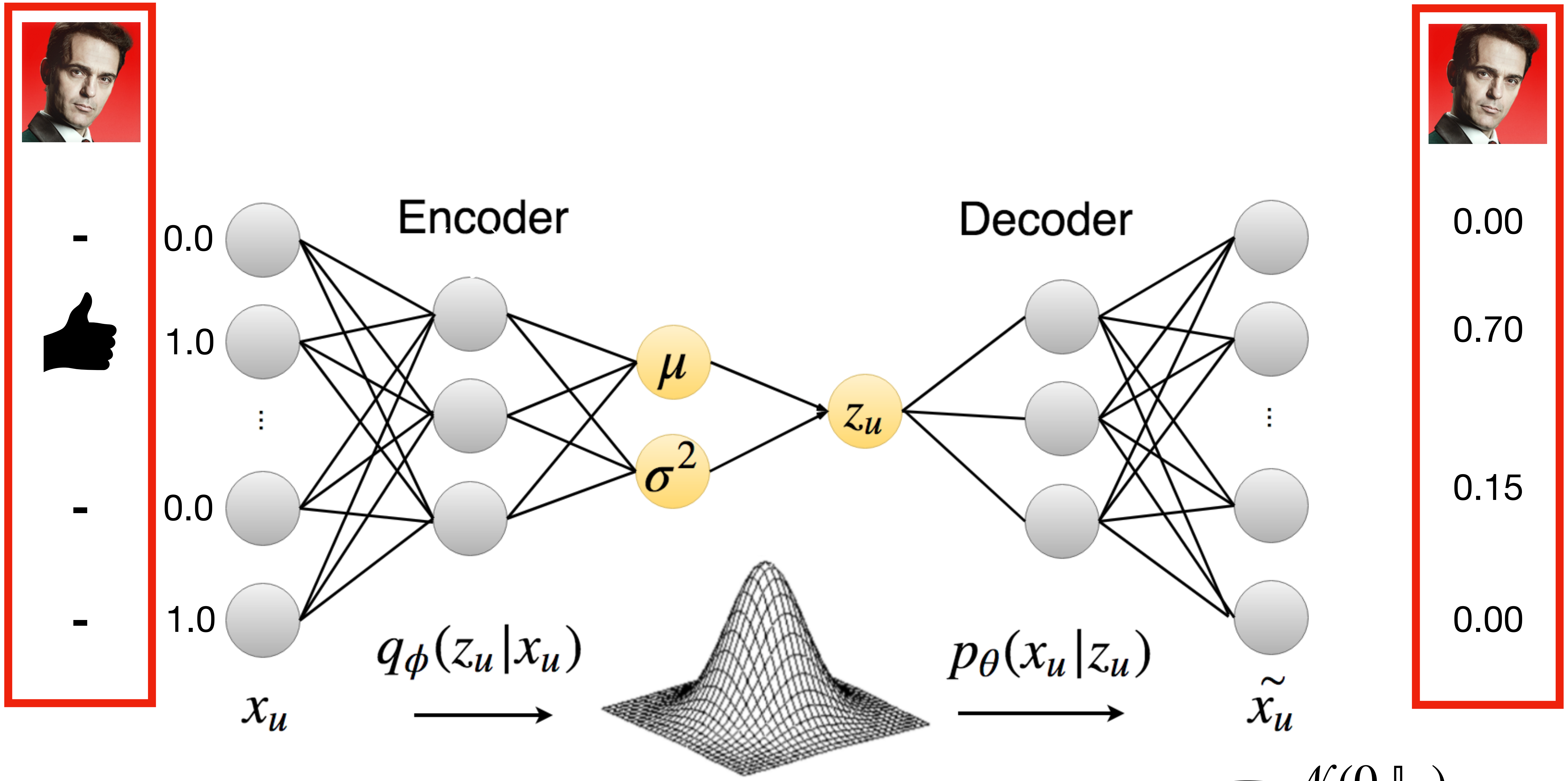
Training Objective: $\mathcal{L}_\beta \equiv \underline{\mathbb{E}_q[\log p_\theta(x_u|z_u)]} - \beta \cdot \underline{KL(q_\phi(z_u|x_u) \,||\, p(z_u))}$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

“Variational Autoencoders for Collaborative Filtering” D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

VAEs for Collaborative Filtering



$\mathcal{N}(0, \mathbb{I}_K)$

Training Objective: $\mathcal{L}_\beta \equiv \mathbb{E}_q[\log p_\theta(x_u|z_u)] - \beta \cdot KL(q_\phi(z_u|x_u) \,||\, p(z_u))$

(Negative) reconstruction error **Regularization term (Kullback-Leibler Divergence)**

"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

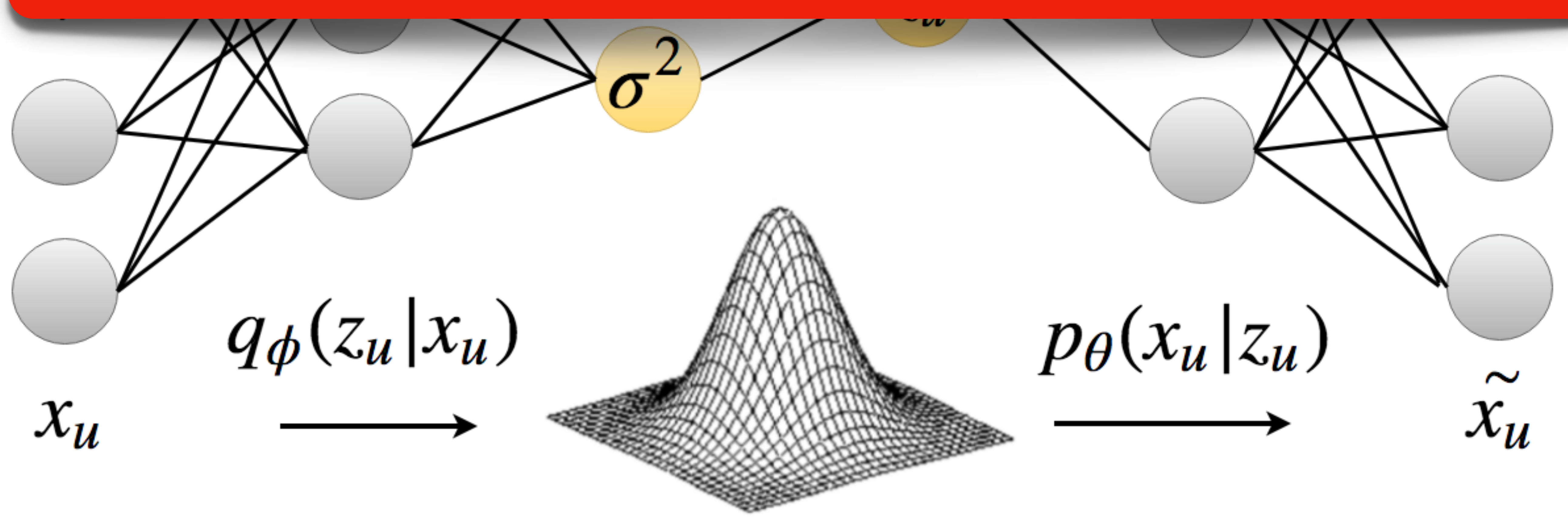
VAEs for Collaborative Filtering

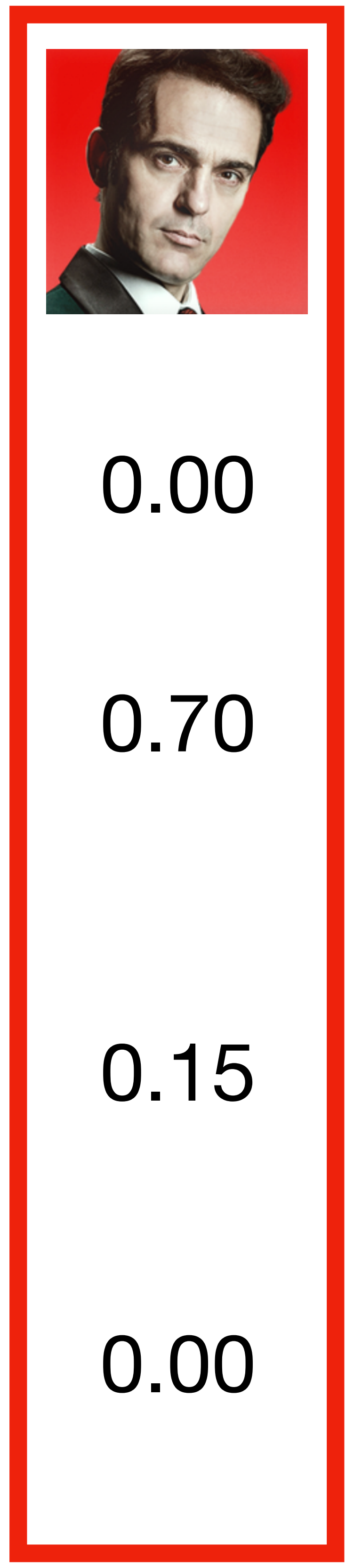
My variational posterior:

$$q_\phi(z_{u1} \,|\, x_{u1})$$

should be near my prior:

$$p(z_{u1}) \sim \mathcal{N}(0, \mathbb{I}_K)$$



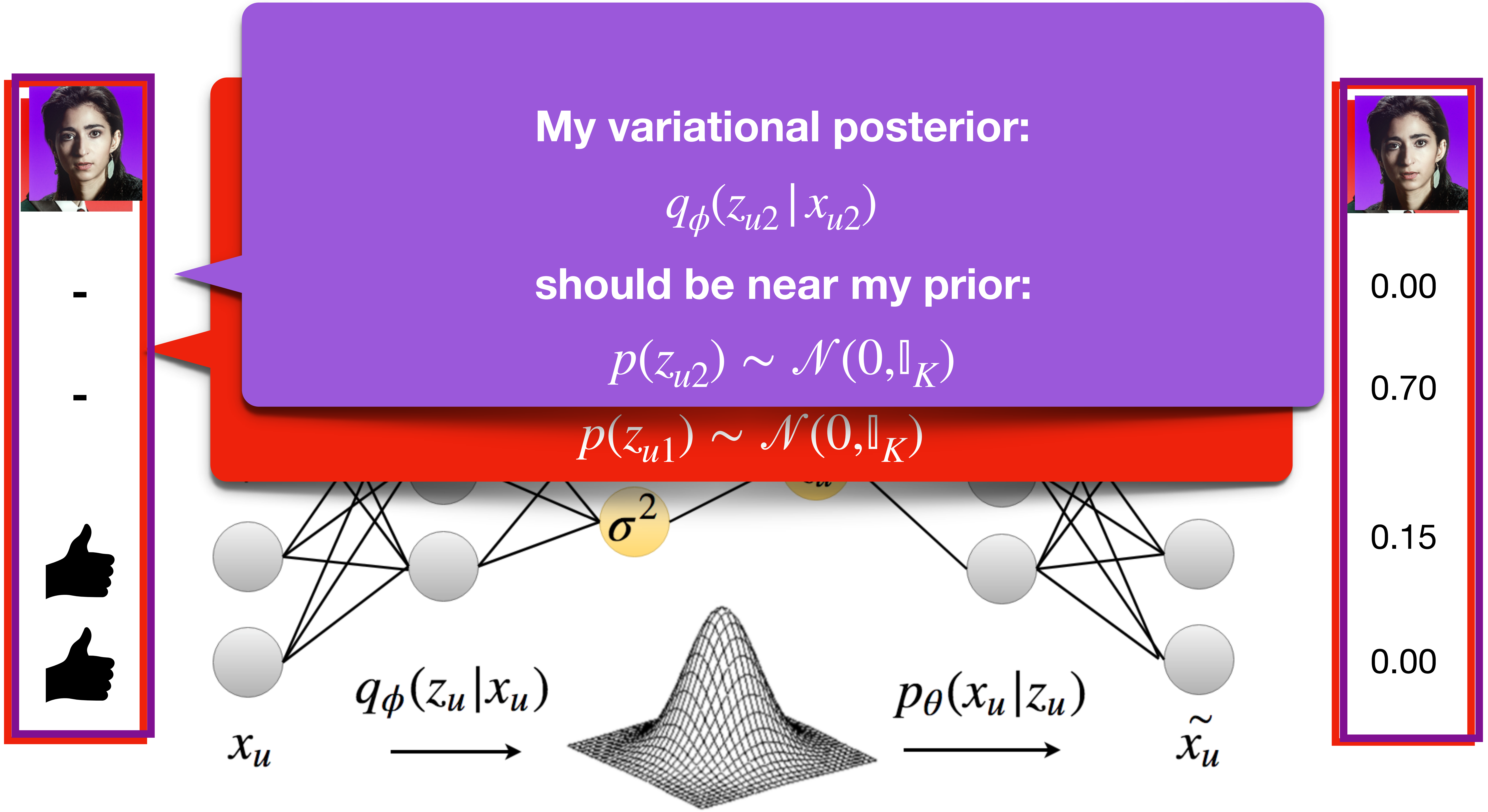


Training Objective: $\mathcal{L}_\beta \equiv \mathbb{E}_q[\log p_\theta(x_u|z_u)] - \beta \cdot KL(q_\phi(z_u|x_u) \,||\, p(z_u))$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

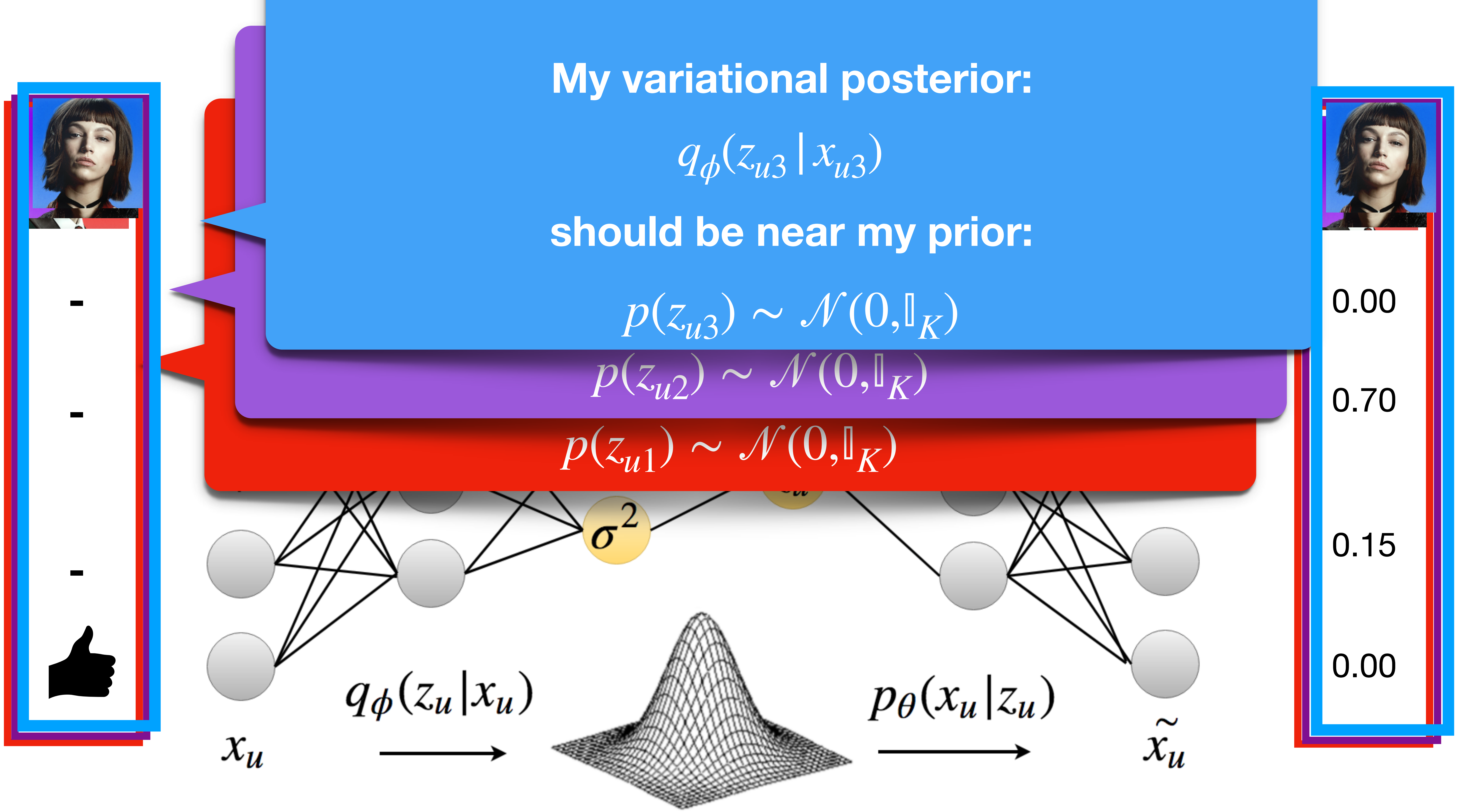


Training Objective: $\mathcal{L}_\beta \equiv \underbrace{\mathbb{E}_q[\log p_\theta(x_u|z_u)]}_{\text{(Negative) reconstruction error}} - \beta \cdot \underbrace{KL(q_\phi(z_u|x_u) \,||\, p(z_u))}_{\text{Regularization term (Kullback-Leibler Divergence)}}$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

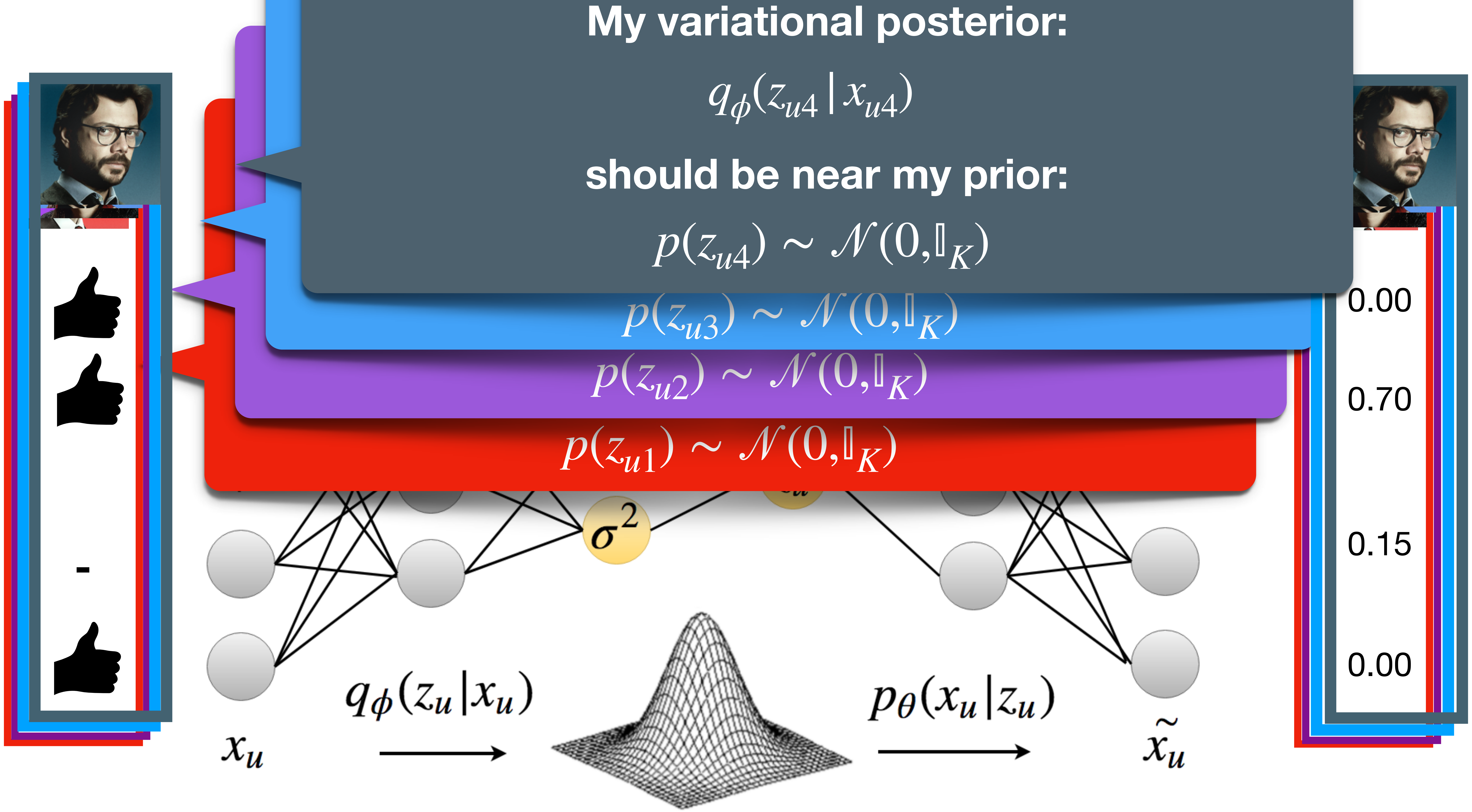


Training Objective: $\mathcal{L}_\beta \equiv \mathbb{E}_q[\log p_\theta(x_u|z_u)] - \beta \cdot KL(q_\phi(z_u|x_u) \,||\, p(z_u))$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

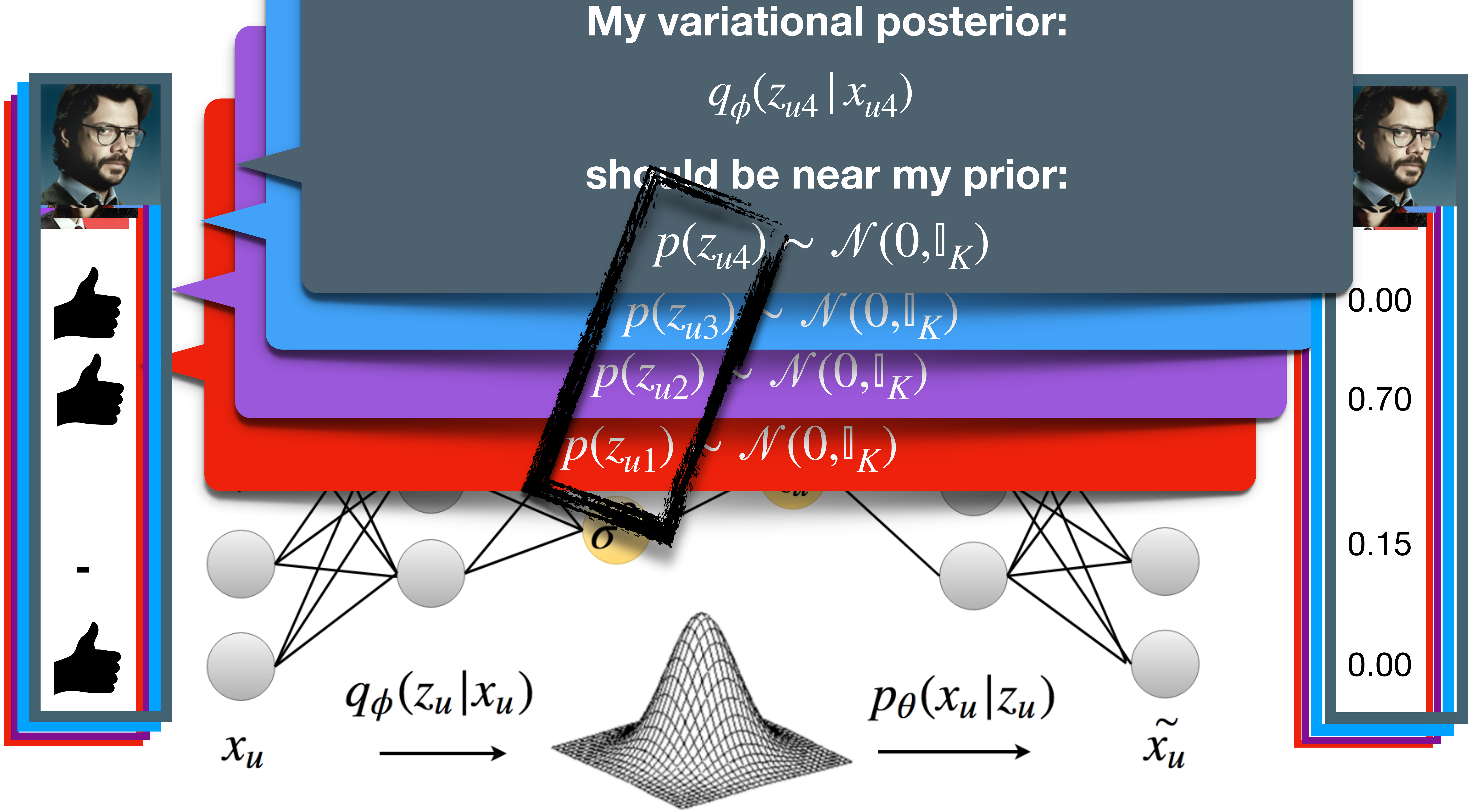


Training Objective: $\mathcal{L}_\beta \equiv \mathbb{E}_q[\log p_\theta(x_u|z_u)] - \beta \cdot KL(q_\phi(z_u|x_u) \,||\, p(z_u))$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

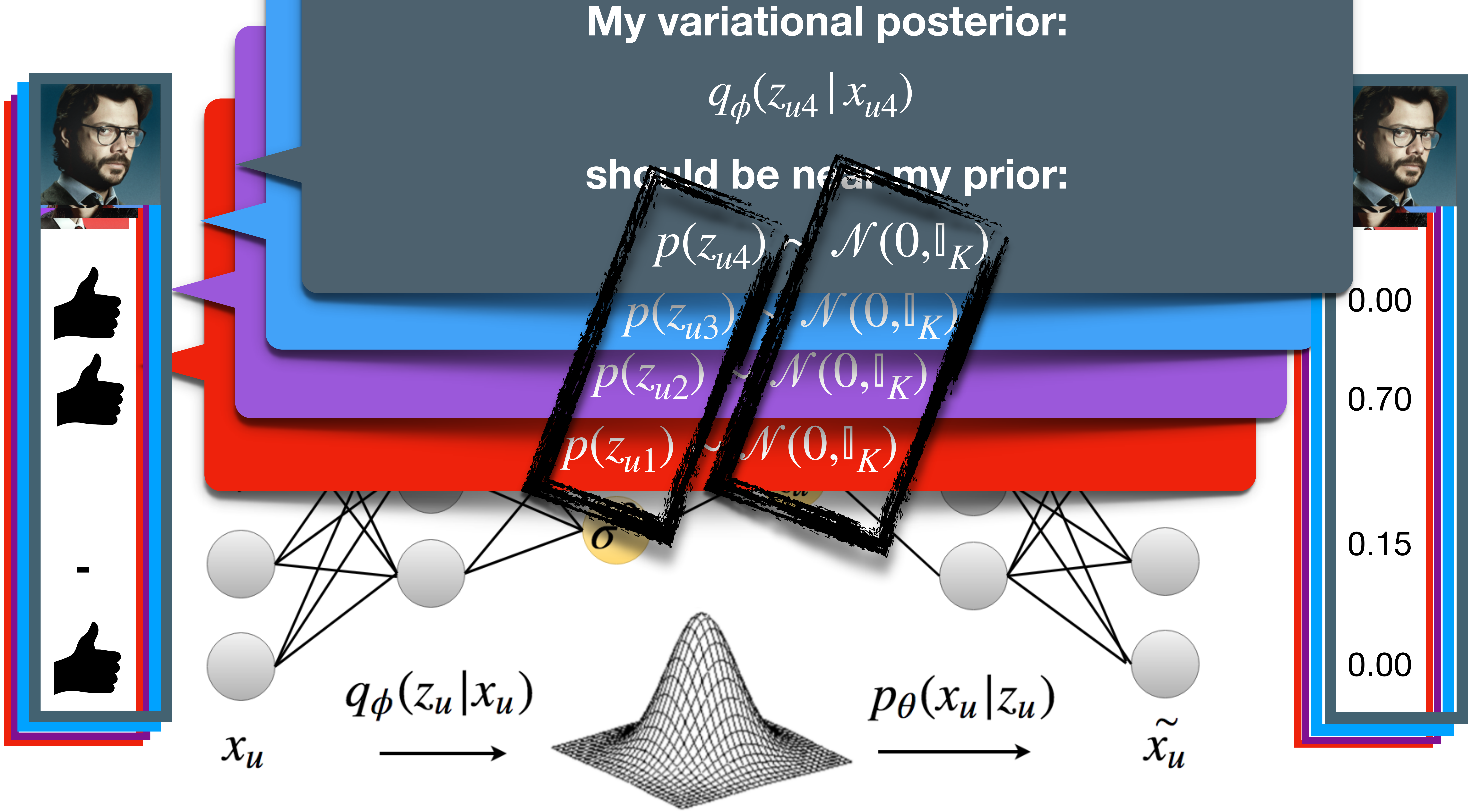


Training Objective: $\mathcal{L}_\beta \equiv \mathbb{E}_q[\log p_\theta(x_u|z_u)] - \beta \cdot KL(q_\phi(z_u|x_u) \; || \; p(z_u))$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

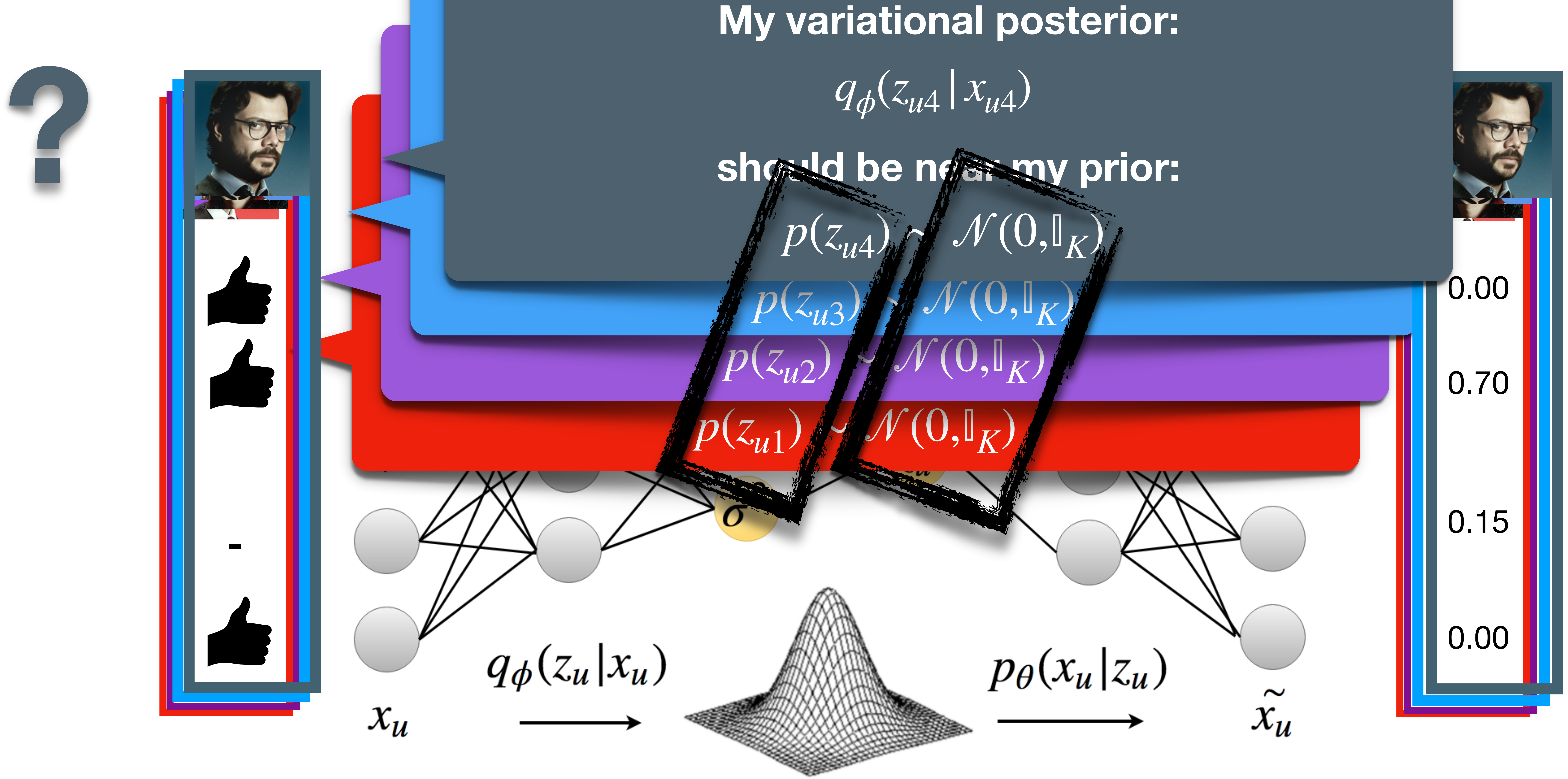
"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.



Training Objective: $\mathcal{L}_\beta \equiv \underbrace{\mathbb{E}_q[\log p_\theta(x_u|z_u)]} - \beta \cdot \underbrace{KL(q_\phi(z_u|x_u) \,||\, p(z_u))}$

(Negative) reconstruction error **Regularization term (Kullback-Leibler Divergence)**

"Variational Autoencoders for Collaborative Filtering" D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

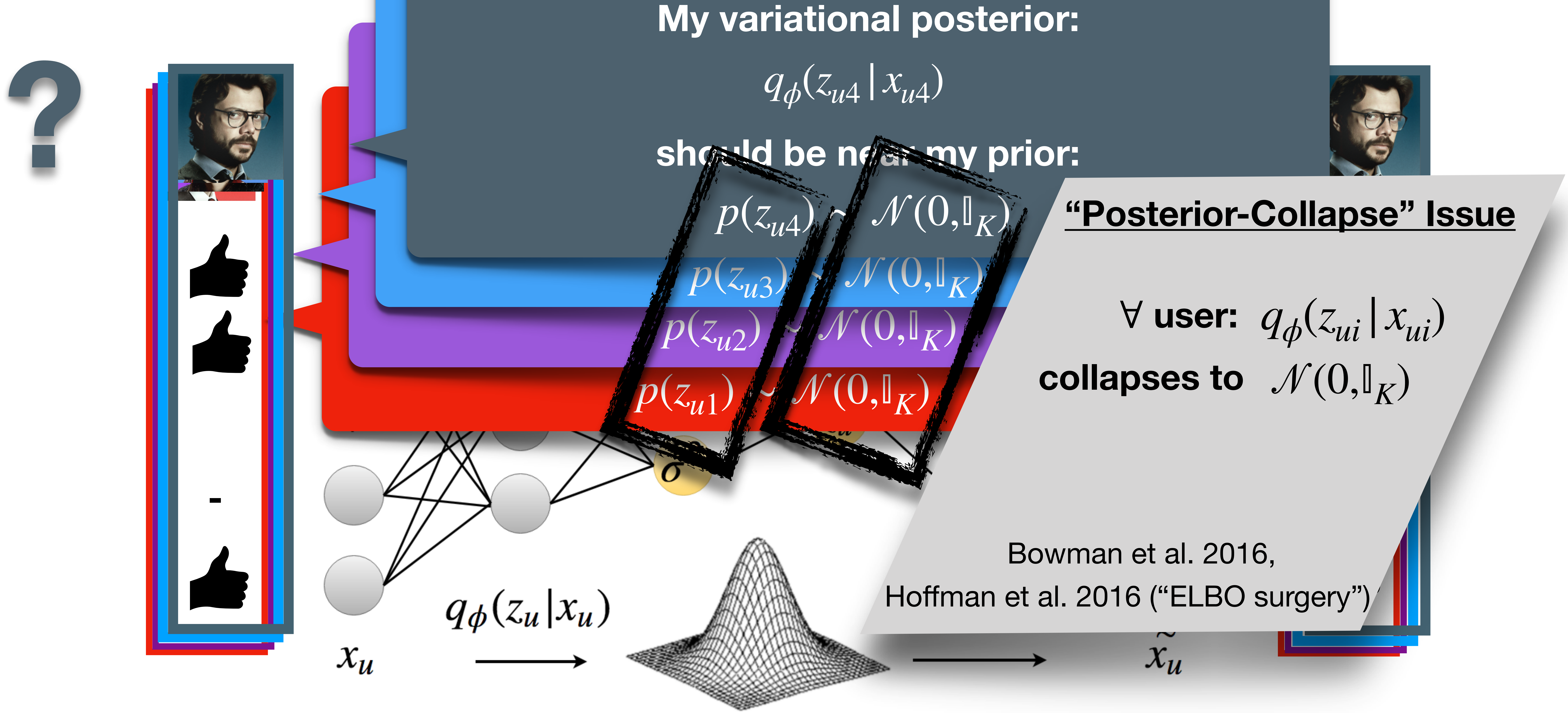


Training Objective: $\mathcal{L}_\beta \equiv \underline{\mathbb{E}_q[\log p_\theta(x_u|z_u)]} - \beta \cdot \underline{KL(q_\phi(z_u|x_u) \,||\, p(z_u))}$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

“Variational Autoencoders for Collaborative Filtering” D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

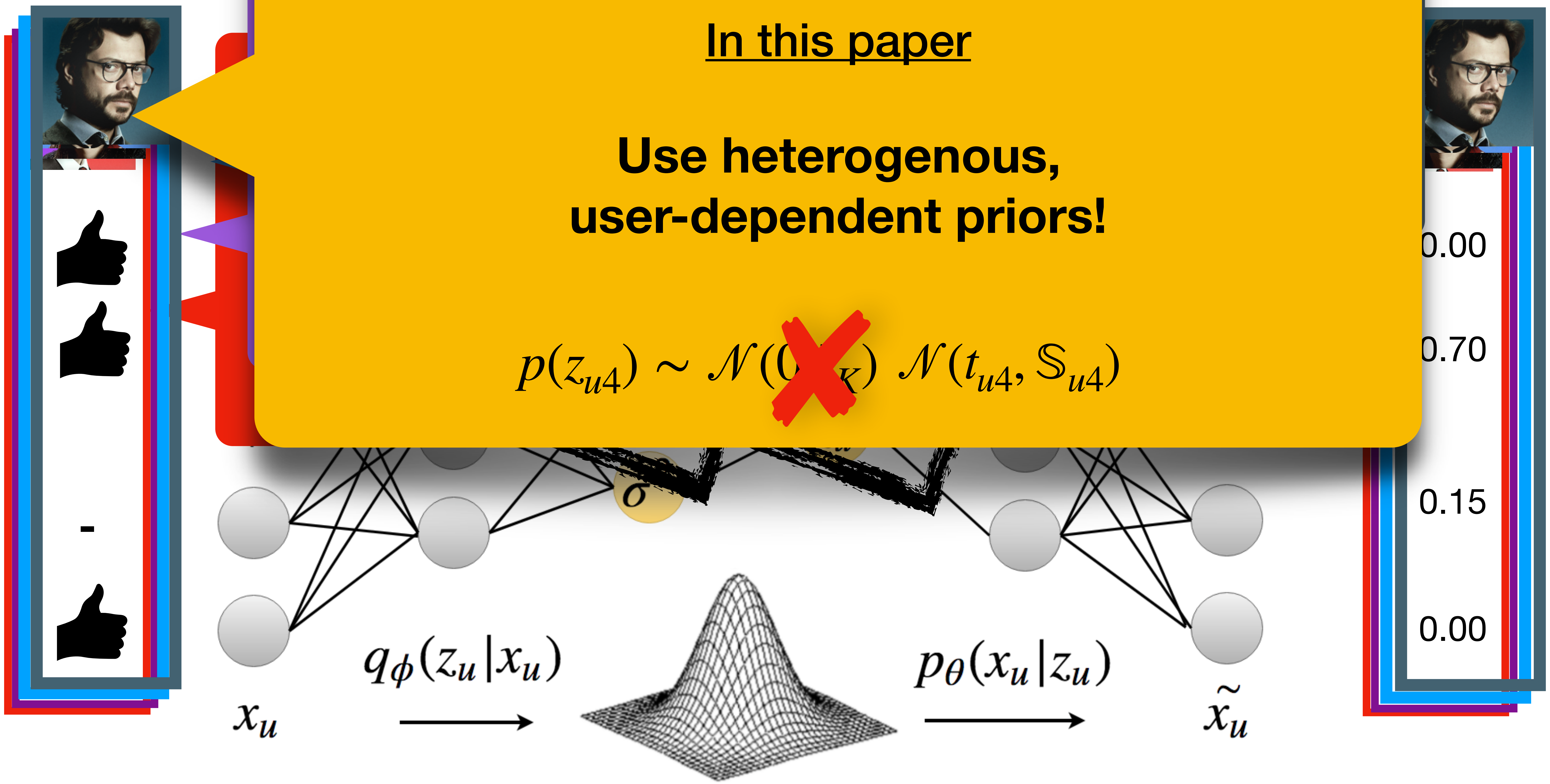


Training Objective: $\mathcal{L}_\beta \equiv \underbrace{\mathbb{E}_q[\log p_\theta(x_u|z_u)]}_{} - \beta \cdot \underbrace{KL(q_\phi(z_u|x_u) \,||\, p(z_u))}_{}$

(Negative) reconstruction error

Regularization term (Kullback-Leibler Divergence)

“Variational Autoencoders for Collaborative Filtering” D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.



<u>In this paper</u>

Use heterogenous, user-dependent priors!

$$p(z_{u4}) \sim \cancel{\mathcal{N}(0, I_K)}\ \mathcal{N}(t_{u4}, \mathbb{S}_{u4})$$

Training Objective: $\mathcal{L}_\beta \equiv \underbrace{\mathbb{E}_q[\log p_\theta(x_u|z_u)]}_{\text{(Negative) reconstruction error}} - \beta \cdot \underbrace{KL(q_\phi(z_u|x_u)\,||\,p(z_u))}_{\text{Regularization term (Kullback-Leibler Divergence)}}$

(Negative) reconstruction error **Regularization term (Kullback-Leibler Divergence)**

“Variational Autoencoders for Collaborative Filtering” D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, WWW 2018.

This work: Item Recommendation with VAEs and Heterogenous Priors

Using heterogenous, user-dependent priors

- For each user u , we replace $z_u \sim \mathcal{N}(0, \mathbb{I}_K)$ by $z_u \sim \mathcal{N}(t_u, \mathbb{S}_u)$
  - Prior parameters $(t_u, \mathbb{S}_u)$ encode user preferences
  - Explicitly encourage user diversity in latent VAE space

$t_u \in \mathbb{R}^K$

$\mathbb{S}_u \in \mathbb{R}^{K \times K}$

This work: Item Recommendation with VAEs and Heterogenous Priors

Using heterogenous, user-dependent priors

- For each user u , we replace $z_u \sim \mathcal{N}(0, \mathbb{I}_K)$ by $z_u \sim \mathcal{N}(t_u, \mathbb{S}_u)$,
  - Prior parameters $(t_u, \mathbb{S}_u)$ encode user preferences
  - Explicitly encourage user diversity in latent VAE space

$t_u \in \mathbb{R}^K$

$\mathbb{S}_u \in \mathbb{R}^{K \times K}$

?

How can you encode my preferences?

Have I revealed them?

This work: Item Recommendation with VAEs and Heterogenous Priors

Using heterogenous, user-dependent priors

- For each user u, we replace $z_u \sim \mathcal{N}(0, \mathbb{I}_K)$ by $z_u \sim \mathcal{N}(t_u, \mathbb{S}_u)$.
  - Prior parameters $(t_u, \mathbb{S}_u)$ encode user preferences
  - Explicitly encourage user diversity in latent VAE space

$t_u \in \mathbb{R}^K$

$\mathbb{S}_u \in \mathbb{R}^{K \times K}$

?

How can you encode my preferences?

Have I revealed them?

Yes, you have!

By writing reviews…

Item Recommendation with VAEs and Heterogenous Priors

Users reveal their preferences in text reviews

Item Recommendation with VAEs and Heterogenous Priors

Users reveal their preferences in text reviews


Pascale H.
Elmont, NY
1 friend
11 reviews
2 photos

3/28/2018
1 check-in

The burgers here are really good and if you're gluten free they offer a lettuce bun instead of the potato bun. As for sides,I'm not in love with the fries and the onion rings. The portion size is good an large enough to share. My friend really enjoyed her milkshake.

Yelp Review

Item Recommendation with VAEs and Heterogenous Priors

Users reveal their preferences in text reviews

Pascale H.
Elmont, NY
1 friend
11 reviews
2 photos

3/28/2018
1 check-in

The burgers here are really good and if you're gluten free they offer a lettuce bun instead of the potato bun. As for sides,I'm not in love with the fries and the onion rings. The portion size is good an large enough to share. My friend really enjoyed her milkshake.

Yelp Review

10/10

One of the best I've seen!
amytgr-1 1 January 2018

This has been a real treat! An amazing series, great acting, direction and such a suspenseful story it's really one of the very best I've seen ever. I love heist movies and I just found this one in Netflix and I literally couldn't stop watching through the night. The characters are simply amazing! Don't miss this!

IMDB Review

Item Recommendation with VAEs and Heterogenous Priors

Users reveal their preferences in text reviews

Pascale H.
Elmont, NY
1 friend
11 reviews
2 photos

3/28/2018
1 check-in

The burgers here are really good and if you're gluten free they offer a lettuce bun instead of the potato bun. As for sides,I'm not in love with the fries and the onion rings. The portion size is good an large enough to share. My friend really enjoyed her milkshake.

likes burgers

cares about portion size

Yelp Review

10/10

One of the best I've seen!
amytgr-1 1 January 2018

This has been a real treat! An amazing series, great acting, direction and such a suspenseful story it's really one of the very best I've seen ever. I love heist movies and I just found this one in Netflix and I literally couldn't stop watching through the night. The characters are simply amazing! Don't miss this!

likes heist movies

likes suspense

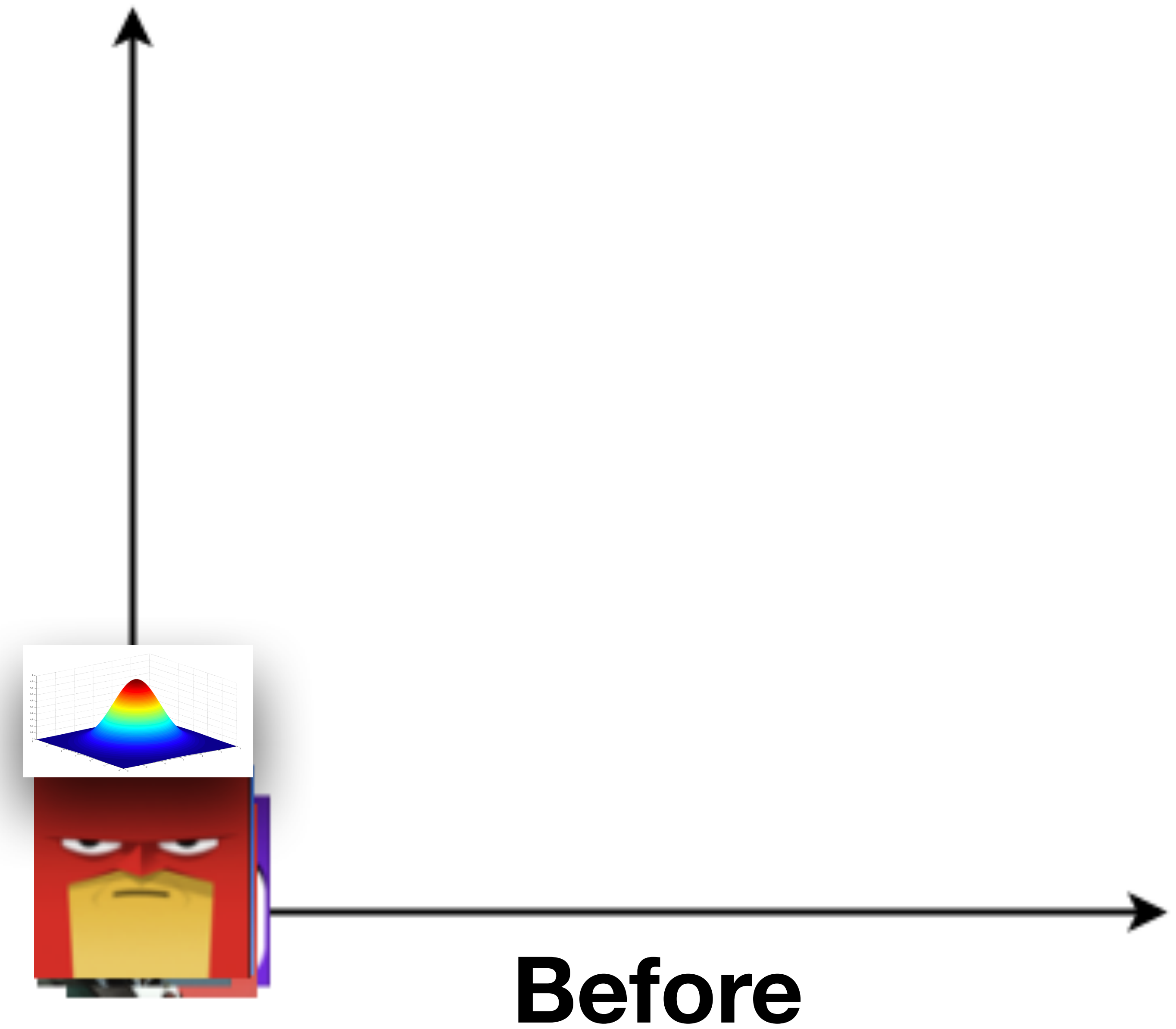
IMDB Review

Item Recommendation with VAEs and Heterogenous Priors

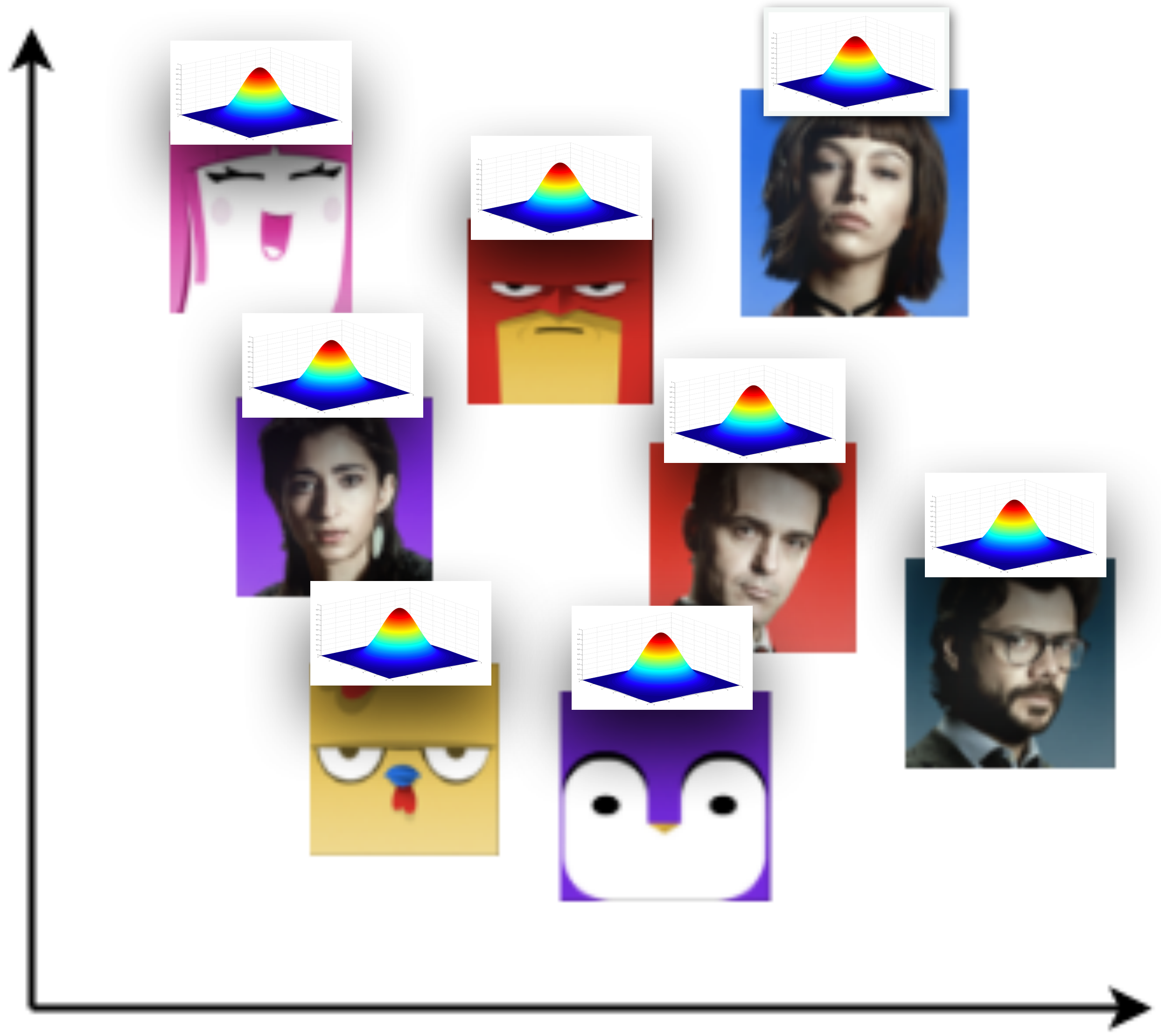
Encoding user preferences from text (2 methods):

- t_u, $\mathbb{S}_u$: functions of the user's review text

$$z_u \sim \mathcal{N}(t_u, \mathbb{S}_u)$$
$$t_u \in \mathbb{R}^K$$
$$\mathbb{S}_u \in \mathbb{R}^{K \times K}$$



Before



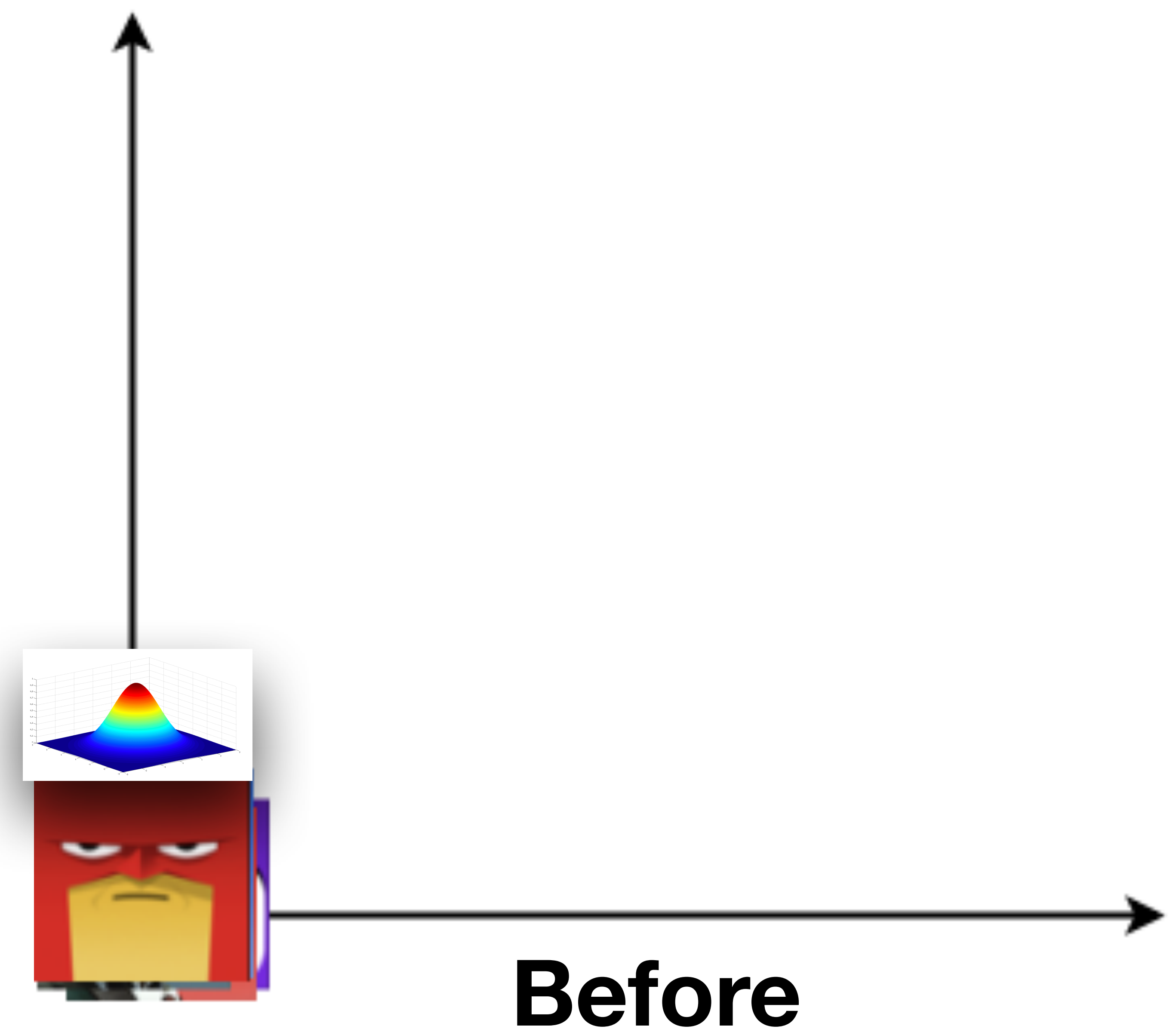
After

Item Recommendation with VAEs and Heterogenous Priors

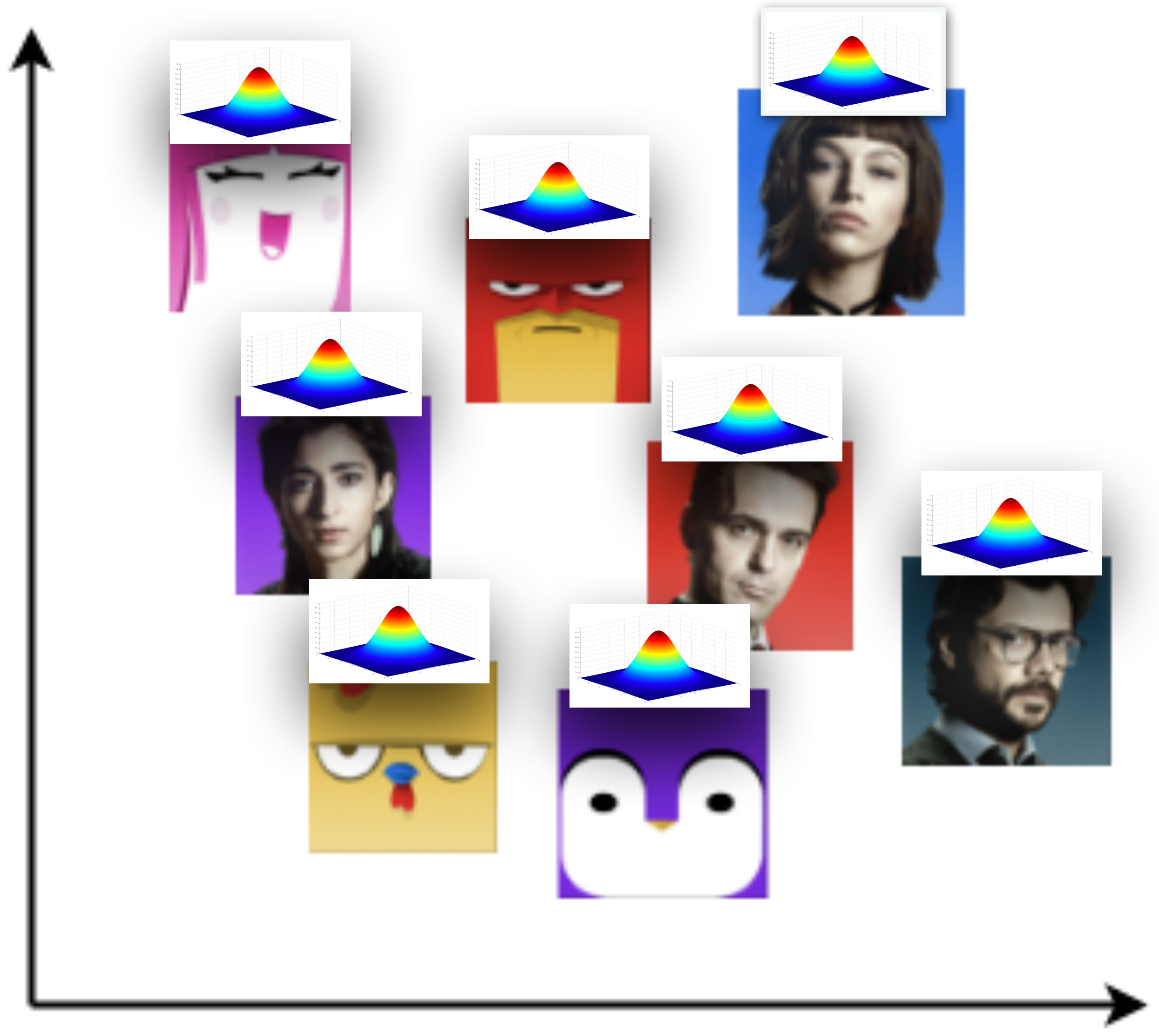
Encoding user preferences from text (2 methods):

- t_u, $\mathbb{S}_u$: functions of the user's review text
  - **Method 1:** Word Embeddings (word2vec)
  - **Method 2:** Probabilistic Topic Models (Latent Dirichlet Allocation - LDA)

$$z_u \sim \mathcal{N}(t_u, \mathbb{S}_u),$$
$$t_u \in \mathbb{R}^K$$
$$\mathbb{S}_u \in \mathbb{R}^{K \times K}$$



Before



After

Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 1: Word Embeddings (word2vec)**

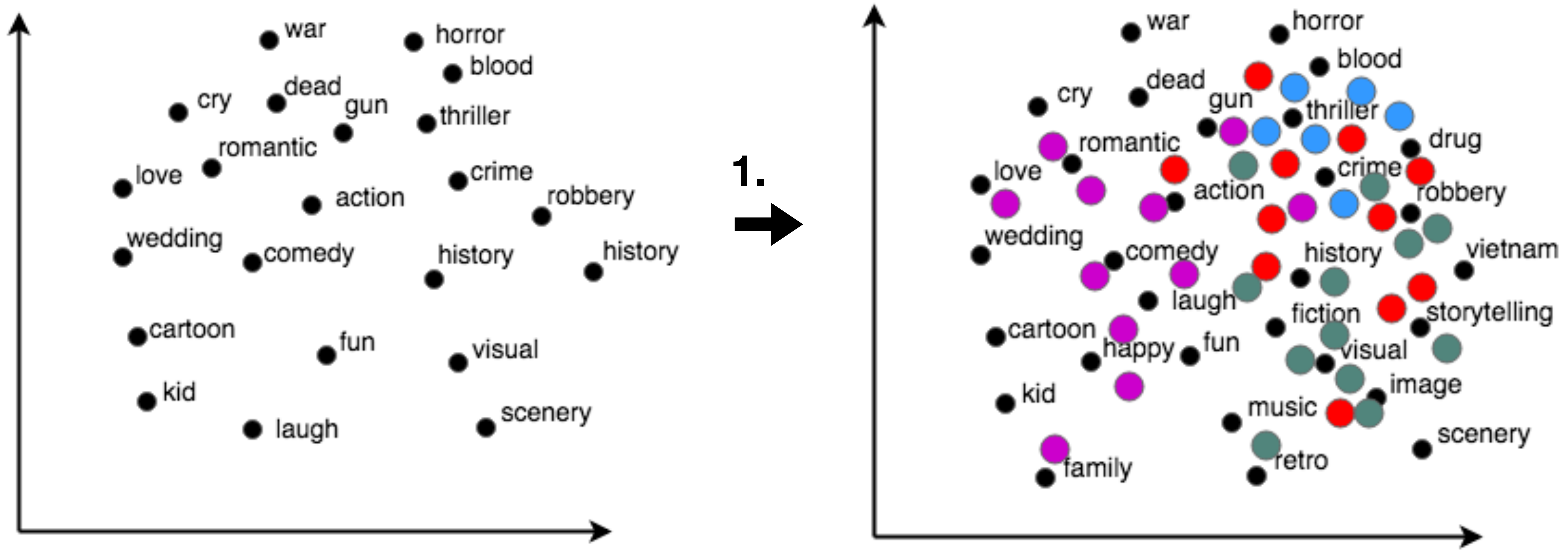
K**-dim Word Embeddings**

Mikolov et. al. 2013: "Efficient estimation of word representations in vector space"

Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 1: Word Embeddings (word2vec)**
    1. Create **review** embeddings: avg of word embeddings



K-dim Word Embeddings

Review Embeddings

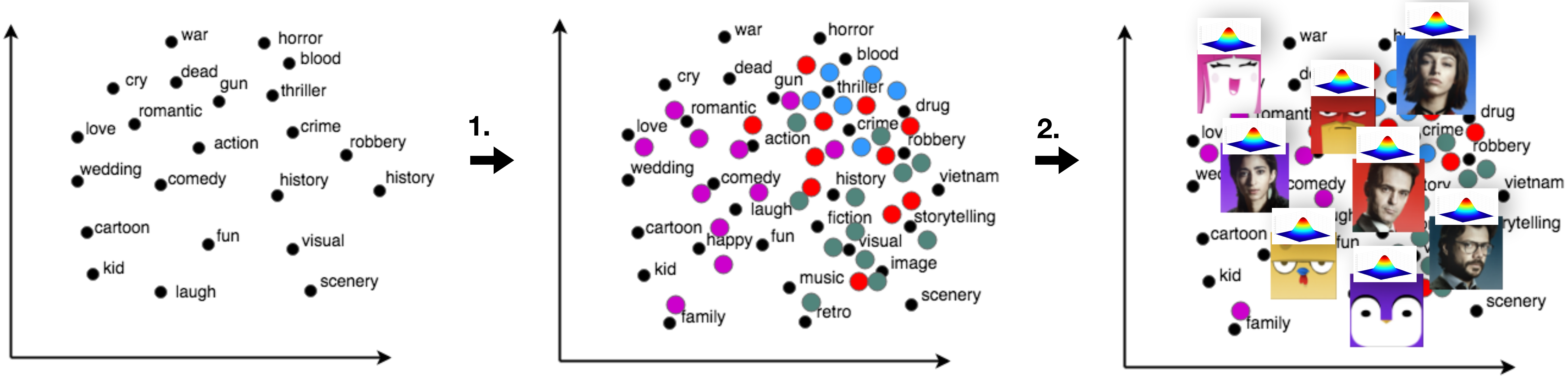
Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 1: Word Embeddings (word2vec)**
    1. Create **review** embeddings: avg of word embeddings
    2. Represent each **user**: Gaussian distribution $z_u \sim \mathcal{N}(t_u, \mathbb{S}_u)$
        - t_u : avg of review embeddings (written by u)
        - $\mathbb{S}_u$: diagonal covariance matrix
          diagonal values $s_1, \ldots, s_K \in \mathbb{R}$: std of review embeddings

$t_u \in \mathbb{R}^K$

$\mathbb{S}_u \in \mathbb{R}^{K \times K}$



K-dim Word Embeddings | Review Embeddings | User Distributions

Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 2: Probabilistic Topic Models (Latent Dirichlet Allocation - LDA)**

Topic 1	Topic 2	…	Topic K
horror	romance		action
blood	love		robbery
crime	kiss		kill
gun	wedding		police
…	…		…

David Blei, Andrew Y Ng, Michael I Jordan, 2003: "Latent Dirichlet Allocation"

Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 2: Probabilistic Topic Models (Latent Dirichlet Allocation - LDA)**
    1. Train LDA to extract K topics

Topic 1	Topic 2	…	Topic K
horror	romance		action
blood	love		robbery
crime	kiss		kill
gun	wedding		police
…	…		…

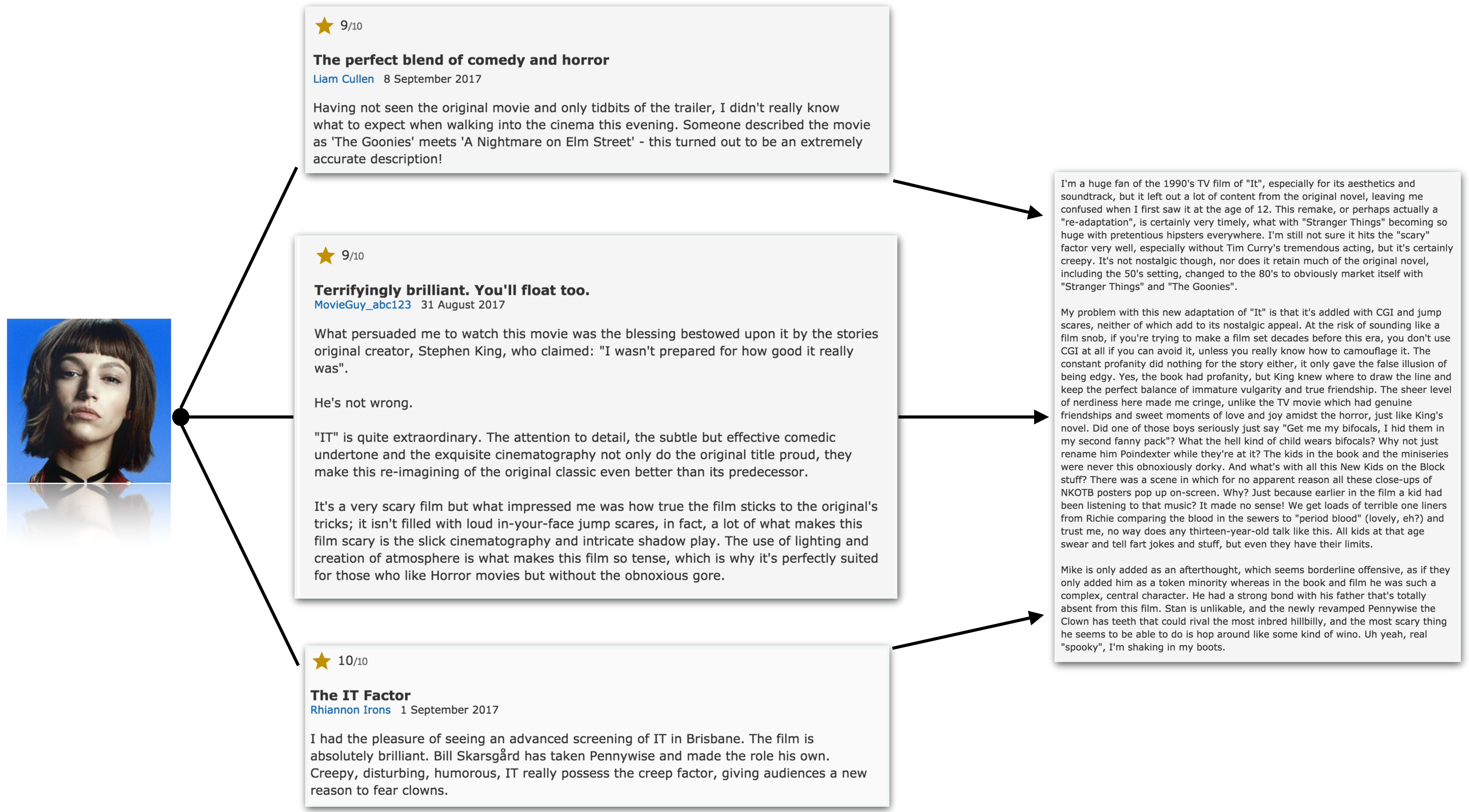
David Blei, Andrew Y Ng, Michael I Jordan, 2003: "Latent Dirichlet Allocation"

Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 2: Probabilistic Topic Models (Latent Dirichlet Allocation - LDA)**
    1. Train LDA to extract K topics
    2. For each user u :
        - 2a: concatenate all of the user's reviews in one document

<u>Topic 1</u>	<u>Topic 2</u>	⋯	<u>Topic K</u>
horror	romance		action
blood	love		robbery
crime	kiss		kill
gun	wedding		police
…	…		…

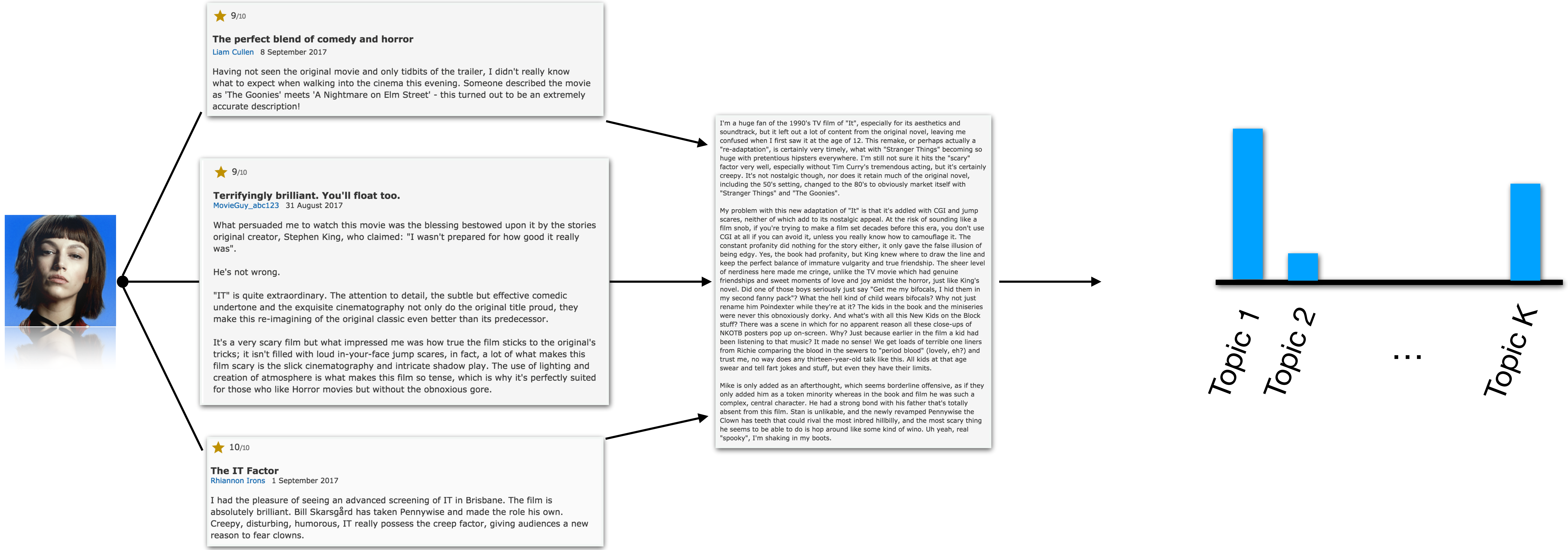


Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 2: Probabilistic Topic Models (Latent Dirichlet Allocation - LDA)**
    1. Train LDA to extract K topics
    2. For each user u :
        - 2a: concatenate all of the user's reviews in one document
        - 2b: represent u as the distribution over the K topics

Topic 1	**Topic 2**	...	**Topic K**
horror	romance		action
blood	love		robbery
crime	kiss		kill
gun	wedding		police
...	...		...

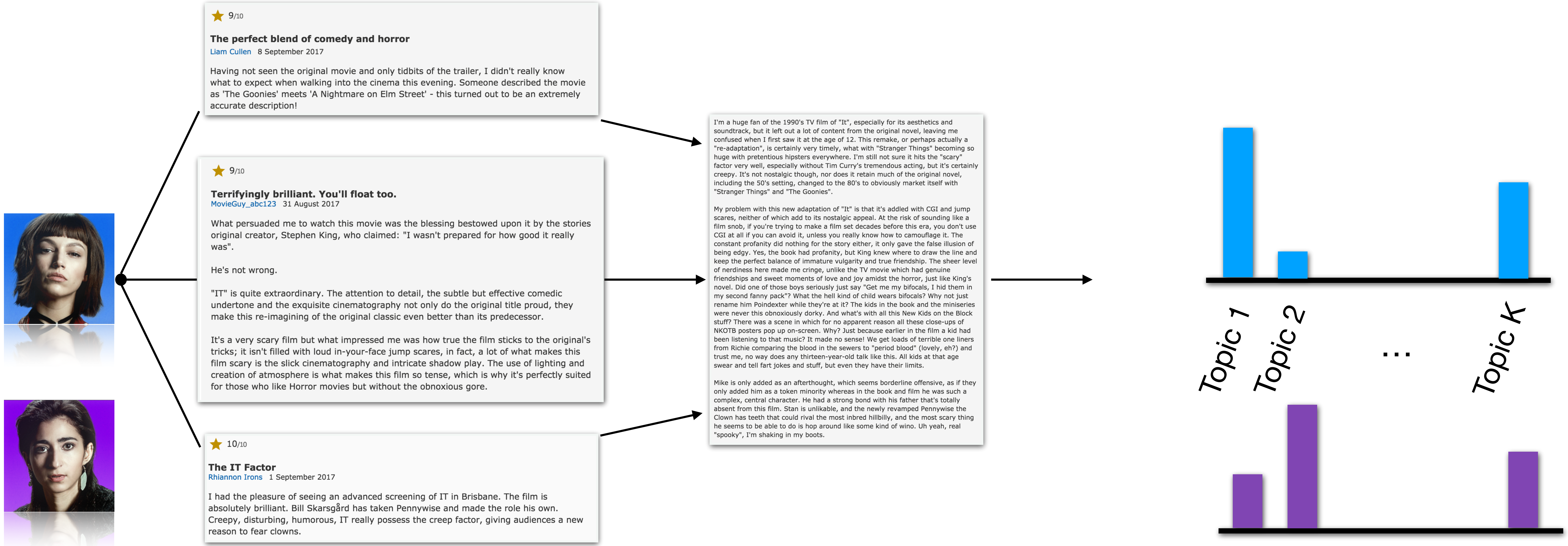


Item Recommendation with VAEs and Heterogenous Priors

Encoding user preferences from text:

- **Method 2: Probabilistic Topic Models (Latent Dirichlet Allocation - LDA)**
    1. Train LDA to extract K topics
    2. For each user u :
        - 2a: concatenate all of the user's reviews in one document
        - 2b: represent u as the distribution over the K topics

<u>Topic 1</u>	<u>Topic 2</u>	…	<u>Topic K</u>
horror	romance		action
blood	love		robbery
crime	kiss		kill
gun	wedding		police
…	…		…



Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline

Online Reviews

Item Recommendation with VAEs and Heterogenous Priors

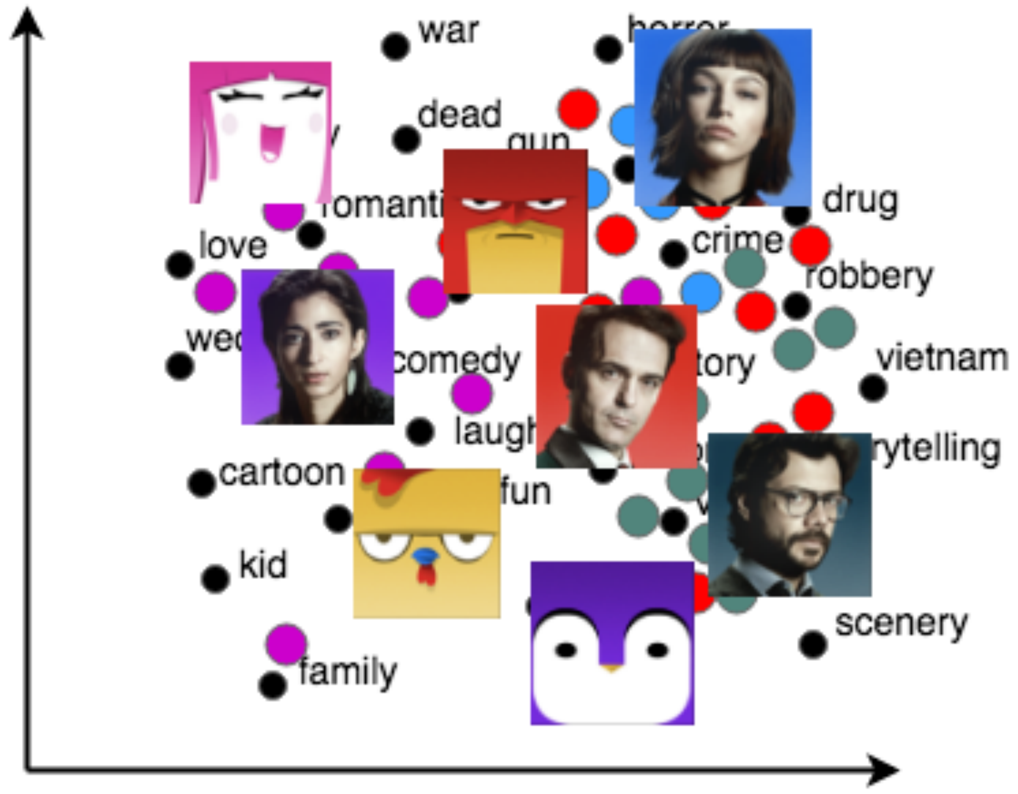
Item Recommendation Pipeline

Online Reviews

$\xrightarrow{\text{Text}}$

1. word2vec
2. LDA

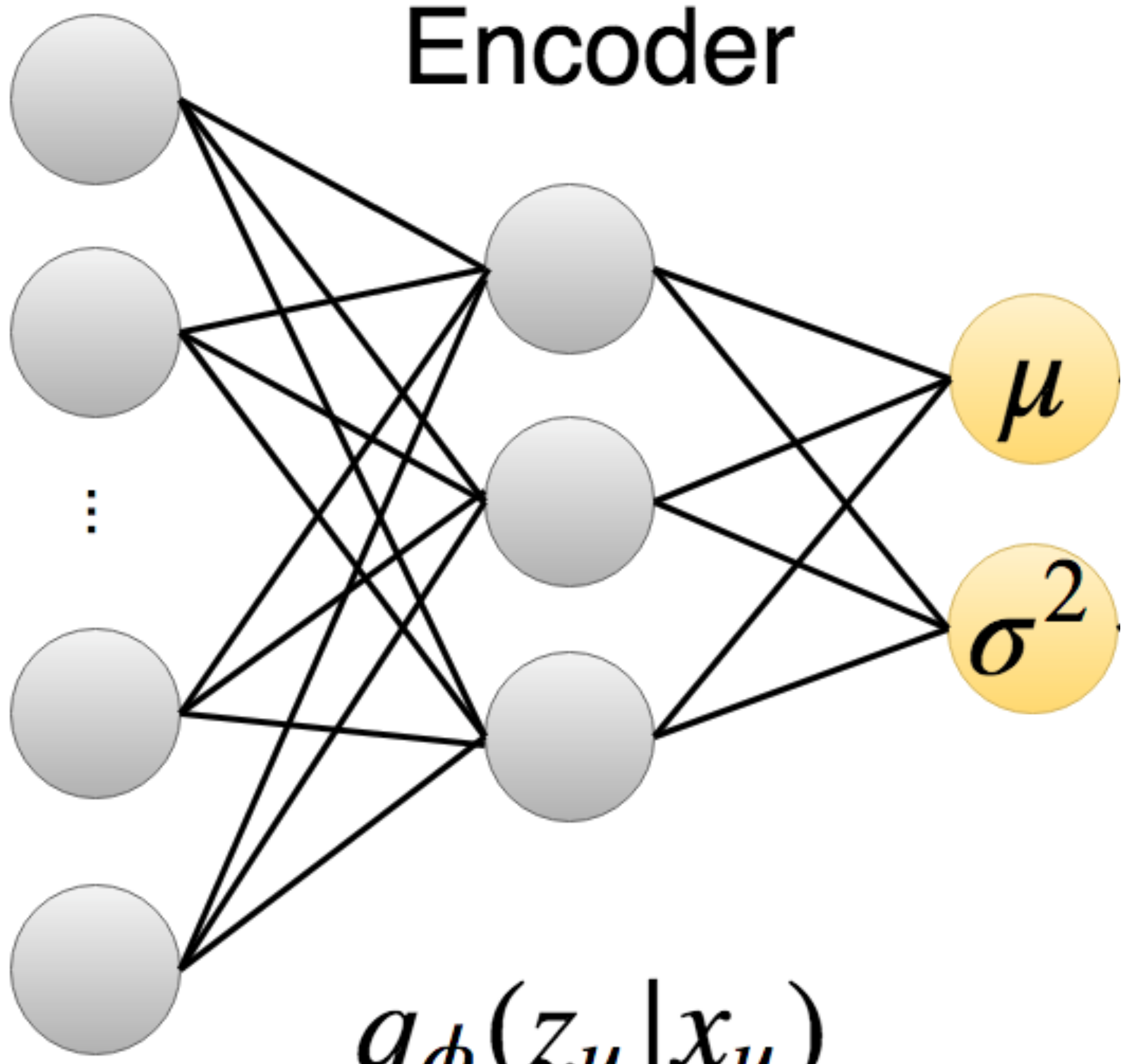
$p(z_u)$



Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline

Encoder



$q_\phi(z_u|x_u)$

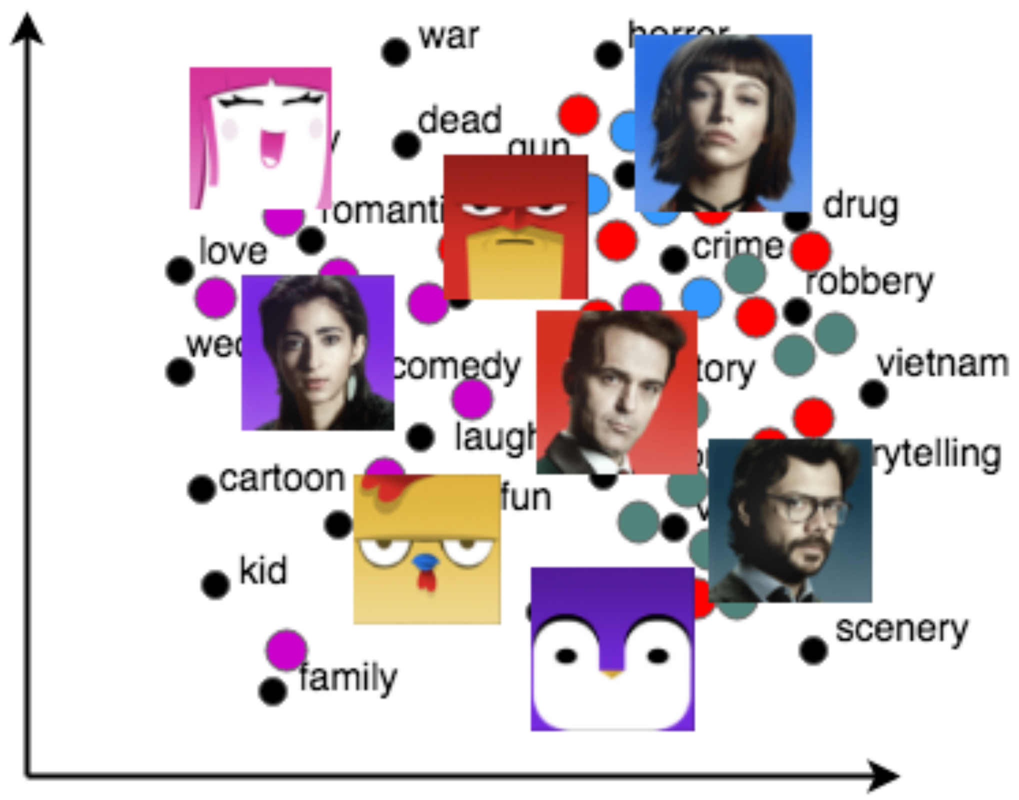
x_u

$\longrightarrow$

Ratings

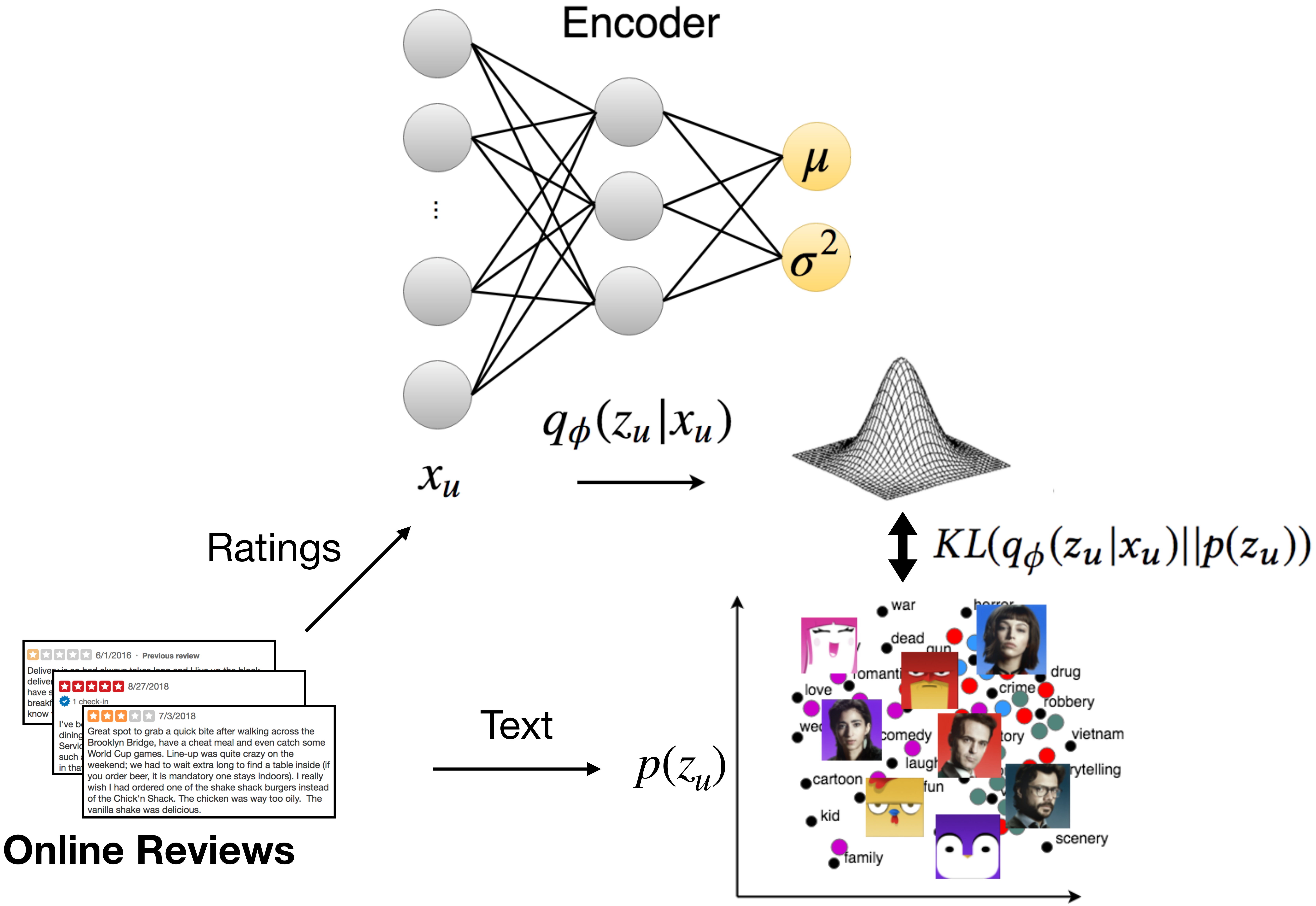
Online Reviews

Text $\longrightarrow$ $p(z_u)$



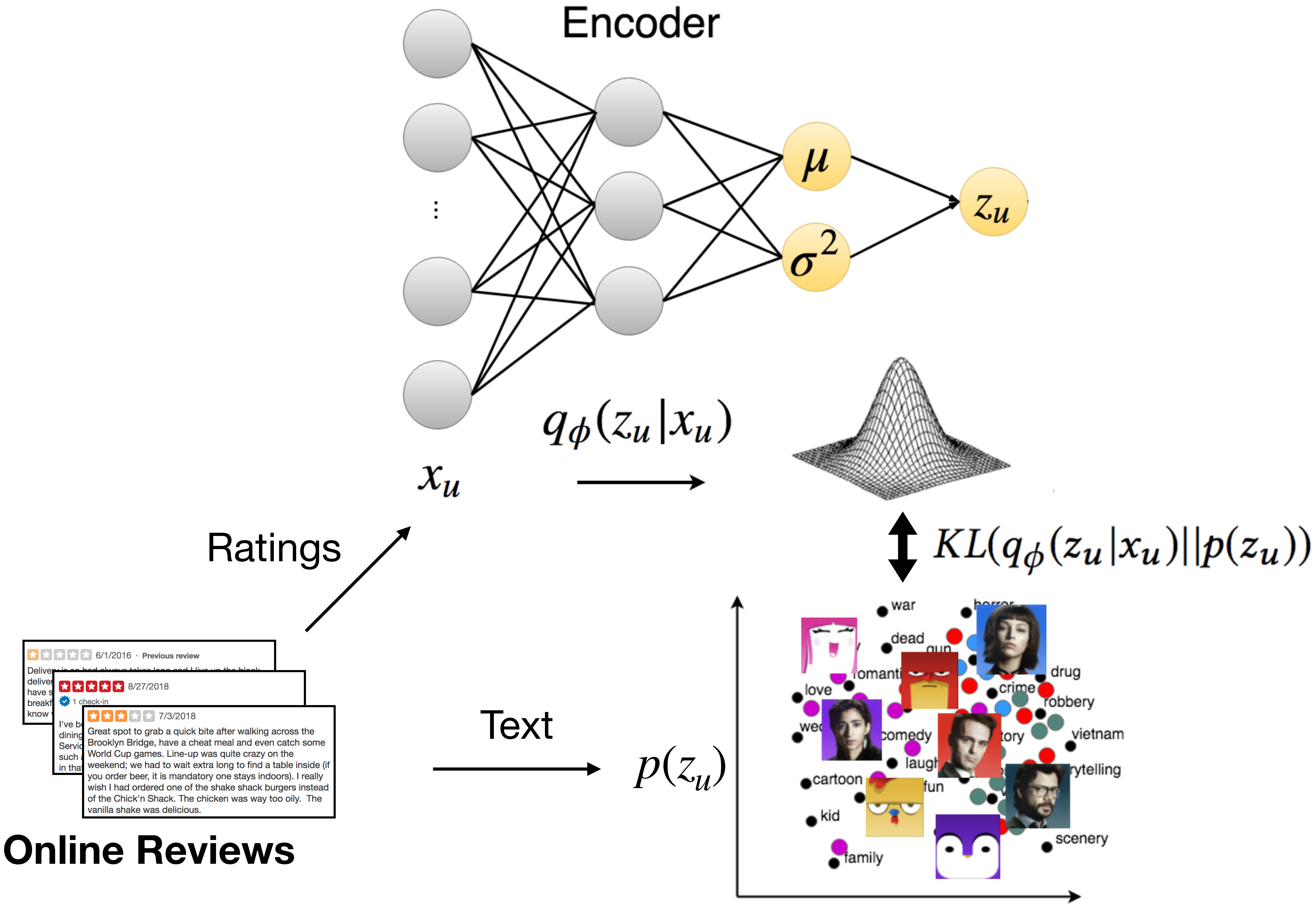
Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline



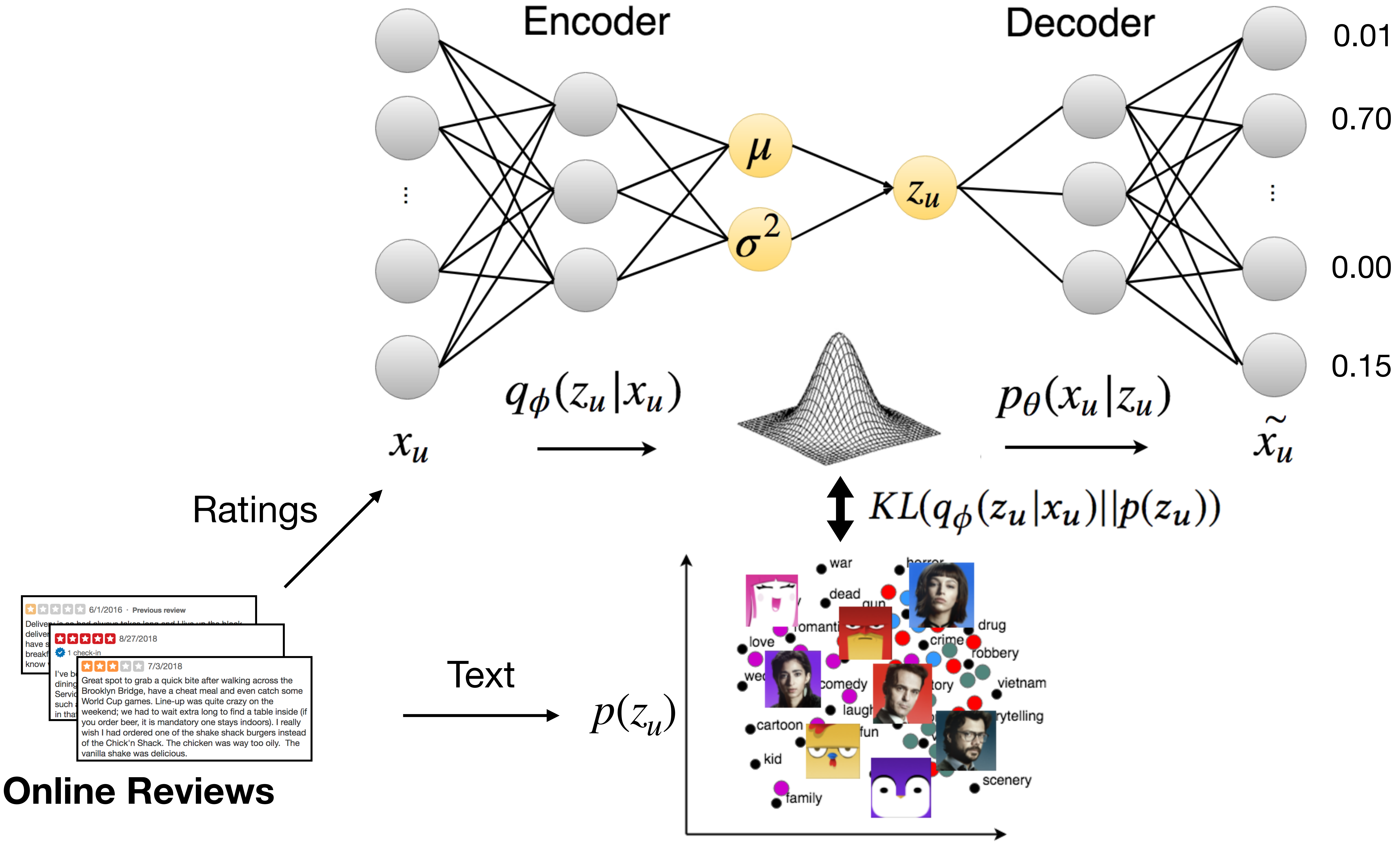
Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline



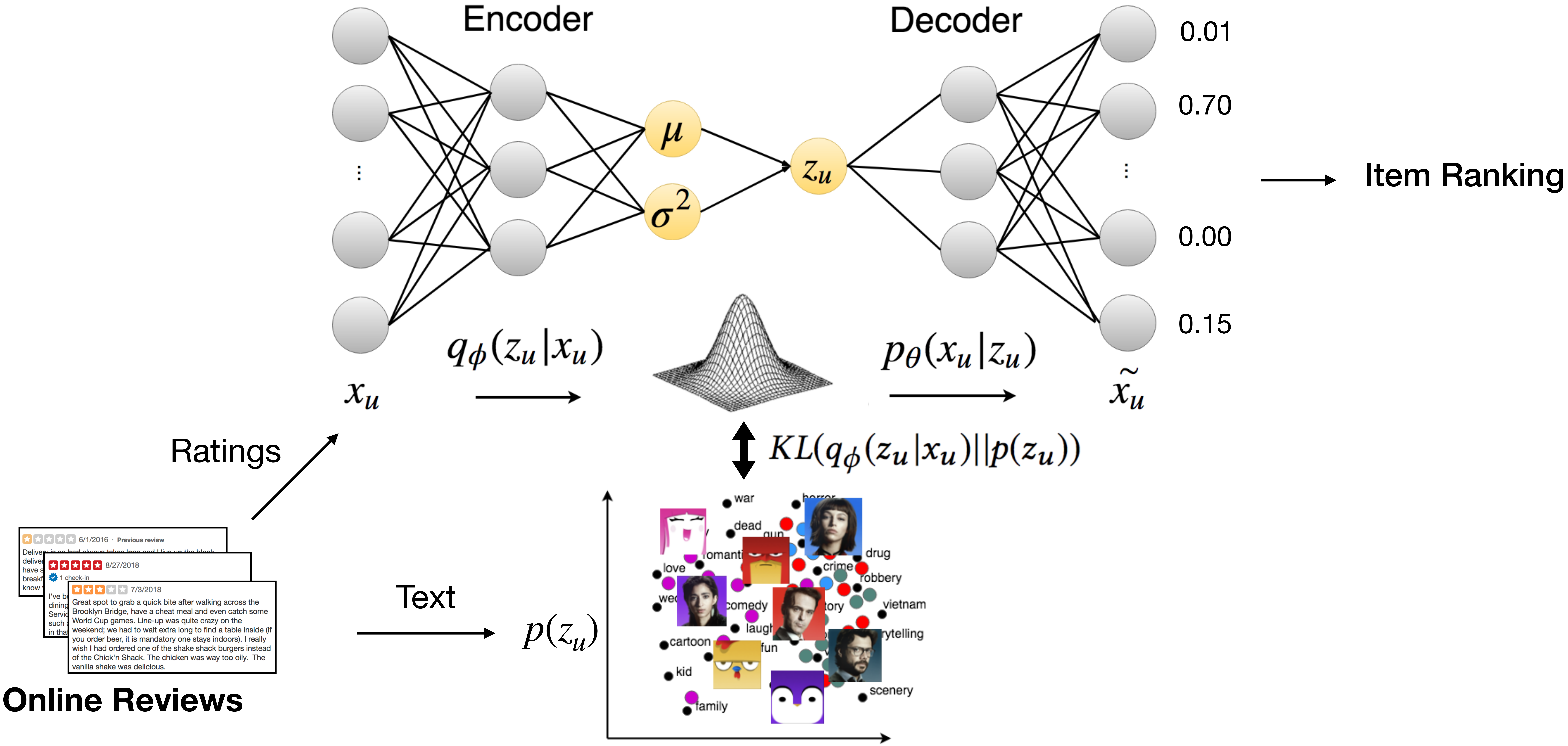
Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline



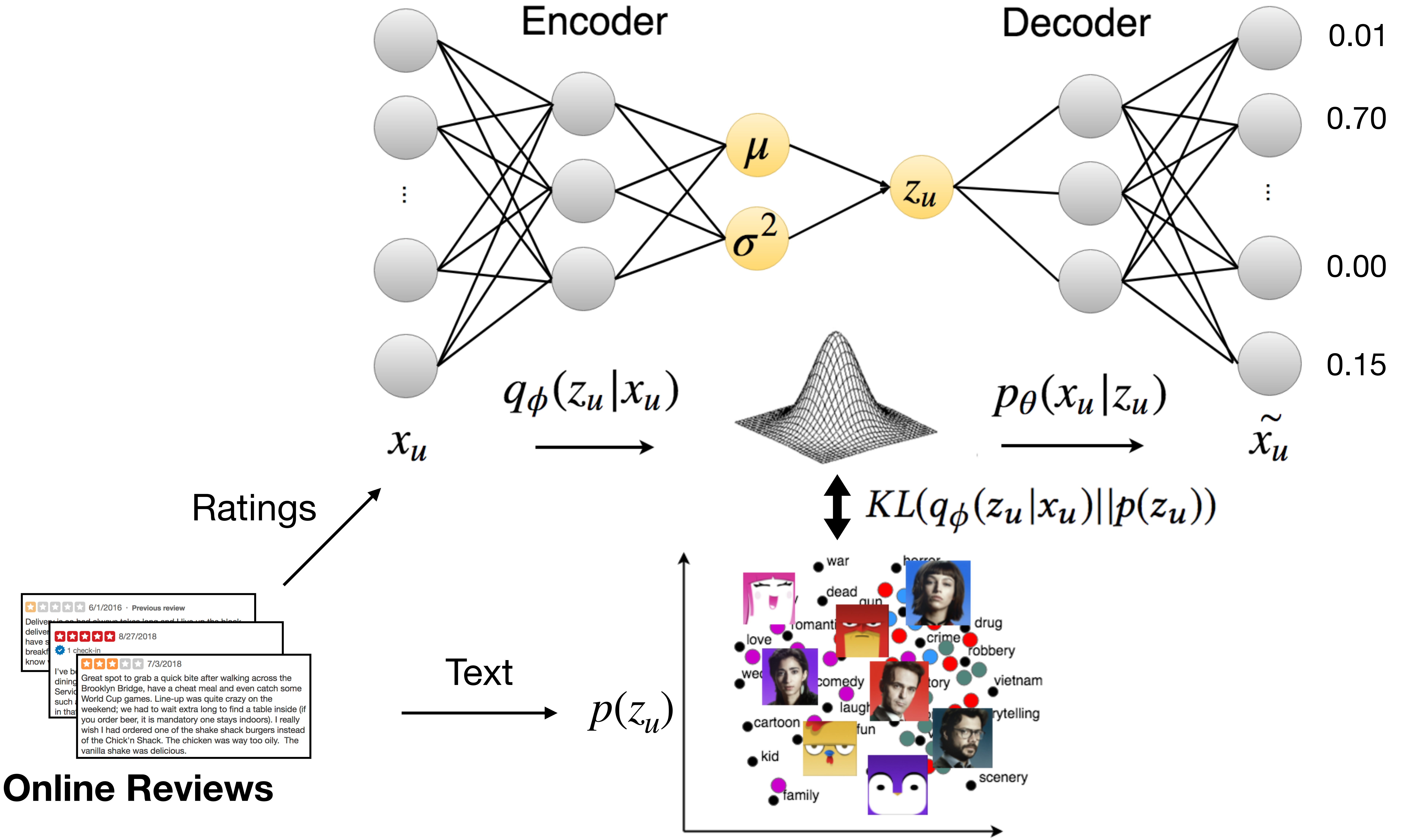
Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline



Item Recommendation with VAEs and Heterogenous Priors

Item Recommendation Pipeline



Item Ranking

Evaluate on held-out ratings

Evaluation Metrics

- NDCG@100
- Recall@20
- Recall@50

Item Recommendation with VAEs and Heterogenous Priors

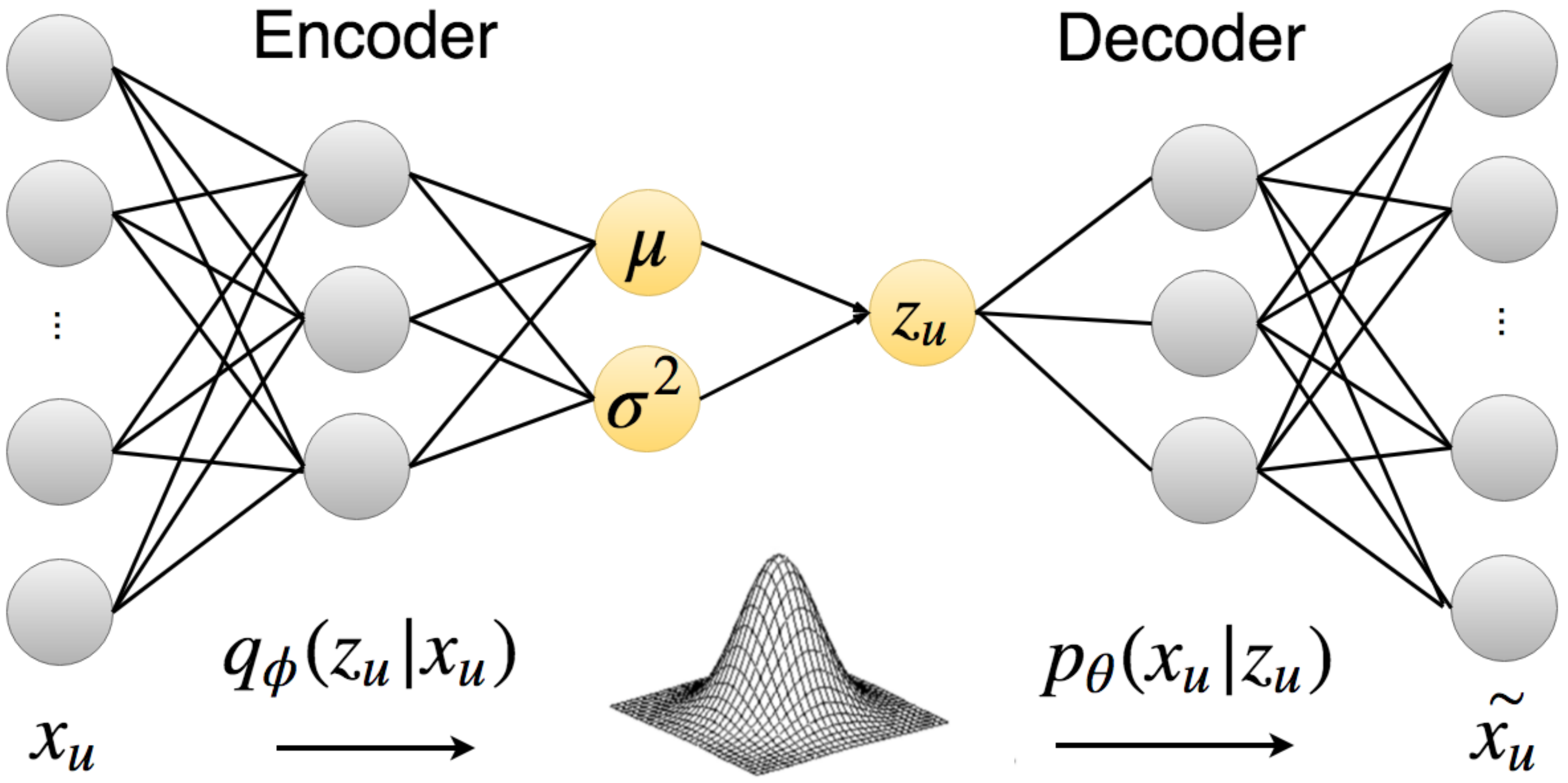
Item Recommendation Pipeline

Encoder

MLP: $I \times 600 \times 300$

Decoder

MLP: $300 \times 600 \times I$



→ Item Ranking

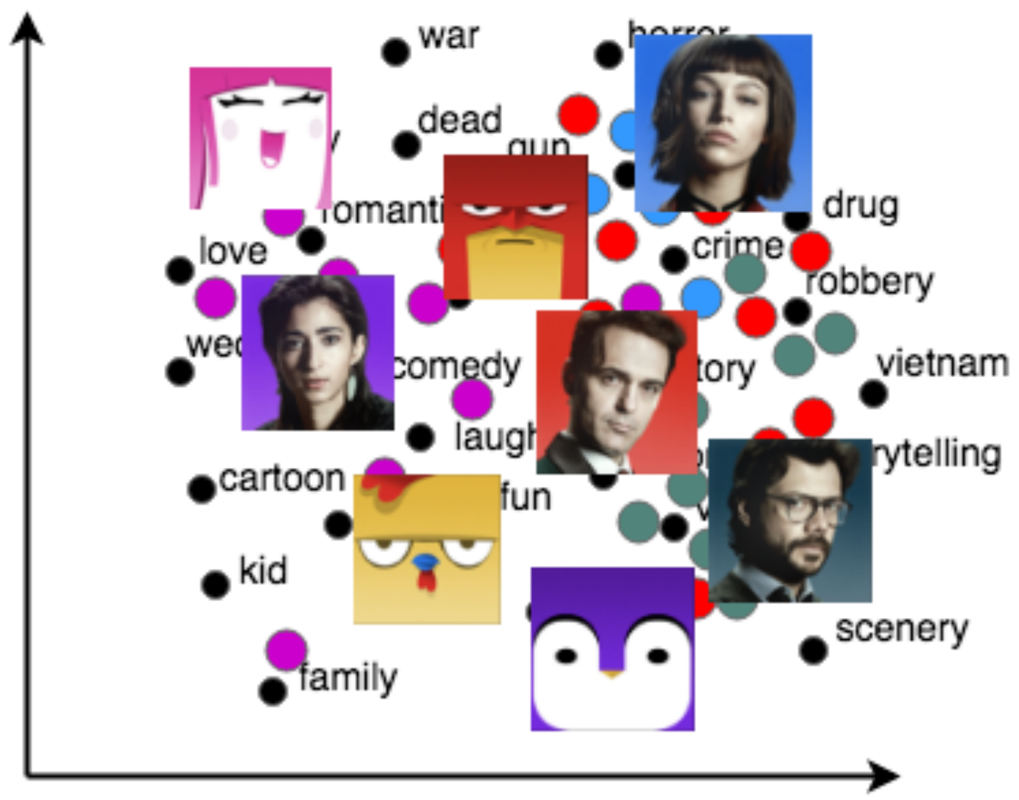
↓

Evaluate on held-out ratings

Ratings

Online Reviews

Text → $p(z_u)$



Evaluation Metrics

- NDCG@100
- Recall@20
- Recall@50

Item Recommendation with VAEs and Heterogenous Priors

Evaluation Datasets: Online Reviews (Rating & Text)

- Yelp Challenge Dataset
- IMDB Corpus of Movie Reviews

Preprocessing:

- Binarize ratings
  - Yelp: 1-2 stars $\rightarrow$ 0, 3-5 stars $\rightarrow$ 1
  - IMDB: 1-4 stars $\rightarrow$ 0, 5-10 stars $\rightarrow$ 1
- Reduce sparsity (cutoff)
  - Yelp: discard businesses < 30 reviews, users < 5 reviews
  - IMDB: discard movies < 5 reviews, users < 5 reviews

% non-empty entries

Dataset	#users	#items	#ratings	sparsity
Yelp	930496	65536	20000263	0.053e-3%
Yelp cutoff	92208	13085	1257420	0.104%
IMDB	50331	21740	278907	0.025%
IMDB cutoff	8080	8357	167593	0.248%

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Evaluation Results

	Model	Text Feat	IMDB	Yelp
Ranking the items in random order	RAND	-		
Matrix Factorization	MF	-		
Text-only: Ranking items according to $cos(t_u, t_i)$	Text-kNN	word2vec		
VAE (Liang et al. 2018)	Mult-VAE	-		
*VAE with **random** user-dependent priors*	VAE-RP	-		
VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$	VAE-TR	word2vec		
	VAE-TR	LDA		
*VAE with **heterogenous** user-dependent priors*	VAE-HPrior	word2vec		
	VAE-HPrior	LDA		

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Ranking the items in random order

Matrix Factorization

Text-only: Ranking items according to $cos(t_u, t_i)$

VAE (Liang et al. 2018)

*VAE with **random** user-dependent priors*

VAE with Text Regularization

$\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$

*VAE with **heterogenous** user-dependent priors*

Evaluation Results

Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
RAND	-	0.006	0.001
MF	-	0.066	0.070
Text-kNN	word2vec	0.026	0.003
Mult-VAE	-	0.147	0.104
VAE-RP	-	0.148	0.106
VAE-TR	word2vec	0.149	0.106
VAE-TR	LDA	0.145	0.107
VAE-HPrior	word2vec	**0.174**	0.114
VAE-HPrior	LDA	**0.174**	**0.119**
std of scores		~ 0.007	~ 0.003

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Ranking the items in random order — RAND

Matrix Factorization — MF

Text-only: Ranking items according to $cos(t_u, t_i)$ — Text-kNN

VAE (Liang et al. 2018) — Mult-VAE

VAE with ***random*** *user-dependent priors* — VAE-RP

VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$ — VAE-TR

VAE with ***heterogenous*** *user-dependent priors* — VAE-HPrior

Evaluation Results

Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
RAND	-	0.006	0.001
MF	-	0.066	0.070
Text-kNN	word2vec	0.026	0.003
Mult-VAE	-	0.147	0.104
VAE-RP	-	0.148	0.106
VAE-TR	word2vec	0.149	0.106
VAE-TR	LDA	0.145	0.107
VAE-HPrior	word2vec	**0.174**	0.114
VAE-HPrior	LDA	**0.174**	**0.119**
std of scores		~ 0.007	~ 0.003

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Ranking the items in random order

Matrix Factorization

Text-only: Ranking items according to $cos(t_u, t_i)$

VAE (Liang et al. 2018)

*VAE with **random** user-dependent priors*

VAE with Text Regularization

$$\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$$

*VAE with **heterogenous** user-dependent priors*

Evaluation Results

Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
RAND	-	0.006	0.001
MF	-	0.066	0.070
Text-kNN	word2vec	0.026	0.003
Mult-VAE	-	0.147	0.104
VAE-RP	-	0.148	0.106
VAE-TR	word2vec	0.149	0.106
VAE-TR	LDA	0.145	0.107
VAE-HPrior	word2vec	**0.174**	0.114
VAE-HPrior	LDA	**0.174**	**0.119**
std of scores		~ 0.007	~ 0.003

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Evaluation Results

	Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
Ranking the items in random order	RAND	-	0.006	0.001
Matrix Factorization	MF	-	0.066	0.070
Text-only: Ranking items according to $cos(t_u, t_i)$	Text-kNN	word2vec	0.026	0.003
VAE (Liang et al. 2018)	Mult-VAE	-	0.147	0.104
*VAE with **random** user-dependent priors*	VAE-RP	-	0.148	0.106
VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \text{dist}(z_u, t_u)$	VAE-TR	word2vec	0.149	0.106
	VAE-TR	LDA	0.145	0.107
*VAE with **heterogenous** user-dependent priors*	VAE-HPrior	word2vec	**0.174**	0.114
	VAE-HPrior	LDA	**0.174**	**0.119**

std of scores ~ 0.007 (IMDB), ~ 0.003 (Yelp)

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Evaluation Results

	Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
Ranking the items in random order	RAND	-	0.006	0.001
Matrix Factorization	MF	-	0.066	0.070
Text-only: Ranking items according to $cos(t_u, t_i)$	Text-kNN	word2vec	0.026	0.003
VAE (Liang et al. 2018)	Mult-VAE	-	0.147	0.104
*VAE with **random** user-dependent priors*	VAE-RP	-	0.148	0.106
VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$	VAE-TR	word2vec	0.149	0.106
	VAE-TR	LDA	0.145	0.107
*VAE with **heterogenous** user-dependent priors*	VAE-HPrior	word2vec	**0.174**	0.114
	VAE-HPrior	LDA	**0.174**	**0.119**
	std of scores		~ 0.007	~ 0.003

Mult-VAE ⟶ VAE-HPrior

Relative Performance Improvement:

	IMDB NDCG@100	IMDB Recall@20	IMDB Recall@50	Yelp NDCG@100	Yelp Recall@20	Yelp Recall@50
Relative Performance Improvement	**+18.4%**	**+29.4%**	**+17.7%**	**+14.4%**	**+18.7%**	**+12.3%**

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Evaluation Results

	Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
Ranking the items in random order	RAND	-	0.006	0.001
Matrix Factorization	MF	-	0.066	0.070
Text-only: Ranking items according to $cos(t_u, t_i)$	Text-kNN	word2vec	0.026	0.003
VAE (Liang et al. 2018)	Mult-VAE	-	0.147	0.104
*VAE with **random** user-dependent priors*	VAE-RP	-	0.148	0.106
VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \text{dist}(z_u, t_u)$	VAE-TR	word2vec	0.149	0.106
	VAE-TR	LDA	0.145	0.107
*VAE with **heterogenous** user-dependent priors*	VAE-HPrior	word2vec	**0.174**	0.114
	VAE-HPrior	LDA	**0.174**	**0.119**
	std of scores		~ 0.007	~ 0.003

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Evaluation Results

	Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
Ranking the items in random order	RAND	-	0.006	0.001
Matrix Factorization	MF	-	0.066	0.070
Text-only: Ranking items according to $cos(t_u, t_i)$	Text-kNN	word2vec	0.026	0.003
VAE (Liang et al. 2018)	Mult-VAE	-	0.147	0.104
VAE with ***random*** *user-dependent priors*	VAE-RP	-	0.148	0.106
VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$	VAE-TR	word2vec	0.149	0.106
	VAE-TR	LDA	0.145	0.107
VAE with ***heterogenous*** *user-dependent priors*	VAE-HPrior	word2vec	**0.174**	0.114
	VAE-HPrior	LDA	**0.174**	**0.119**

std of scores ~ 0.007 ~ 0.003

Item Recommendation with VAEs and Heterogenous Priors

Model Comparison

Evaluation Results

Model Comparison	Model	Text Feat	IMDB NDCG@100	Yelp NDCG@100
Ranking the items in random order	RAND	-	0.006	0.001
Matrix Factorization	MF	-	0.066	0.070
Text-only: Ranking items according to $cos(t_u, t_i)$	Text-kNN	word2vec	0.026	0.003
VAE (Liang et al. 2018)	Mult-VAE	-	0.147	0.104
VAE with ***random*** *user-dependent priors*	VAE-RP	-	0.148	0.106
VAE with Text Regularization $\mathscr{L}_\gamma = \mathscr{L}_\beta - \gamma \cdot \mathrm{dist}(z_u, t_u)$	VAE-TR	word2vec	0.149	0.106
	VAE-TR	LDA	0.145	0.107
VAE with ***heterogenous*** *user-dependent priors*	VAE-HPrior	word2vec	**0.174**	0.114
	VAE-HPrior	LDA	**0.174**	**0.119**
Denoising autoencoder (Liang et al. 2018)	Mult-DAE	-	**0.178**	**0.121**
	std of scores		~ 0.007	~ 0.003

Item Recommendation with VAEs and Heterogenous Priors

Conclusions

- Extend VAEs to Collaborative Filtering with side information (ratings + text)
  - User-agnostic ⟶ user-dependent priors
  - Prior parameters as functions of the users' review text
  - User representations in a multimodal latent space (encoding ratings + text)
- Outperform the existing Mult-VAE model (up to 29.41% relative improvement in Recall@20)
- Perform comparably to a denoising autoencoder (Mult-DAE).

Ongoing & Future work

- Experiments: VAE-HPrior vs Mult-DAE on different levels of sparsity
- Models: more effective aspect-based methods for extracting user preferences from text reviews
- Data: extra side-information available (e.g., geolocation)

Thank you!

<u>Contact</u>

Giannis Karamanolakis
gkaraman@cs.columbia.edu

Columbia University

Questions?

Contact
Giannis Karamanolakis
gkaraman@cs.columbia.edu

COLUMBIA UNIVERSITY