Relational Deep Learning: A Deep Latent Variable Model for Link Prediction

Hao Wang, Xingjian Shi, Dit-Yan Yeung

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

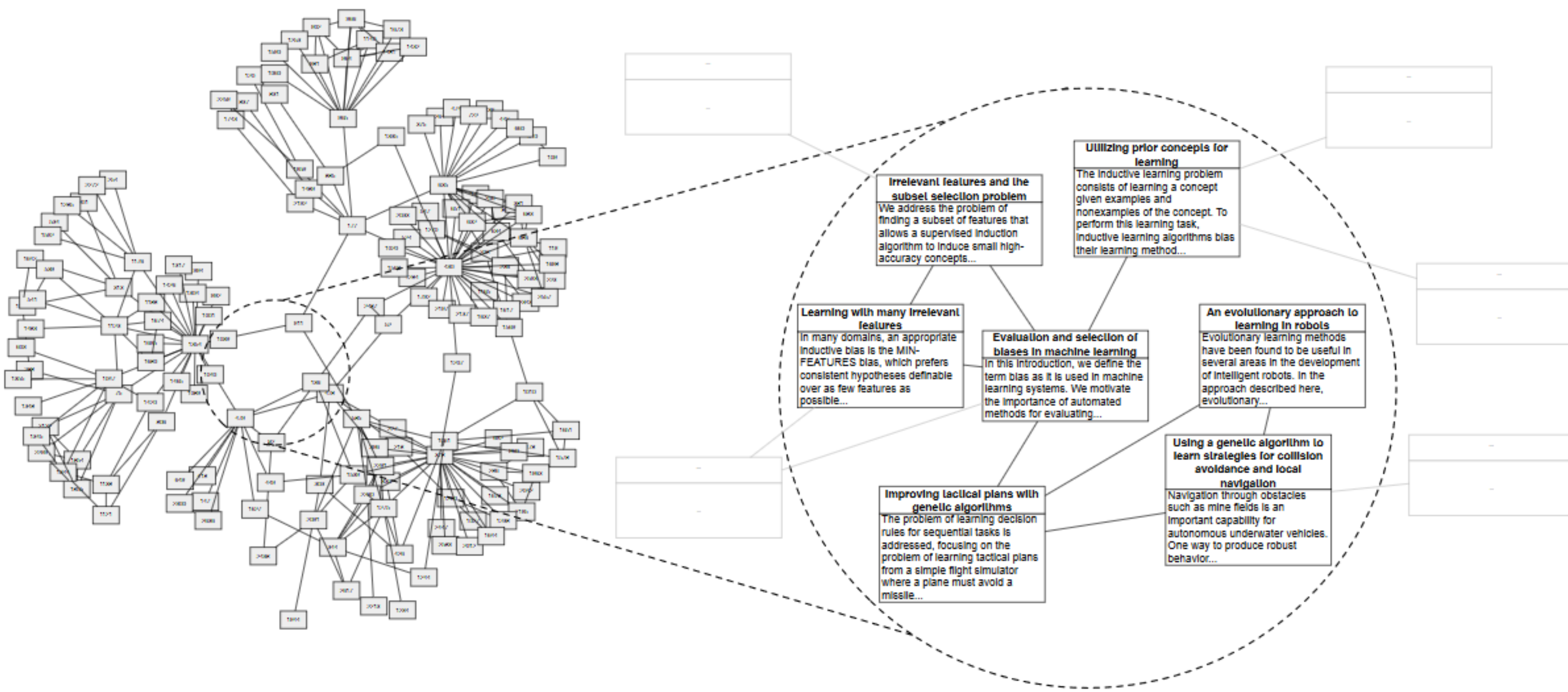
Carnegie Mellon University

- Motivation
- Bayesian Deep Learning
- Relational Deep Learning
- Parameter Learning
- Experiments
- Conclusion

Motivation: Link Prediction

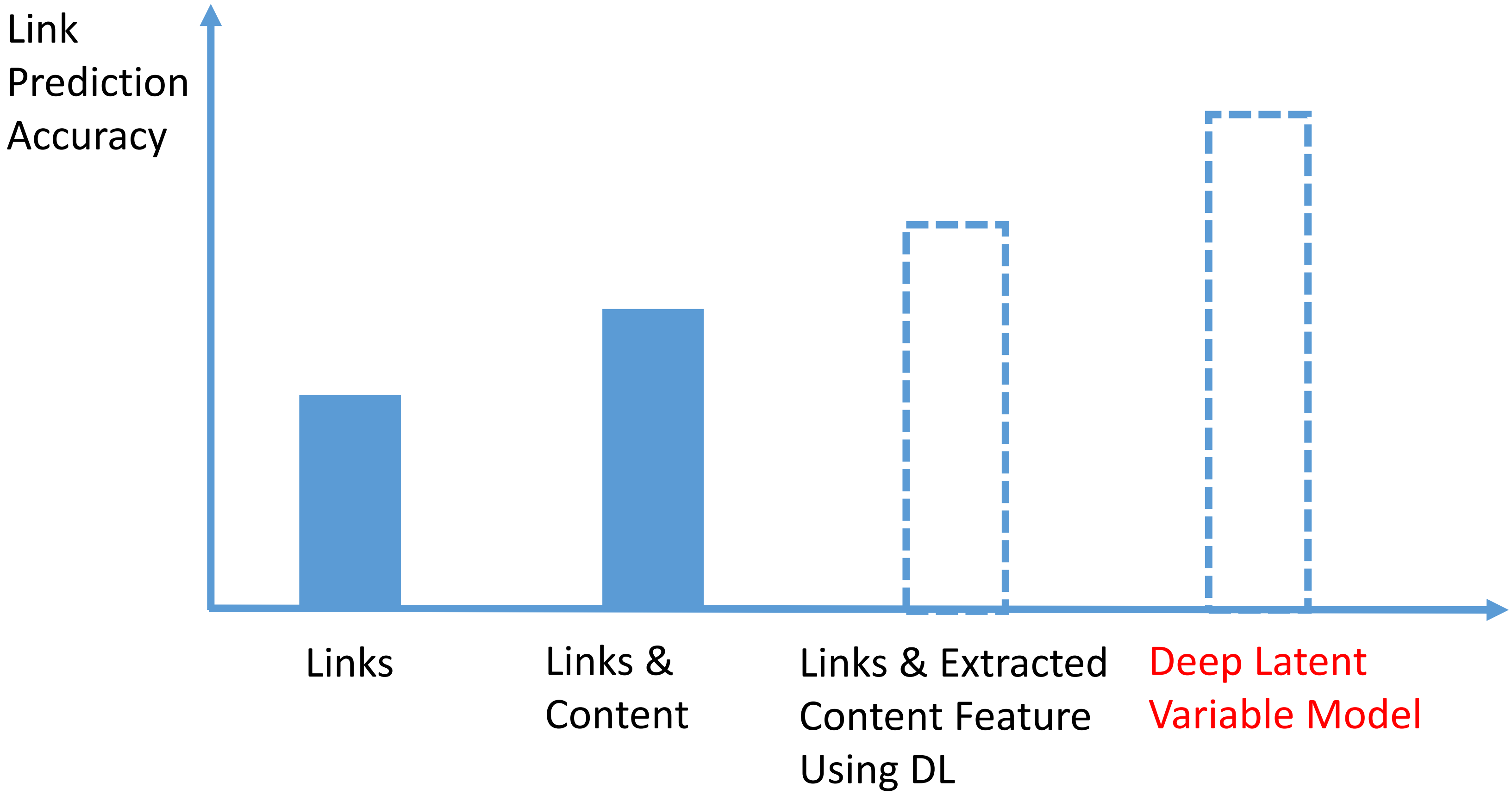
Social Network Analysis (e.g., prediction friendship in Facebook)

Motivation: Link Prediction

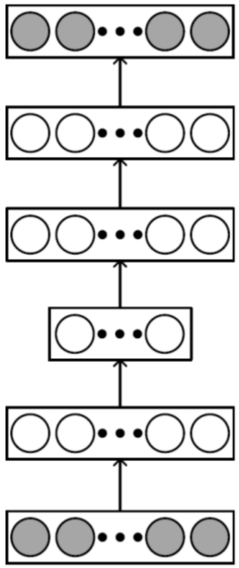


Document Networks (e.g., citation networks, co-author networks)

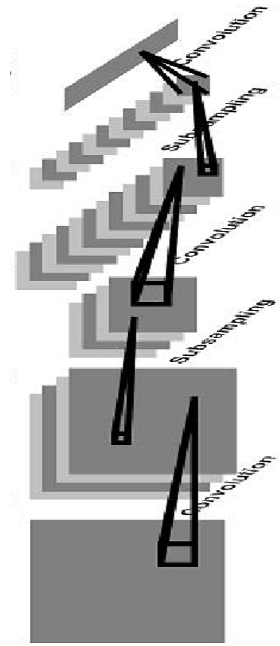
Motivation: Deep Latent Variable Models



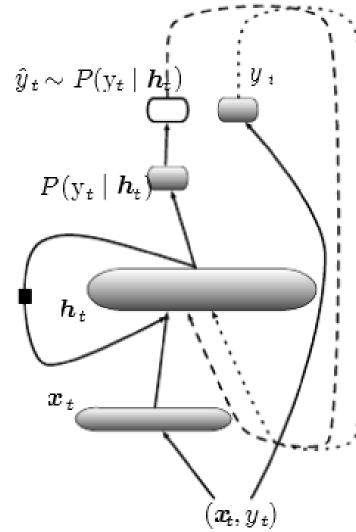
Motivation: Deep Latent Variable Models



Stacked denoising autoencoders



Convolutional neural networks

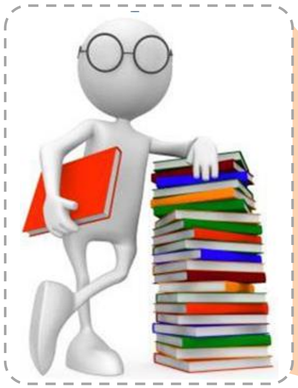


Recurrent neural networks

Typically for i.i.d. data

- Motivation
- Bayesian Deep Learning
- Relational Deep Learning
- Parameter Learning
- Experiments
- Conclusion

Bayesian Deep Learning



Perception component

Content understanding

Posts by users

Text in articles

Task-Specific component

Target task

Link prediction

Bayesian deep learning (BDL)

- Maximum a posteriori (MAP)
- Markov chain Monte Carlo (MCMC)
- Variational inference (VI)

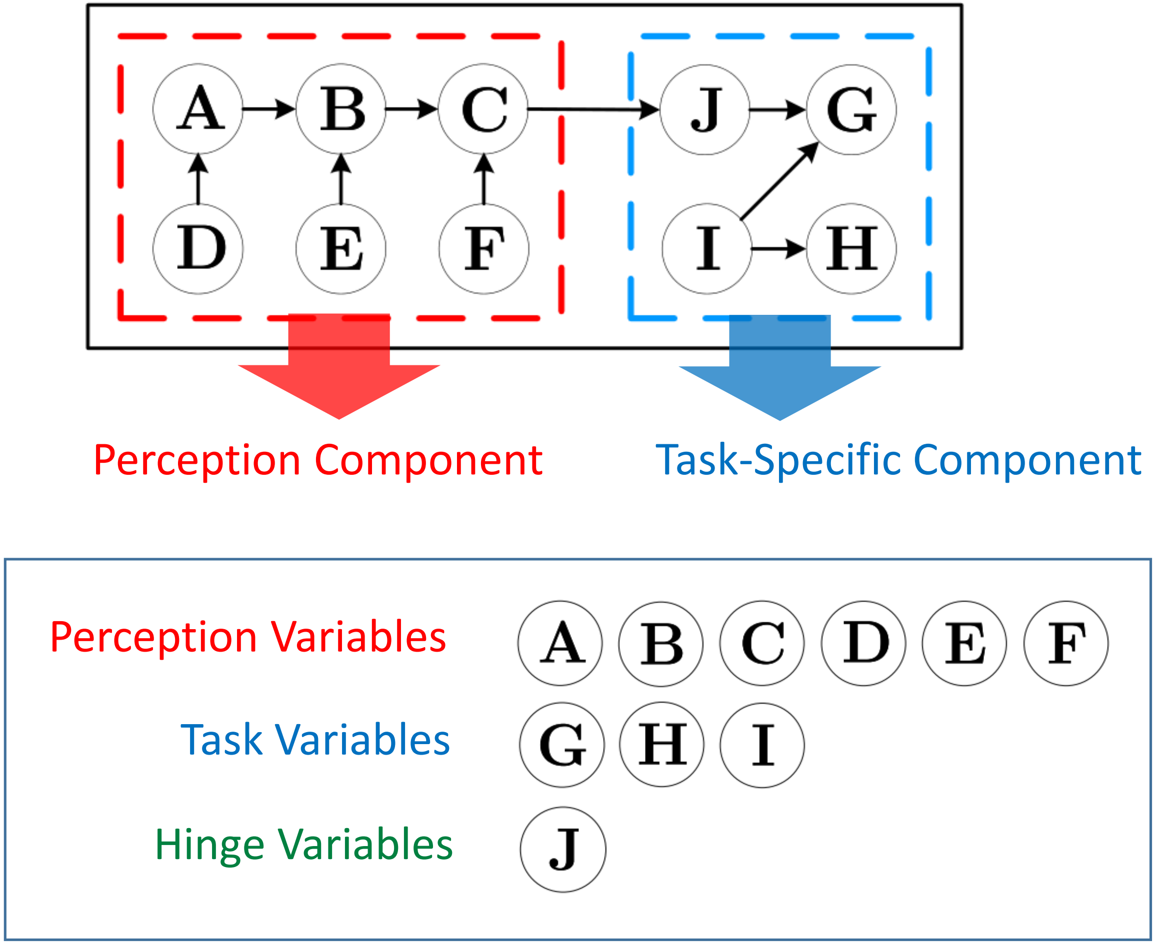
[Wang et al. 2016]

Bayesian Deep Learning

Applications	Models	Hinge Variables	Learning
Recommender Systems	CDL [Wang et al.]	$\{\mathbf{V}\}$	MAP
	Bayesian CDL [Wang et al.]	$\{\mathbf{V}\}$	Gibbs Sampling
	Marginalized CDL [Li et al.]	$\{\mathbf{V}\}$	MAP
	Symmetric CDL [Li et al.]	$\{\mathbf{V}, \mathbf{U}\}$	MAP
	Collaborative Deep Ranking [Ying et al.]	$\{\mathbf{V}\}$	MAP
Topic Models	Relational SDAE [Wang et al.]	$\{\mathbf{S}\}$	MAP
	DPFA-SBN [Gan et al.]	$\{\mathbf{X}\}$	Hybrid MC
	DPFA-RBM [Gan et al.]	$\{\mathbf{X}\}$	Hybrid MC
Control	Embed to Control [Watter et al.]	$\{\mathbf{z}_t, \mathbf{z}_{t+1}\}$	Variational Inference

[Wang et al. 2016]

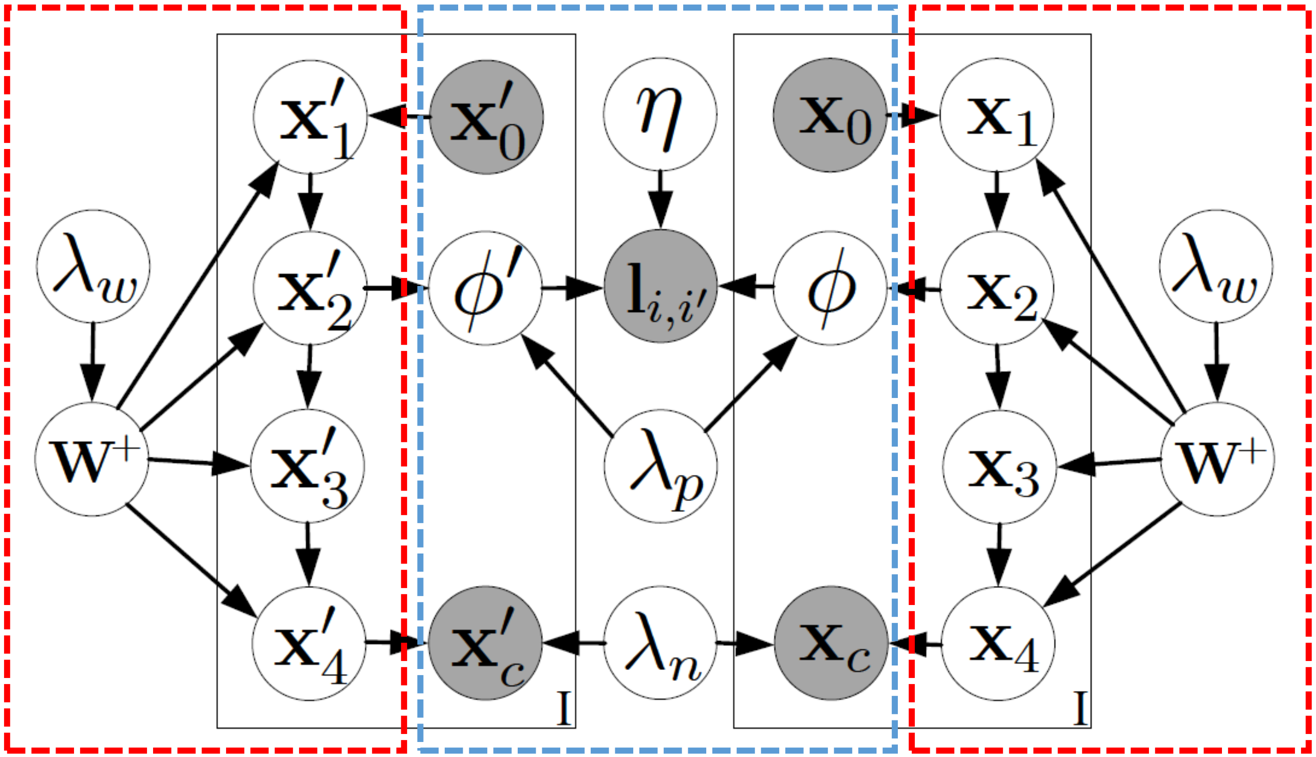
A Principled Probabilistic Framework



[Wang et al. 2016]

- Motivation
- Bayesian Deep Learning
- Relational Deep Learning
- Parameter Learning
- Experiments
- Conclusion

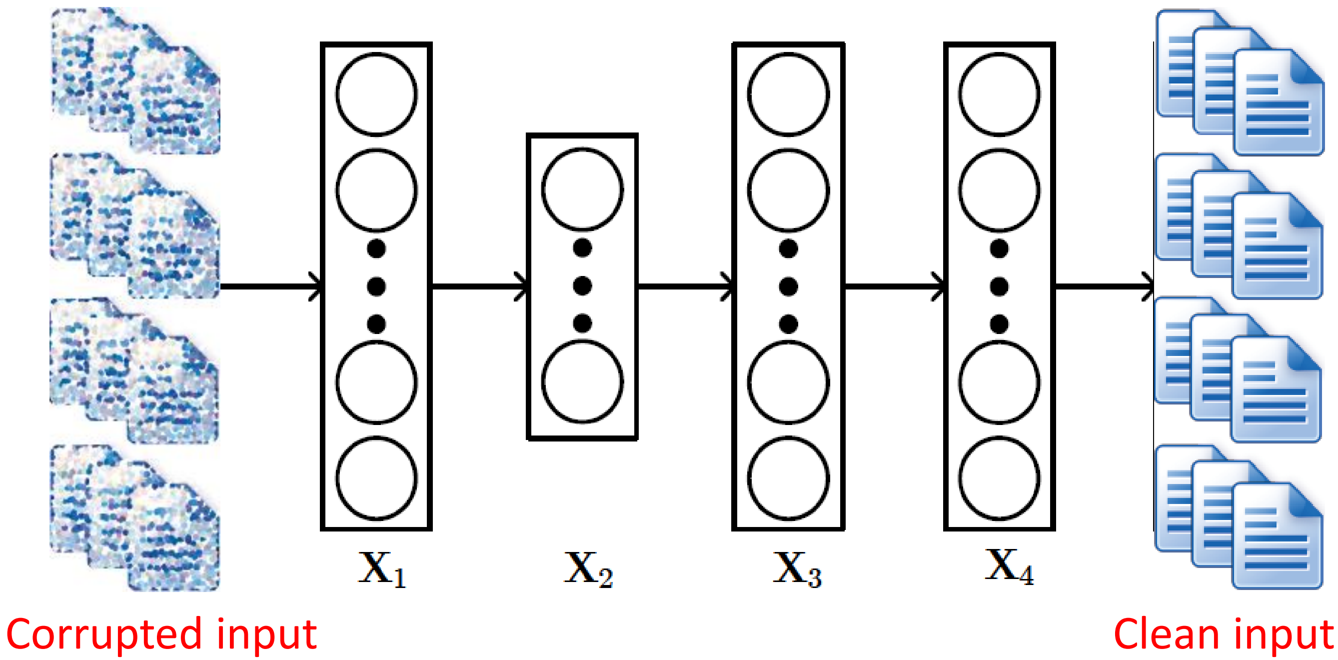
Relational Deep Learning: Graphical Model



Perception component: relational and deep representation learning

Task-specific component: link prediction

Stacked Denoising Autoencoders (SDAE)



SDAE solves the following optimization problem:

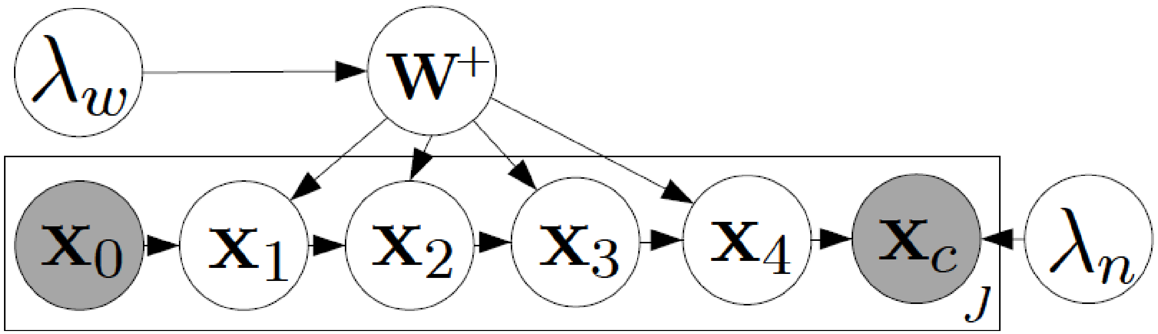
$$\min_{\{\mathbf{W}_l\},\{\mathbf{b}_l\}} \|\mathbf{X}_c - \mathbf{X}_L\|_F^2 + \lambda \sum_l \|\mathbf{W}_l\|_F^2,$$

where λ is a regularization parameter and $\|\cdot\|_F$ denotes the Frobenius norm.

[Vincent et al. 2010]

Probabilistic SDAE

Graphical model:



Generative process:

$$\mathbf{W}_{l,*n} \sim \mathcal{N}(0, \lambda_w^{-1}\mathbf{I}_{K_l})$$

$$\mathbf{b}_l \sim \mathcal{N}(0, \lambda_w^{-1}\mathbf{I}_{K_l})$$

$$\mathbf{X}_{l,j*} \sim \mathcal{N}(\sigma(\mathbf{X}_{l-1,j*}\mathbf{W}_l + \mathbf{b}_l), \boxed{\lambda_s^{-1}\mathbf{I}_{K_l}})$$

$$\mathbf{X}_{c,j*} \sim \mathcal{N}(\mathbf{X}_{L,j*}, \lambda_n^{-1}\mathbf{I}_B)$$

Generalized SDAE

Notation:

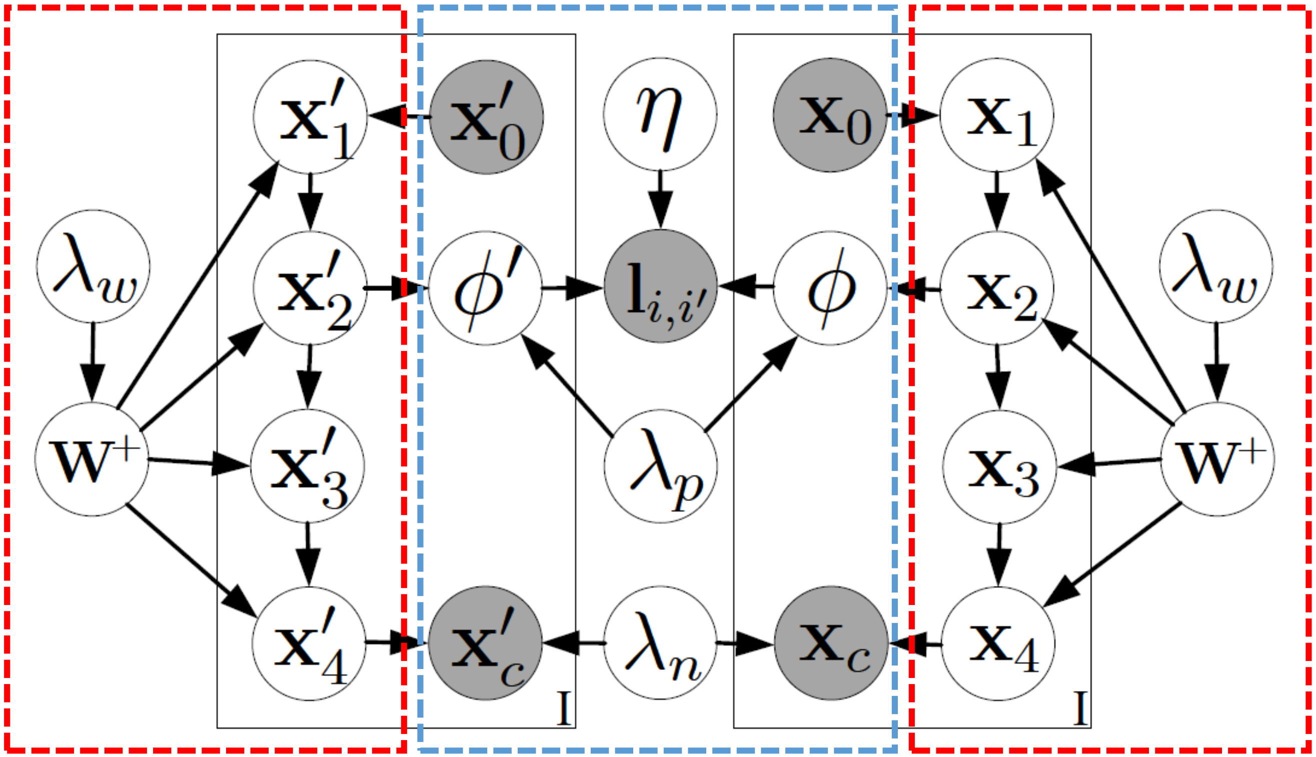
$\mathbf{X}_0$ corrupted input

$\mathbf{X}_c$ clean input

$\mathbf{W}^+$ weights and biases

[Wang et al. 2015]

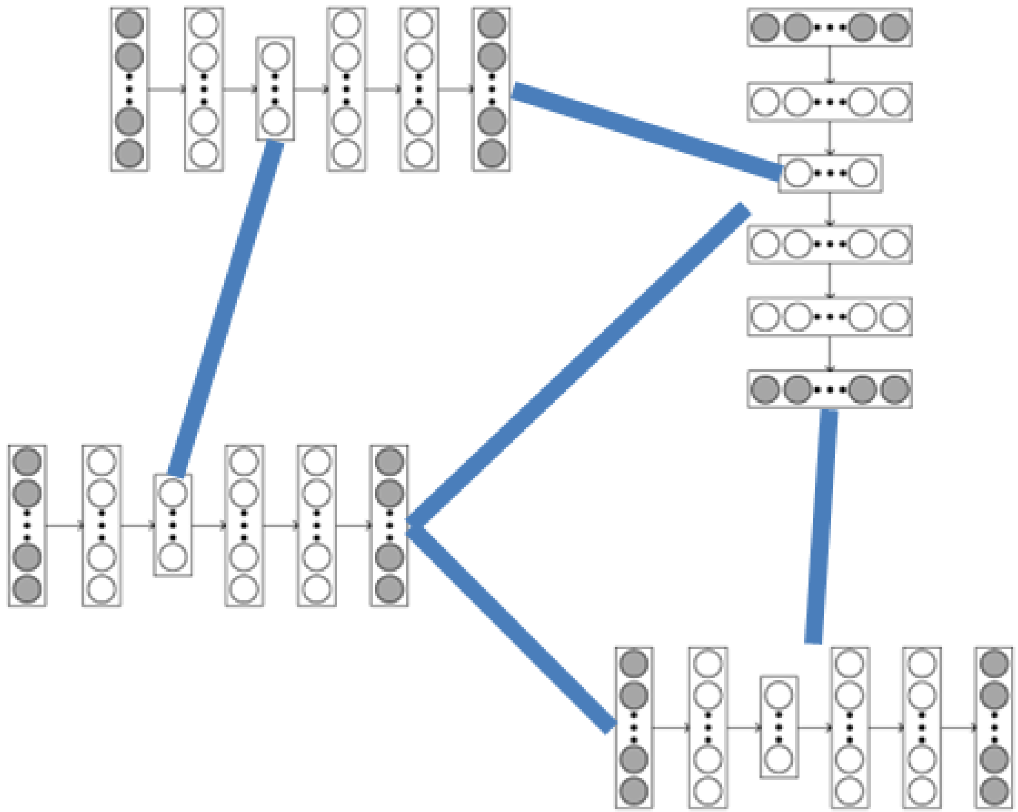
Relational Deep Learning



Probabilistic SDAE

Modeling relation among nodes

Network of Probabilistic SDAE



Many interconnected probabilistic SDAEs with shared weights

- Motivation
- Bayesian Deep Learning
- Relational Deep Learning
- Parameter Learning
- Experiments
- Conclusion

MAP Inference

maximizing the posterior probability is equivalent to maximizing the joint log-likelihood

$$
\begin{aligned}
\mathscr{L} = &-\frac{\lambda_w}{2}\sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) \\
&-\frac{\lambda_p}{2}\sum_i \|\boldsymbol{\phi}_i - \mathbf{X}_{\frac{L}{2},i*}^T\|_2^2 - \frac{\lambda_n}{2}\sum_i \|\mathbf{X}_{L,i*} - \mathbf{X}_{c,i*}\|_2^2 \\
&-\frac{\lambda_s}{2}\sum_l \sum_i \|\sigma(\mathbf{X}_{l-1,i*}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_{l,i*}\|_2^2 \\
&-\frac{\lambda_e}{2}\|\boldsymbol{\eta}\|_2^2 + \sum_{l_{i,i'}=1} \log \sigma(\boldsymbol{\eta}^T(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_{i'})).
\end{aligned}
$$

MAP Inference

Prior (regularization) for link prediction parameters, weights, and biases

$$
\begin{aligned}
\mathscr{L} = & \boxed{-\frac{\lambda_w}{2}\sum_l \left(\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2\right)} \\
& - \frac{\lambda_p}{2}\sum_i \|\boldsymbol{\phi}_i - \mathbf{X}^T_{\frac{L}{2},i*}\|_2^2 - \frac{\lambda_n}{2}\sum_i \|\mathbf{X}_{L,i*} - \mathbf{X}_{c,i*}\|_2^2 \\
& - \frac{\lambda_s}{2}\sum_l\sum_i \|\sigma(\mathbf{X}_{l-1,i*}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_{l,i*}\|_2^2 \\
& \boxed{-\frac{\lambda_e}{2}\|\boldsymbol{\eta}\|_2^2} + \sum_{l_{i,i'}=1} \log \sigma(\boldsymbol{\eta}^T(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_{i'})).
\end{aligned}
$$

MAP Inference

Generating node features from content representation with Gaussian offset

$$
\begin{aligned}
\mathscr{L} = &-\frac{\lambda_w}{2}\sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) \\
&\boxed{-\frac{\lambda_p}{2}\sum_i \|\boldsymbol{\phi}_i - \mathbf{X}^T_{\frac{L}{2},i*}\|_2^2} - \frac{\lambda_n}{2}\sum_i \|\mathbf{X}_{L,i*} - \mathbf{X}_{c,i*}\|_2^2 \\
&-\frac{\lambda_s}{2}\sum_l \sum_i \|\sigma(\mathbf{X}_{l-1,i*}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_{l,i*}\|_2^2 \\
&-\frac{\lambda_e}{2}\|\boldsymbol{\eta}\|_2^2 + \sum_{l_{i,i'}=1} \log \sigma(\boldsymbol{\eta}^T(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_{i'})).
\end{aligned}
$$

MAP Inference

'Generating' clean input from the output of probabilistic SDAE with Gaussian offset

$$\begin{aligned}\mathscr{L} = &-\frac{\lambda_w}{2}\sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) \\ &-\frac{\lambda_p}{2}\sum_i \|\boldsymbol{\phi}_i - \mathbf{X}_{\frac{L}{2},i*}^T\|_2^2 - \boxed{\frac{\lambda_n}{2}\sum_i \|\mathbf{X}_{L,i*} - \mathbf{X}_{c,i*}\|_2^2} \\ &-\frac{\lambda_s}{2}\sum_l\sum_i \|\sigma(\mathbf{X}_{l-1,i*}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_{l,i*}\|_2^2 \\ &-\frac{\lambda_e}{2}\|\boldsymbol{\eta}\|_2^2 + \sum_{l_{i,i'}=1} \log\sigma(\boldsymbol{\eta}^T(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_{i'})).\end{aligned}$$

MAP Inference

Generating the input of Layer l from the output of Layer l-1 with Gaussian offset

$$
\begin{aligned}
\mathscr{L} = & -\frac{\lambda_w}{2}\sum_l\left(\|\mathbf{W}_l\|_F^2+\|\mathbf{b}_l\|_2^2\right) \\
& -\frac{\lambda_p}{2}\sum_i\|\boldsymbol{\phi}_i-\mathbf{X}^T_{\frac{L}{2},i*}\|_2^2-\frac{\lambda_n}{2}\sum_i\|\mathbf{X}_{L,i*}-\mathbf{X}_{c,i*}\|_2^2 \\
& \boxed{-\frac{\lambda_s}{2}\sum_l\sum_i\|\sigma(\mathbf{X}_{l-1,i*}\mathbf{W}_l+\mathbf{b}_l)-\mathbf{X}_{l,i*}\|_2^2} \\
& -\frac{\lambda_e}{2}\|\boldsymbol{\eta}\|_2^2+\sum_{l_{i,i'}=1}\log\sigma(\boldsymbol{\eta}^T(\boldsymbol{\phi}_i\circ\boldsymbol{\phi}_{i'})).
\end{aligned}
$$

MAP Inference

Generating links from Bernoulli distributions parameterized by η and ϕ

$$
\begin{aligned}
\mathscr{L} = &-\frac{\lambda_w}{2}\sum_l \left(\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2\right) \\
&-\frac{\lambda_p}{2}\sum_i \|\boldsymbol{\phi}_i - \mathbf{X}^T_{\frac{L}{2},i*}\|_2^2 - \frac{\lambda_n}{2}\sum_i \|\mathbf{X}_{L,i*} - \mathbf{X}_{c,i*}\|_2^2 \\
&-\frac{\lambda_s}{2}\sum_l\sum_i \|\sigma(\mathbf{X}_{l-1,i*}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_{l,i*}\|_2^2 \\
&-\frac{\lambda_e}{2}\|\boldsymbol{\eta}\|_2^2 + \boxed{\sum_{l_{i,i'}=1} \log\sigma\left(\boldsymbol{\eta}^T(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_{i'})\right)}
\end{aligned}
$$

Bayesian Treatment:
Generalized Variational Inference

$$\log q_j^*(\mathbf{Z}_j) = \mathrm{E}_{i \neq j}[\log p(\mathbf{X}_0, \mathbf{X}_c, \mathbf{Z})] + const$$

- $q_1(\mathbf{Z}_1) = q(\mathbf{W}^+)$: Variational distributions for weights/biases.
- $q_2(\mathbf{Z}_2) = q(\boldsymbol{\phi}_i)$: Variational distributions for generated node features.
- $q_3(\mathbf{Z}_3) = q(\boldsymbol{\eta})$: Variational distributions for parameters of the link prediction model.
- $q_4(\mathbf{Z}_4) = q(\xi)$: Variational parameters to approximate the sigmoid function.

Use Laplace approximation rather than variational inference for weights/biases.

Example: Updating ϕ as a Product of Gaussians

Update ϕ for node i as a product of two Gaussians

$$q(\boldsymbol{\phi}_i|\mathbf{X}_{0,i*}) \approx \mathcal{N}(\boldsymbol{\phi}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

$$\boldsymbol{\mu}_i = \boldsymbol{\Sigma}_i(\mathbf{S}_i^{-1}\mathbf{m}_i + \mathbf{S}_i'^{-1}\mathbf{m}_i')$$

$$\boldsymbol{\Sigma}_i^{-1} = \mathbf{S}_i^{-1} + \mathbf{S}_i'^{-1}.$$

First Gaussian

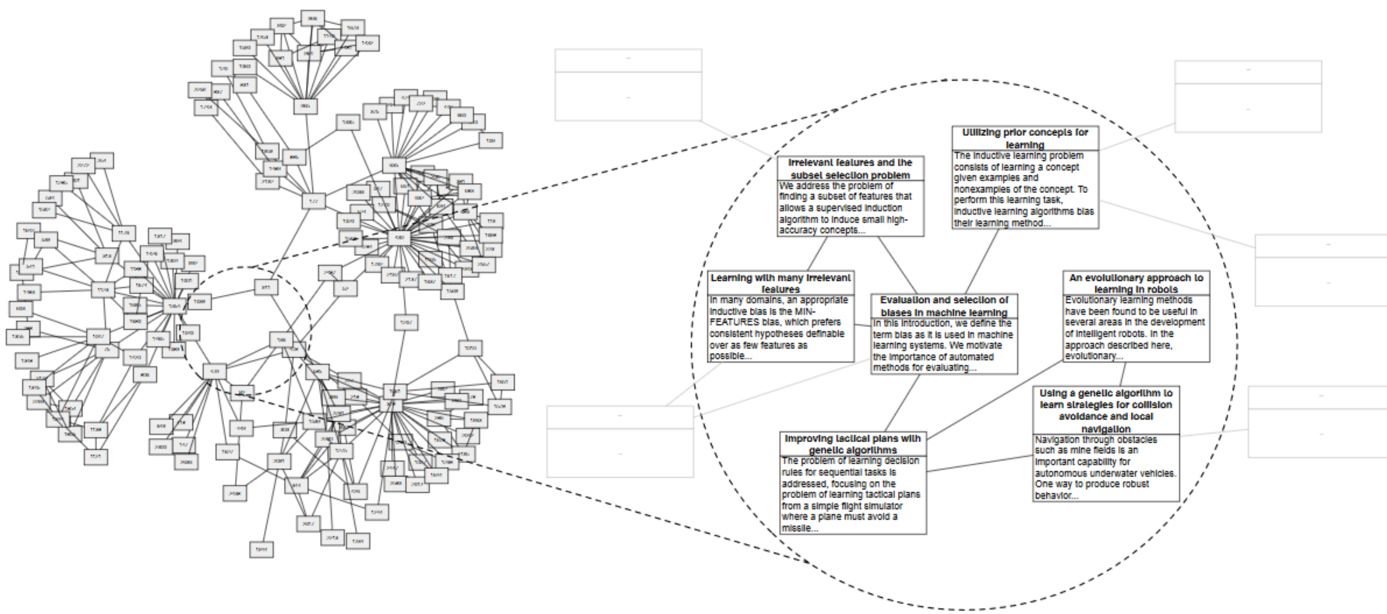
- $\mathbf{m}_i$: Encoding generated by probabilistic SDAE.
- $\mathbf{S}_i$: Variance of probabilistic SDAE.

Second Gaussian

- $\mathbf{m}_i'$: Weighted average of all $\boldsymbol{\eta} \circ \boldsymbol{\phi}_{i'}$.
- $\mathbf{S}_i'$: Variance of all $\boldsymbol{\eta} \circ \boldsymbol{\phi}_{i'}$.

- Motivation
- Bayesian Deep Learning
- Relational Deep Learning
- Parameter Learning
- Experiments
- Conclusion

Experiments: Settings



Document Networks (e.g., citation networks)

datasets	# nodes	# links
citeulike-a	16,980	44,709
citeulike-t	25,975	32,565
arXiv	27,770	352,807

Experiments: Link Rank and AUC

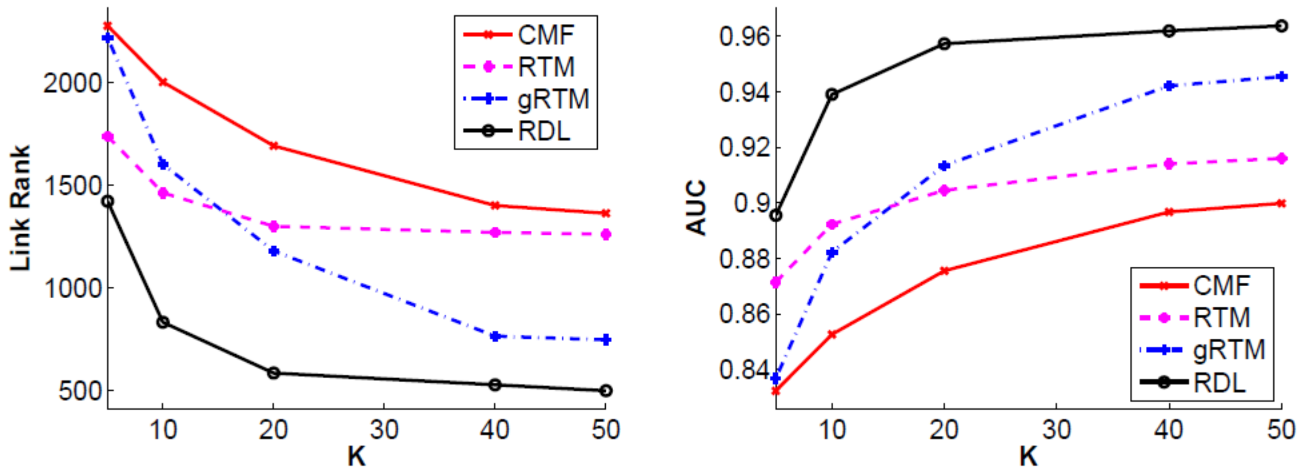


Figure 2: Link rank and AUC of compared models for *citeulike-a*. A 2-layer RDL is used.

Link rank: how high our predicted links rank in the ground truth
AUC: area under curve

Experiments: Link Rank and AUC

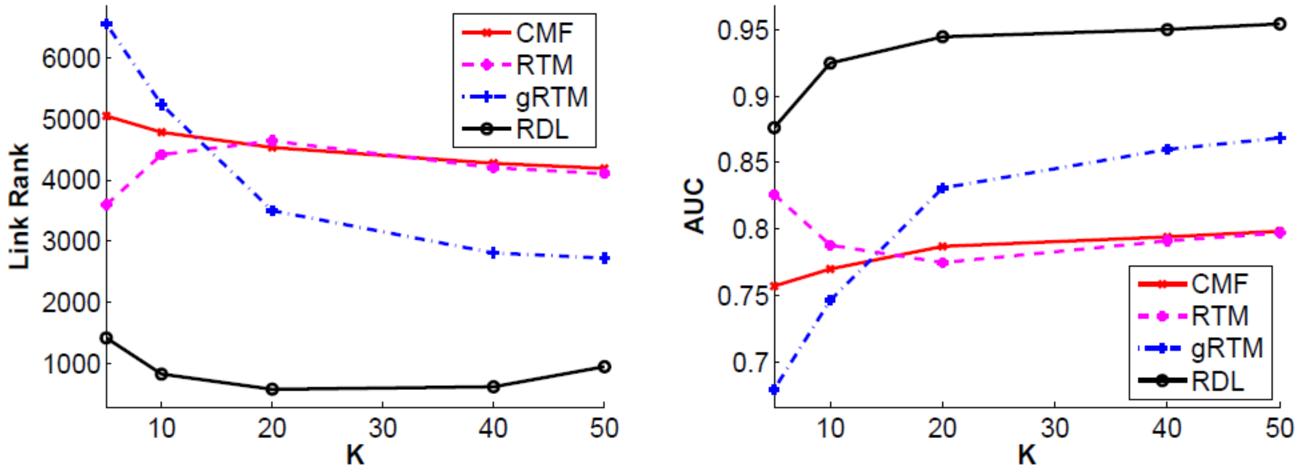


Figure 3: Link rank and AUC of compared models for *citeulike-t*. A 2-layer RDL is used.

Link rank: how high our predicted links rank in the ground truth
AUC: area under curve

Experiments: Link Rank and AUC

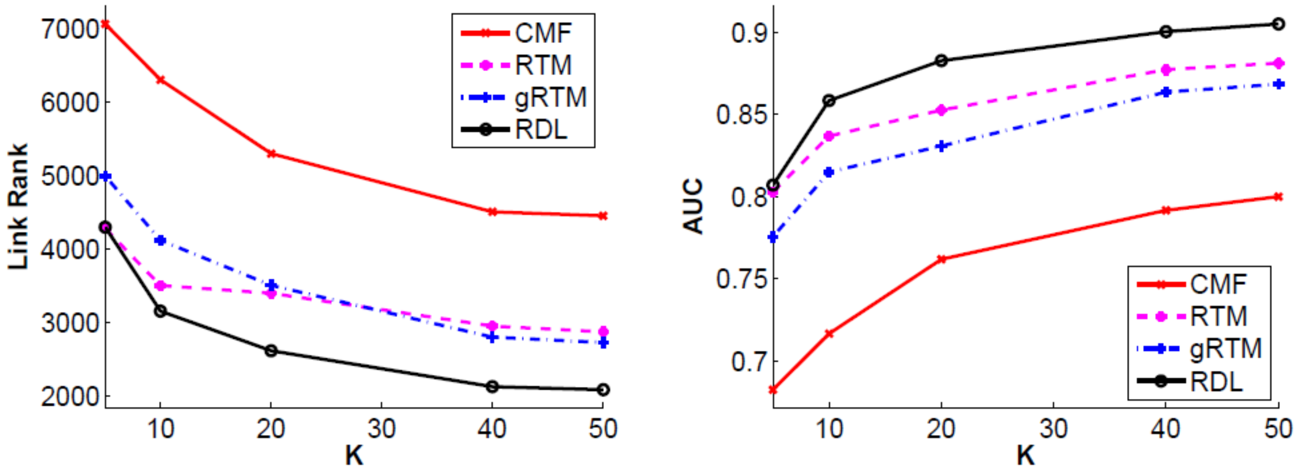


Figure 4: Link rank and AUC of compared models for *arXiv*. A 2-layer RDL is used.

Link rank: how high our predicted links rank in the ground truth
AUC: area under curve

Experiments: RDL Variants

Link rank of baselines (the first 3 columns) and
RDL variants (the last 4 columns) on three datasets (L = 4)

Method	VAE+BLR	VFAE+BLR	SDAE+BLR	MAPRDL	BSDAE1+BLR	BSDAE2+BLR	BayesRDL
citeulike-a	980.81	960.15	992.48	495.97	849.02	761.57	**473.59**
citeulike-t	1599.62	1531.16	1356.85	951.31	1341.15	1310.12	**911.31**
arXiv	3367.25	3316.29	2916.18	2028.72	2947.79	2708.17	**1982.84**

VAE: Variational Autoencoder
VRAE: Variational Fair Autoencoder
BLR: Bayesian Logistic Regression
BSDAE1: Bayesian treatment of probabilistic SDAE (mean only)
BSDAE2: Bayesian treatment of probabilistic SDAE (mean and variance)
MAPRDL: RDL with MAP inference
BayesRDL: RDL with full Bayesian treatment

Experiments: Depth

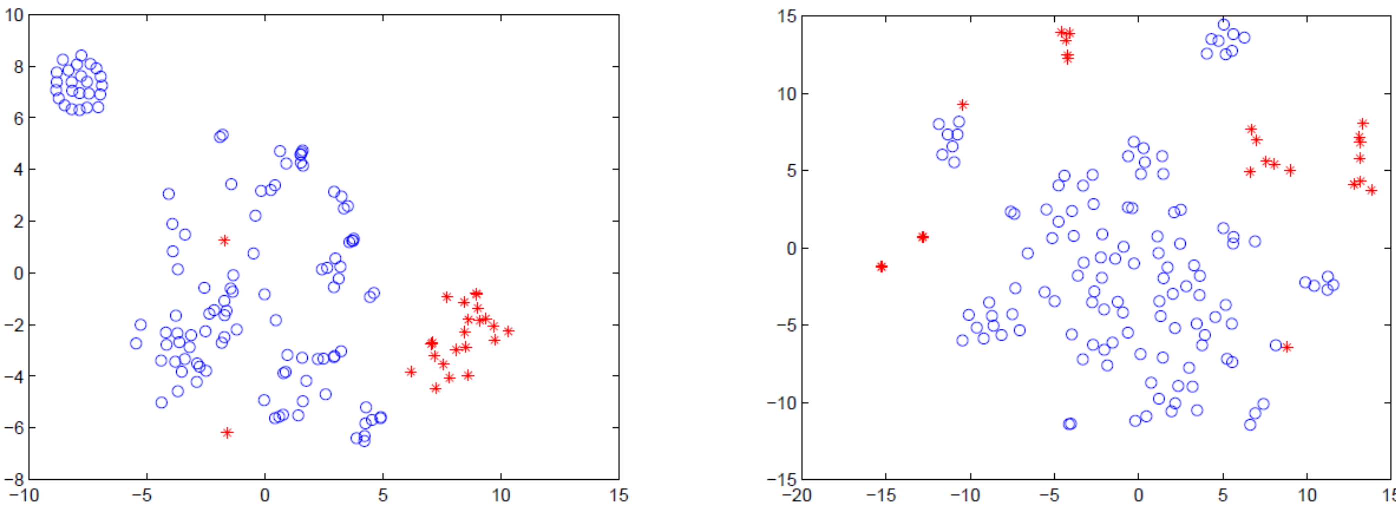
Performance of RDL with different number of layers (MAP)

	Link Rank			AUC		
	RDL-1	RDL-2	RDL-3	RDL-1	RDL-2	RDL-3
citeulike-a	825.74	495.97	**488.41**	0.939	**0.964**	0.963
citeulike-t	2060.17	951.31	**912.43**	0.894	0.954	**0.955**
arXiv	5241.97	**2080.72**	2730.08	0.755	**0.905**	0.855

Performance of RDL with different number of layers (Bayesian treatment)

	Link Rank			AUC		
	RDL-1	RDL-2	RDL-3	RDL-1	RDL-2	RDL-3
citeulike-a	789.85	473.59	**471.47**	0.946	**0.971**	0.970
citeulike-t	1904.83	911.31	**867.78**	0.906	0.956	**0.960**
arXiv	4965.01	**1982.84**	2612.12	0.801	**0.914**	0.866

Case Study: RDL and RTM



t-SNE visualization of latent factors learned by RDL (left) and RTM (right).

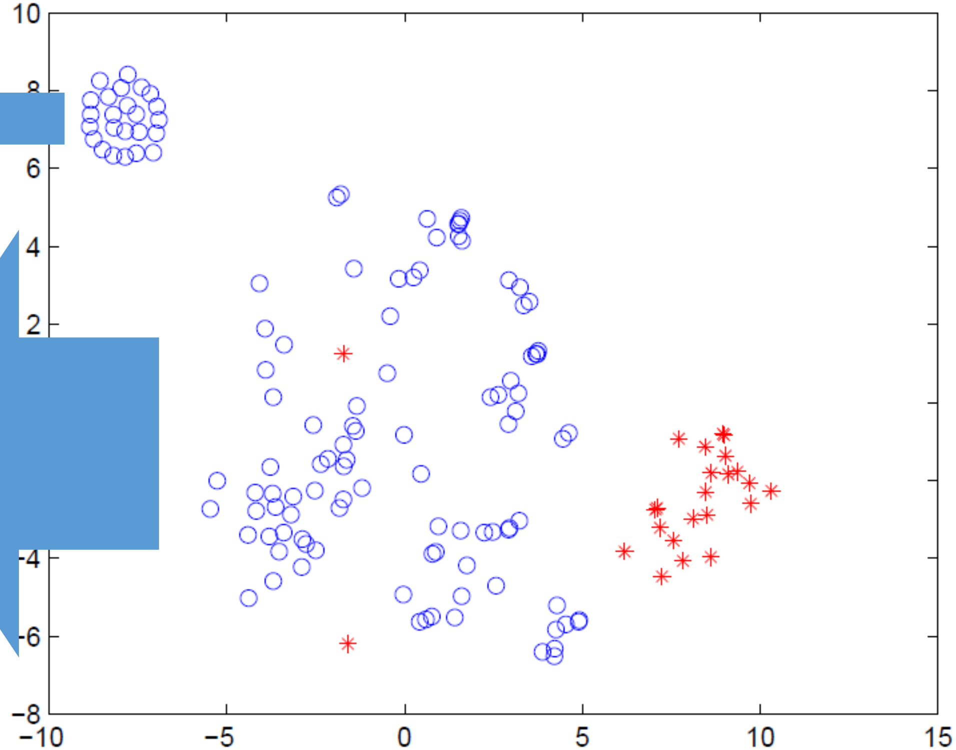
Target article:
From DNA sequence to transcriptional behaviour: a quantitative approach

∗ (red): articles with links to the target article
○ (blue): articles without links to the target article

Case Study: RDL

Articles written in **German**, which are rare in the datasets

Some **bestselling books**:
The 4-Hour Work Week
Mary Bell's Complete Dehydrator Cookbook



t-SNE visualization of latent factors learned by RDL.

Target article:
From DNA sequence to transcriptional behaviour: a quantitative approach

Case Study: RDL ang gRTM

Top 10 link predictions made by gRTM and RDL for two articles from citeulike-a

	Query: Object class recognition by unsupervised scale-invariant learning
gRTM	Layered depth images
	Using spin images for efficient object recognition in cluttered 3D scenes
	Snakes: active contour models
	Visual learning and recognition of 3-D objects from appearance
	Contextual priming for object detection
	Visual categorization with bags of keypoints
	Non-parametric model for background subtraction
	Alignment by maximization of mutual information
	Rapid object detection using a boosted cascade of simple features
	W4: real-time surveillance of people and their activities
RDL	**Distinctive image features from scale-invariant keypoints**
	visual learning and recognition of 3-D objects from appearance
	Object recognition with features inspired by visual cortex
	Unsupervised learning of models for recognition
	Robust object recognition with cortex-like mechanisms
	Generative versus discriminative methods for object recognition
	Using spin images for efficient object recognition in cluttered 3D scenes
	Learning generative visual models from few training examples
	3D object modeling and recognition using affine-invariant patches
	A Bayesian approach to unsupervised one-shot learning of object categories

Key Concepts

Object recognition

Unsupervised learning

Scale-invariant learning

Case Study: RDL ang gRTM

Top 10 link predictions made by gRTM and RDL for two articles from citeulike-a

	Query: SCOP database in 2004: refinements integrate structure and sequence family data
gRTM	Pfam: multiple sequence alignments and HMM-profiles of protein domains
	Structure, function and evolution of multidomain proteins
	Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB
	Nature of the protein universe
	The CATH domain structure database and related resources
	Phylogenetic classification of short environmental DNA fragments
	The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes
	LGA: a method for finding 3D similarities in protein structures
	Amino acid substitution matrices from protein blocks
	Multiple protein sequence alignment
RDL	The universal protein resource (UniProt)
	E-MSD: an integrated data resource for bioinformatics
	Gene3D: comprehensive structural and functional annotation of genomes
	The universal protein resource (UniProt) in 2010
	Gene3D: modelling protein structure, function and evolution
	The universal protein resource (UniProt): an expanding universe of protein information
	Pfam: clans, web tools and services
	The Pfam protein families database
	The protein data bank
	SCOP: a structural classification of proteins database

Key Concepts
Protein structures
Protein databases

- Motivation
- Bayesian Deep Learning
- Relational Deep Learning
- Parameter Learning
- Experiments
- Conclusion

Conclusion

- First Bayesian DL model for link prediction
- Joint Bayesian DL is beneficial
- Significant improvement on the state of the art
- RDL as representation learning

Future Work

- **Multi-relational data (co-author & citation networks)**
  - Boost predictive performance
  - Discover relationship between different networks
- **GVI for other neural nets (CNN/RNN) and BayesNets**
  - pSDAE + link prediction
  - pCNN + recommendation
  - pRNN + community detection
- **Replace probabilistic SDAE with other Bayesian neural nets**
  - Variational autoencoders
  - Natural-parameter networks

www.wanghao.in
hwangaz@connect.ust.hk