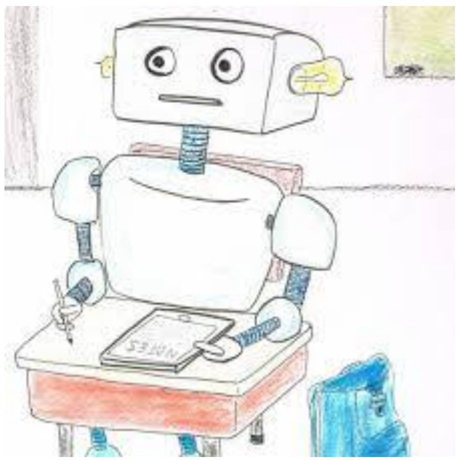
Bayesian deep learning: bridging the gap between probabilistic graphical models and deep learning

Hao Wang
The Hong Kong University of Science and Technology

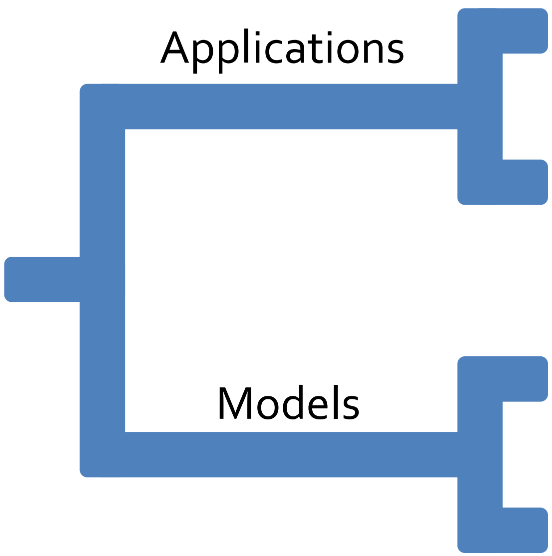
My Research Interest



Machine Learning

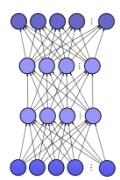
Data Mining

My Research Interest

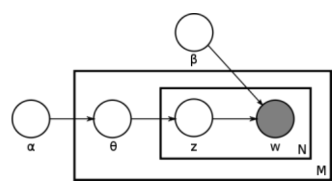


Recommender Systems

Social Network Analysis

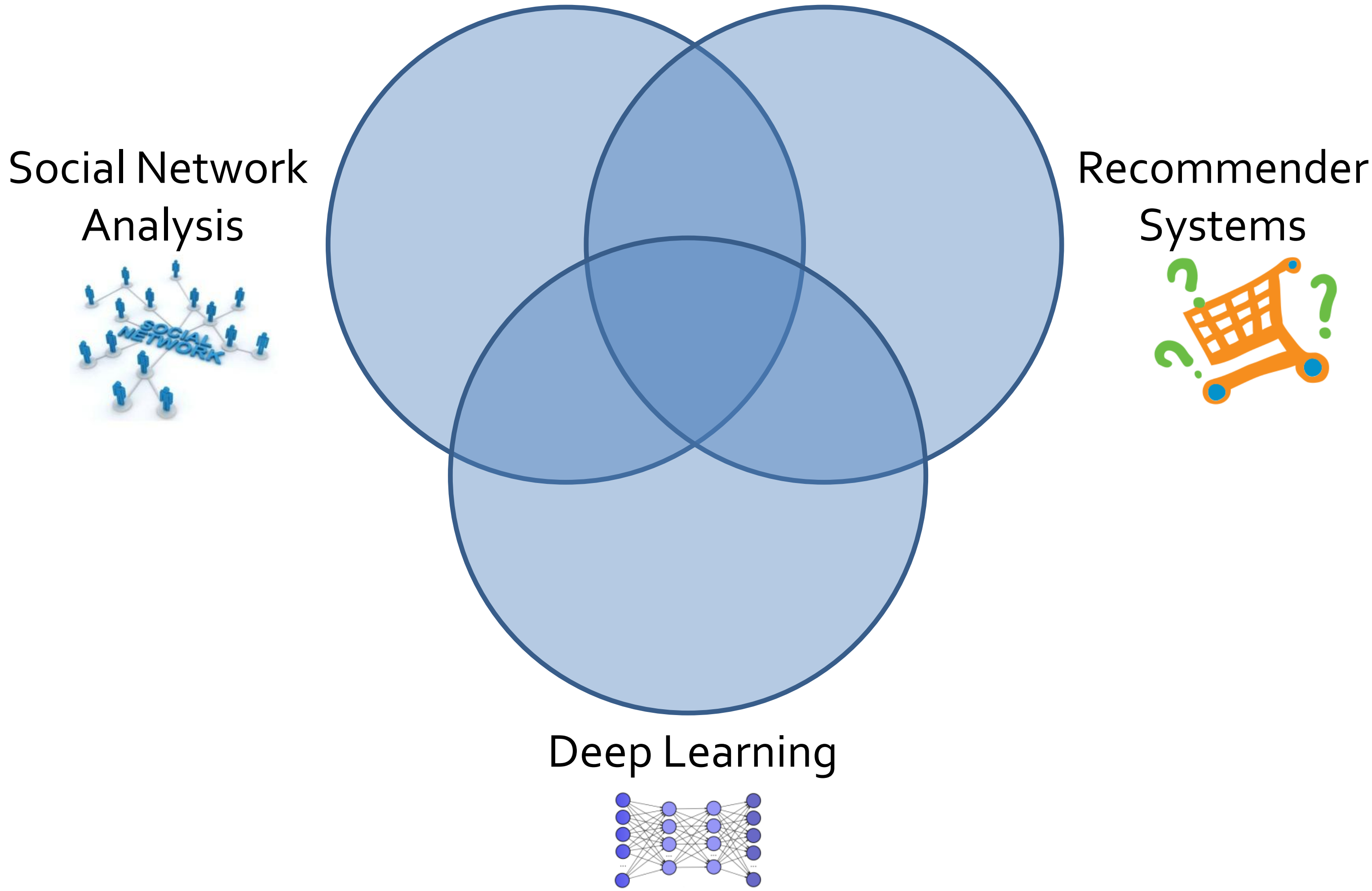


Deep Learning

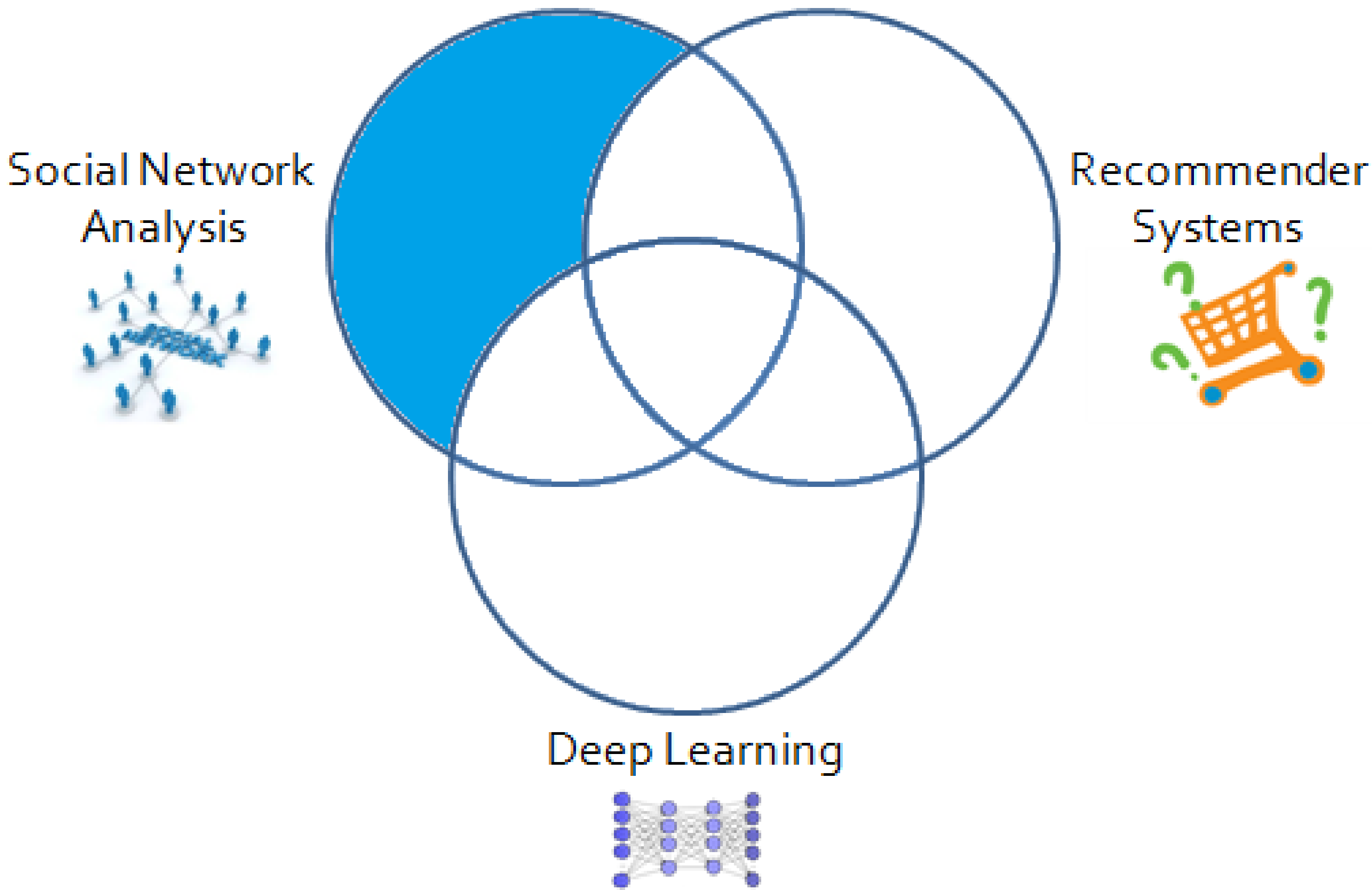


Probabilistic Graphical Models

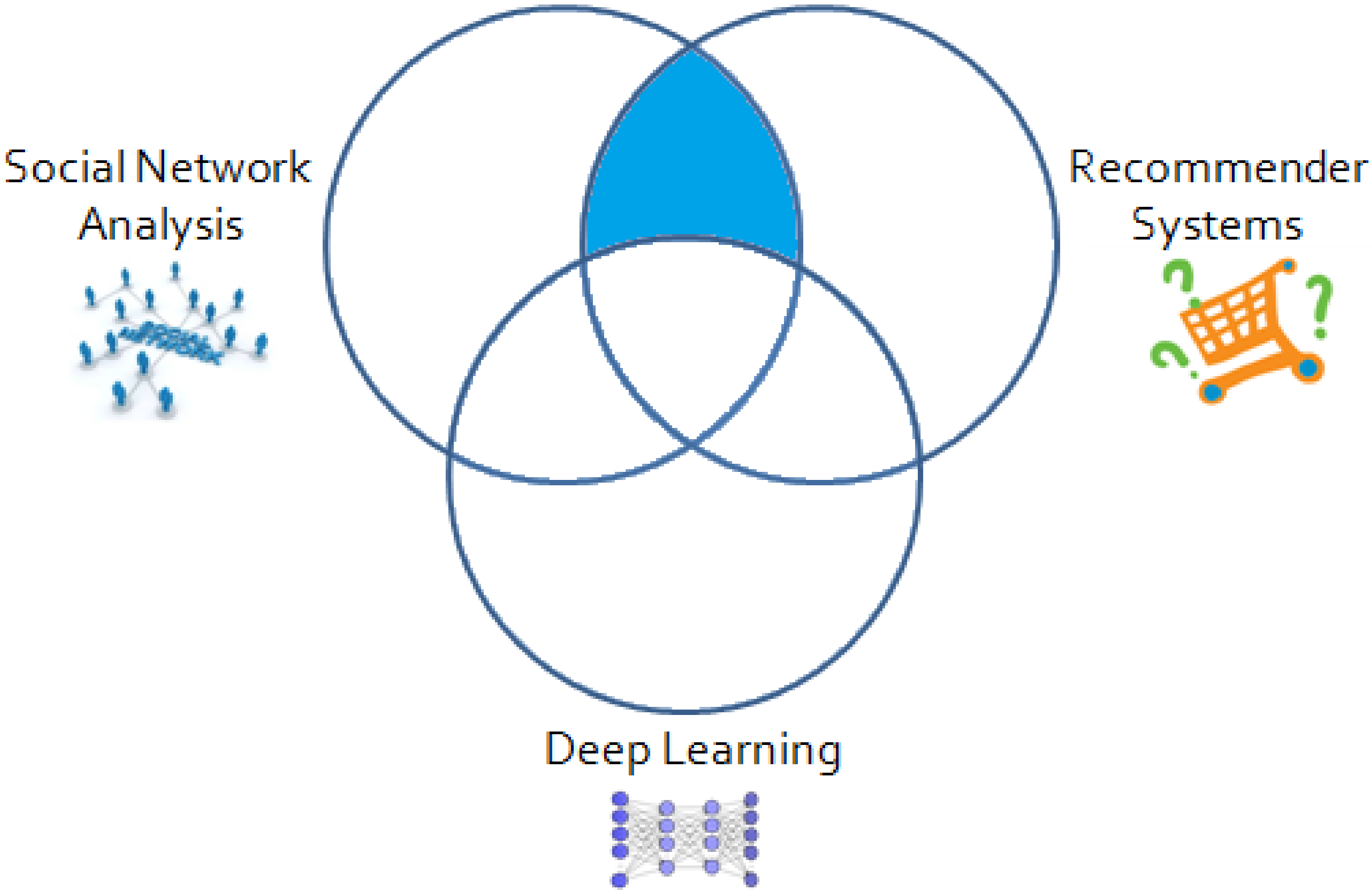
My Research Interest



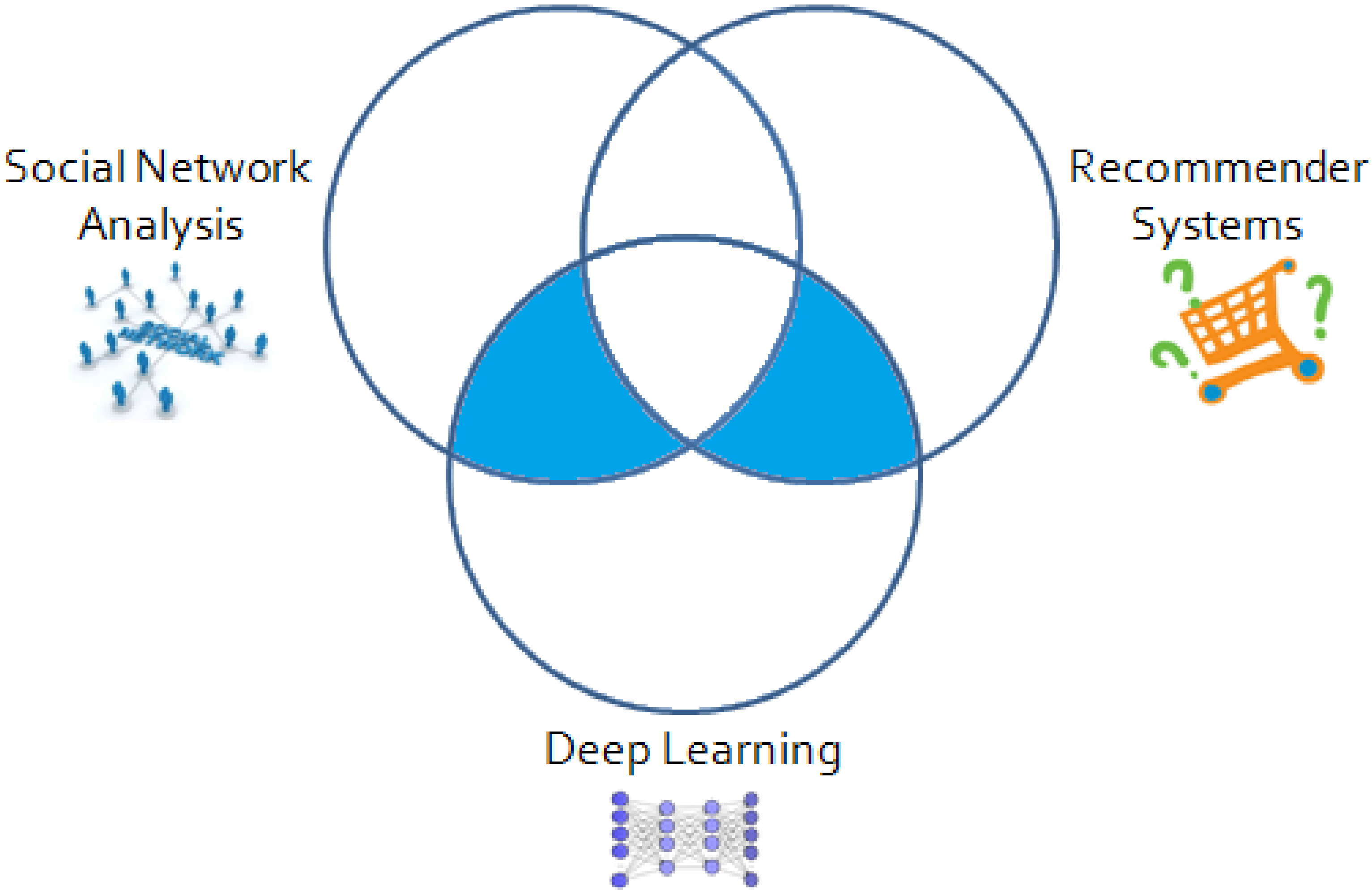
My Research Interest



My Research Interest



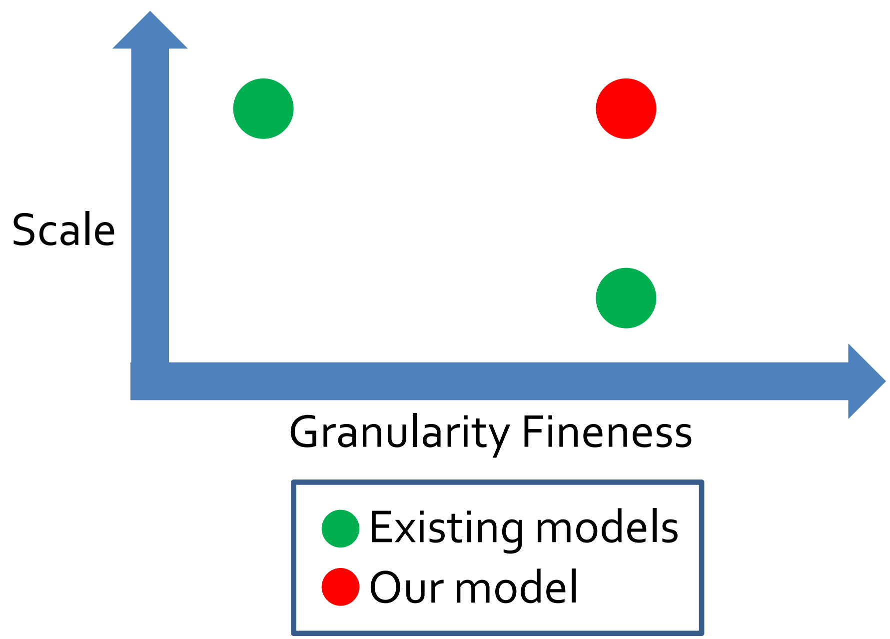
My Research Interest



Social Network Analysis

"Online Egocentric Models for Citation Network"
[Wang & Li, IJCAI 2013]

Motivation:



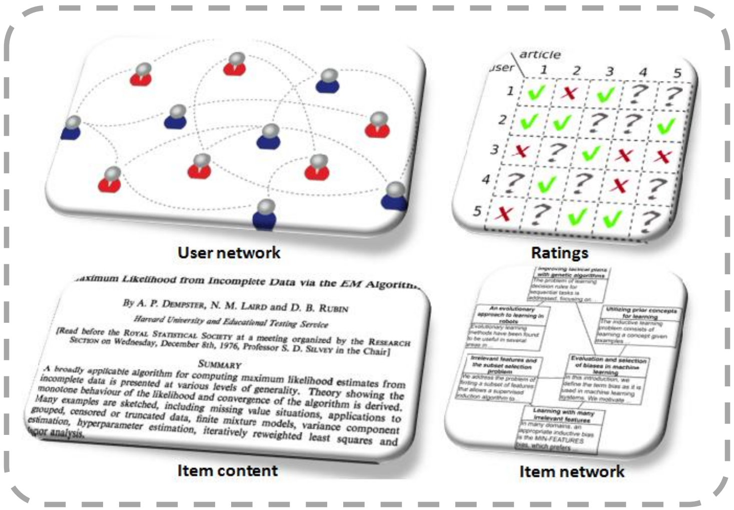
Social Network Analysis

"Online Egocentric Models for Citation Network"
[Wang & Li, IJCAI 2013]

Main Idea:

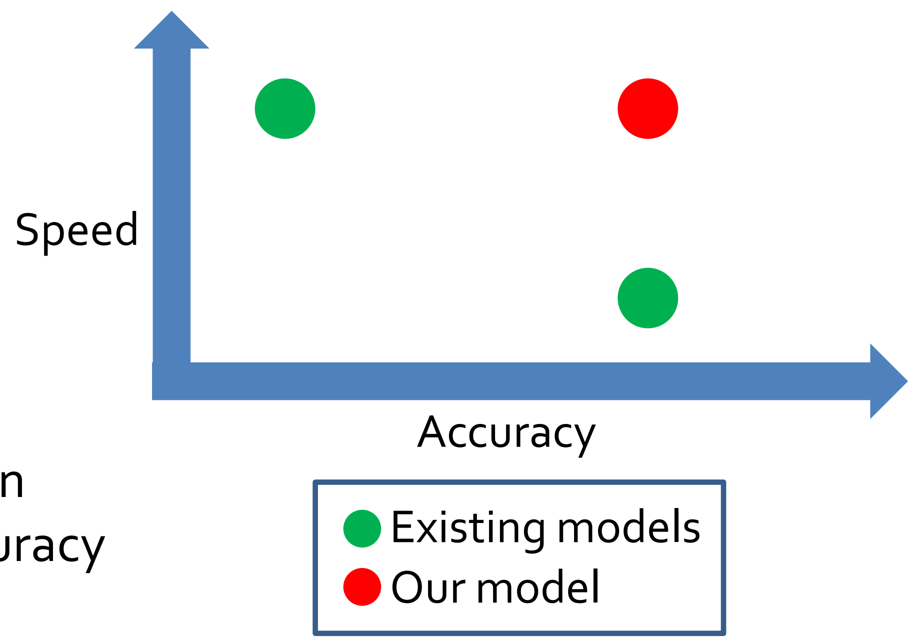
1. Convex formulation for online updates of models
2. Truncate insignificant terms to approximate optimal solutions and speed up training

Social Network Analysis Meets Recommender Systems

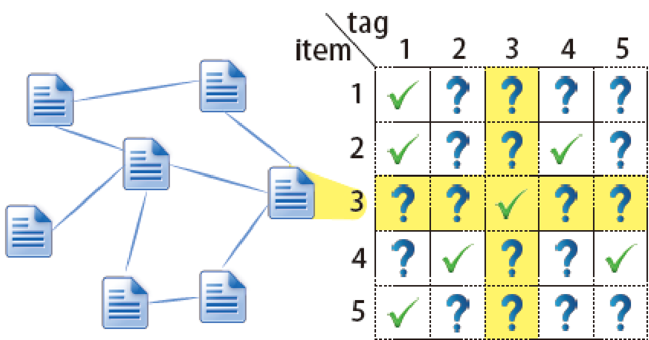


- Networks as auxiliary information
- Speed up training and boost accuracy

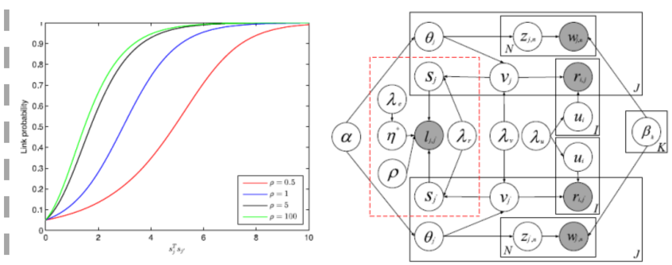
Motivation:



Social Network Analysis Meets Recommender Systems

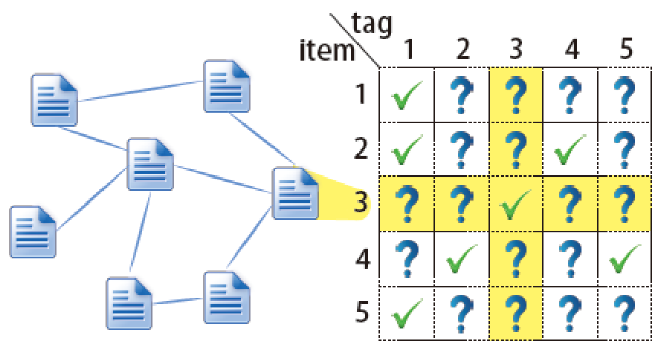


"Collaborative Topic Regression with Social Regularization for Tag Recommendation"
[Wang et al., IJCAI 2013]



"Relational Collaborative Topic Regression for Recommender Systems"
[Wang & Li, TKDE 2015]

Social Network Analysis Meets Recommender Systems



"Collaborative Topic Regression with Social Regularization for Tag Recommendation"

[Wang et al., IJCAI 2013]

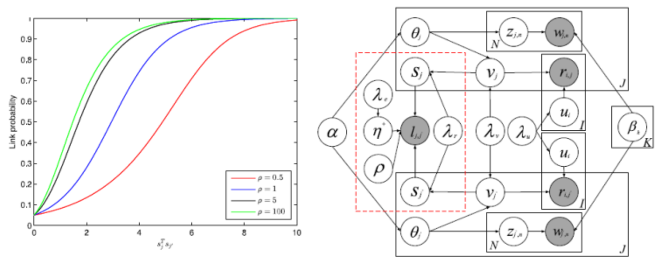
Main Idea:

1. Use network information as a prior to regularize the model
2. Use product of Gaussians to bridge heterogeneous information

	Title	**How much can behavioral targeting help online advertising?**		
Article information	Top topic 1	web, search, engine, pages, keyword		
	Top topic 2	mobile, phones, attitudes, advertising, consumer		
	7 true tags	behavioral_targeting, advertising, ads, user_profile, computational_advertising, recommend, user-behavior		
	CTR (baseline)	TRUE	CTR-SR (our method)	TRUE
Top 10 recommended tags	1. random-walks	no	1. behavioral_targeting	**yes**
	2. page-rank	no	2. ads	**yes**
	3. computational_advertising	**yes**	3. computational_advertising	**yes**
	4. citizen-science	no	4. random-walks	no
	5. natural_history	no	5. page-rank	no
	6. search_engine	no	6. developing	no
	7. engine	no	7. recommend	**yes**
	8. searchengine	no	8. advertising	**yes**
	9. what	no	9. what	no
	10. re-ranking	no	10. need	no

[Wang et al., IJCAI 2013]

Social Network Analysis Meets Recommender Systems

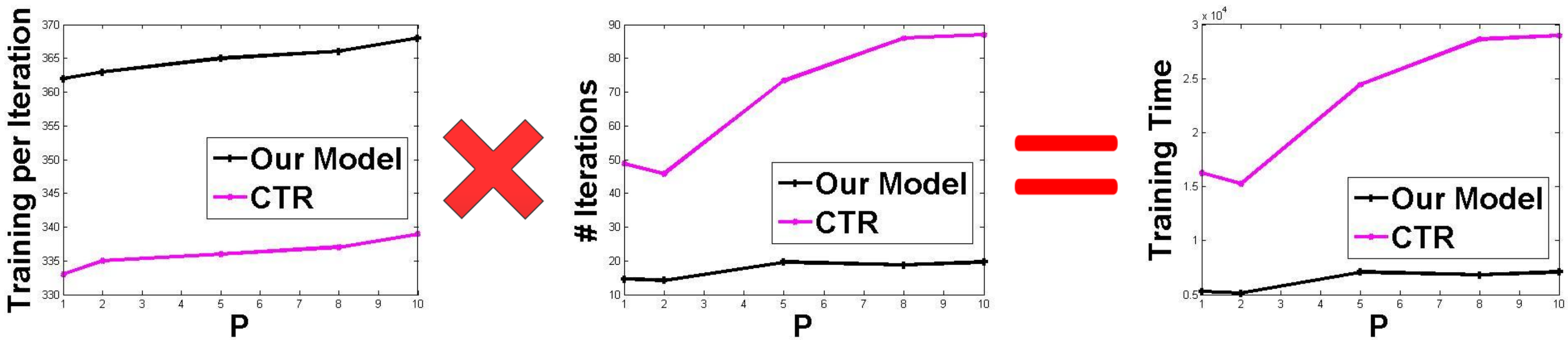


"Relational Collaborative Topic Regression for Recommender Systems"
[Wang & Li, TKDE 2015]

Main Idea:

1. Use network information as observed variables
2. A continuous family of link probability functions
3. Use auxiliary information to speed up convergence and cut training time

Cut Training Time

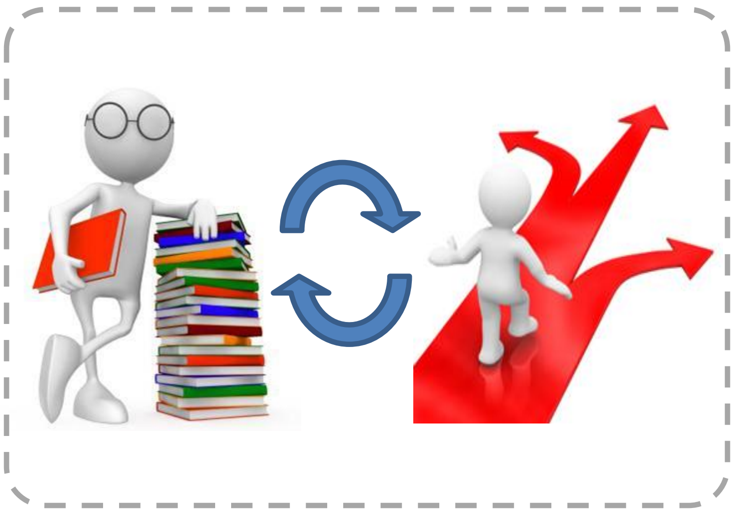


Training Time per Iteration X Number of Iterations = Total Training Time

(P: Density of ratings in the training set)

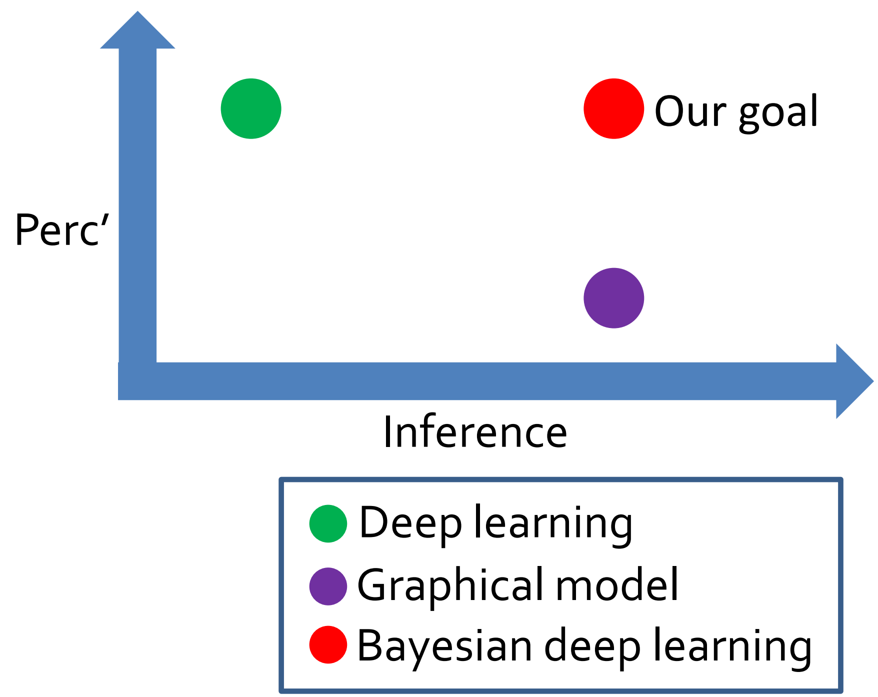
[Wang & Li, TKDE 2015]

Bayesian Deep Learning

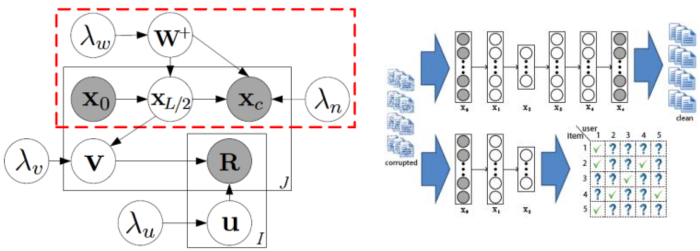


Deep Learning & Graphical Models
Perception & Inference

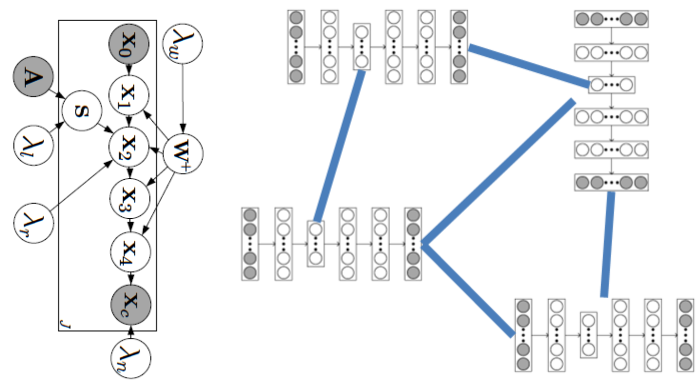
Motivation:



Bayesian Deep Learning

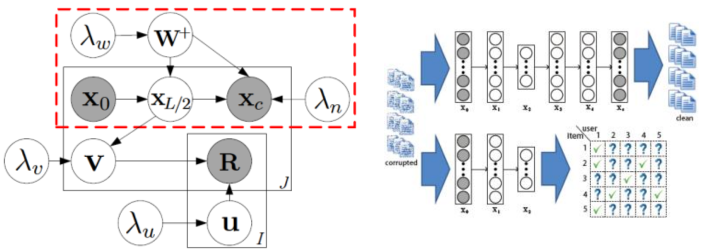


"Collaborative Deep Learning for Recommender Systems"
[Wang et al., KDD 2015]



"Relational Stacked Denoising Auto-encoder for Tag Recommendation"
[Wang et al., AAAI 2015]

Deep Learning
Meets Recommender Systems

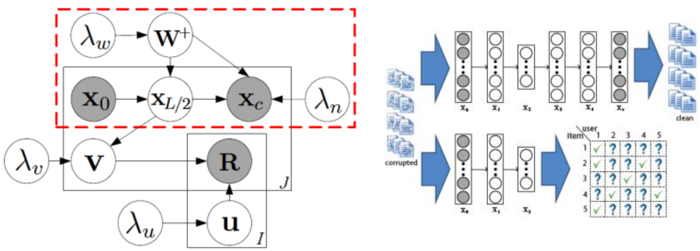


"Collaborative Deep Learning for Recommender Systems"
[Wang et al., KDD 2015]

Motivation:

- Deep learning is good at perception, not recommendation
- Collaborative Filtering is good at recommendation, not perception
- Combine the power of the two

Deep Learning
Meets Recommender Systems

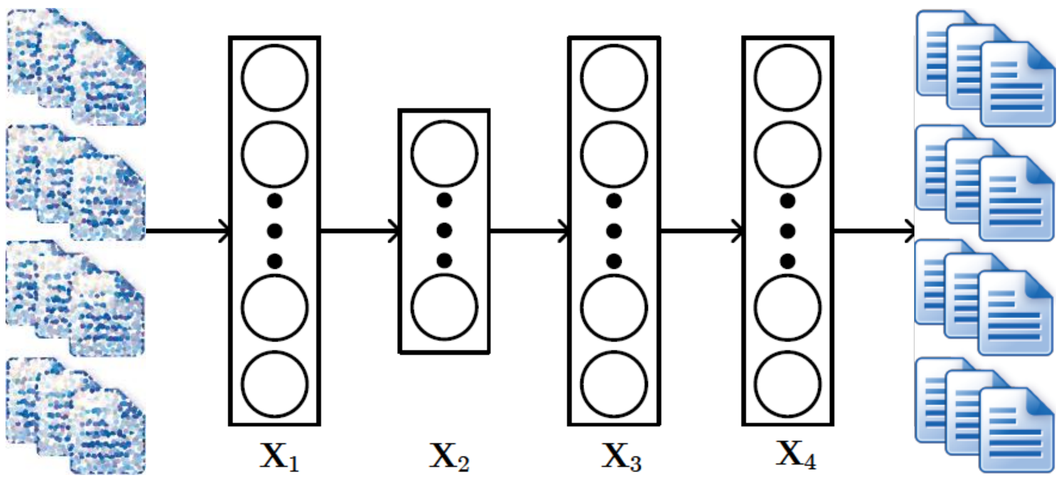


"Collaborative Deep Learning for Recommender Systems"
[Wang et al., KDD 2015]

Main Idea:

1. A unified probabilistic graphical model
2. Break the i.i.d. assumption
3. Easy to incorporate auxiliary information

Stacked Denoising Autoencoders (SDAE)



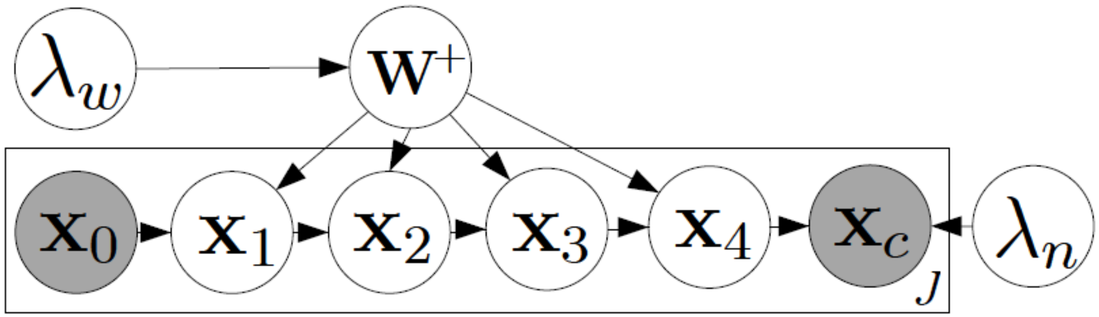
Corrupted data

Clean data

[Vincent et al. 2010]

Probabilistic SDAE

Graphical model:



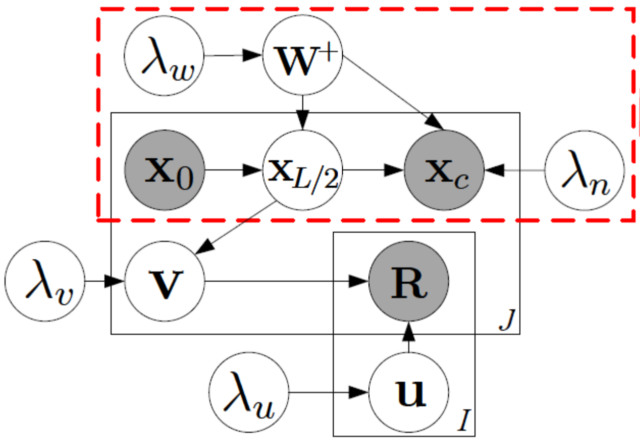
Notation:

- $\mathbf{x}_0$ corrupted data
- $\mathbf{x}_c$ clean data
- $\mathbf{W}^+$ weights and biases

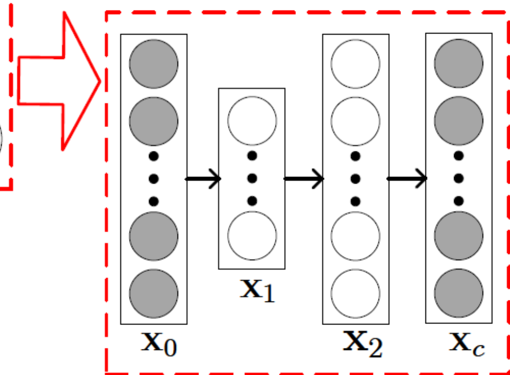
[Wang et al., KDD 2015]

Collaborative Deep Learning

Graphical model:



Collaborative deep learning



SDAE

Two-way interaction

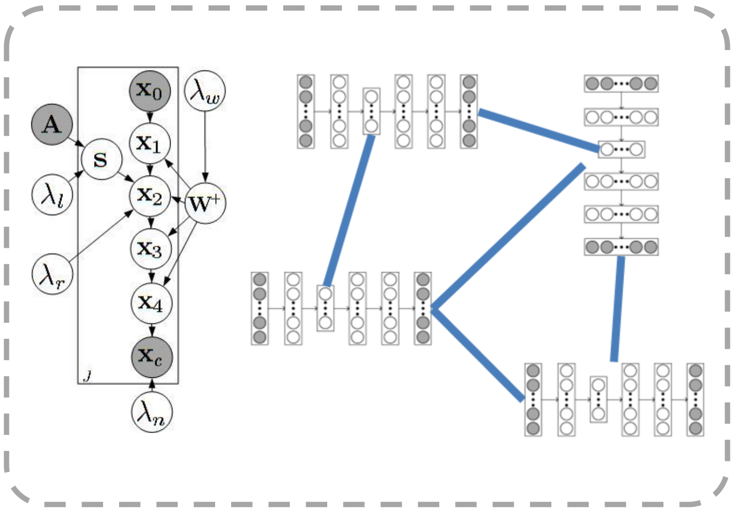
- Powerful representation
- Infer missing ratings
- Infer missing content

Notation:

- $\mathbf{R}$ rating of item j from user i
- $\mathbf{v}$ latent vector of item j
- $\mathbf{u}$ latent vector of user i
- $\mathbf{x}_0$ corrupted data
- $\mathbf{x}_c$ clean data
- $\mathbf{W}^+$ weights and biases
- $\mathbf{x}_{L/2}$ content representation

[Wang et al., KDD 2015]

Deep Learning
Meets Social Network Analysis

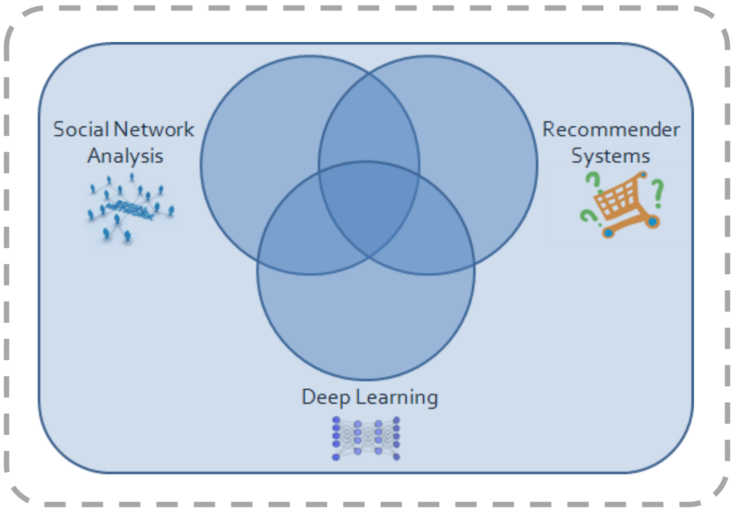


"Relational Stacked Denoising Auto-encoder for Tag Recommendation"
[Wang et al., AAAI 2015]

Main Idea:

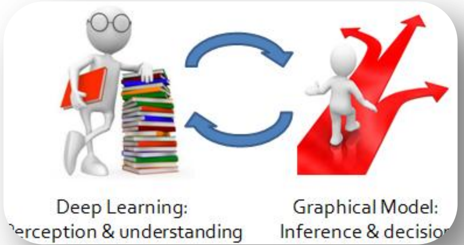
1. Connected items have similar features
2. Design a graphical model to incorporate network information
3. Can be extended for multiple networks

Future Goal



General Framework:

1. Ability of understanding text, images, and videos
2. Ability of inference and planning under uncertainty
3. Close the gap between human intelligence and artificial intelligence



Thanks!
Q&A

September 11th , 2015