Uncertainty Aware Semi-Supervised Learning on Graph Data

Xujiang Zhao[1], Feng Chen[1], Shu Hu[2], Jin-Hee Cho[3]

[1]The University of Texas at Dallas, [2]University at Buffalo, SUNY, [3]Virginia Tech

Why is predicting uncertainty important?

Uncertainty vs. Misclassification

When our model only knows `car’ and `road,’

1 *Incorrect prediction*

High Uncertainty

Misclassify “car” as “road”

2 *Out-of-distribution*

High Uncertainty

?

Misclassify “deer” (OOD object) as “car”

Is it important to know:

- ✓ why we don’t know?
- ✓ how much we don’t know?

So how can we predict the uncertainty based on its root cause?

Would it really help for our decision making?

Types of uncertainty

1. *Epistemic* uncertainty (a.k.a. model/parameter uncertainty)
   - Measures what model doesn't know
   - Due to **limited data and knowledge**

   Aleatoric uncertainty (a.k.a. data uncertainty)
   - Measures what you can't understand from the data
   - Due to **randomness**

   → Probabilistic Uncertainty

2. ***Vacuity* uncertainty** (a.k.a. ignorance)
   - Measures uncertainty due to **a lack of evidence**

   ***Dissonance* uncertainty**
   - Measures uncertainty due to **conflicting evidence**

   → Evidential Uncertainty

Evidential Uncertainty

Task: 3 class image classification

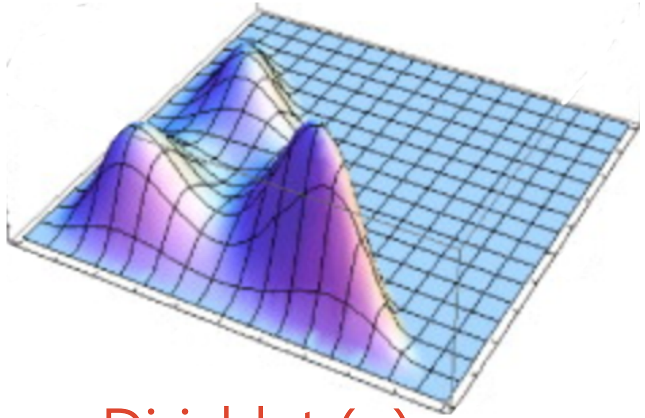
Training Data: Dog (e_1 = 10 images) Cat (e_2 = 10 images) Pig (e_3 = 10 images)

$e = [e_1, \ldots, e_K]$ Evidence (Historical observations)

Subjective Opinion

$\omega = (\boldsymbol{b}, u, \boldsymbol{a})$

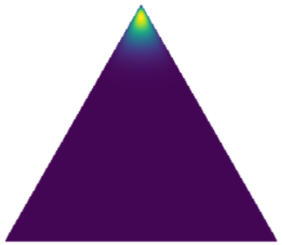
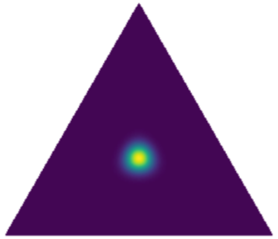
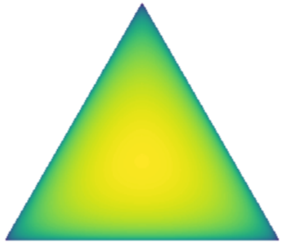
$\alpha = e + 1$



Dirichlet (α)

A subjective opinion modeled based on `Subjective Logic' which uses Dirichlet distribution to measure multiple dimensions of uncertainty in classification tasks

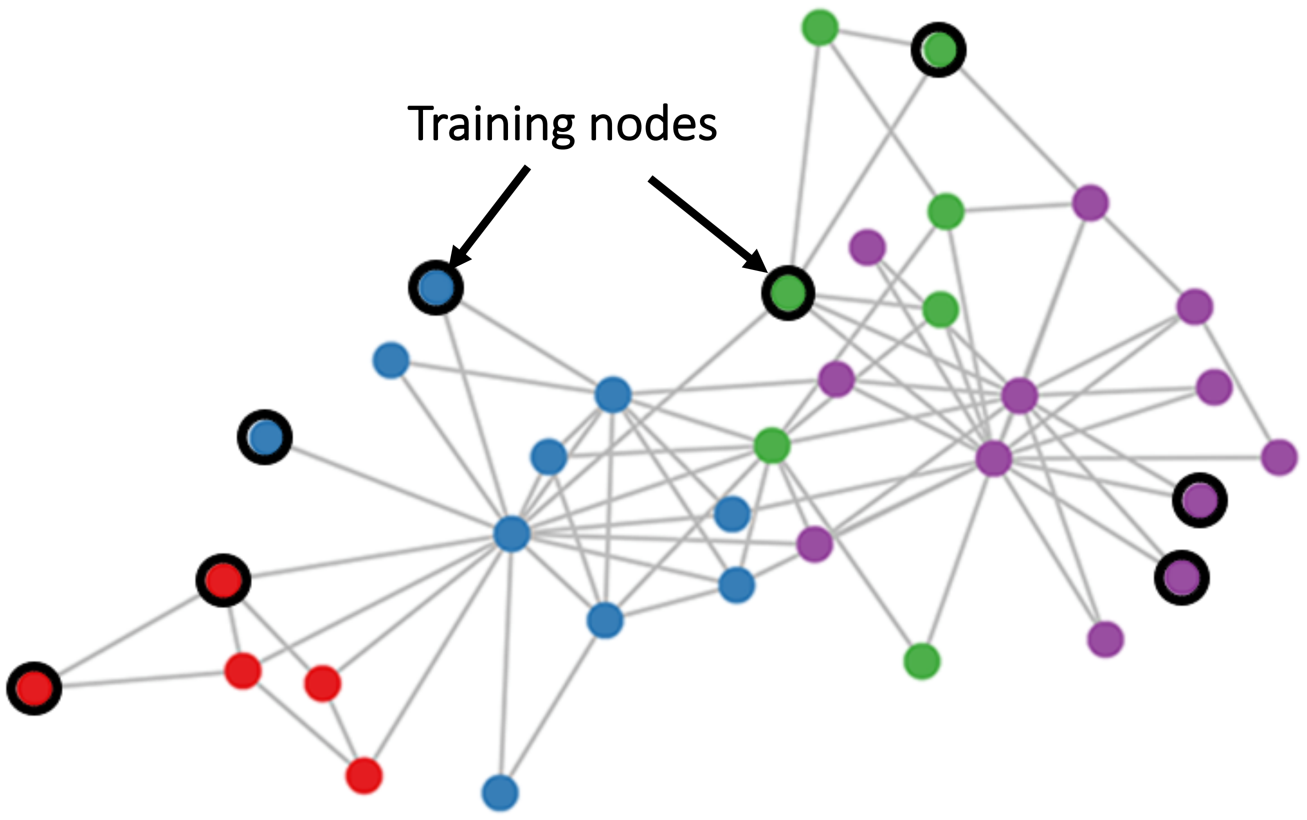
Why Evidential Uncertainty?

	Confidence Prediction	Conflict Prediction	Out-of-Distribution
			
Dirichlet Distribution	α = [11, 1, 1]	α = [11, 11, 11]	α = [1, 1, 1]
Expected Probability	p = [0.83, 0.083, 0.083]	p = [1/3, 1/3, 1/3]	p = [1/3, 1/3, 1/3]
	Low Uncertainty	High Dissonance (conflicting evidence)	High Vacuity (lack of evidence)
Test image			

Different Vacuity

Sample probability

Problem Formulation



Semi-supervised learning (SSL)

Given: graph $\mathcal{G} = (\mathbb{V}, \mathbb{E}, \mathbf{r}, \mathbf{y}_{\mathbb{L}})$

$\mathbb{V}$: Node

$\mathbb{E}$: Edge

$\mathbf{r}$: node-level feature

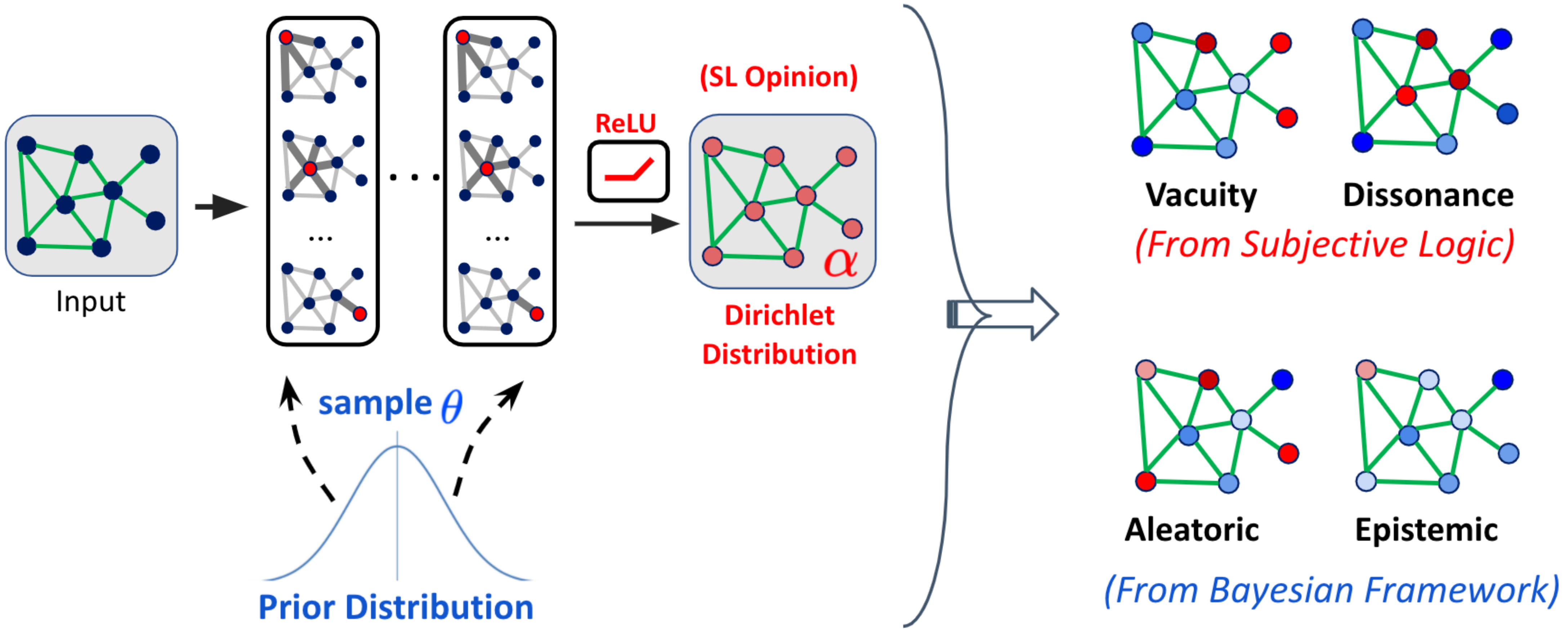
$\mathbf{y}_{\mathbb{L}}$: training labels (*K* classes)
a small set of training node (black circle)

Goal:

- class probabilities $\boldsymbol{p}$
- **multidimensional uncertainty $\boldsymbol{u}$**

Vacuity, dissonance, epistemic, aleatoric

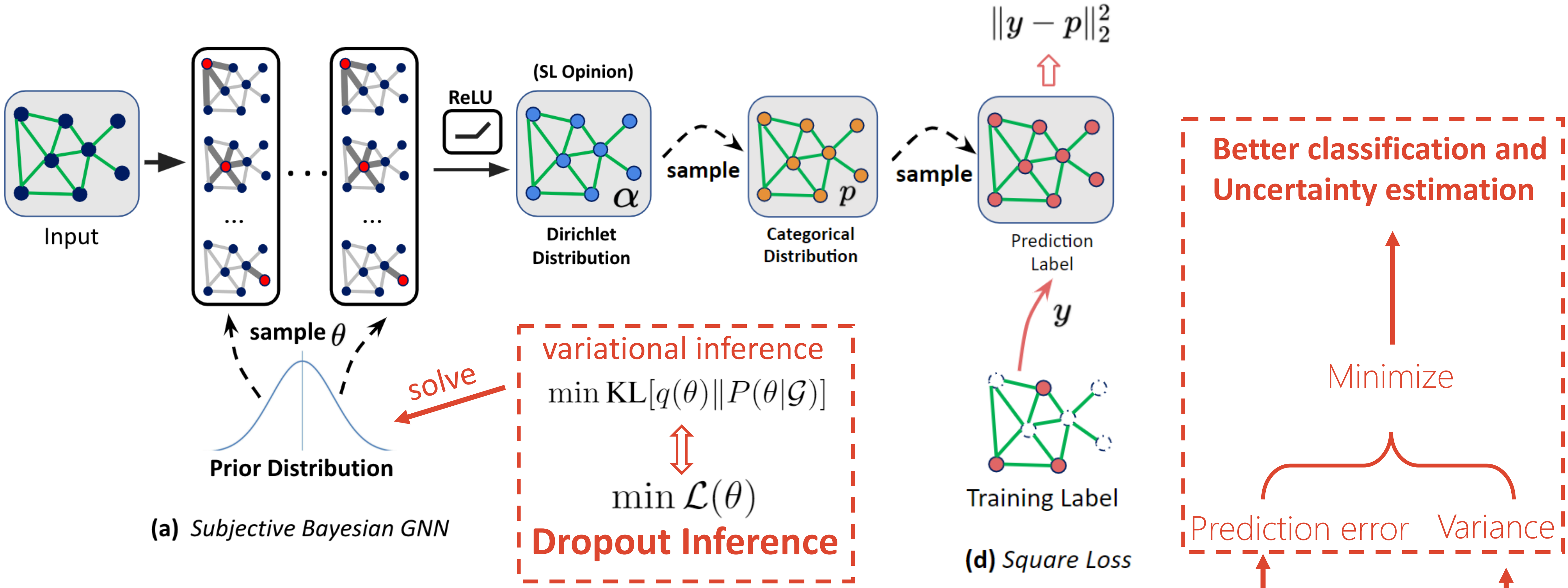
Uncertainty Aware Framework



Subjective Bayesian GNN

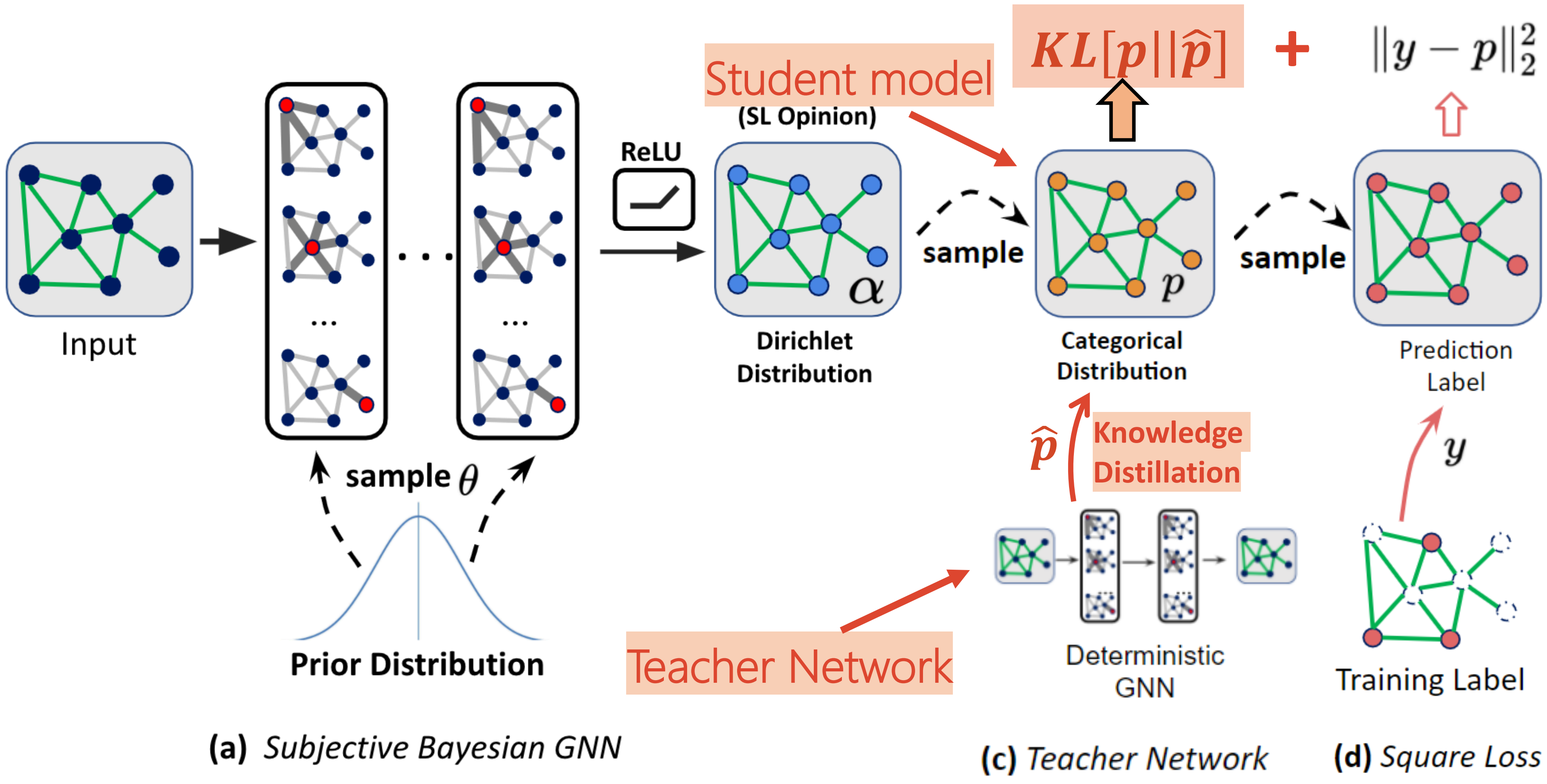
Multiple Uncertainties

Training Uncertainty Framework

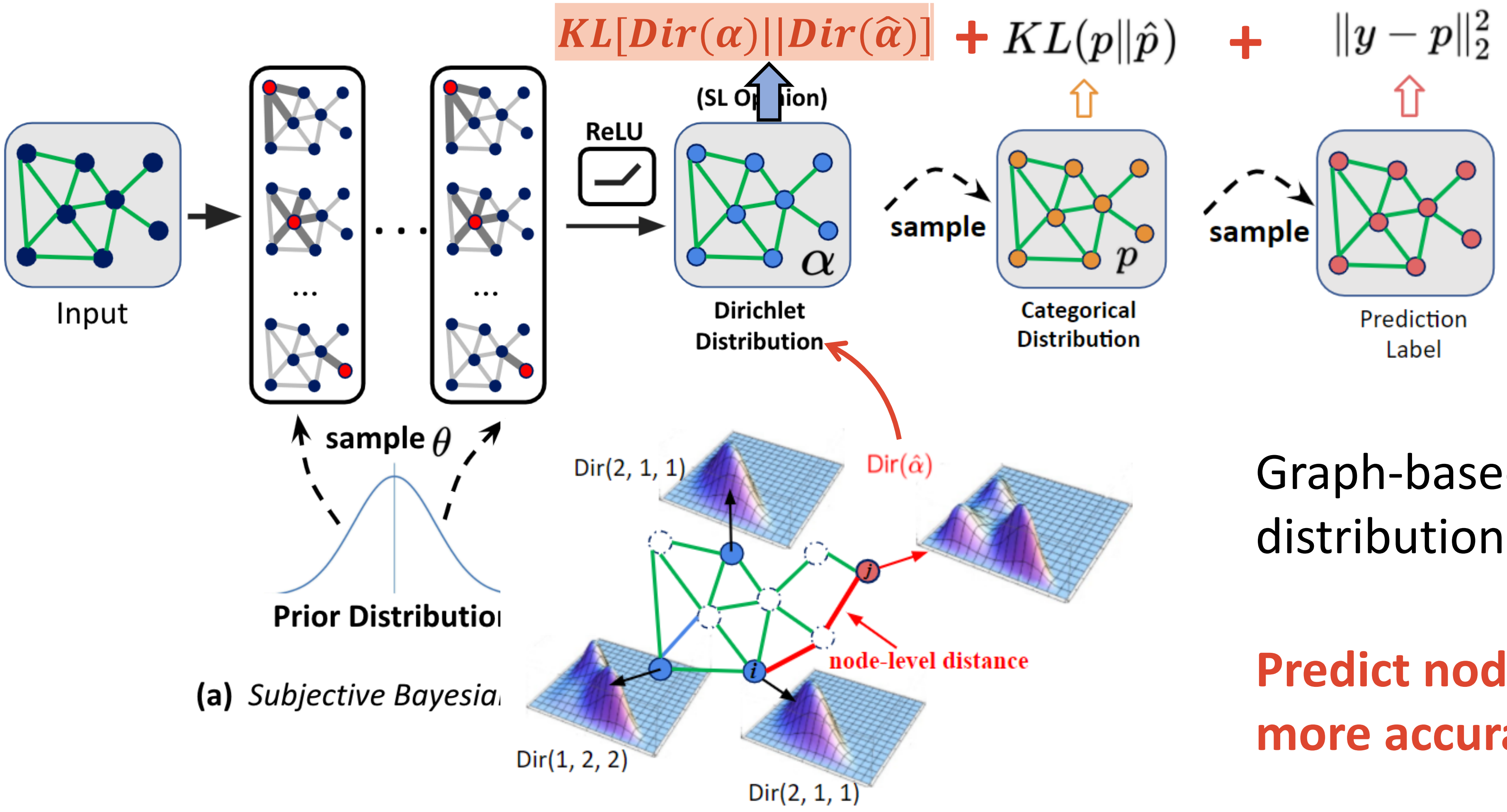


$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i\in\mathbb{L}} \int \|\mathbf{y}_i - \mathbf{p}_i\|_2^2 \cdot P(\mathbf{p}_i|A,\mathbf{r};\boldsymbol{\theta})d\mathbf{p}_i = \sum_{i\in\mathbb{L}} \sum_{k=1}^{K} \left(y_{ik} - \mathbb{E}[p_{ik}]\right)^2 + \mathrm{Var}(p_{ik}),$$

Training Uncertainty Framework



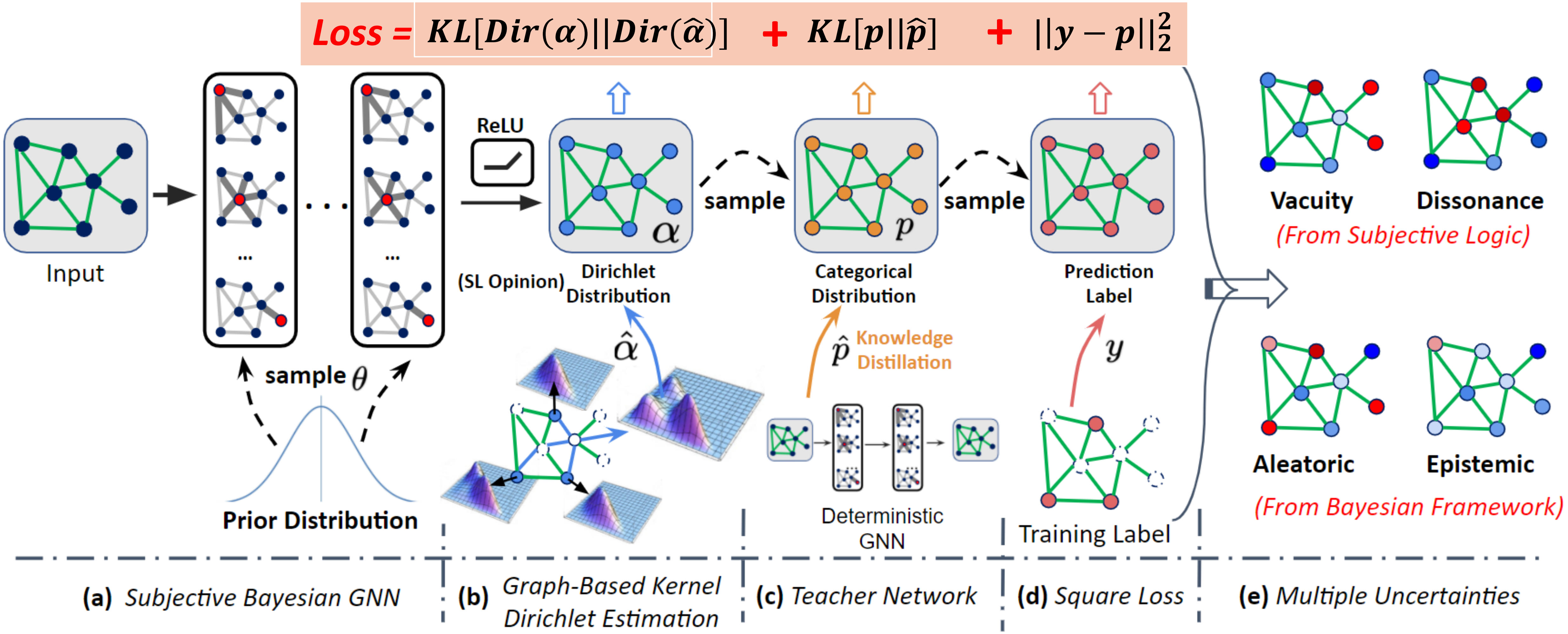
Training Uncertainty Framework



Graph-based Kernel Dirichlet distribution Estimation (GKDE)

Predict node-level Dirichlet more accurately

Training Uncertainty Framework



Key Theoretical Results

- Relations between multiple uncertainties
- Impact of Graph-based Kernel Dirichlet distribution Estimation

Relationships Between Multiple Uncertainties

Consider a simplified scenario:

y: a multinomial random variable

y ~ Cal(p): follows a K-class categorical distribution

p ~ Dir(α): the class probabilities p follow a Dirichlet distribution

Probabilistic Uncertainty

$$\underbrace{\mathcal{I}[y, \mathbf{p}|\boldsymbol{\alpha}]}_{\text{Epistemic}} = \underbrace{\mathcal{H}\Big[\mathbb{E}_{\mathrm{Prob}(\mathbf{p}|\boldsymbol{\alpha})}[P(y|\mathbf{p})]\Big]}_{\text{Entropy}} - \underbrace{\mathbb{E}_{\mathrm{Prob}(\mathbf{p}|\boldsymbol{\alpha})}\Big[\mathcal{H}[P(y|\mathbf{p})]\Big]}_{\text{Aleatoric}}.$$

Relationships Between Multiple Uncertainties

Consider a simplified scenario:

y: a multinomial random variable

y ~ **Cal(p)**: follows a K-class categorical distribution

p ~ **Dir(α)**: the class probabilities p follow a Dirichlet distribution

Evidential Uncertainty

$$\mathrm{Dir}(\alpha) \Longleftrightarrow \omega = (\boldsymbol{b}, u, \boldsymbol{a})$$

$$vac(\omega) = K / \sum_{k=1}^{K} \alpha_k$$

Vacuity (lack of evidence)

$$diss(\omega) = \sum_{k=1}^{K} \left(\frac{b_k \sum_{j \neq k} b_j \mathrm{Bal}(b_j, b_k)}{\sum_{j \neq k} b_j} \right)$$

Dissonance (conflict evidence)

Relationships Between Multiple Uncertainties

General relations on all prediction scenarios.

vacuity + dissonance ≤ 1

- High vacuity → Low dissonance
- High dissonance → Low vacuity

Would not increase at same time!

Evidence	$e = [1, 2]$	$e = [2, 2]$	$e = [2, 200]$	$e = [200, 200]$
	lack of evidence	lack of evidence	confidence	conflicting evidence
Vacuity	**High**	**High**	Low	Low
dissonance	Low	Low	Low	**High**

Relationships Between Multiple Uncertainties

General relations on all prediction scenarios.

vacuity > epistemic

- vacuity is an upper bound of epistemic uncertainty
- On a sufficiently large amount of evidence available, vacuity and epistemic would close to zero.

Relationships Between Multiple Uncertainties

Special relations on the OOD

1 = vacuity = entropy > aleatoric > epistemic > dissonance = 0

Special relations on the Conflicting Prediction

entropy = 1, dissonance → 1, aleatoric → 1, vacuity → 0, epistemic → 0

entropy > aleatoric > dissonance > vacuity > epistemic

- entropy cannot distinguish different types of uncertainty due to different root causes.

Relationships Between Multiple Uncertainties

Special relations on the OOD

1 = vacuity = entropy > aleatoric > epistemic > dissonance = 0

Special relations on the Conflicting Prediction

entropy = 1, dissonance → 1, aleatoric → 1, vacuity → 0, epistemic → 0

entropy > aleatoric > dissonance > vacuity > epistemic

- a high aleatoric uncertainty value and a low epistemic uncertainty value are observed under both cases.

Relationships Between Multiple Uncertainties

Special relations on the OOD

1 = vacuity = entropy > aleatoric > epistemic > dissonance = 0

Special relations on the Conflicting Prediction

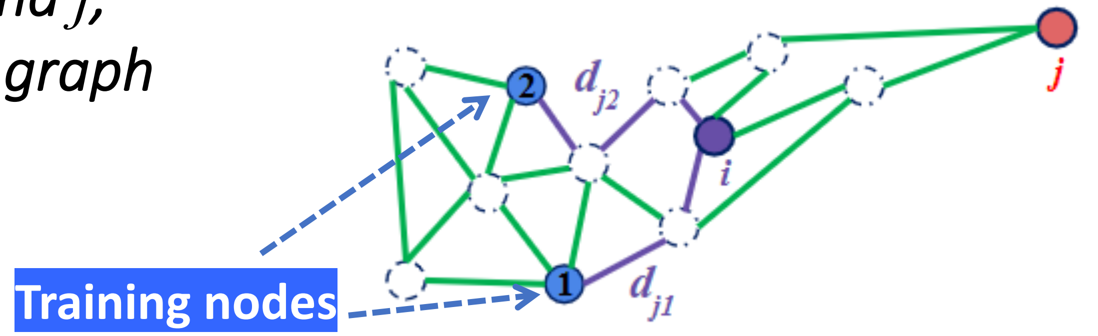
entropy = 1, dissonance → 1, aleatoric → 1, vacuity → 0, epistemic → 0

entropy > aleatoric > dissonance > vacuity > epistemic

- **vacuity and dissonance can clearly distinguish OOD from a conflicting prediction.**

Impact of Graph-based Kernel Dirichlet distribution Estimation

Given L training nodes and two testing nodes i and j,
Let $\boldsymbol{d}_i = [d_{i1}, \ldots, d_{iL}]$, $\boldsymbol{d}_i = [d_{j1}, \ldots, d_{jL}]$ *be the graph distance from training nodes.*



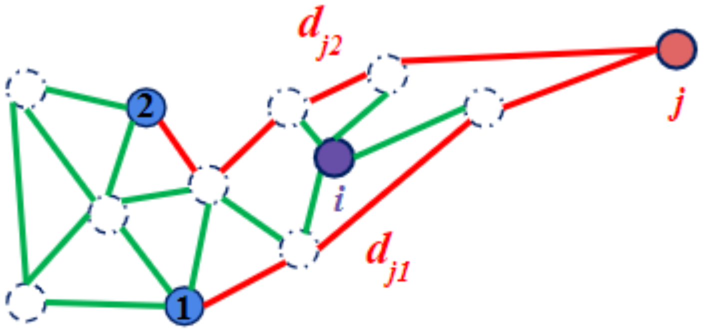
If for all $l \in \{1, \ldots, L\}, d_{il} \leq d_{jl}$, *we have*

$$\widehat{\boldsymbol{vacuity}}_i \leq \widehat{\boldsymbol{vacuity}}_j$$

$$d_{i1} < d_{j1}$$
$$d_{i2} < d_{j2}$$

estimated based on GKDE.

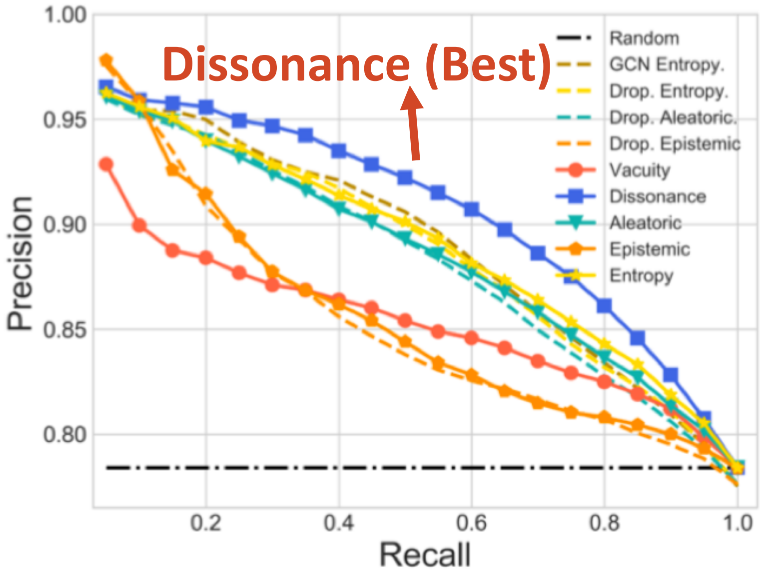
High vacuity occurs when testing node far away from training nodes.



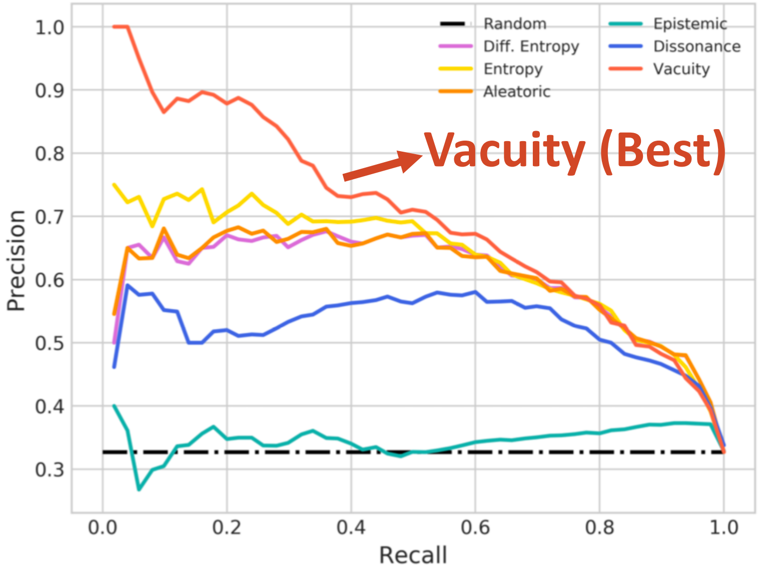
Key Experiment Result

Misclassification Detection

OOD Detection



(c) PR curves on Pubmed



(b) PR curves on Amazon Computers

Misclassification ← High Uncertainty → Out-of-distribution

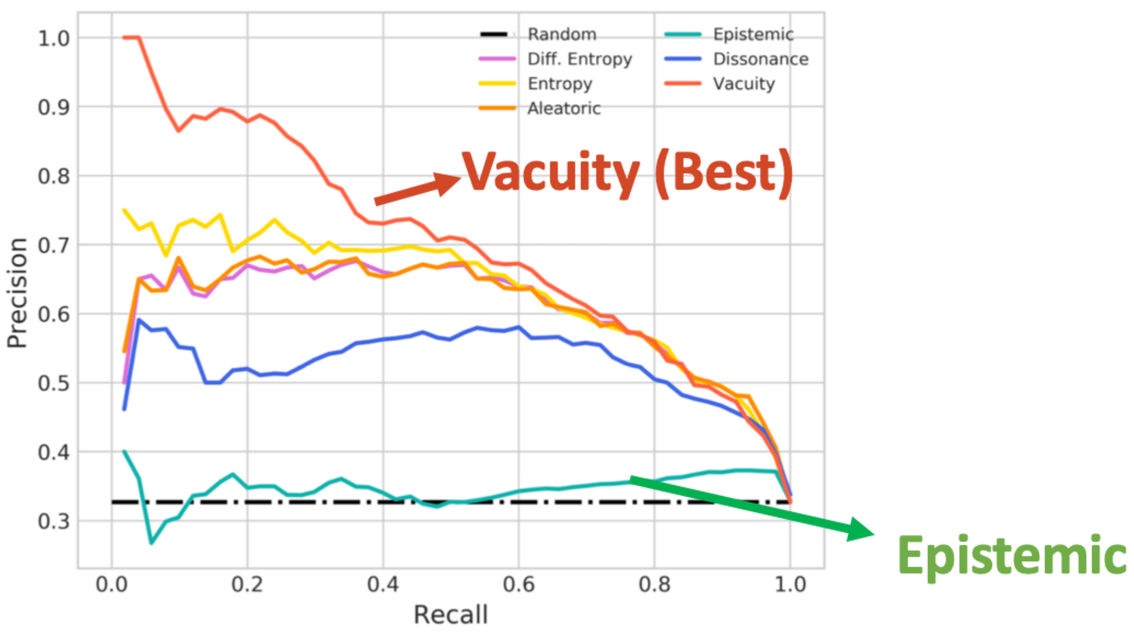
Correct Prediction ← Low Uncertainty → In distribution

Why is Epistemic Uncertainty Less Effective than Vacuity?

Epistemic uncertainty works well in CV applications (supervised learning for OOD detection).

How epistemic uncertainty performances in SSL? **Not good.**

OOD Detection



(b) PR curves on Amazon Computers

Why is Epistemic Uncertainty Less Effective than Vacuity?

Epistemic uncertainty works well in CV applications (supervised learning for OOD detection).

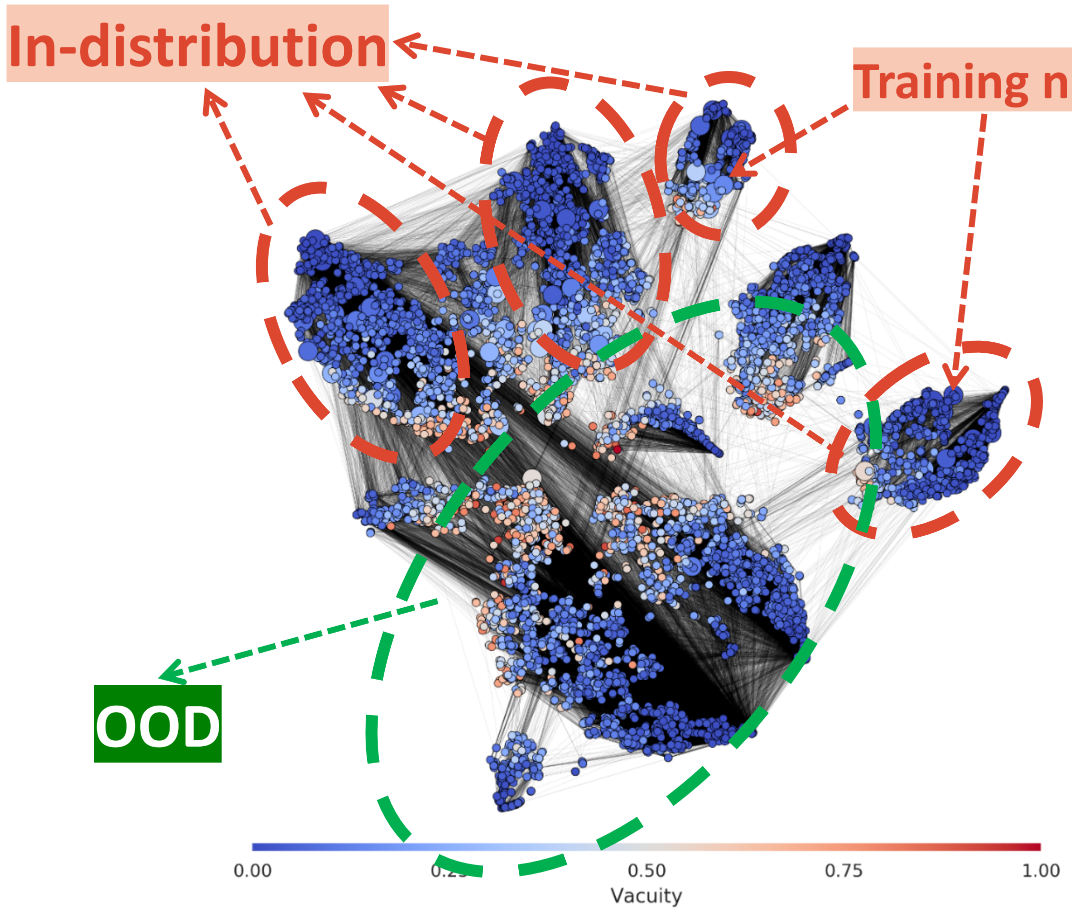
How epistemic uncertainty performances in SSL? **Not good.**

Semi-supervised image classification (unlabeled set contains OOD)

	Epistemic	Supervised	VAT	Mean Teacher	Pseudo Label
(MNIST)	**In-Distribution**	0.140	**0.116**	**0.105**	**0.041**
(FashionMNIST)	**Out-of-Distribution**	**0.249**	0.049	0.076	0.020

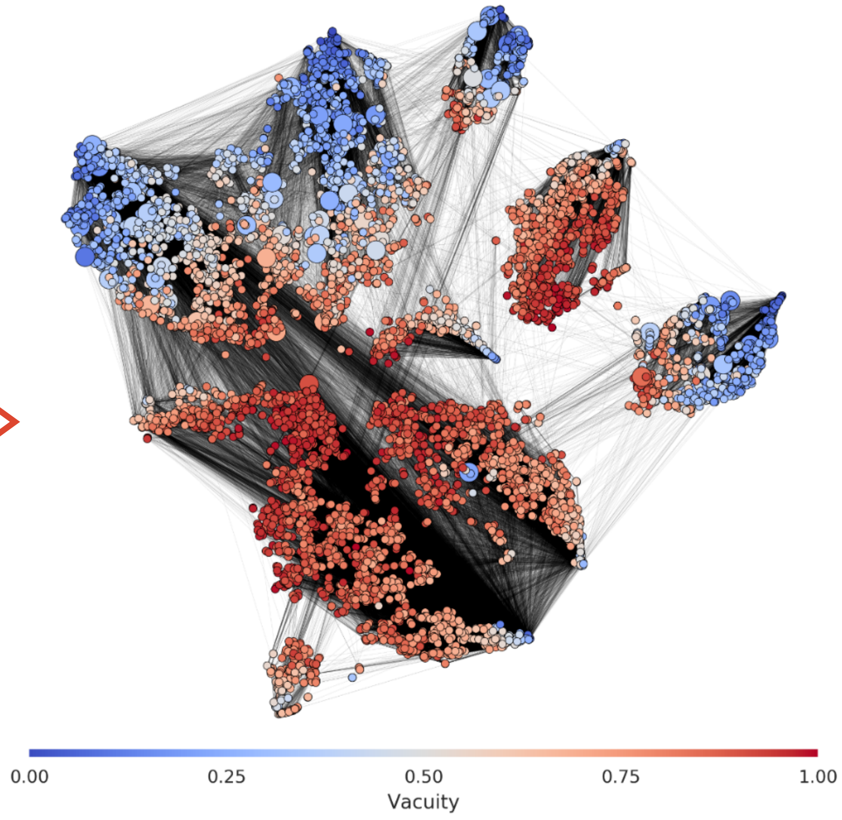
In semi-supervised learning setting, the features of unlabeled nodes are also fed to a model for training process to provide the model with a high confidence on its output.

Beneficial of GKDE



OOD Detection
AUROC: 64%

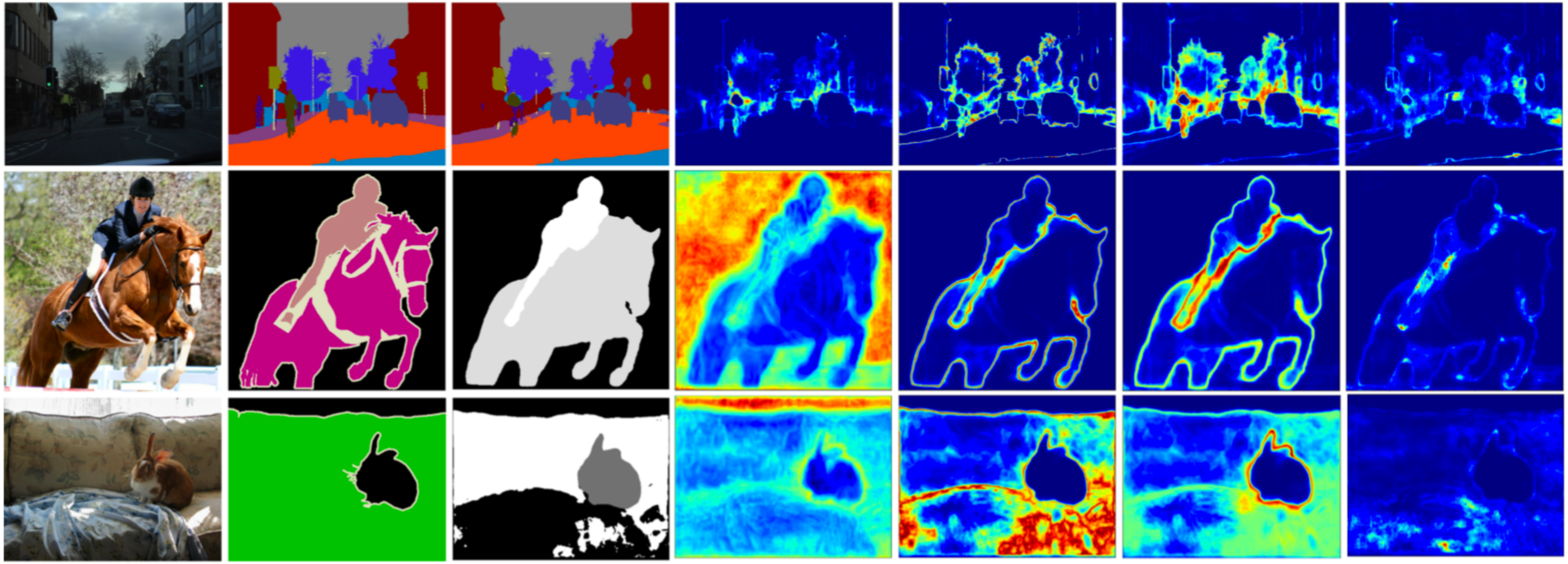
add GKDE



OOD Detection
AUROC: 93%

Extension to other Deep Learning Model (CNN)

Replace GKDE to other method to estimation prior Dirichlet Distribution.



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Vacuity Uncertainty (e) Dissonance Uncertainty (f) Aleatoric Uncertainty (g) Epistemic Uncertainty

Thank you!

Any Question & Comments?

Xujiang Zhao, CS@UT-Dallas
Xujiang.zhao@utdallas.edu