We thank all reviewers for their valuable comments. We addressed your major comments as below:

[Reviewer 2]: **Q1: Provide the rationales about the choice of Dirichlet distribution**. Our proposed framework is designed to predict the subjective opinions about the classification of testing nodes, such that a variety of uncertainty types, such as vacuity and dissonance, can be quantified based on the estimated subjective opinions. As a subjective opinion can be equivalently represented by a Dirichlet distribution about the class probabilities (See Sections 3.1-3.2), we proposed a way to predict the node-level subjective opinions in the form of node-level Dirichlet distributions. Figure 1 illustrates the effect of Dirichlet distribution for distinguishing each uncertainty type from others. There are two scenarios: The left scenario has 10 observations for each of the three classes while the right scenario has 1 observation for each class. The left scenario has a high dissonance while the right scenario has a high vacuity. These two uncertainty types can be distinguished by using the estimated Dirichlet distributions ($\boldsymbol{\alpha}$ refers to the Dirichlet parameters), which is not possible by using the estimated class probabilities ($\mathbf{p}$).

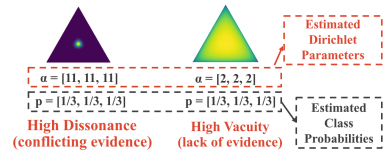
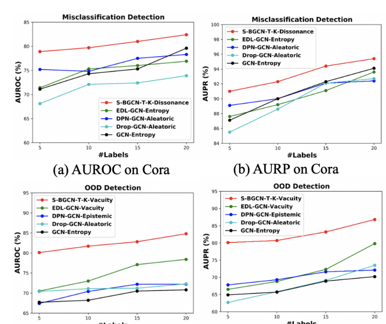


Figure 1: An illustration of high dissonance and high vacuity based on Dirichlet distribution.

Q2: Clarify the significance and novelty of the contribution. The significance and novelty of our proposed work lies in the following unique contributions: (1) Our work is the first that developed the multi-source uncertainty quantification framework that estimates various types of uncertainties by taking a hybrid approach from both DL and evidence/belief theory domains for graph data. (2) We provided the first theoretical basis by demonstrating the mathematical proof that clarifies the relationships between the four important uncertainty types: vacuity, dissonance, aleatoric, and epistemic uncertainty. (3) We proposed the first-known graph-based Kernel Dirichlet distribution Estimation (GKDE) approach that estimates node-level Dirichlet distributions based on graph-structural information. We validated its performance via a theoretical analysis of GKDE as shown in Proposition 1 and an empirical experiment to demonstrate an extensive performance analysis. We proved the theoretical relationships between the four uncertainty types (the second contribution) via the mathematical proof and developed the GKDE approach (the third contribution) to support our proposed multi-source uncertainty quantification framework for graph data (the first contribution).

[Reviewer 3]: **Q1: Discuss and differentiate the different approaches for uncertainty estimation in graphs**. Existing approaches for graph data have focused on estimating an overall predictive uncertainty for node classification based on the entropy of the predicted class probabilities for each test node. Some of the recent methods have explicitly modeled the uncertainty of graph structure in order to better predict the overall predictive uncertainty, such as the methods based on edge-level dropouts [DropEdge in Rong et. al, 2019], graph Gaussian processes (GPs) [13], and Bayesian GNNs + MMSBM (mixed membership stochastic block model) [23]. However, no prior work has explored the decomposition of overall uncertainty into the multiple dimensions as considered in our paper for GNNs, which require the prediction of second-order uncertainty information about the class probabilities based on Dirichlet distribution. Our proposed framework can be readily extended to support prediction of the additional uncertainty dimension on graph structure by integrating graph GPs or MMSBM into our proposed framework.

Q2: Investigate the performance change under varying a number of labeled nodes. An empirical analysis of the performance change for misclassification and OOD detection is shown in Fig 2. The results on the Cora and Citseer datasets demonstrate that our proposed method (S-BGCN-T-K) consistently outperformed the four competitive methods in terms of AUROC and AUPR for varying numbers of labeled nodes. We also observed worse performance in the AUROC and AUPR of all the methods under a lower number of labeled nodes. This makes sense as the less labels we have the higher uncertainty to train the models, resulting in higher the misclassification rate in overall. **Q3: Clarify the additional KL term for theta.** [5] showed that minimizing the KL term can be well approximated by minimizing the cross entropy (or squared error) loss function (See Appendix B.8 for detail). Therefore, we used the squared error loss function instead of the KL term. **Q4: Compare with a (Bayesian) GCN baseline Dropout+DropEdge in Rong et. al, 2019.** As shown in the table below, our proposed method performed better than Dropout+DropEdge on the Cora and Citeer datasets for misclassificaiton detection. A similar trend was observed for OOD detection.



(a) AUROC on Cora (b) AURP on Cora (c) AUROC on Citeseer (d) AUPR on Citeseer

Figure 2: Performance change with different number of labeled nodes.

[Reviewer 4]: **Q1: Clarify notations**. (1) The semicolon in Equation 5 is a typo and should be replaced by a comma. (2) About the entropy term in Equation 5, 'P' is missing before '$(y|x;\boldsymbol{\theta})$' and the the correct entropy term should be $\mathcal{H}\big[\mathbb{E}_{P(\boldsymbol{\theta}|\mathcal{G})}[P(y|x,\boldsymbol{\theta})]\big]$, referring to the entropy of expected distribution. (3) We agree with you that asymptotic complexity is a more meaningful metric to represent the efficiency of an algorithm, so we will use it in a revised paper.

Dataset	Model	AUROC					AUPR				
		Va.	Dis.	Al.	Ep.	En.	Va.	Dis.	Al.	Ep.	En.
Cora	S-BGCN-T-K	70.6	**82.4**	75.3	68.8	77.7	90.3	**95.4**	92.4	87.8	93.4
	DropEdge	-	-	76.6	56.1	76.6	-	-	93.2	85.4	93.2
Citeseer	S-BGCN-T-K	65.4	**74.0**	67.2	60.7	70.0	79.8	**85.6**	82.2	75.2	83.5
	DropEdge	-	-	71.1	51.2	71.1	-	-	84.0	70.3	84.0

Va.: Vacuity, Dis.: Dissonance, Al.: Aleatoric, Ep.: Epistemic, En.: Entropy