We gratefully appreciate all reviewers' suggestions. Below please find the responses to the specific comments.

▷ **General responses. Q: Existing contrastive learning (CL) framework & graph data augmentation methods.** *Reply:* As reviewers #1&2 point out, (i) it is nontrivial to develop the CL framework that shifts from images to graphs, since most components need to be adapted; and (ii) though present in previous work, data augmentation for graphs is under-explored, whereas our thorough experiments revealed insights into the role of augmentation combinations in GraphCL for various graph data. Furthermore, GraphCL generalizes previous works in what and how structural invariance in graphs is captured, which will be detailed in responses to reviewer #2. **Q: Theoretical analysis.** *Reply:* (i) A major stream of theoretical analysis for CL is currently based on mutual information (MI). For instance, we can view the contrastive objective in GraphCL as maximizing the MI between graph views by generalizing the Deep Graph Infomax (DGI) approach (see **Q*** of reviewer #2), which can be included in the final version for completeness. (ii) However, we also emphasize that the MI-view of CL has been challenged recently and more theoretical development is needed. For instance, [1] shows that the success of the Infomax approaches is only loosely connected to MI but strongly depends on the bias in encoders & estimators. We instead may explore some newer perspectives (e.g. [2]) in the future work. We believe that our work provides comprehensive guide to the practice of graph representation learning and inspirations for new theoretical development. **Q: Node-level representation learning and tasks.** *Reply:* We implemented GraphCL for a node-level task based on the experiment setting in DGI [3], where Subgraph and NodeDrop are respectively chosen for Cora and Citeseer based on validation. The preliminary results below are promising and more tasks/datasets will be included during revision.

Dataset	DGI	GraphCL	Dataset	DGI	GraphCL
Cora	82.3%	82.4%	Citeseer	71.8%	73.1%

▷ **Reviewer #1. Q: Strengthening the writing of related work, experiment settings and broader impact.** *Reply:* Thanks for detailed comments. We will revise to improve the contents and the structure as suggested.

▷ **Reviewer #2. Q*: Insights into the GraphCL motivation and methodology** *Reply:* Recalling the contrastive objective (Eq. (3) in the paper) in GraphCL, we can rewrite the loss function in the expectation form for each batch as:

$$
\begin{aligned}
l &= -\mathbb{E}_{\mathbb{P}_{(\hat{\mathbb{G}}_i,\hat{\mathbb{G}}_j)}}\{\mathrm{sim}(g(f(\hat{\mathcal{G}}_i)), g(f(\hat{\mathcal{G}}_j)))/\tau\} + \mathbb{E}_{\mathbb{P}_{\hat{\mathbb{G}}_i}\times\mathbb{P}_{\hat{\mathbb{G}}_j}}\{\log\sum_{\hat{\mathcal{G}}_j}\exp(\mathrm{sim}(g(f(\hat{\mathcal{G}}_i)), g(f(\hat{\mathcal{G}}_j)))/\tau)\} \\
&= -\mathbb{E}_{\mathbb{P}_{\hat{\mathbb{G}}_i}}\{\mathbb{E}_{\mathbb{P}_{(\hat{\mathbb{G}}_j|\hat{\mathbb{G}}_i)}} T_w(f(\hat{\mathcal{G}}_i), f(\hat{\mathcal{G}}_j)) - \mathbb{E}_{\mathbb{P}_{\hat{\mathbb{G}}_j}}\log\sum_{\hat{\mathcal{G}}_j}\exp(T_w(f(\hat{\mathcal{G}}_i), f(\hat{\mathcal{G}}_j)))\}, \qquad (1)
\end{aligned}
$$

where $\mathbb{P}_{(\hat{\mathbb{G}}_i,\hat{\mathbb{G}}_j)}, \mathbb{P}_{(\hat{\mathbb{G}}_j|\hat{\mathbb{G}}_i)}, \mathbb{P}_{\hat{\mathbb{G}}_i}$ are respectively the joint, conditional and marginal distribution, and $T_w: \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is a learnable score function that we parametrize with the similarity function $\mathrm{sim}(\cdot,\cdot)$ and the projection head $g(\cdot)$. Notice that Eq. (1) fits the formulation of the InfoNCE loss [4, 2] and therefore minimizing the contrastive objective (1) is explicitly maximizing a lower bound of the mutual information between the latent representations of two views of graphs $f(\hat{\mathcal{G}}_i), f(\hat{\mathcal{G}}_j)$. Therefore, we treat our contrastive objective (Eq. (1)) as a general formulation for graph contrastive learning, that can be instantiated as a specific algorithm by designing how to construct different views of graphs, where augmentation is adopted as a general method for view constructions. **Q: Related work.** *Reply:* We will include and discuss the suggested papers in revision. We emphasize that our GraphCL framework generalizes previous works for graphs in what to contrast (various augmentation options, unified as a general method for view constructions) and how to contrast (contrastive objective (Eq. (1)) as a general formulation). For instance, the Infomax approaches [3] only assume fixed views of original graphs and subgraphs in the objective (1) for approximation. Thus, previous works can be regarded as special cases of GraphCL that is a more general framework to learn structural invariance from diverse graph data without making specialized assumptions.

▷ **Reviewer #3.** Please refer to the general responses for the answers.

▷ **Reviewer #4. Q: Missing citations and prior work** *Reply:* Thanks for bringing some missing related work to our attention. We will add them during revision. **Q: More sophisticated graph data augmentation.** *Reply:* Thanks for bringing [5] to our attention for designing potential more complicated augmentation. Although the ECCV acceptance is after the NeurIPS submission, we find the paper interesting proposing domain specific augmentation (for scene graph in computer vision), and will definitely explore it in our final version.

[1] Michael Tschannen, et al. On Mutual Information Maximization for Representation Learning. ICLR, 2020.

[2] Yonglong Tian, et al. What Makes for Good Views for Contrastive Learning. ECCV, 2020.

[3] Petar Veličković, et al. Deep Graph Infomax. ICLR, 2019.

[4] Aaron van den Oord, et al. Representation Learning with Contrastive Predictive Coding. arXiv, 2018.

[5] Roei Herzig, et al., Learning Canonical Representations for Scene Graph to Image Generation. ECCV, 2020.