Author feedback for paper "Diffusion Improves Graph Learning"

R1: Learning θ_k, computational complexity. We have recently investigated exactly the question of how to directly learn θ_k in a way that lets the model leverage the full space of graph diffusions. While we did manage to directly train meaningful diffusion weights, we also found that so far the diffusion schemes described in the paper (PPR and heat kernel) are among the best possible. Please also note that we already discuss GDC's computational complexity when introducing sparsification on page 3. We will make this more explicit in the camera-ready version.

R2: GNN + diffusion. Papers [29] and, to some extent, [65] design diffusion-based GNNs for node classification. In our work we systematize this approach and show that it is applicable to a much wider class of models and tasks. Moreover, we explain why this approach works by analyzing its close relationship to spectral methods.

R2: Discussion of limitations. While lines 111-114 and 234-238 already discuss the approach's limitations, we will improve the paragraphs further and make our approach's area of applicability clearer in the camera-ready version. We suspect GDC not to perform well in settings with more complex edges (e.g. knowledge graphs) or graph reconstruction tasks such as link prediction (Q5, as already described in lines 234-238 of the paper). When investigating the suggested PPI (protein-interaction) dataset we found that the underlying data used for graph construction already includes graph diffusion-like mechanisms (e.g. regulatory interactions, protein complexes, and metabolic enzyme-coupled interactions). This provides further evidence for the effectiveness of diffusion and also indicates why using GDC does not yield further improvement on this dataset. Furthermore, we recently performed preliminary experiments on graph classification, which showed that GDC can help with this task as well (e.g. $+2.5$ percentage points with GCN on DD).

R2: Homophilious datasets. There are certainly many different settings to which GNNs have been applied. In this context, graphs with homophily are still one of the most important use cases for graph learning. Many real world datasets fall into this category, including all kinds of social networks, messaging networks, and purchasing networks, and a whole body of literature is focused solely on this domain. The vast majority of classic and deep-learning based methods for graphs either explicitly or implicitly assume homophily, e.g. graph cuts, spectral clustering, DeepWalk, DGI, or GCN. Furthermore, we are not aware of a single publication in the graph learning space that is as thorough in their experimental evaluation as we are. We rigorously evaluate our method on 9 diverse models and 6 datasets, i.e. in 54 use cases, which is way beyond the 5 use cases usually considered (e.g. in DGI).

R2: Short questions. (Q2) New experiments show that GDC performs well for various label rates, with the performance difference improving further for sparser settings, as shown in Fig. 1. We will include the full experiment in the final paper. Please note that a label rate of $80\,\%$ is not realistic for semi-supervised classification. (Q1) The difference in DGI performance is due to us using a consistent 128-dim. embedding size for all models and k-means clustering instead of a supervised linear classifier. We will incorporate your comments regarding citations, Fig. 7 (Q4), and the statement on MPNNs vs. spectral methods (Q3) in the final version.

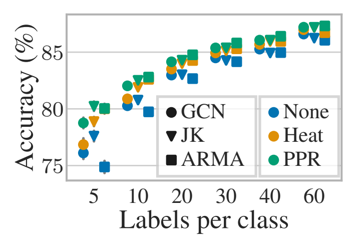

Figure 1: Accuracy on Cora with different label rates.

R3: Spectral analysis. We agree that the connection between graph diffusion and low-pass filtering has been studied before; and cite relevant works in the paper. However, no previous work has proven the direct relationship between the coefficients in spectral methods and graph diffusion that we give in Eq. 4 or the coefficients for PPR given in Eq. 5. These let you switch freely between the two formulations and show how similar spatial and spectral methods are, untangling some confusion we have seen in recent literature (e.g. classifying GCN as a spectral method). Moreover, in our spectral analysis of the full GDC pipeline this connection is only one step out of three.

R3: Eigenvalue perturbation. Using Steward and Sun 1990, Corollary 4.13 we are able to derive a tighter bound on the eigenvalues than Weyl's inequality would provide, i.e. $\sqrt{\sum_{i=1}^{N}(\tilde{\lambda}_i - \lambda_i)^2} \leq ||E||_{\mathrm{F}} \leq N||E||_{\max} \leq N\epsilon$. This bound still significantly overestimates the perturbation since PPR and the heat kernel both exhibit strong localization on real-world graphs and hence the change in eigenvalues empirically does not scale with N (or, rather, $\sqrt{N}$). However, without imposing strong assumptions on the graph's generative process we cannot derive tighter bounds – and these would then not be generally applicable. Since we are interested in real-world graphs, we only included the empirical analysis in the paper. Still, we will add the worst-case bound to the camera-ready version.

R3: Simplicity. The most elegant, effective, and successful solutions are often also the simplest (e.g. SGD, dropout, residual connections, batch normalization, ReLU, ...). Simplicity is a key factor for practical applications and ease of implementation. Further reasons why diffusion is not ad-hoc but indeed very well motivated (beyond our evaluation and spectral analysis) have been given in previous works as described in lines 93-97 and our related work section.