

Figure 1: **Upper:** GRACLUS; **Lower:** COMPLEMENT.

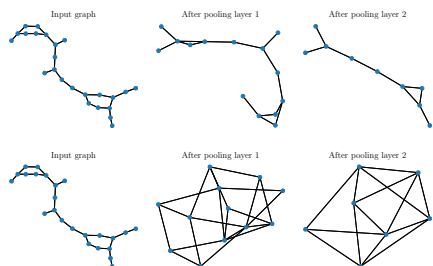


Figure 2: Numbers denote the Graclus performance on ZINC. Colors indicate the performance gap between Graclus and Complement. Blue tones denote cases on which Complement obtains lower MAE.

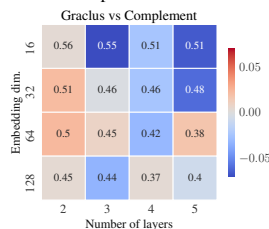


Table 1: Avg. results (AUC for molhiv and Accuracy for others) for the cleaned version of IMDB and two additional larger datasets.

	IMDB-B	DD	molhiv
GRACLUS	70.5	72.1	75.2
COMPLEMENT	70.1	72.0	74.7
DIFFPOOL	70.9	77.0	70.5
U DIFFPOOL	70.2	77.5	71.0
GMN	68.4	74.5	73.9
Rand. GMN	70.2	73.6	74.1
Dist. GMN	70.0	74.9	73.9
Mincut	70.4	76.8	71.7
N-Mincut	71.0	76.7	73.5

We thank the reviewers for their valuable feedback. We are glad the reviewers found we are addressing an important problem (R1, R4) for which we carried a rigorous study (R3) supported by evidence (R5) and that our work sheds new light onto the role of pooling (R4). We have made our best effort to address all the questions given the limited space.

@R1 @R5 Difference between GRACLUS and COMPLEMENT. Thank you for raising this issue. We recognize that our explanation might mislead the reader to believe that COMPLEMENT fully operates on complement graphs. In fact, we only employ the **complement graph to compute cluster assignments** $S^{(l)}$. With the assignments in hand, we apply the **pooling operation** (Eqs. 2 and 3) **using the original graph structure** ($A^{(l)}, X^{(l)}$). That being said, there is no clear reason to believe there is a 1-to-1 correspondence between the representations achieved by GRACLUS and COMPLEMENT. As an example, the Figure 1 shows how different these coarsened structures can be. Note that the paired representations are not complementary. We will improve the description of COMPLEMENT in the manuscript.

@R1 @R3 Additional experiments with larger datasets and graphs. We initially considered datasets that have been broadly used to validate novel pooling methods. Importantly, we have also strived to avoid common yet misleading evaluation protocols and datasets recently reported in benchmarking papers. For completeness, we have also run new experiments (Table 1) on a dataset with **larger graphs (DD, ≈ 250 nodes)** and an **OGB dataset (molhiv, $\approx 50K$ samples)**. These additional results **reinforce our initial findings**. We now have results for a total of eight datasets to support our claims. We will include these new results in the appendix and briefly discuss them in the main paper.

@R1 "Please be more precise" (about low-frequency graph filters). Agreed! We tried to convey that, in analogy with 2D convolutions, d -dimensional node features of a graph correspond to d image channels. And the filtering operation (convolution) acts on signals defined over the nodes. We will rewrite the sentence more precisely.

@R1 "...results only hold for the specific set of hyperparameters". We have followed general guidelines from the original (or benchmarking) papers and used the same hyperparameters for our variants. We believe that this methodology promotes a fair assessment of the role of local pooling. Also, we have shown results for four methods and (now) eight datasets to support our main claims. As additional evidence, Figure 2 reports the performance from Graclus and Complement for different hyperparameter settings, for all of which the performance gap is < 0.05 MAE.

@R1 "oversmoothing was only shown for the GCN layer...". Note that we only employ GCN layers (i.e., mean-based operators) in DiffPool. For the remaining methods, we adopted sum-based operators, which are known to be more expressive than mean-based ones. Thus, **we believe our results are not limited to GCN convolutions**.

@R4 Effect of more/less convolutions. We do not expect local pooling to be more effective with more convolutions. As an example, Figure 2 shows the performance gap between Graclus and Complement as a function of the number of convolutions. Both models obtain very similar performance as we increase the number of layers.

@R4 "A minor concern ... considering more pooling techniques". We also provide **results for MinCutPool (ICML 2020) in Appendix B**, which corroborates our findings. We will mention this in the main text.

@R5 @R1 "Please use a more polite tone". Good point; we fully agree and will remove that slip.

@R5 Removing isomorphisms from IMDB. Thanks for pointing this out. We have rerun the experiments with the cleaned version of the dataset (see Table 1). Gladly, this change **had no impact on our conclusions**.

@R5 "...more representations in the appendix". We agree that this will add illustrative value to our paper. We have saved a number of these representations from which we have chosen only a few to illustrate our claims. We will add more embeddings to the Appendix and also adjust the Figures 5 and 6 to make a direct comparison easier.

@R5 Are Graclus representations smooth? We agree that the representations learned by Graclus are not as smooth as those from GMN and DiffPool. However, Complement produces much smoother representations while maintaining the same performance. As an example, for ZINC graphs, we computed the avg std deviation of their embeddings before and after the 1st pooling layer, for which Complement achieves one order of magnitude lower compared to Graclus.