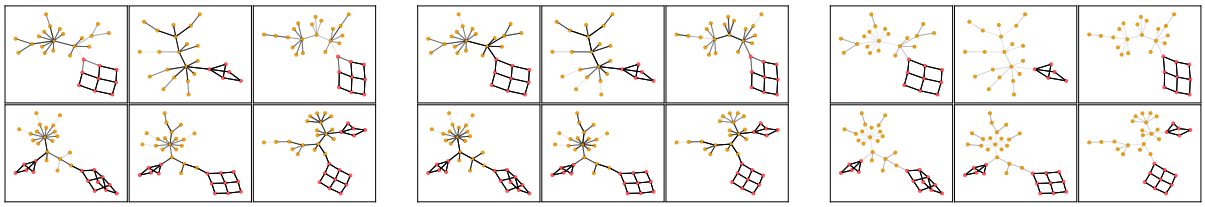
1 More Details about the Experiments

1.1 Interpretation Visualization



(a) GSAT (b) GMT-LIN (c) TOPING

Figure 1: Learned interpretable subgraphs by GSAT, GMT-LIN and TOPING on BA-HouseAndGrid. Figures in each row belong to the same class. Nodes colored red are ground-truth explanations.

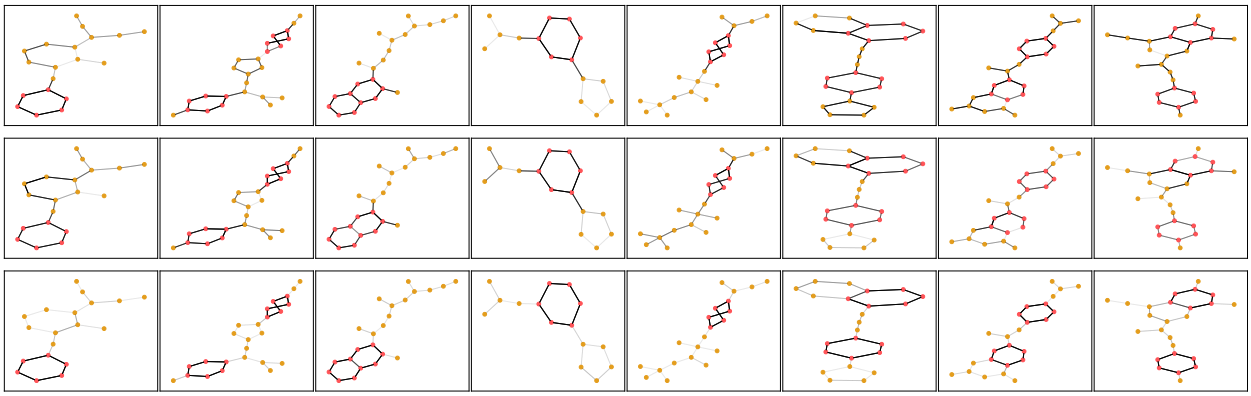


Figure 2: Visualizing attention of GSAT(first row), GMT-LIN(second row) and TOPING (third row) on Benzene. Figures in the same column represent an identical graph. Nodes colored red are ground-truth explanations.

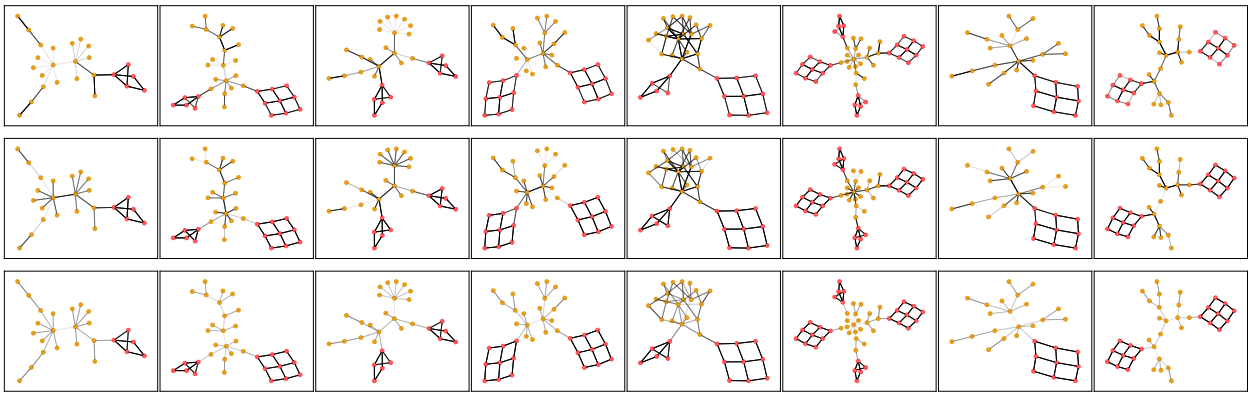


Figure 3: Learned interpretable subgraphs by GSAT (first row), GMT-LIN(second row) and TOPING(third row) on BA-HouseOrGrid-2Rnd. Figures in the same column represent an identical graph. Nodes colored red are ground-truth explanations.

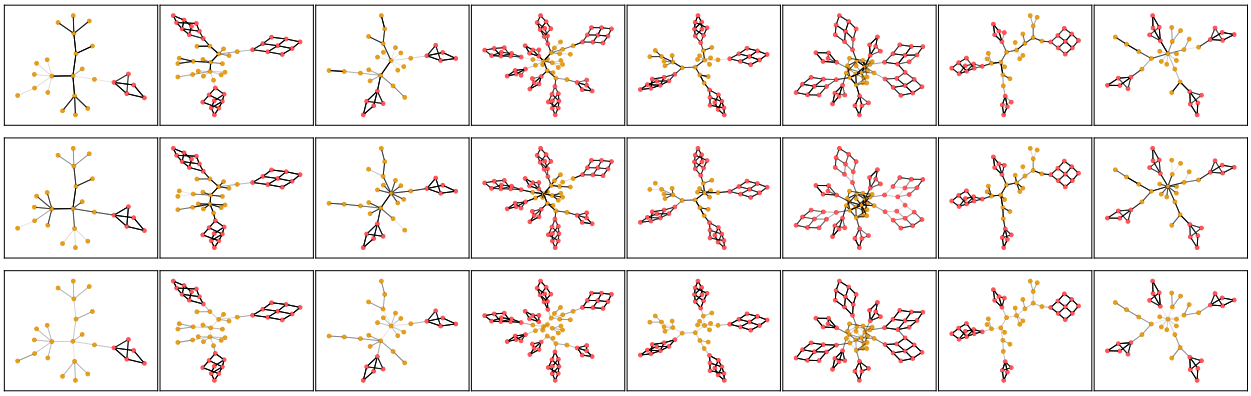


Figure 4: Learned interpretable subgraphs by GSAT (first row), GMT-LIN(second row) and TOPING(third row) on BA-HouseOrGrid-4Rnd. Figures in the same column represent an identical graph. Nodes colored red are ground-truth explanations.

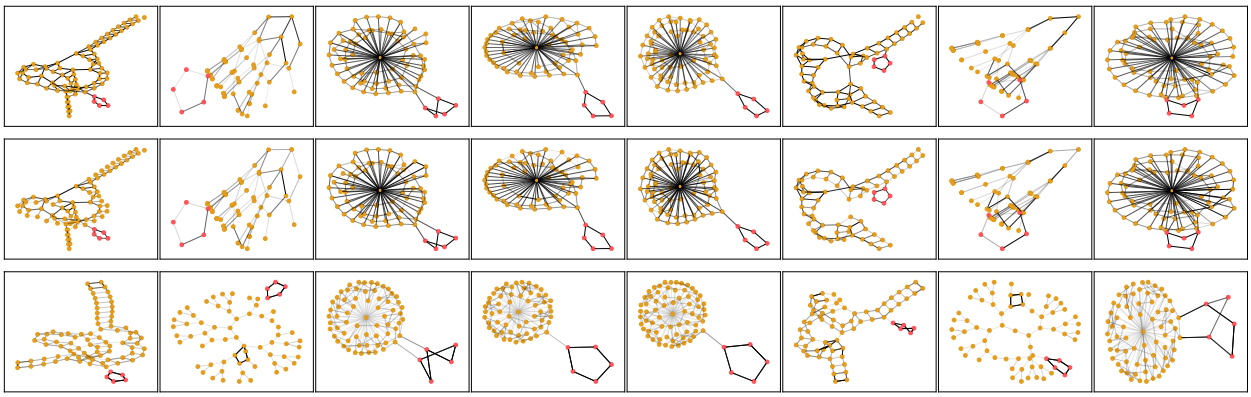


Figure 5: Learned interpretable subgraphs by GSAT (first row), GMT-LIN(second row) and TOPING(third row) on SPmotif0.9 class 0. Figures in the same column represent an identical graph. Nodes colored red are ground-truth explanations.

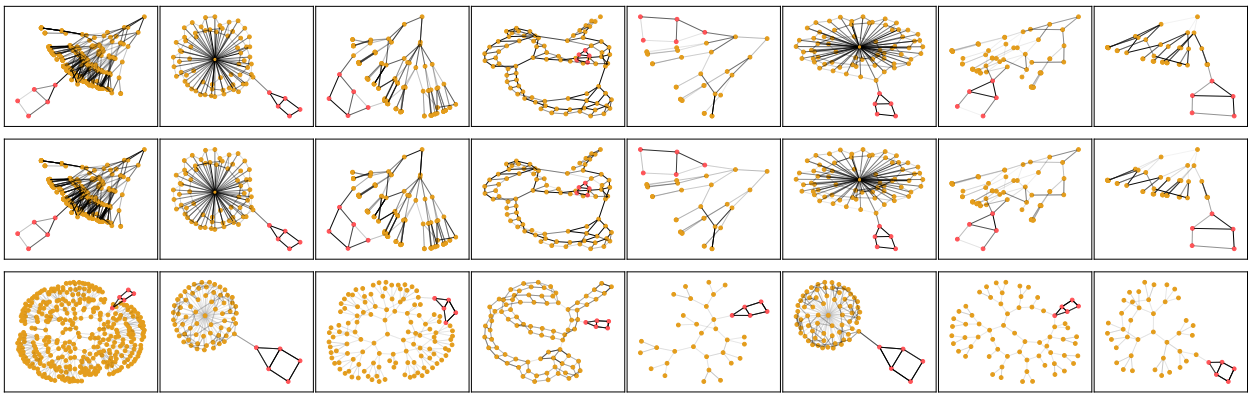


Figure 6: Learned interpretable subgraphs by GSAT (first row), GMT-LIN(second row) and TOPING(third row) on SPmotif0.9 class 1. Figures in the same column represent an identical graph. Nodes colored red are ground-truth explanations.

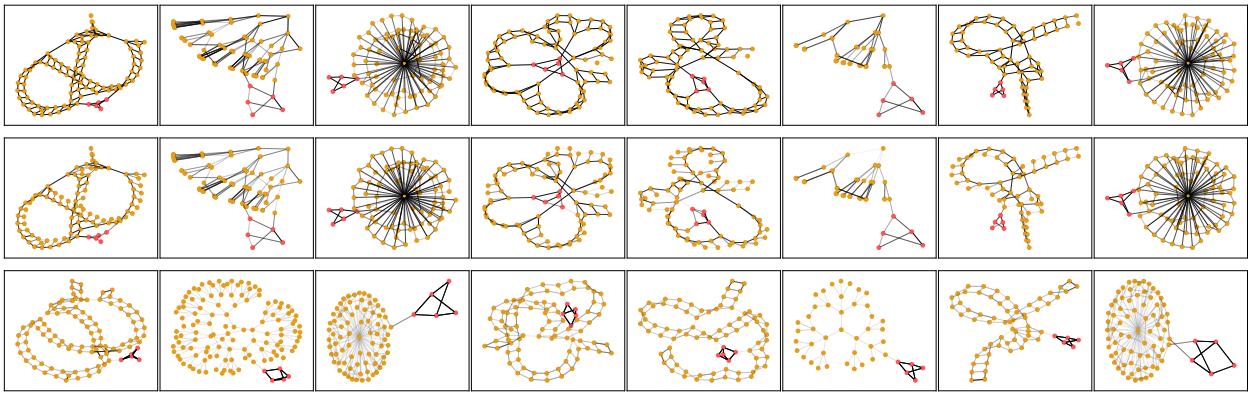


Figure 7: Learned interpretable subgraphs by GSAT (first row), GMT-LIN(second row) and TOPING(third row) on SPmotif0.9 class 2. Figures in the same column represent an identical graph. Nodes colored red are ground-truth explanations.

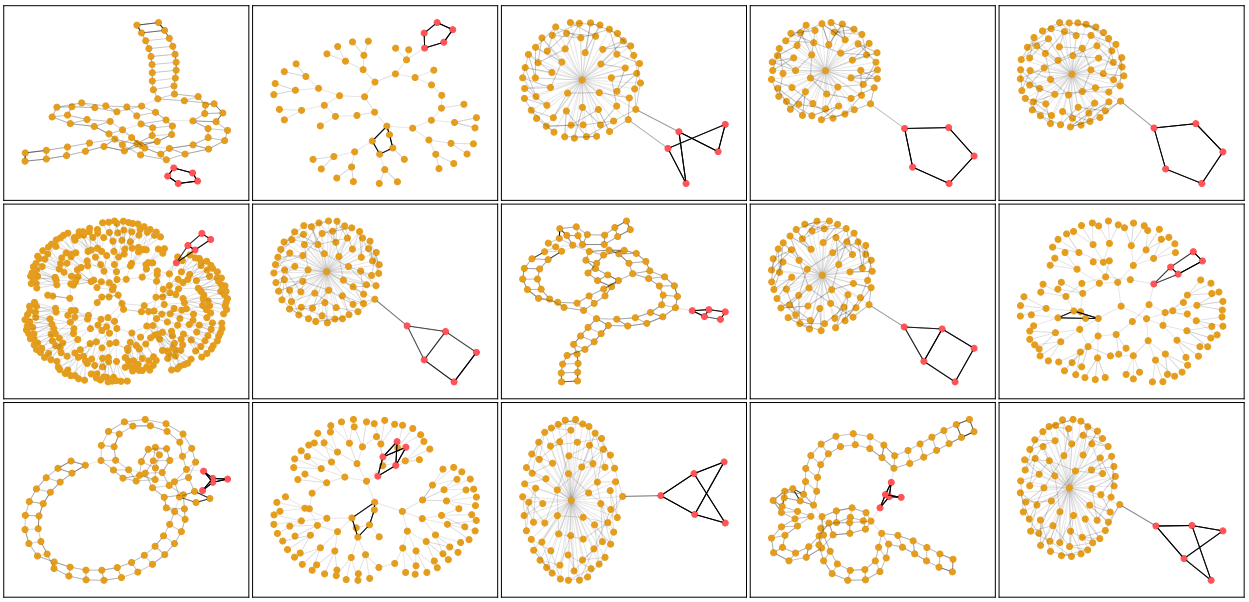


Figure 8: The rationals of SPmotif0.9 learned by TOPING. Figures in each row belong to the same category. Nodes colored red are ground-truth explanations.

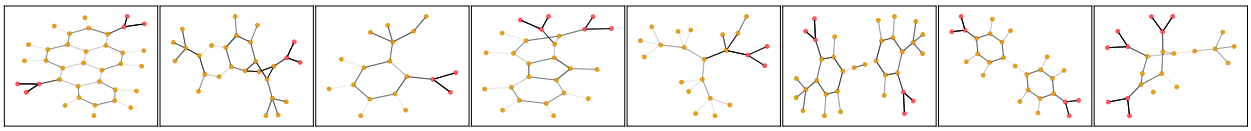


Figure 9: The rationals of Mutag learned by TOPING. Nodes colored red are ground-truth explanations.

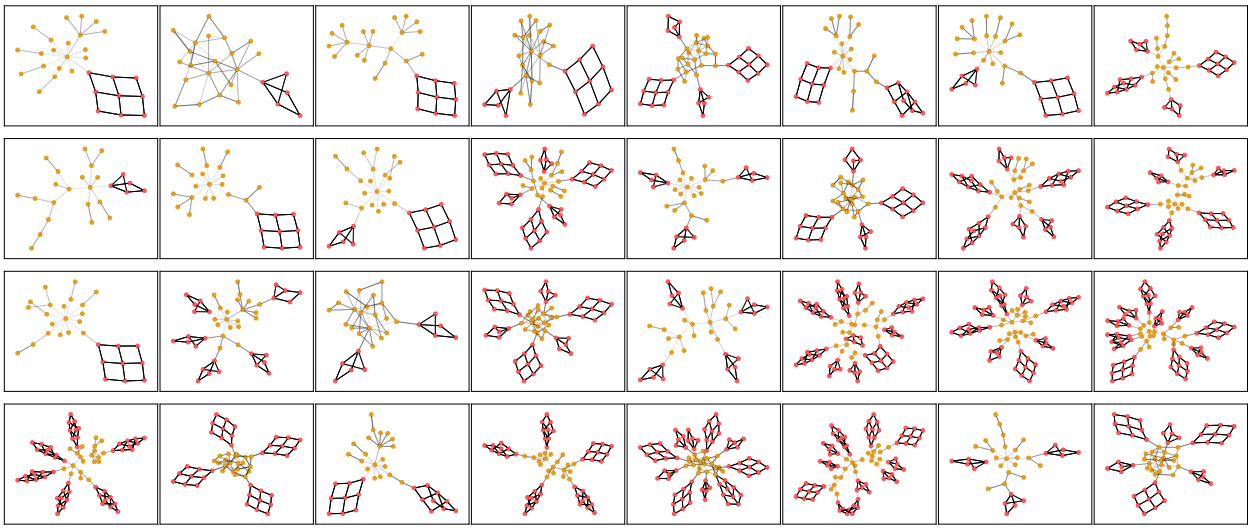


Figure 10: The interpretation visualization results of BA-HouseOrGrid-2Rnd, BA-HouseOrGrid-3Rnd, BA-HouseOrGrid-5Rnd and BA-HouseOrGrid-6Rnd. We use TOPING, which has been well trained on BA-HouseOrGrid-4Rnd, to directly generate the interpretable subgraphs for BA-HouseOrGrid-(2,3,5,6)Rnd. We oberve that this transfer process never hurts the prediction(ACC=100%) and interpretation(AUC=100%) performance on all datasets. Nodes colored red are ground-truth explanations.