Supplementary Information for Discovery of Novel Reticular Materials for Carbon Dioxide Capture using GFlowNets

Flaviu Cipcigan
IBM Research Europe - UK

Jonathan Booth
Science and Technologies Facilities Council

Rodrigo Neumann Barros Ferreira
IBM Research Brazil

Carine Ribeiro dos Santo
IBM Research Brazil

Mathias Steiner
IBM Research Brazil

February 12, 2024

Appendix A: Training details of GFlowNets on all MOF topologies

Figure 1 shows the trajectory balance losses for training a GFlowNet on the ASC topology without edges while Figure 2 shows the logZ. All other training runs on other topologies showed similar behaviour.

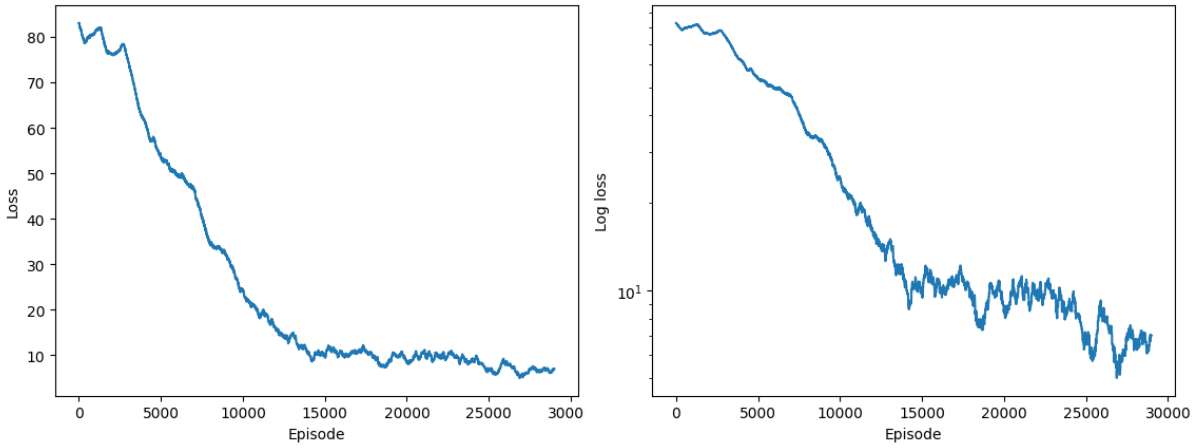

Figure 1: trajectory balance losses for training a GFlowNet on the ASC topology without edges. Losses are smoothed with a 1,000 episode window moving average due to the discovery of a high performing MOF causing a one-episode long spike in the loss.

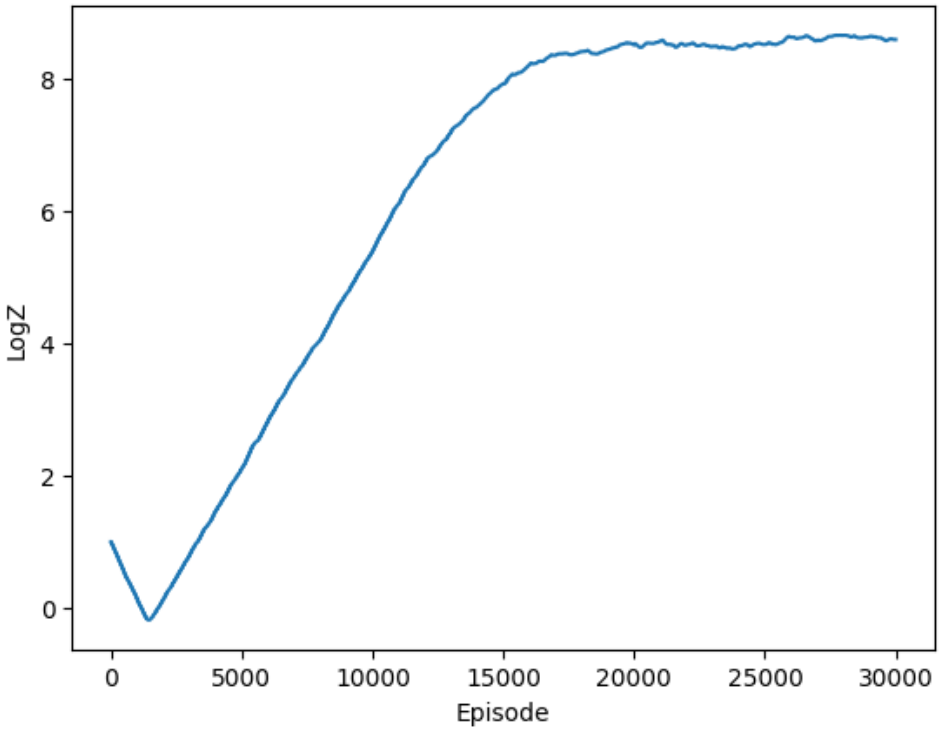

Figure 2: logZ during training for the ASC topology without edges.

The figures below show the performance of the GFlowNet vs random sampling for all eleven topologies with and without edges.

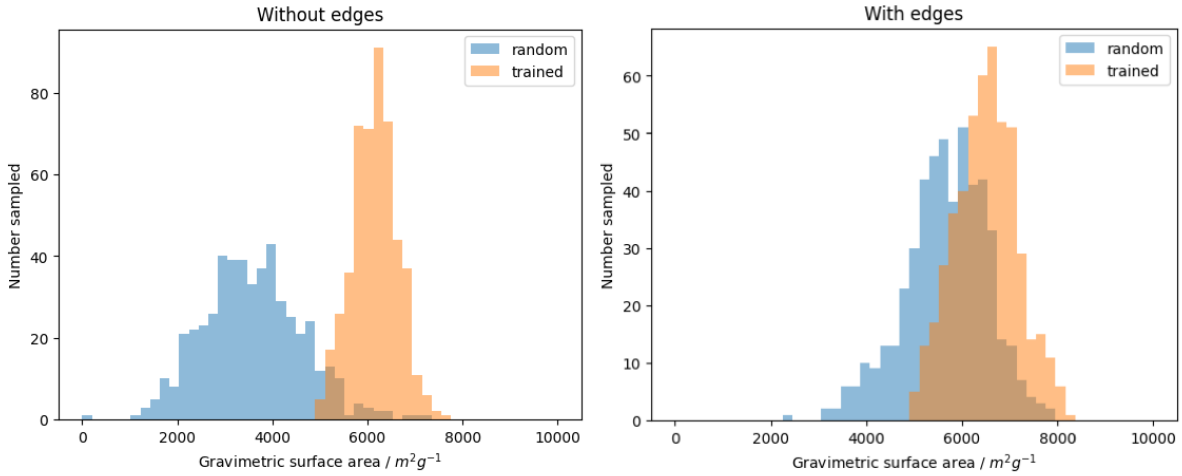

Figure 3: Performance of the GFlowNet trained on the CDZ-E topology.

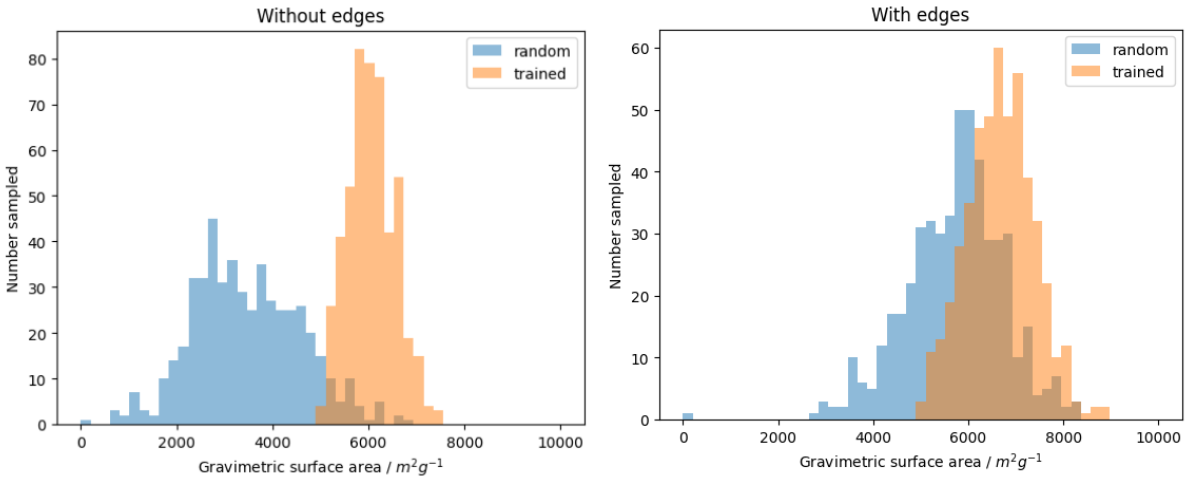

Figure 4: Performance of the GFlowNet trained on the CDL-E topology.

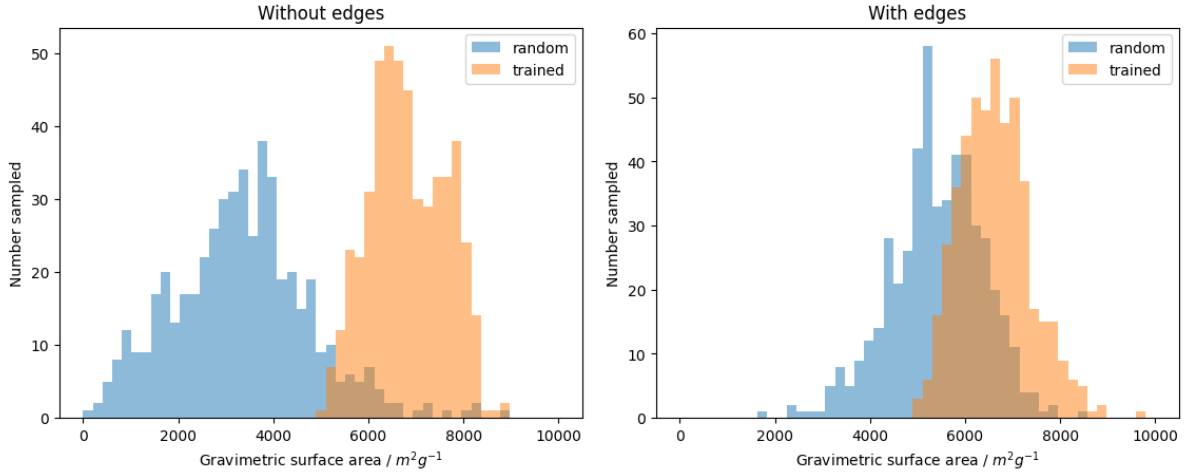

Figure 5: Performance of the GFlowNet trained on the EFT topology.

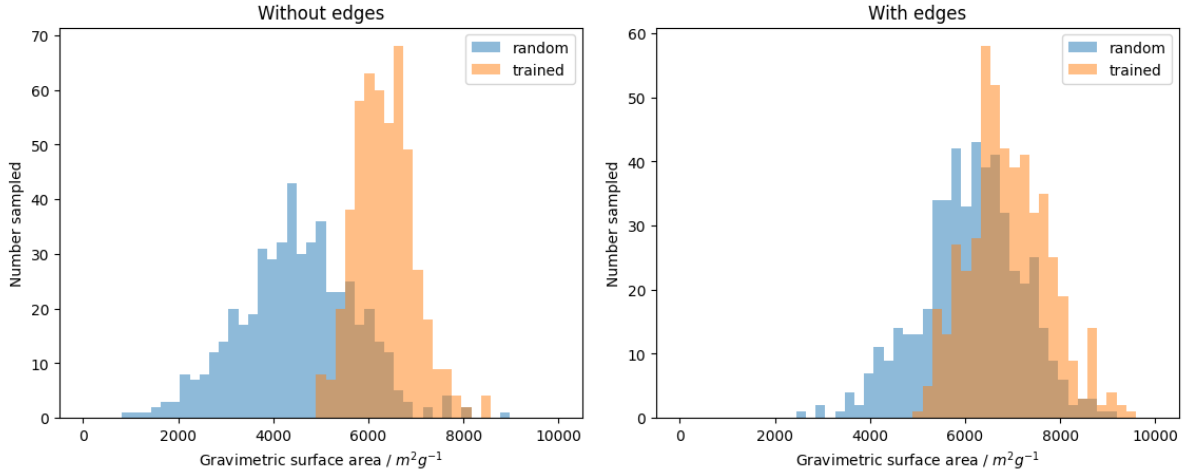

Figure 6: Performance of the GFlowNet trained on the FFC topology.

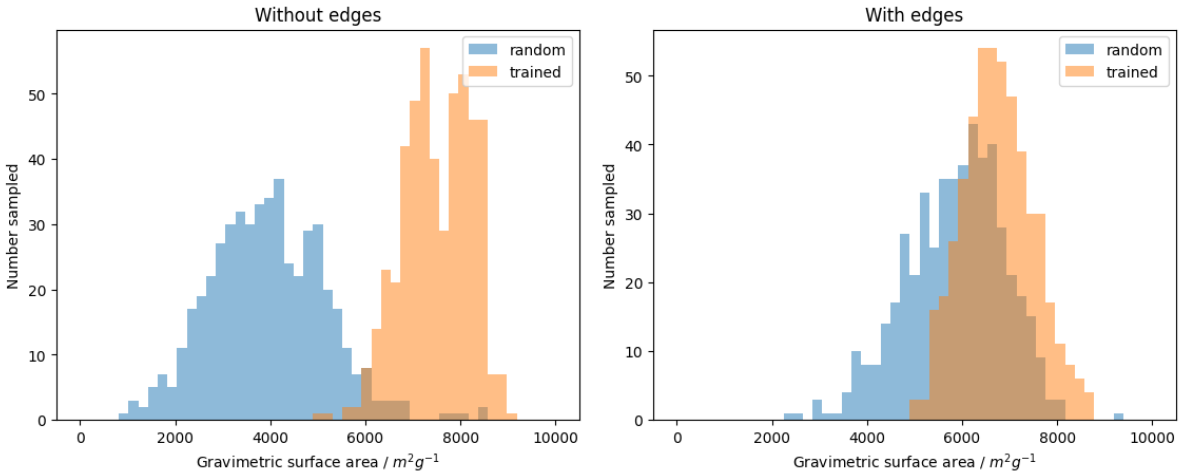

Figure 7: Performance of the GFlowNet trained on the TSG topology.

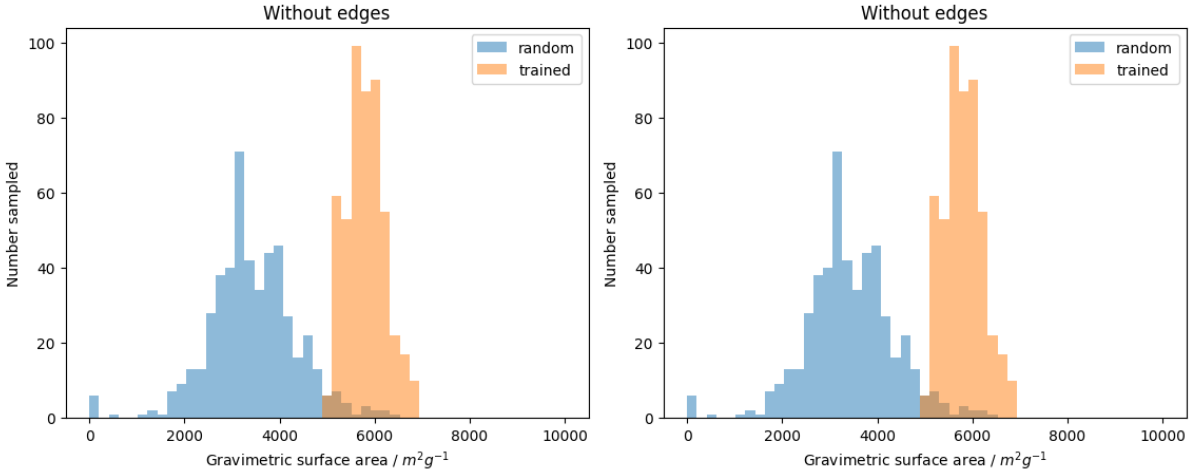

Figure 8: Performance of the GFlowNet trained on the TFF topology.

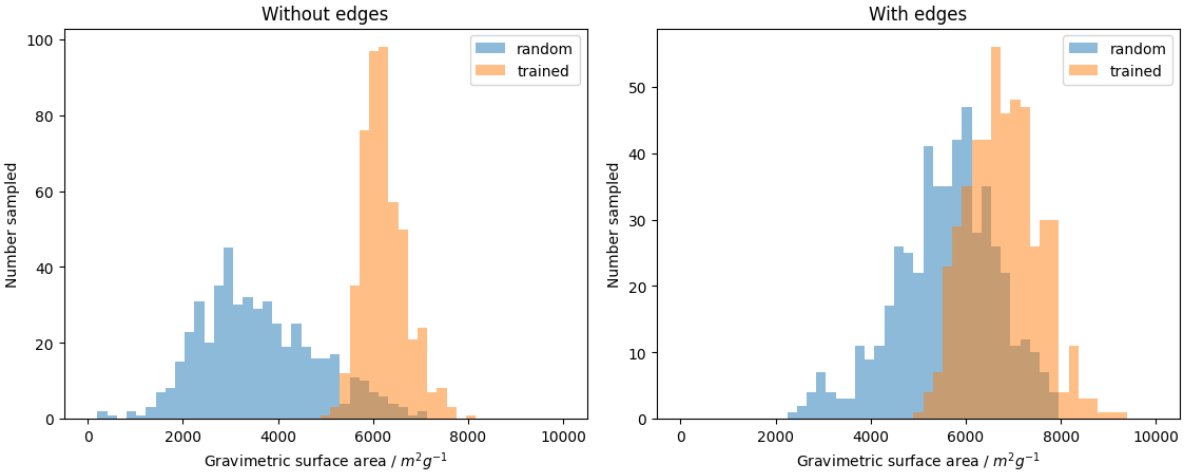

Figure 9: Performance of the GFlowNet trained on the ASC topology.

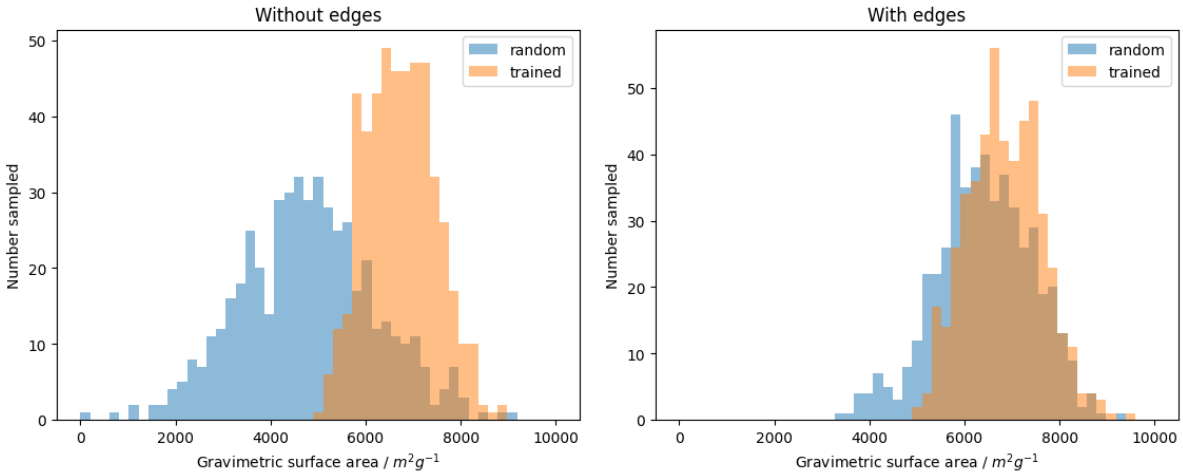

Figure 10: Performance of the GFlowNet trained on the DMG topology.

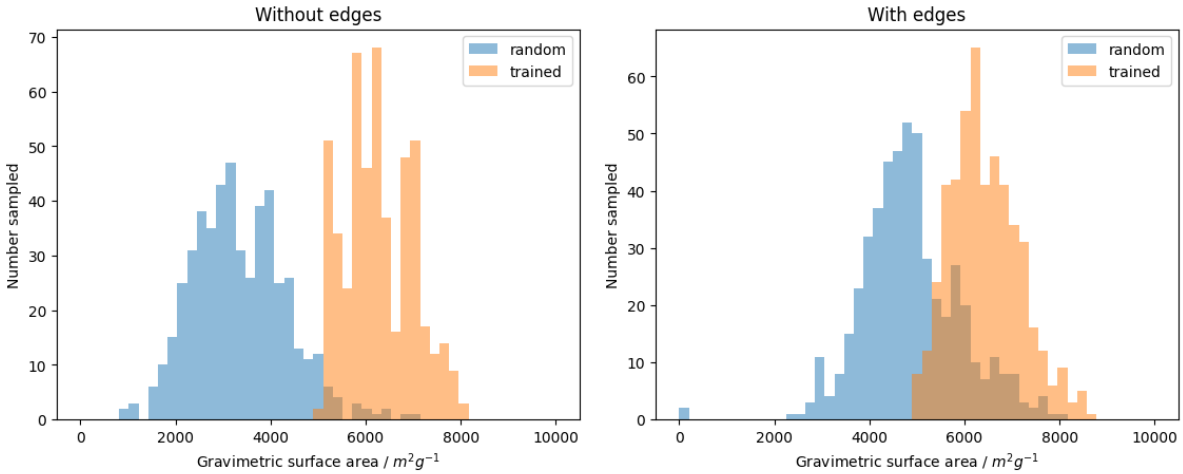

Figure 11: Performance of the GFlowNet trained on the DNQ topology.

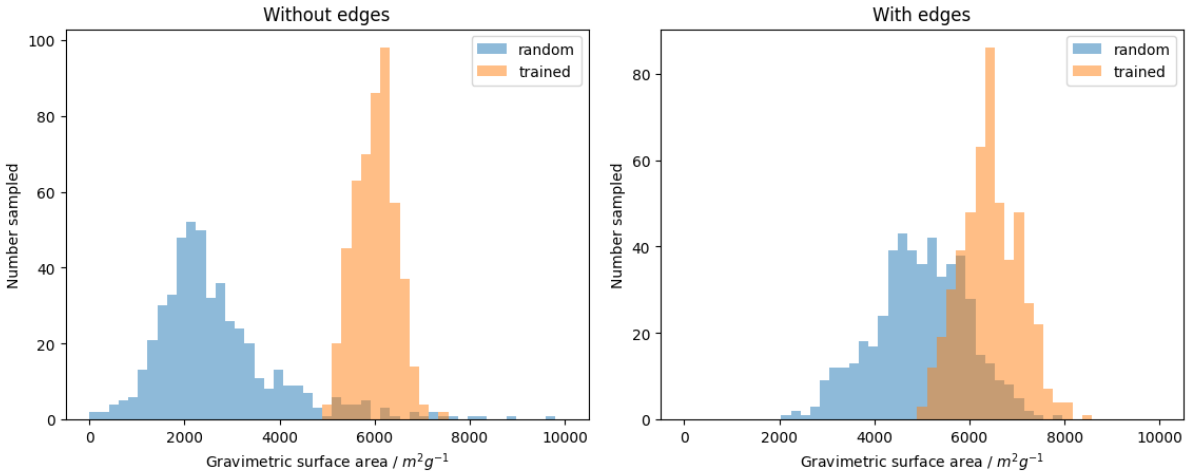

Figure 12: Performance of the GFlowNet trained on the FSO topology.

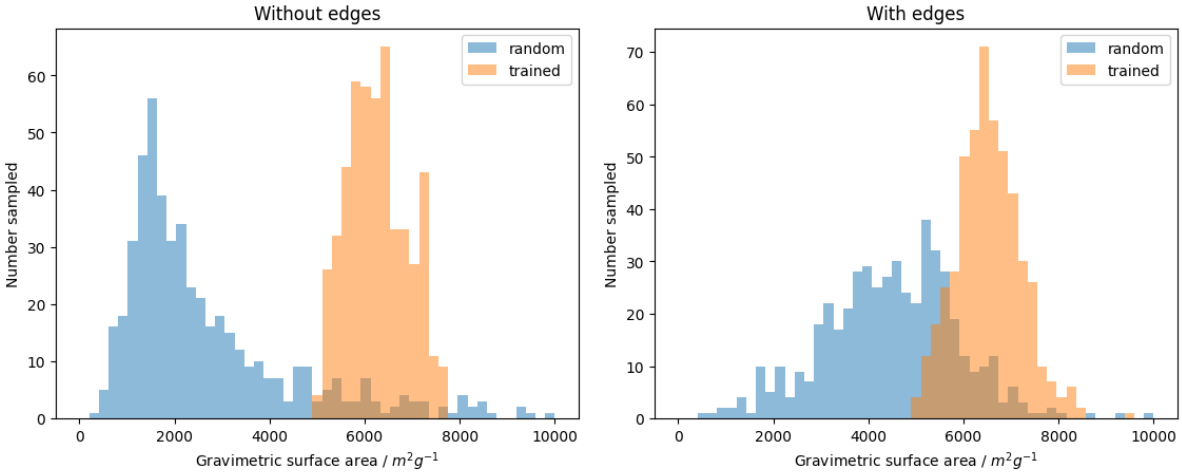

Figure 13: Performance of the GFlowNet trained on the URJ topology.

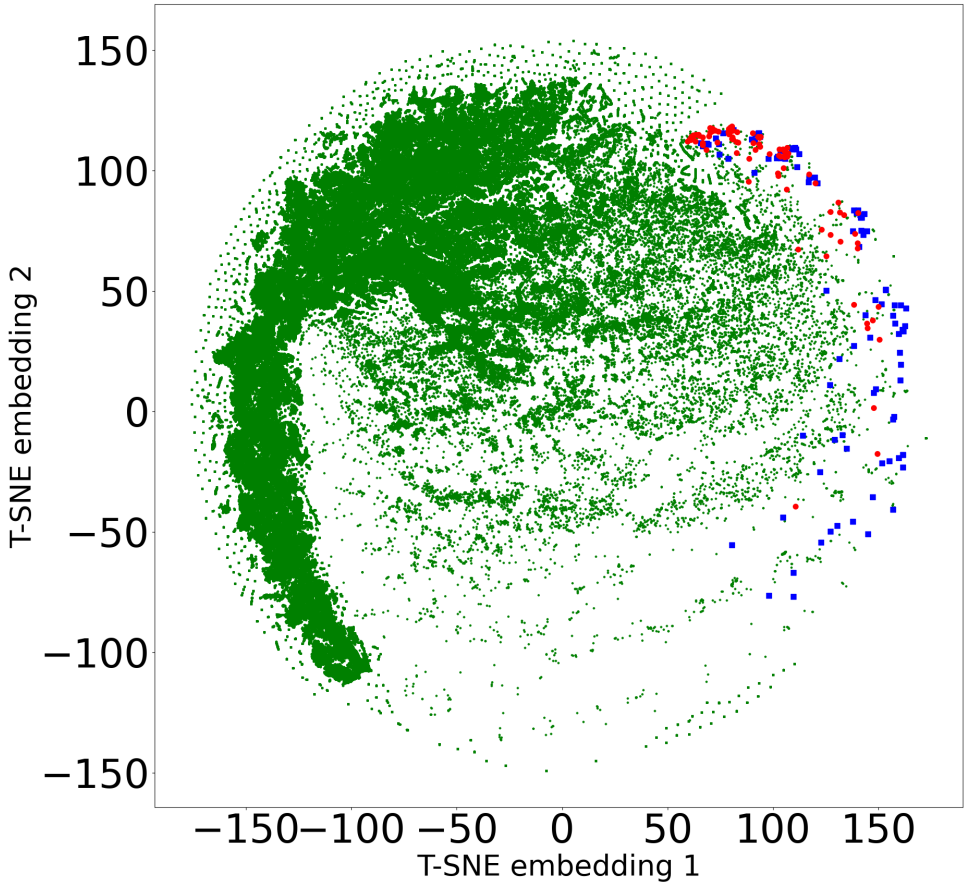

Figure 14: Two dimensional t-SNE embedding of the average minimum distance descriptor of ARC-MOF (green) and the top-100 `matgfn-rm` structures. The red circles are the unrelaxed top-100 structures. The blue squares are the relaxed top-100 structures, with two structures missing due to structural relaxation errors.

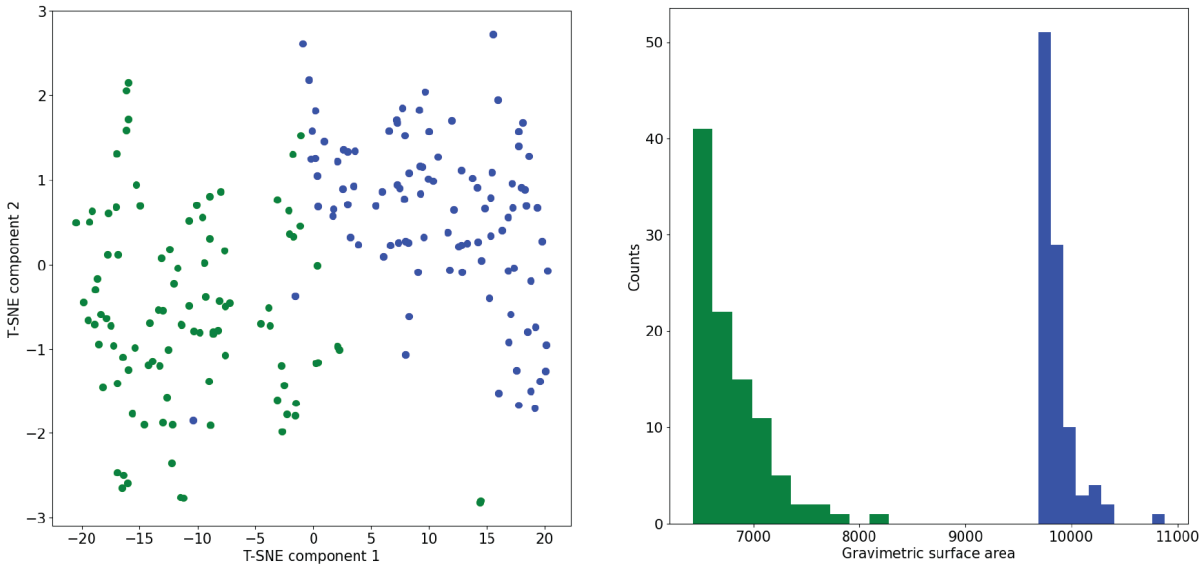

Figure 15: (Left) T-SNE embedding of the average minimum distance (AMD) for the structures with top-100 gravimetric surface area in ARC-MOF (green) and `matgfn-rm` (blue). (Right) Histogram of number of structures with a given gravimetric surface area in ARC-MOF (green) and `matgfn-rm` (blue).