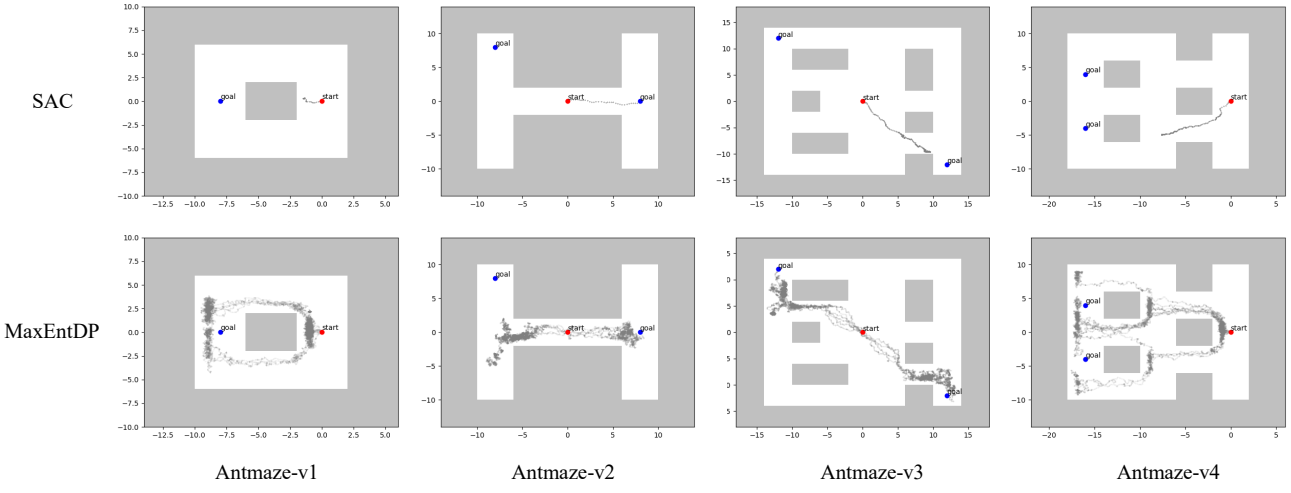

Figure 1. Trajecories generated by MaxEntDP and SAC after 1M environment interactions in Antmaze benchmarks. MaxEntDP can learn diverse behavior modes even in these challenging high-dimensional tasks, while SAC fails to learn different solutions.

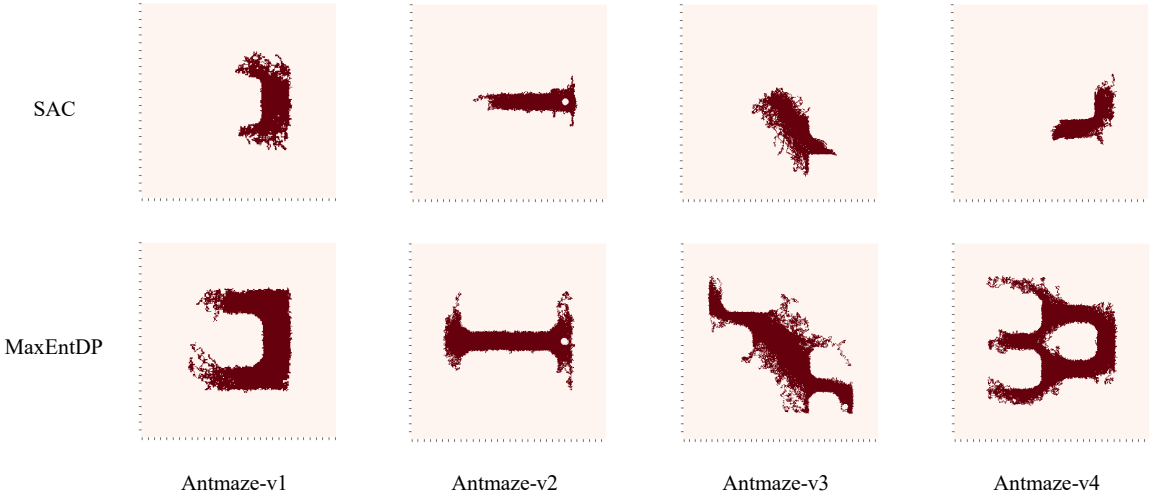

Figure 2. State coverage of MaxEntDP and SAC after 1M environment interactions in Antmaze benchmarks. MaxEntDP can explore different behavior modes at the same time and show broader state coverage than SAC, exhibiting efficient exploration of the high-dimensional state-action space.