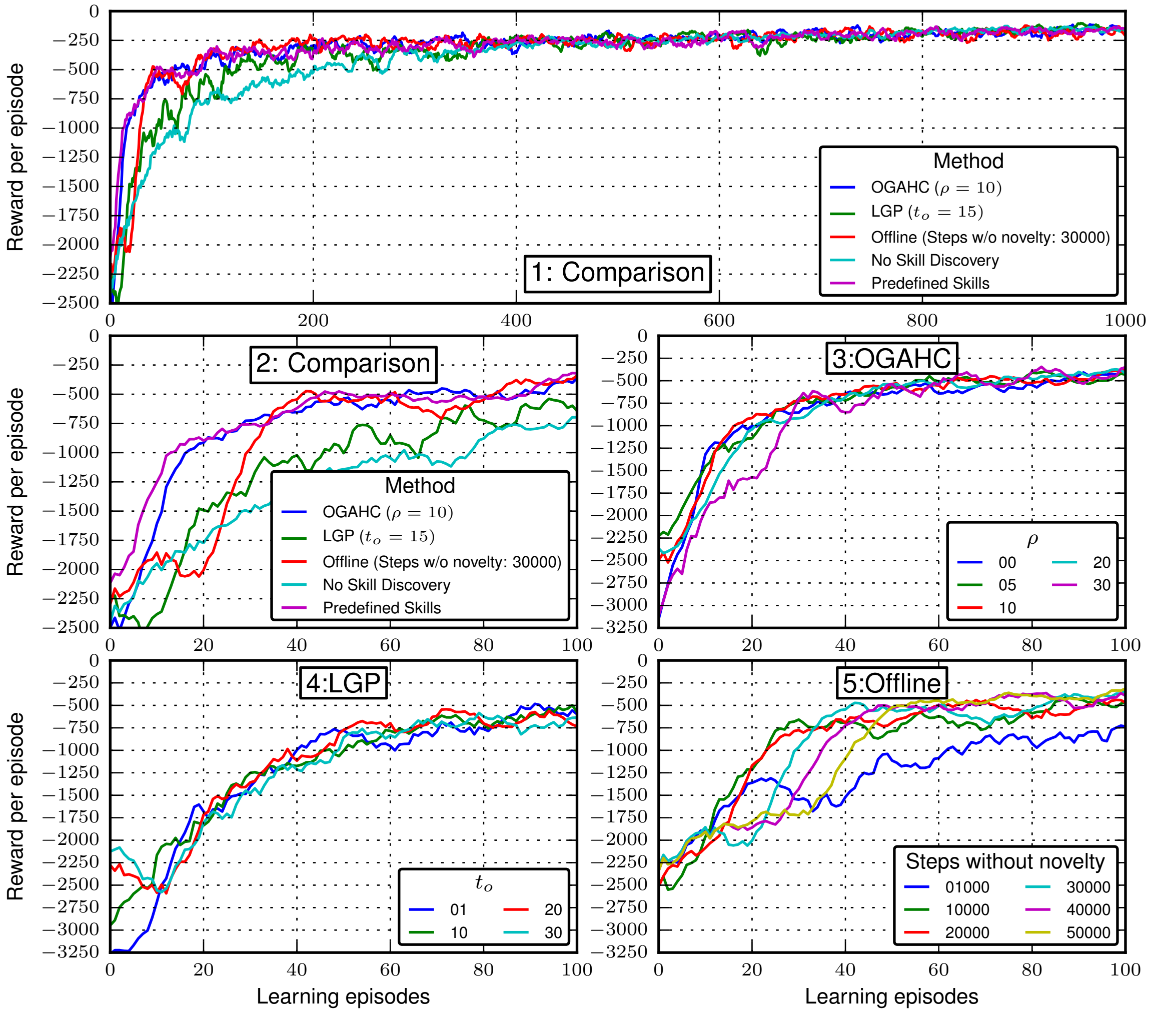
Learning curves (supplementing Figure 4)



Plot 1 shows the learning curves during the first 1000 episodes for the different methods with optimal values for ρ, t_o and "steps without novelty". Plot 2 shows the same comparison zoomed-in to the first 100 episodes. Plot 3, 4, and 5 show the learning curves of OGAHC, LGP, and Offline Clustering for different values of ρ, t_o and "steps without novelty", respectively. All curves are moving-window averages with window-length 10 and averaged over 15 repetitions.

Generation of random graphs and ground truth partitions (supplementing Section 4)

This section describes the generation of random MDPs and the corresponding ground truth partitions that are used in Section 4. These MDPs have been created as follows: the state space of all MDPs has been set to a two dimensional grid of 400 states ($S = \{0, 1, \dots, 19\}^2$) and the action space to contain four discrete actions ($A = \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$). For any state transition, a reward of -1 is given, and an episode starts always in $s_0 = (0, 0)$ and terminates in $s_t = (19, 19)$.

The MDP's state transition probability $P^a_{ss'}$ depends on an implicit connectivity graph G_c. This graph is created based on a partitioning (the later ground-truth partition) of the states that is generated by drawing 7 states (the "centers") uniform randomly from S and assigning each state to the cluster of its closest center (breaking ties randomly). A graph edge is added between any pair of states that are neighbors in the 4-neighborhood and in the same cluster. One additional edge per cluster-pair (p_i, p_j) is added between a randomly drawn pair of neighbors where one is in cluster p_i and the other in p_j. For any state-action pair s, a, let the deterministic successor state be $d(s, a) = s + a$ if $(s, s + a)$ is an edge in G_c and else $d(s, a) = s$. We set $P^a_{ss'} = 1 - (8/9)\chi$ for $s' = d(s, a)$ and $P^a_{ss'} = (1/9)\chi$ for any other state s' from the 9-neighborhood of s. χ is a parameter of the MDP that determines its stochasticity. Note that for $\chi = 0$, the connectivity of G_c controls the connectivity of the sample transition graph, while for $\chi > 0$, it influences only its edge weights.

Visualization of the 50 random graphs