New analysis of “Optimism in the face of uncertainty”.

Regret = Sum of missed rewards (also during training!) compared to an optimal policy.

New complexity parameter for MDPs: The diameter $D$.
How long does it take to travel from one state to another state?

Best known bounds for undiscounted RL with finite state/action space $S \times A$.

Regret after $T$ steps is at most

$$\text{const} \cdot DS \sqrt{TA \log (T)} .$$

PAC-like bound: The average per step regret is at most $\varepsilon$ after

$$T \geq \text{const} \cdot \frac{D^2 S^2 A}{\varepsilon^2} \log \left( \frac{DSA}{\varepsilon} \right)$$
steps.