

**Theorem 6.1** (Informal; Corollaries *F.9* and *F.17*). Consider a 2-HOP task with a token set of size  $n$ .

*For a uniformly randomly sampled train dataset  $D$  of size  $N$ , consider a learner that generalizes within the  $k$ -coverage of  $D$ . Then, for large enough  $n$ , the learner achieves perfect ID generalization with high probability if  $N \gtrsim n^c$  with  $c = 2.5 - \frac{0.5}{k}$ . In contrast, the learner (with  $k \geq 2$ ) does not achieve perfect ID generalization with high probability for some 2-HOP task if  $n^2 \lesssim N \lesssim n^c$ . Here, we ignore the polylogarithmic factors in  $n$ .*