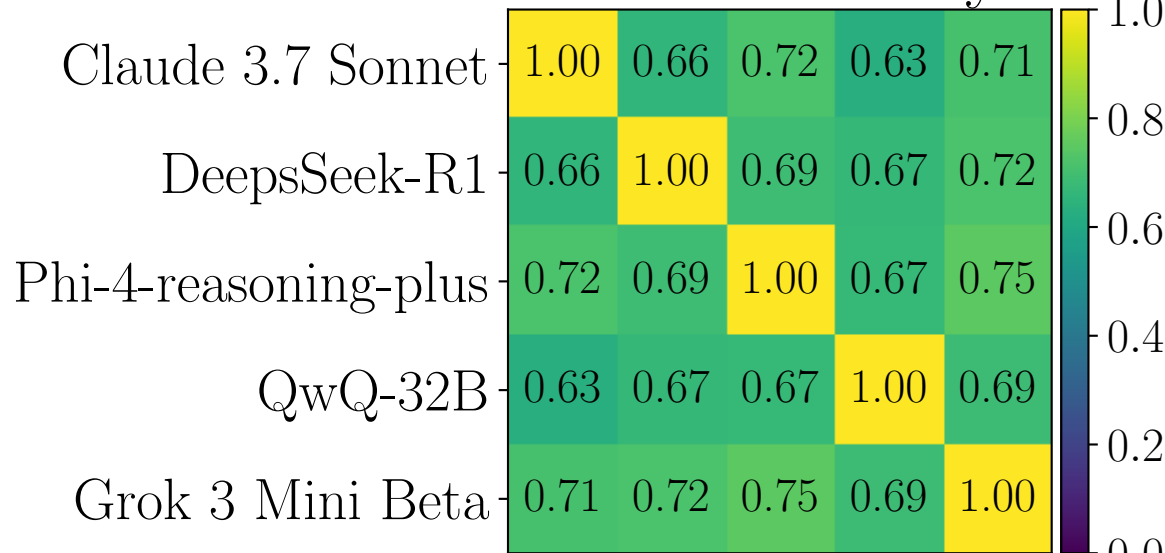
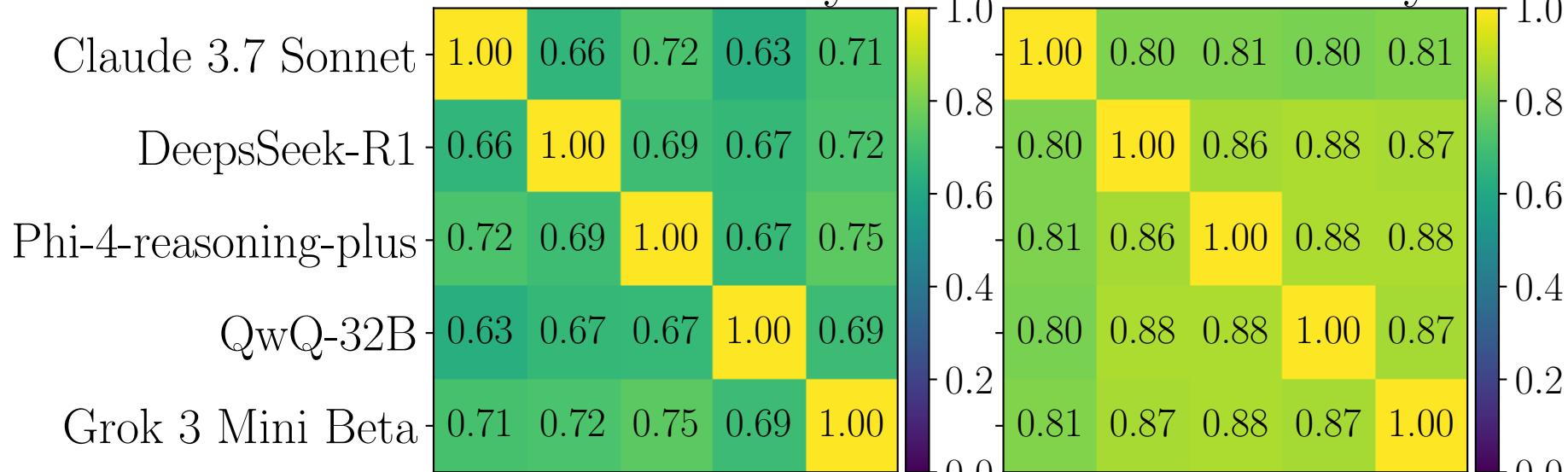


MATH-500

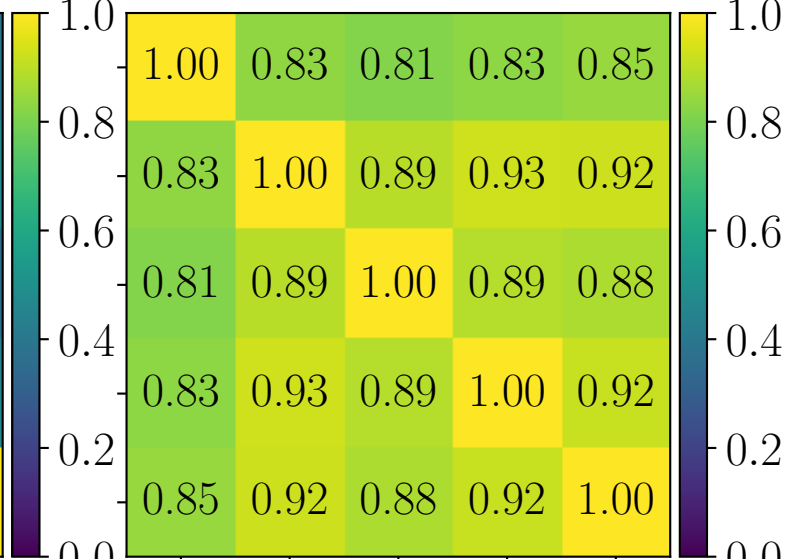
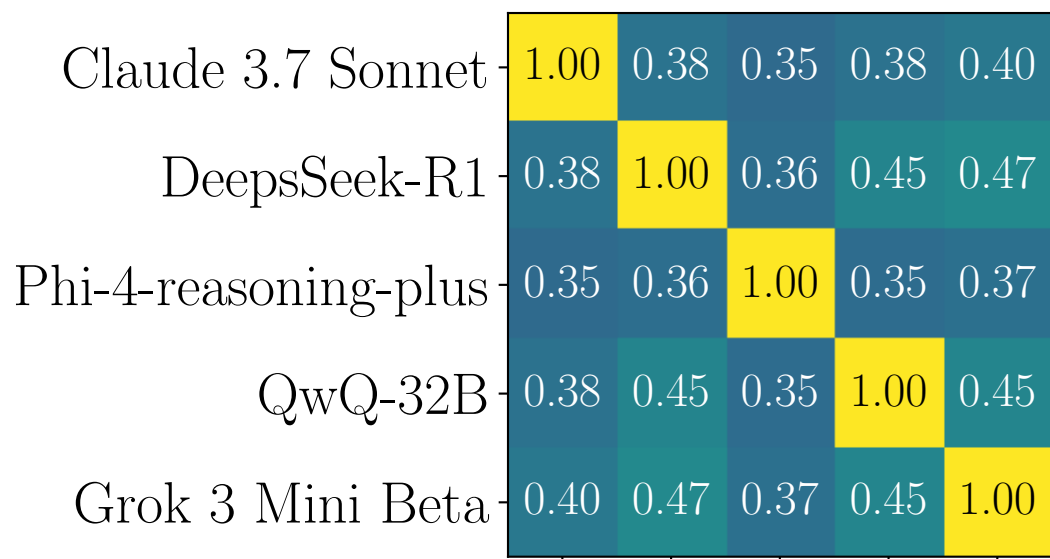
Tree Similarity



Walk Similarity



Game of 24



Claude 3.7 Sonnet

DeepsSeek-R1

Phi-4-reasoning-plus

QwQ-32B

Grok 3 Mini Beta

Claude 3.7 Sonnet

DeepsSeek-R1

Phi-4-reasoning-plus

QwQ-32B

Grok 3 Mini Beta