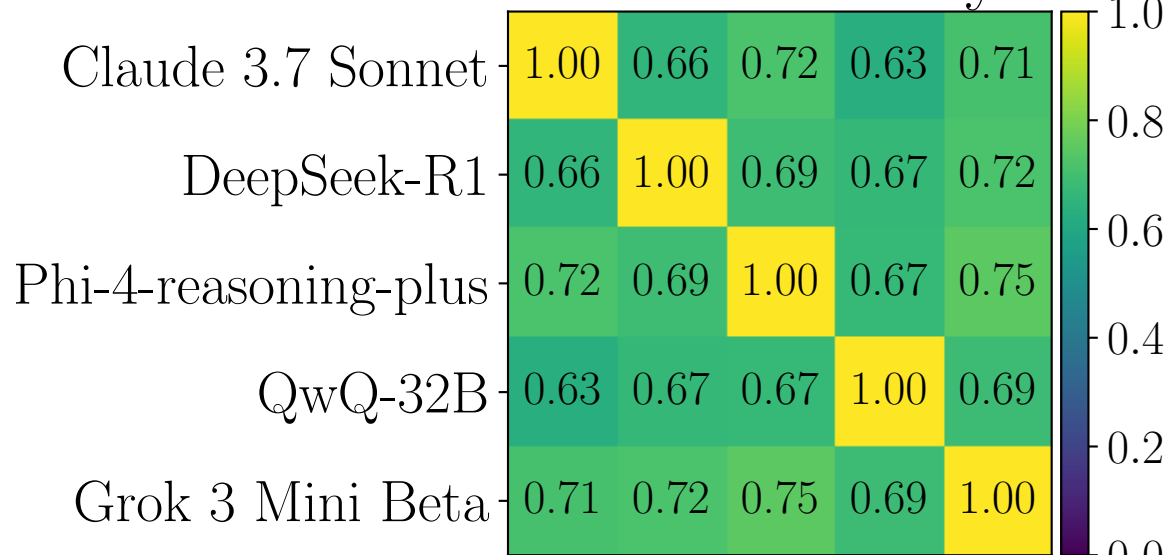
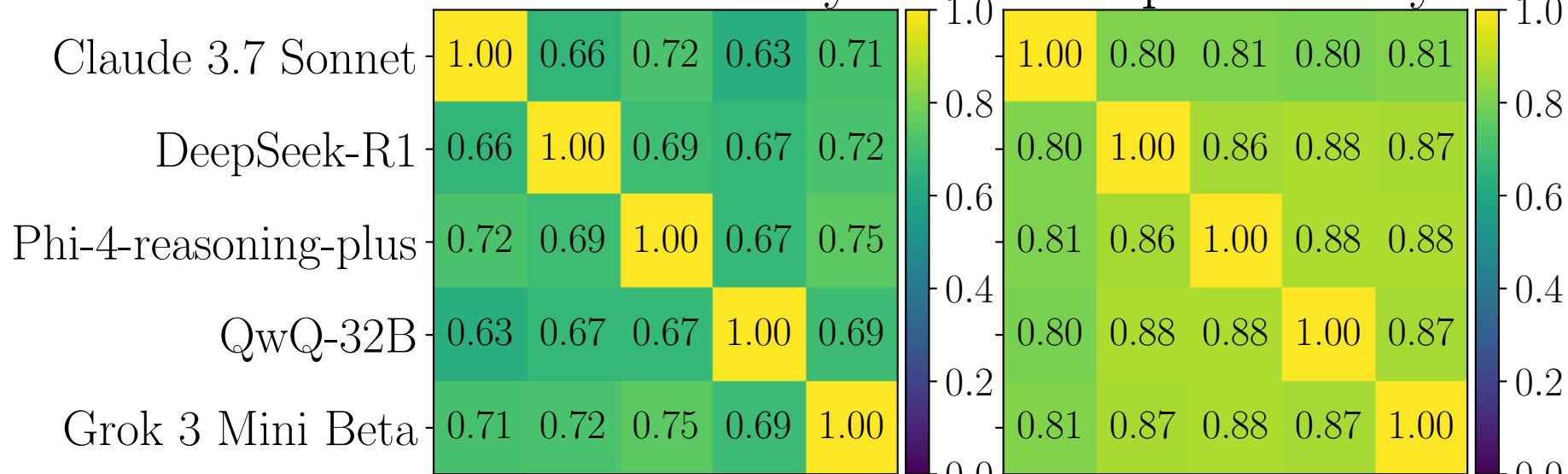


MATH-500

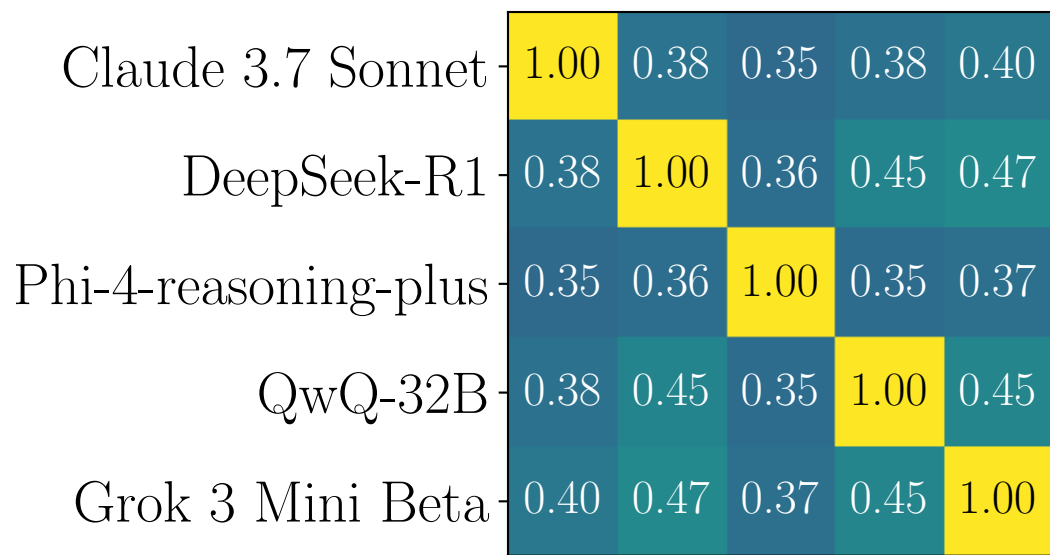
Tree Similarity



Jump Similarity



Game of 24



Claude 3.7 Sonnet

DeepSeek-R1

Phi-4-reasoning-plus

QwQ-32B

Grok 3 Mini Beta

Claude 3.7 Sonnet

DeepSeek-R1

Phi-4-reasoning-plus

QwQ-32B

Grok 3 Mini Beta