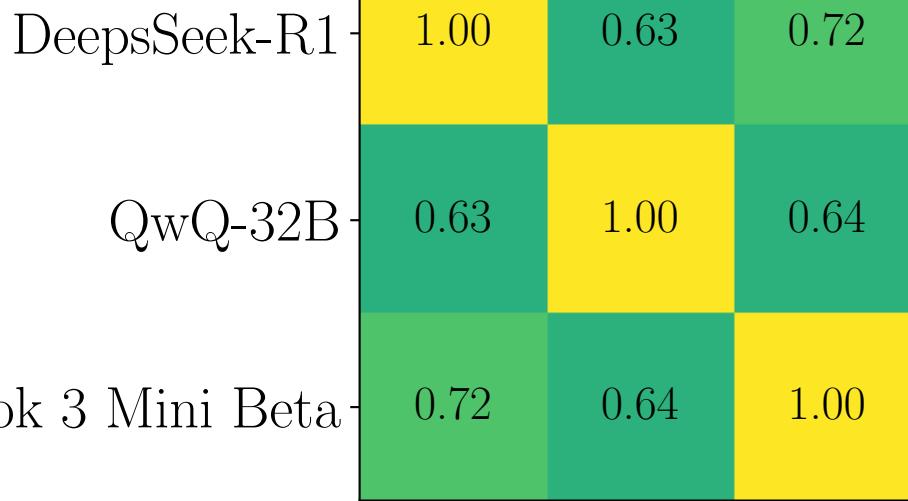
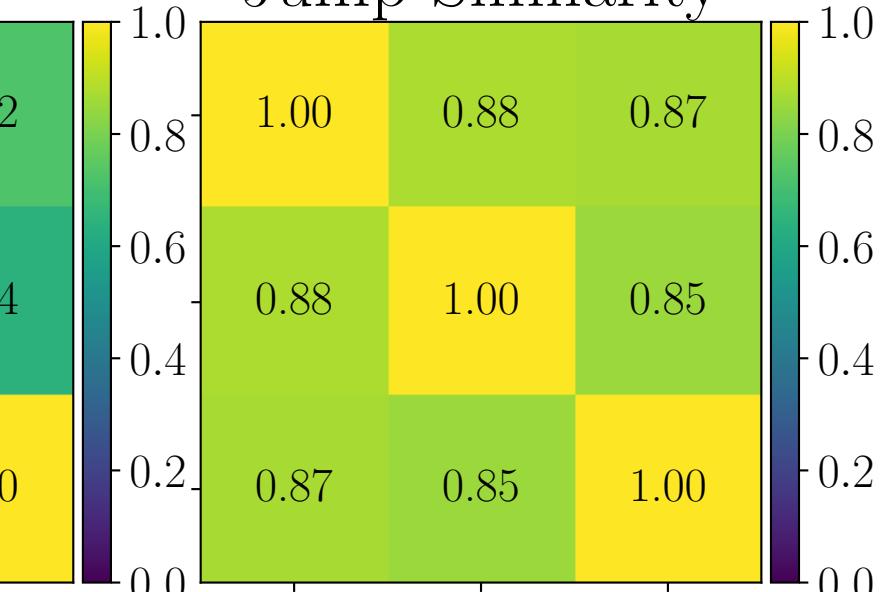


MATH-500

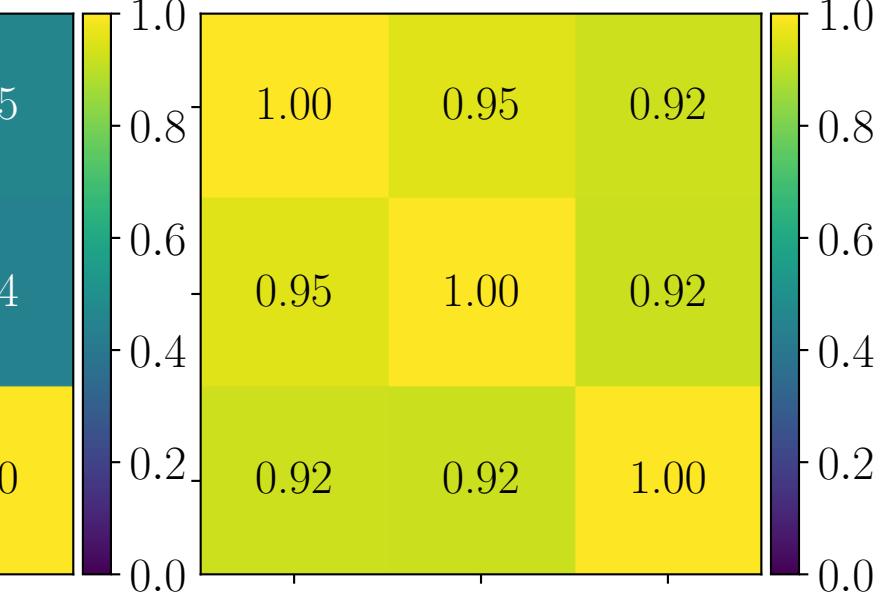
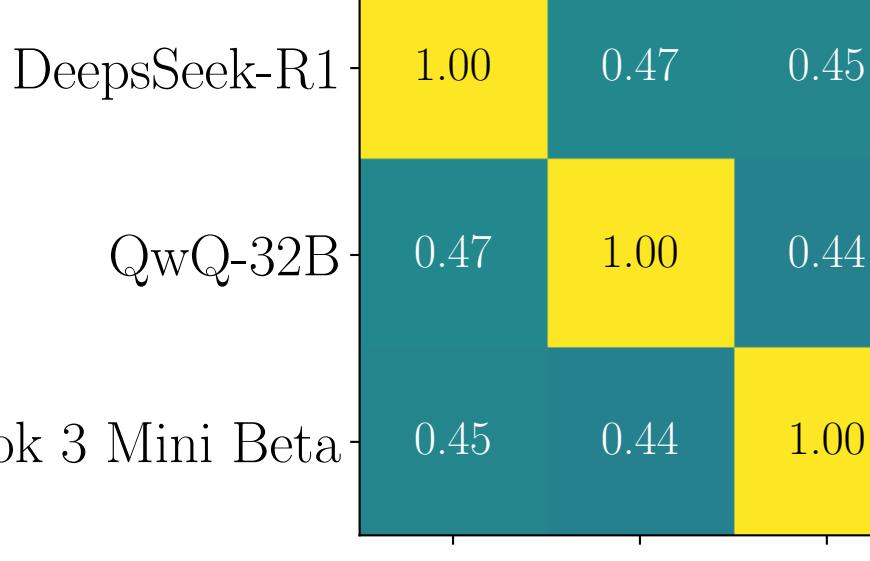
Tree Similarity



Jump Similarity



Game of 24



DeepSeek-R1
QwQ-32B
Grok 3 Mini Beta