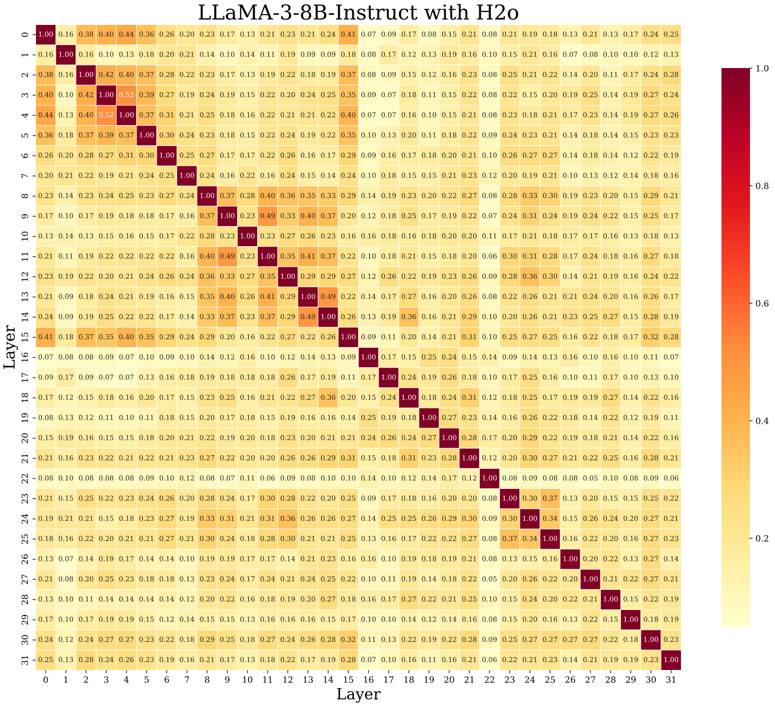

Figure 7: Layer-wise similarity heatmaps of the preserved KV cache indices by H2O on LLaMA-3-8B-Instruct

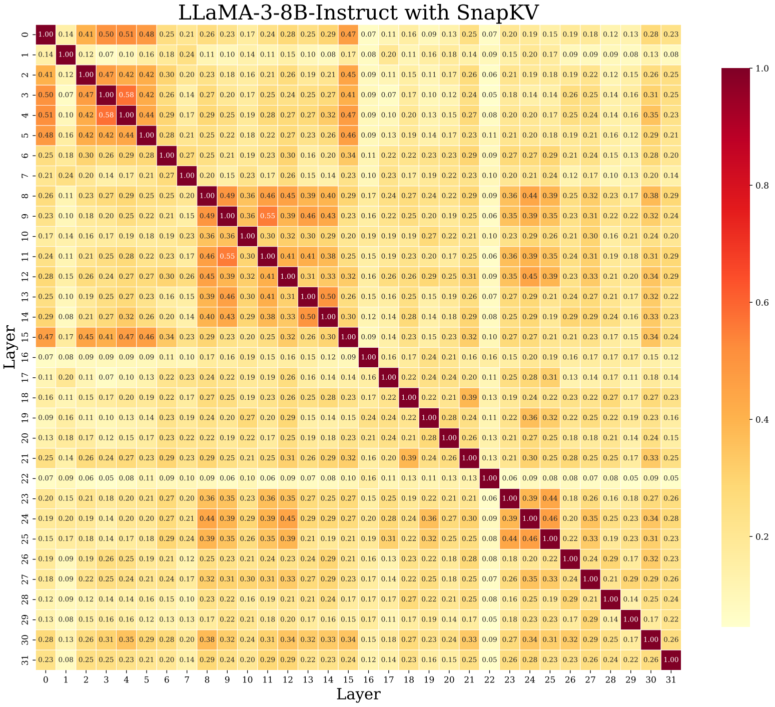

Figure 8: Layer-wise similarity heatmaps of the preserved KV cache indices by SnapKV on LLaMA-3-8B-Instruct