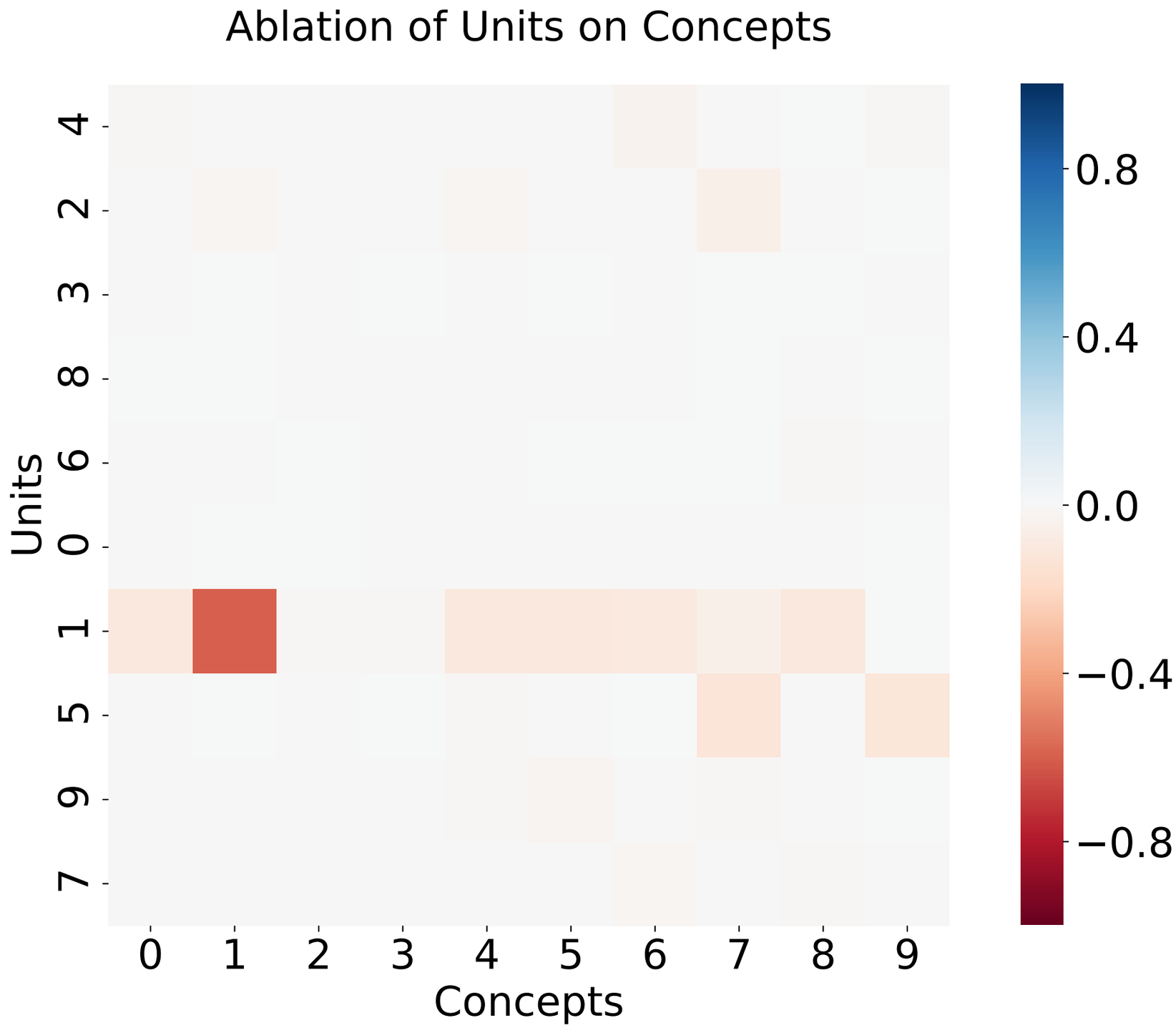
Ablation of Units on Concepts
Units
4
2
3
8
6
0
1
5
9
7
0
1
2
3
4
5
6
7
8
9
Concepts
0.8
0.4
0.0
−0.4
−0.8