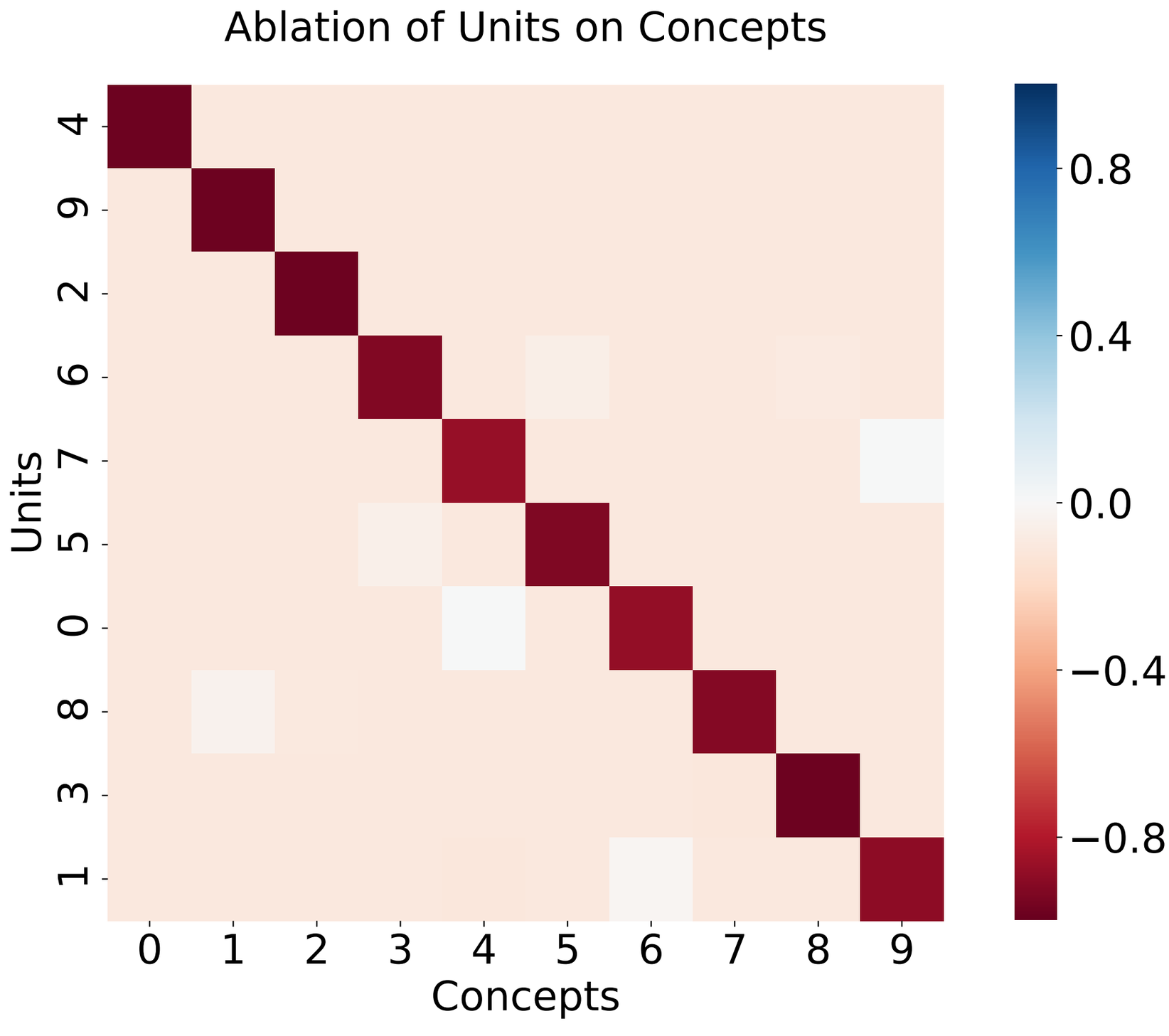
Ablation of Units on Concepts
Units
4
9
2
6
7
5
0
8
3
1
0
1
2
3
4
5
6
7
8
9
Concepts
0.8
0.4
0.0
−0.4
−0.8