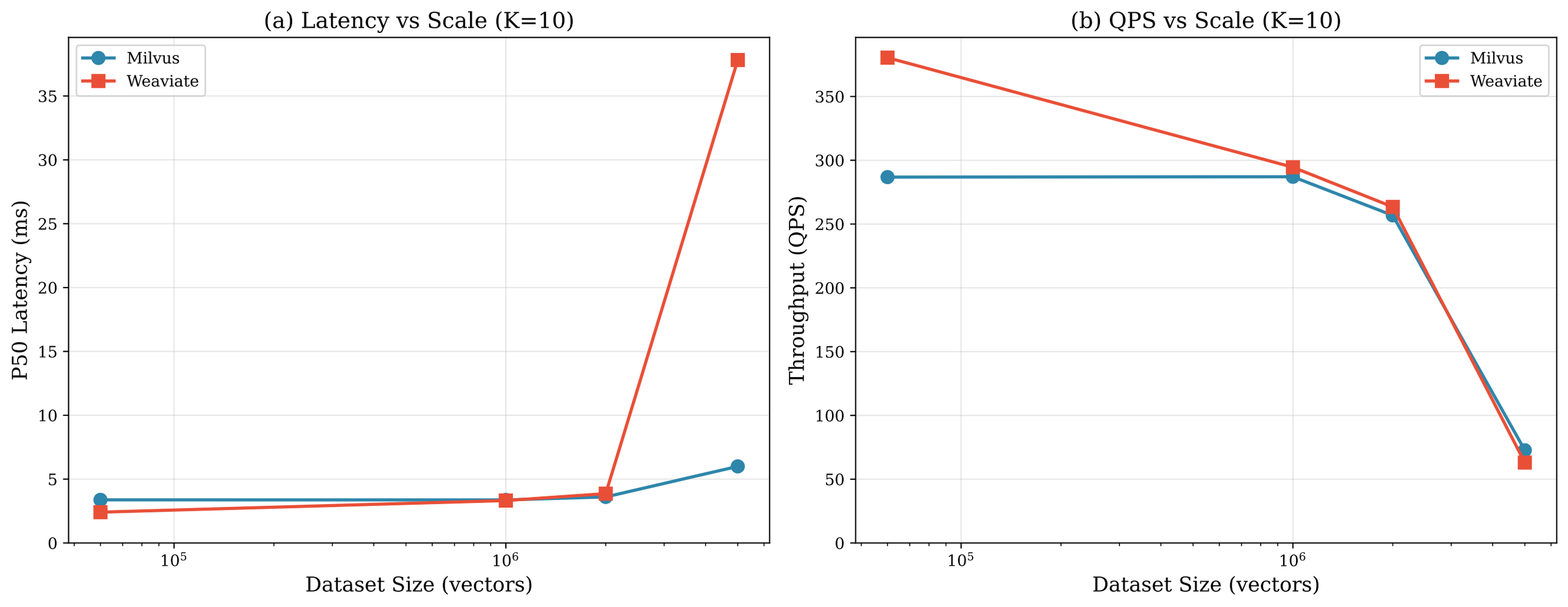
(a) Latency vs Scale (K=10)
Milvus
Weaviate
P50 Latency (ms)
0
5
10
15
20
25
30
35
10^5
10^6
Dataset Size (vectors)
(b) QPS vs Scale (K=10)
Milvus
Weaviate
Throughput (QPS)
0
50
100
150
200
250
300
350
10^5
10^6
Dataset Size (vectors)