

ELPIS: GRAPH-BASED SIMILARITY SEARCH FOR SCALABLE DATA SCIENCE

EXPERIMENTAL RESULTS : QUERY ANSWERING PERFORMANCE

Ilias Azizi
UM6P, Université Paris Cité
ilias.azizi@um6p.ma

Karima Echihabi
UM6P
karima.echihabi@um6p.ma

Themis Palpanas
Université Paris Cité & IUF
themis@mi.parisdescartes.fr

In the figures below, we present the full results of the extensive query answering experimental evaluation comparing Elpis to the state-of-the-art in-memory graph-based ng -approximate similarity search approaches. We evaluate the performance of all approaches on 5 real datasets ranging in size from 1 million to 1 billion vectors. The label of each sub figure points out the dataset name, size, and the number of nearest neighbors retrieved during search, i.e., *Deep25GB_1* refers to the search experiment on Deep dataset, of size 25GB and $k = 1$, except for hard workloads and random datasets experiments where we specify the query set noise percentage and the value of the exponent instead of the dataset size which is 25 GB.

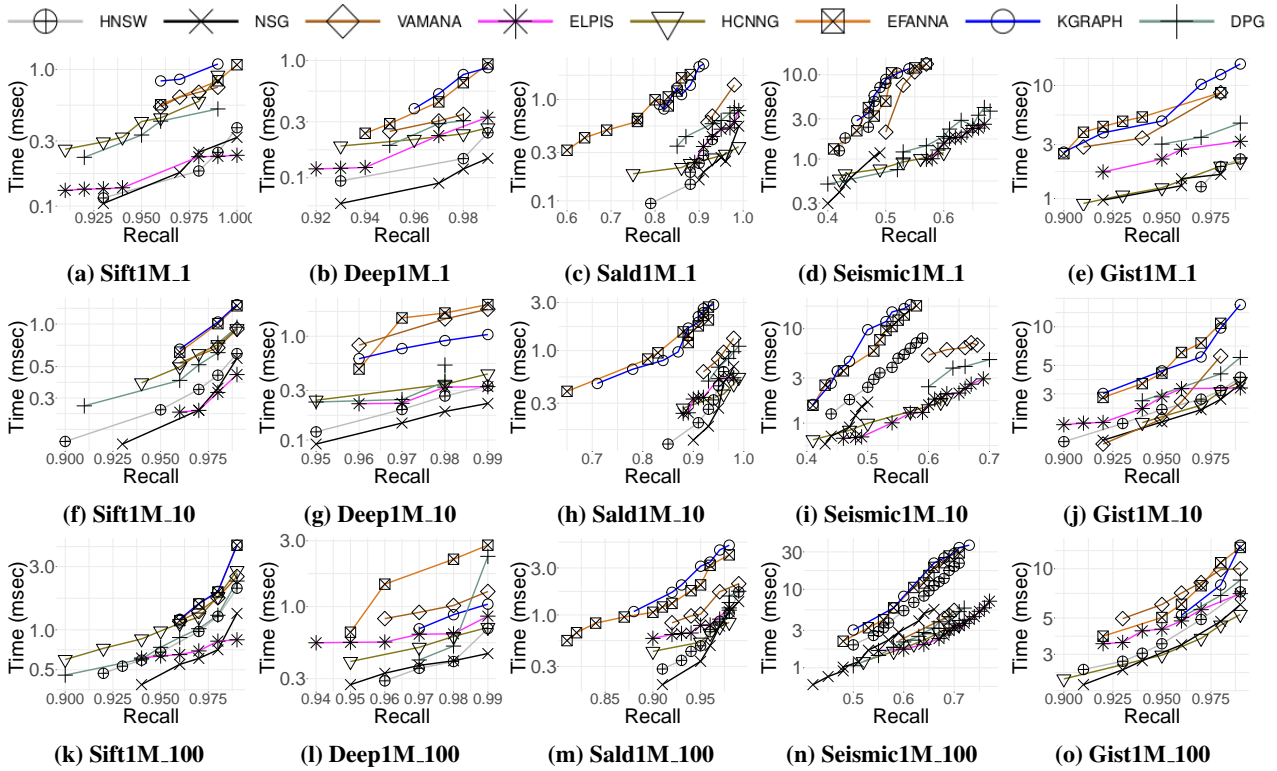


Figure 1: Efficiency vs. accuracy (Dataset = 1M, Queries = 100)

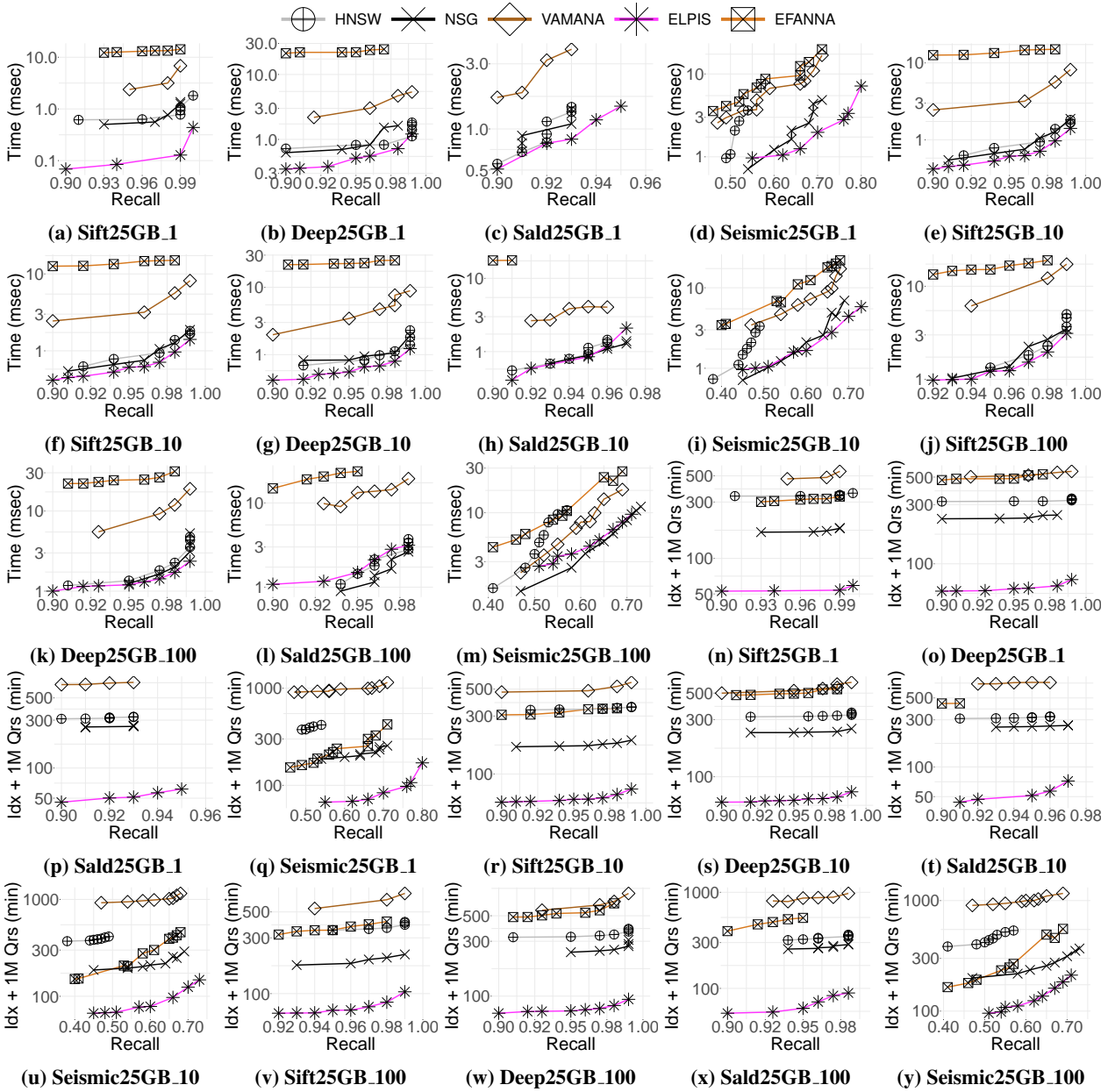


Figure 2: Efficiency vs. accuracy (Dataset = 25GB, Queries = 100)

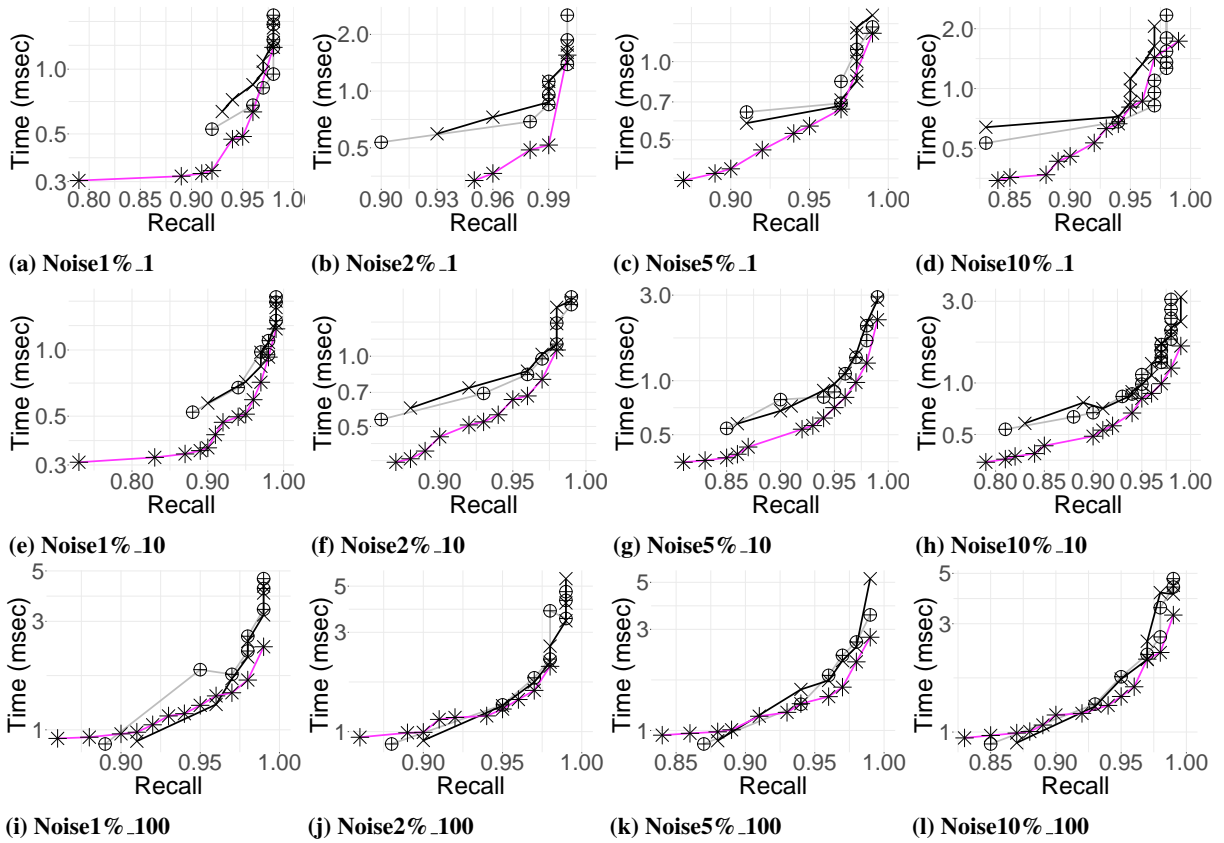


Figure 3: Efficiency vs. accuracy (Dataset = Deep25GB, Queries = 100)

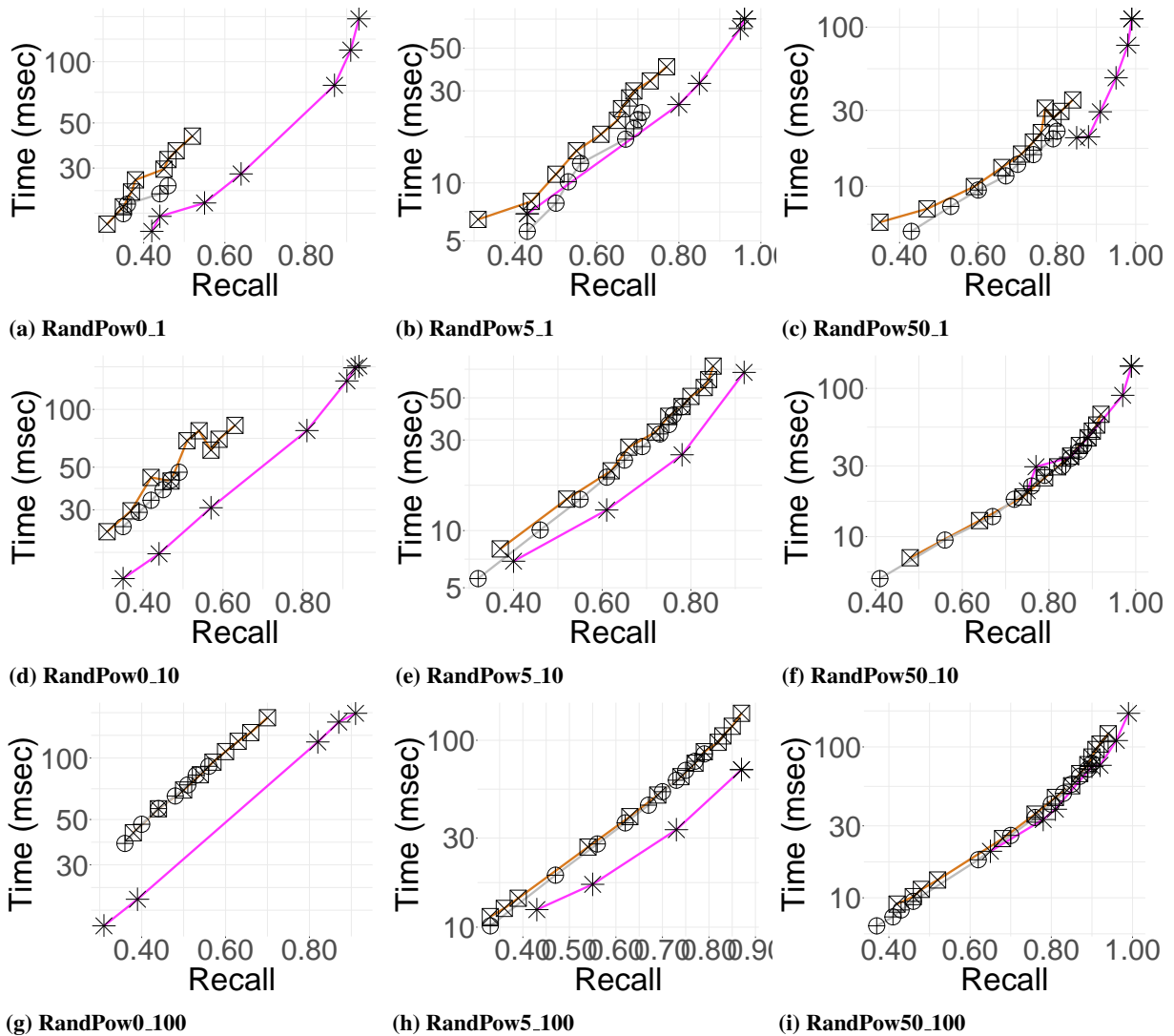


Figure 4: Efficiency vs. accuracy (Dataset = 25GB, Queries = 100)

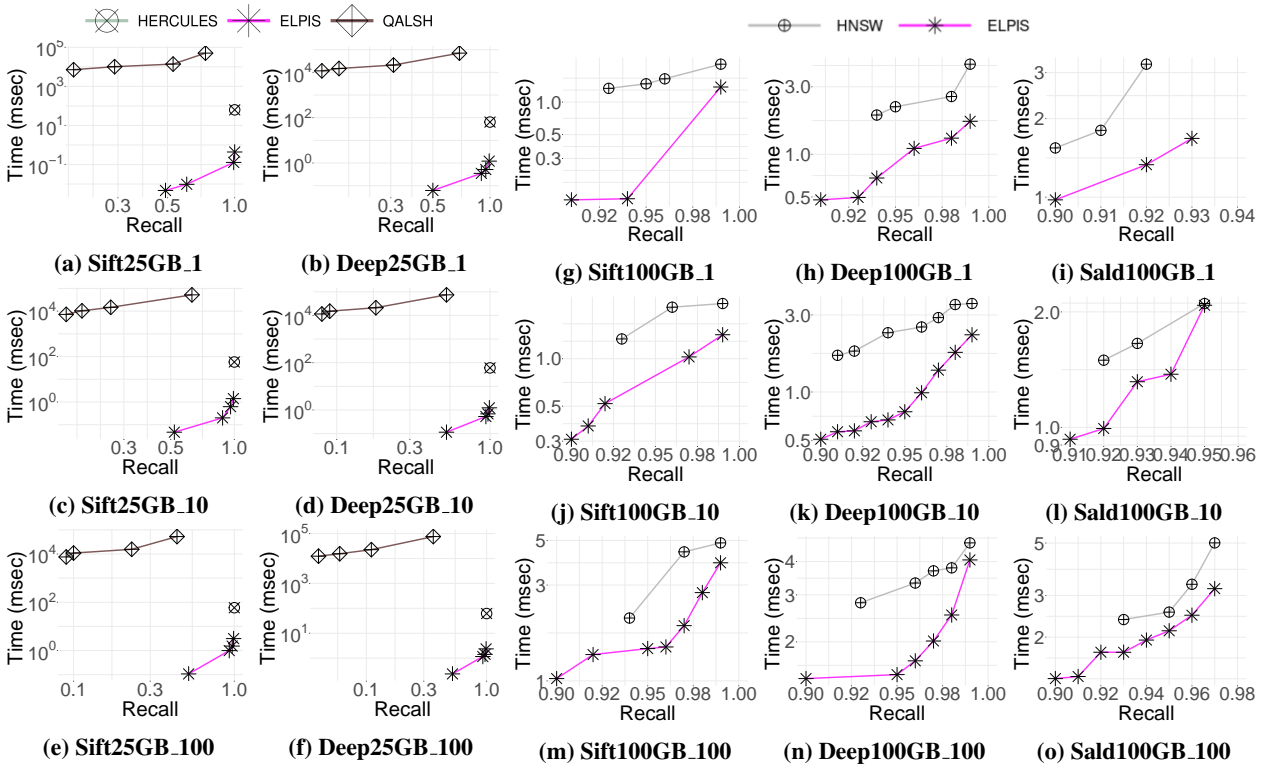


Figure 5: Efficiency vs. accuracy (Dataset = 25GB&100GB, Queries = 100)

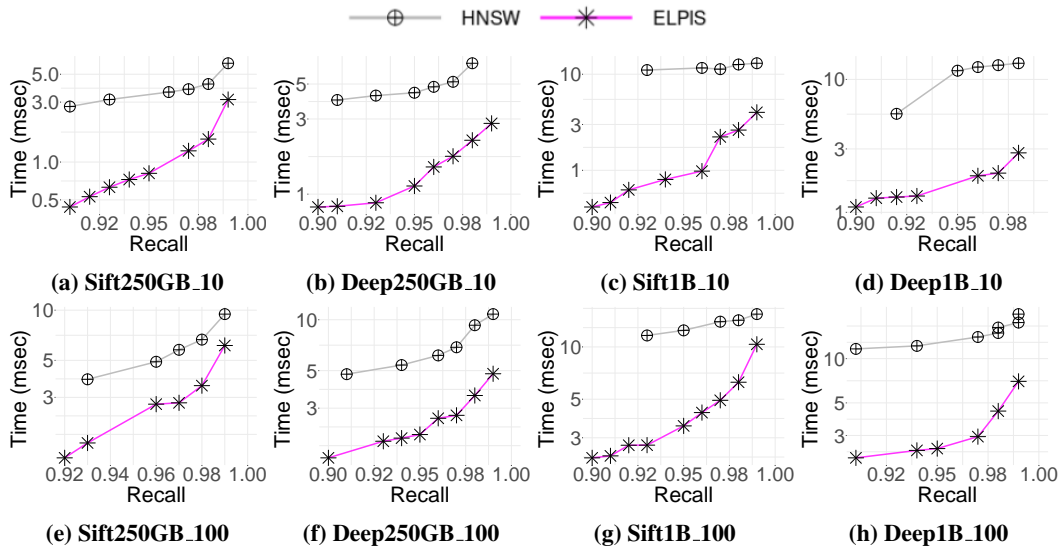


Figure 6: Efficiency vs. accuracy (Dataset = 250GB&1B, Queries = 100)