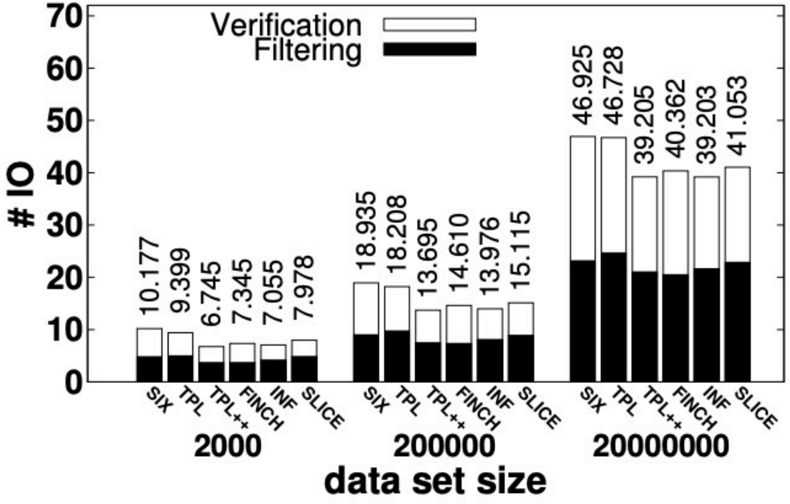
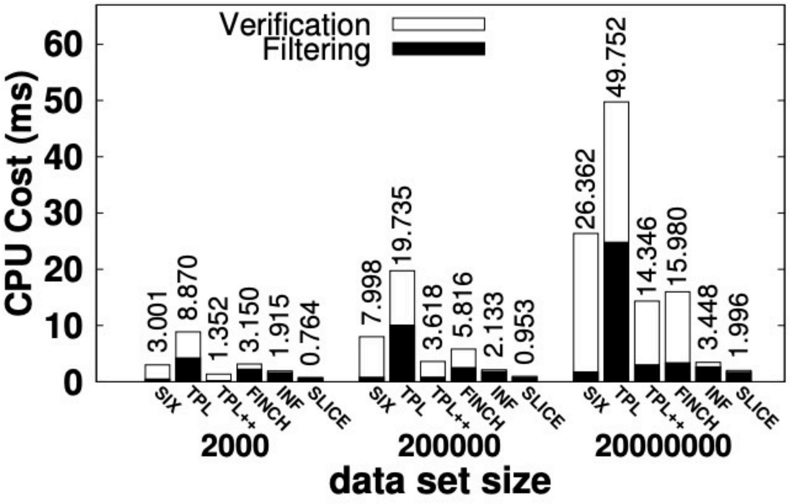
# Figures Plotting

# 1. Bar Charts

# Bar Chart




(a) Effect on I/O cost




(b) Effect on CPU cost

**Figure 15: Effect of data set size (enormous change)**

Shiyu Yang, Muhammad Aamir Cheema, Xuemin Lin, and Wei Wang. Reverse k nearest neighbors query processing: Experiments and analysis. PVLDB 8, no. 5 (2015): 605-616.

**See samples at BarChart/1.1**

# Bar Chart

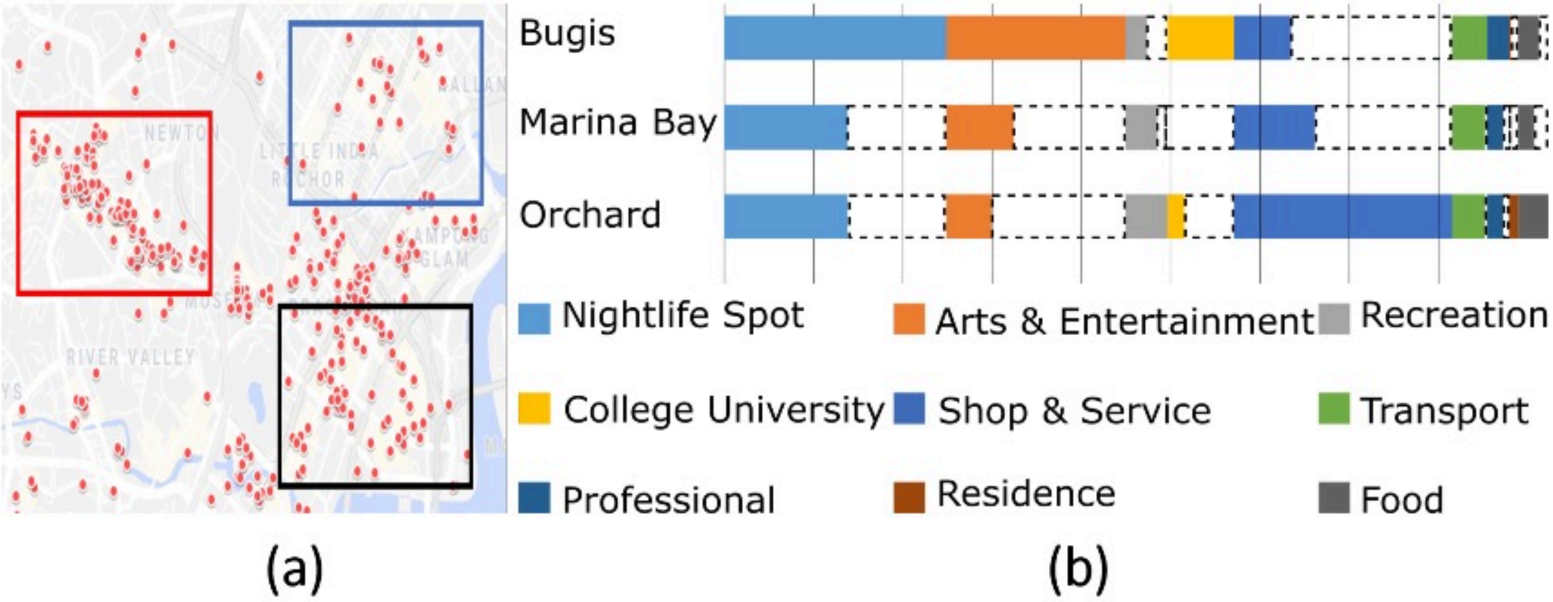
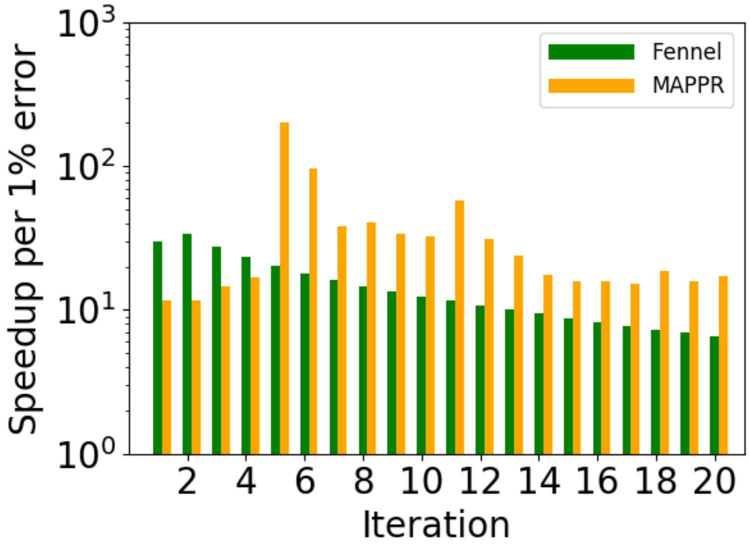
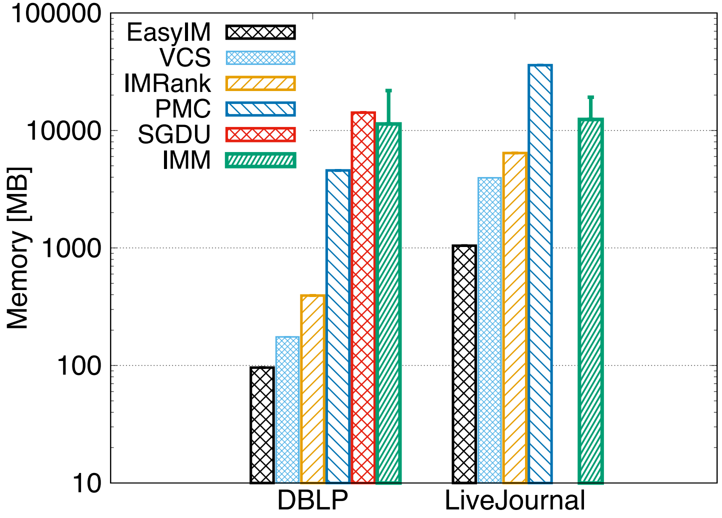



Figure 14: A Case study on Singapore.

Kaiyu Feng, Gao Cong, Christian S. Jensen, and Tao Guo. Finding attribute-aware similar region for data analysis. PVLDB 12, no. 11 (2019): 1414-1426.

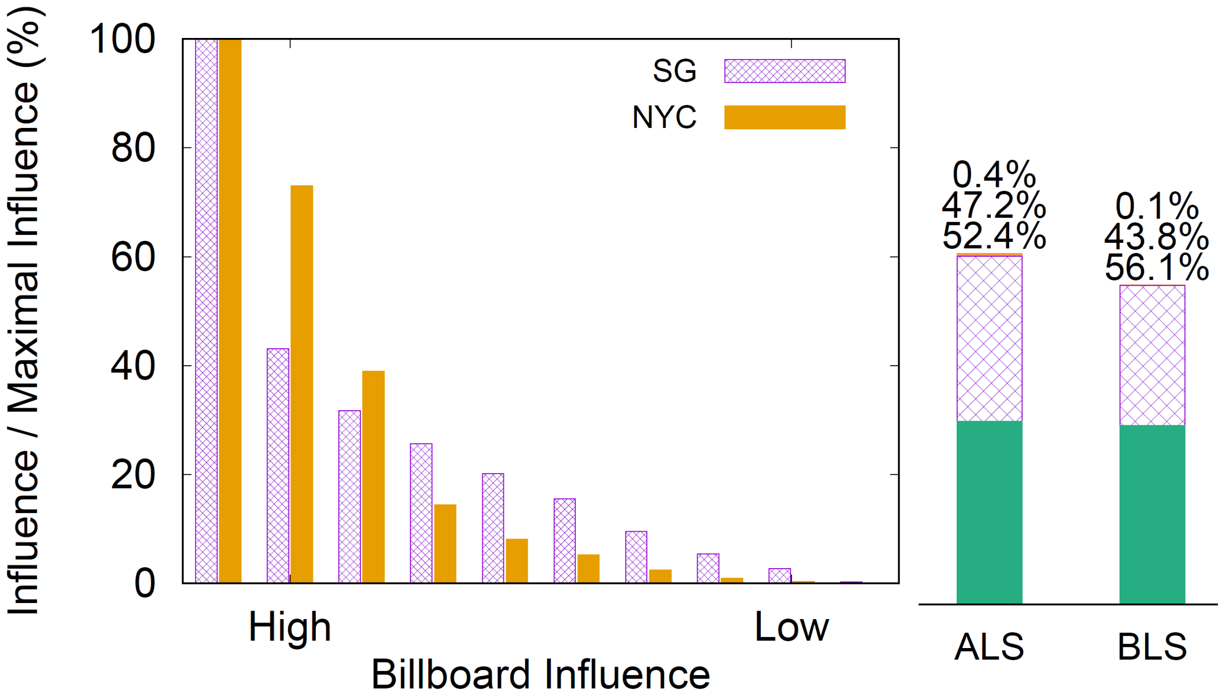
# Bar Chart



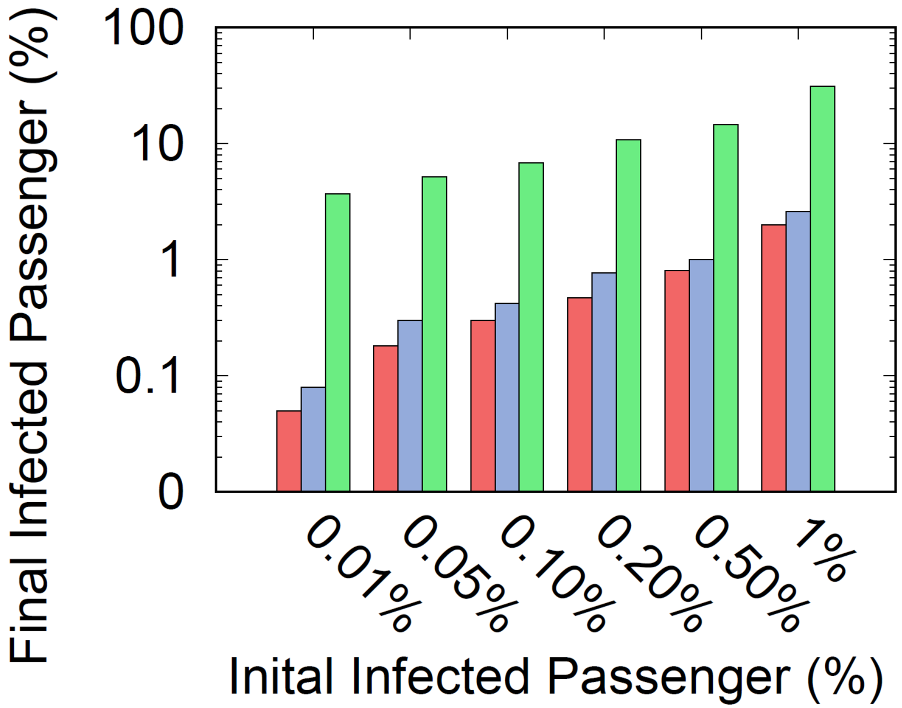

**See samples at BarChart/1.2**

# Bar Chart



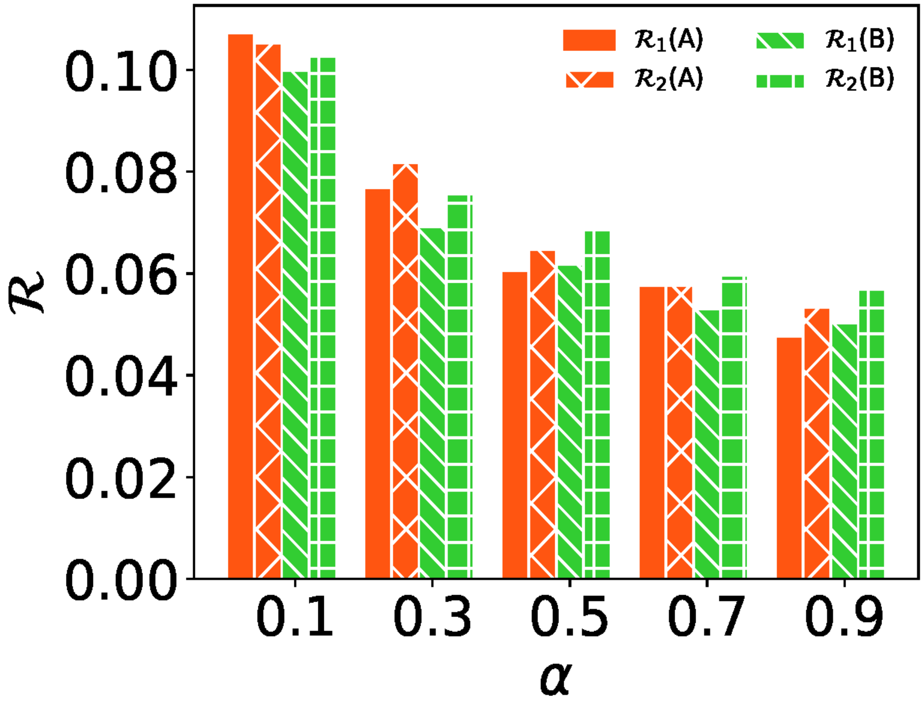

**See samples at BarChart/1.3**

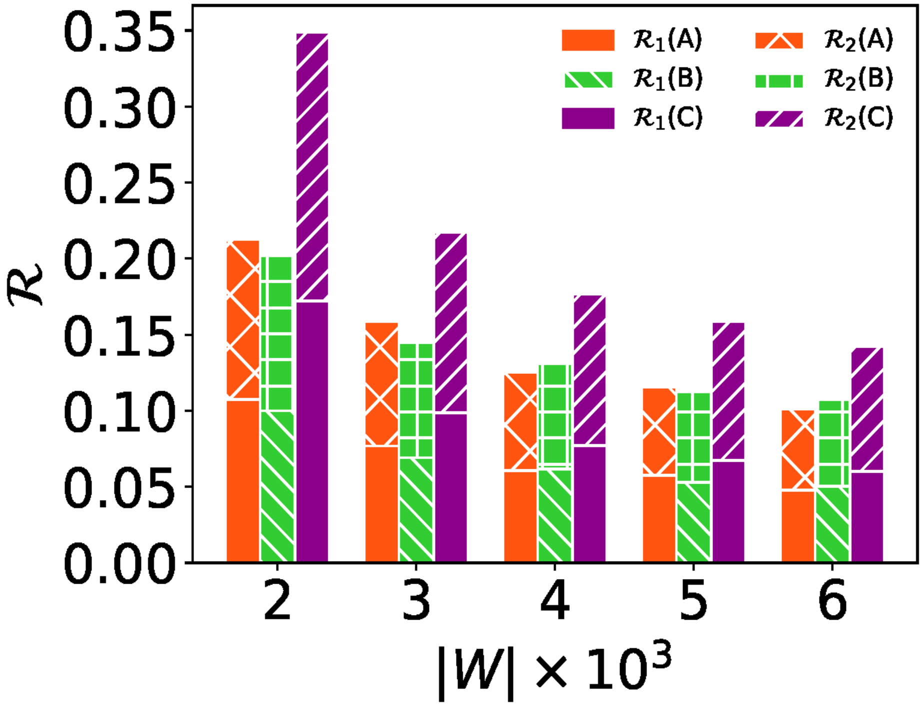
# Bar Chart




**See samples at BarChart/1.3**

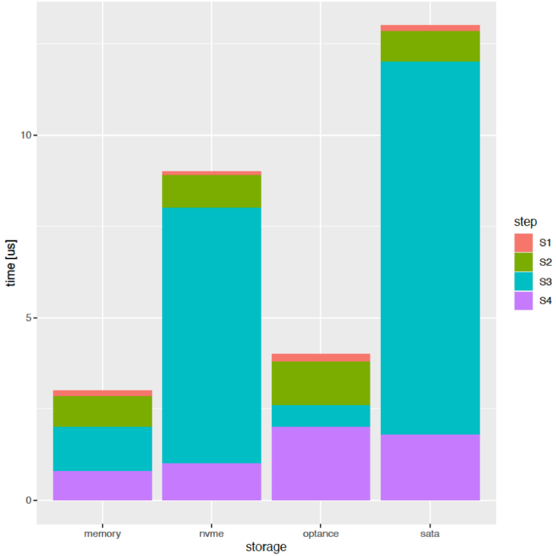
# Bar Chart

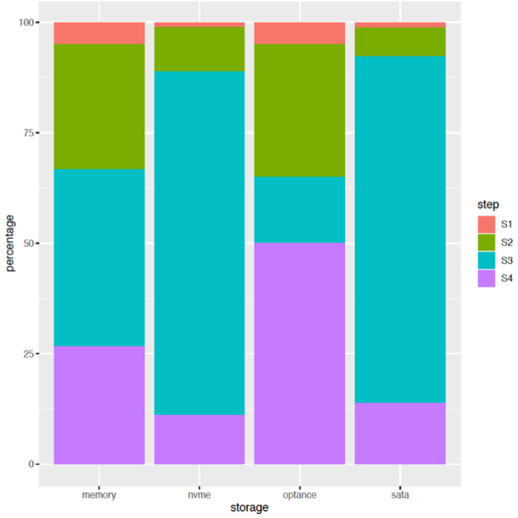


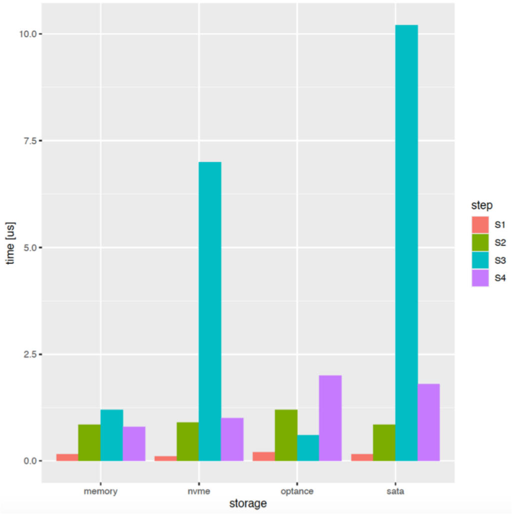



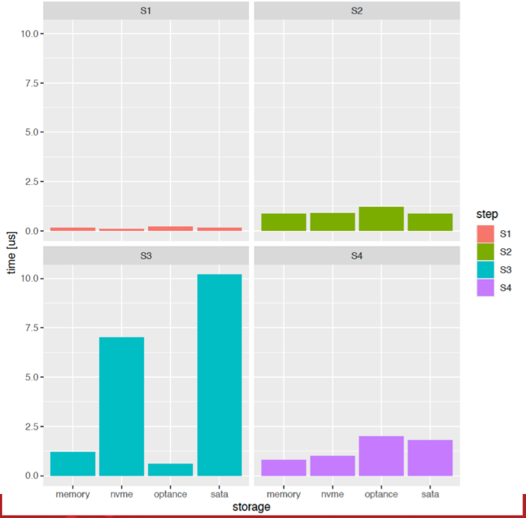

**See samples at BarChart/1.4**

# Bar Chart



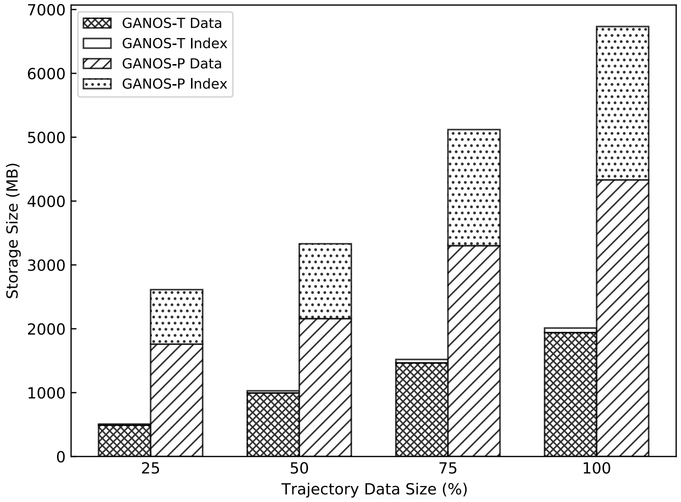




**See samples at BarChart/1.5**

# Bar Chart




**See samples at BarChart/1.6**

[1]

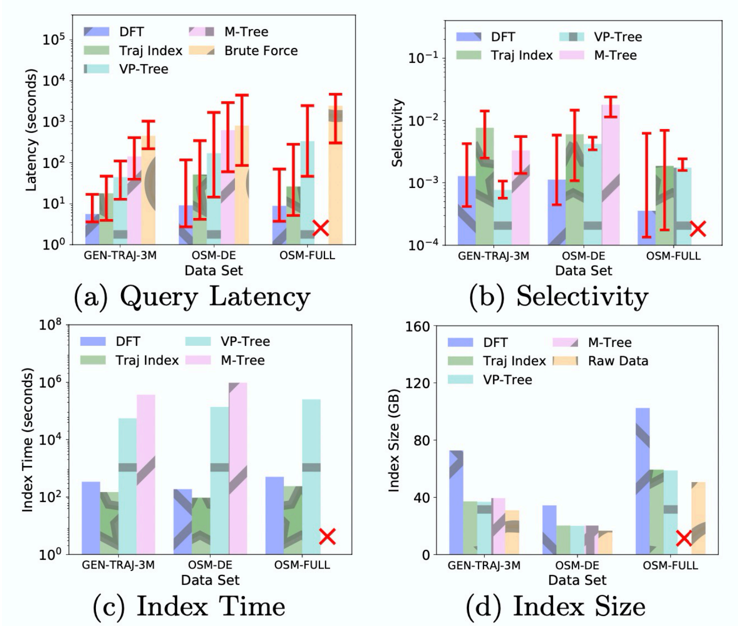



**Figure 8: Comparison against baseline solutions.**

[2]

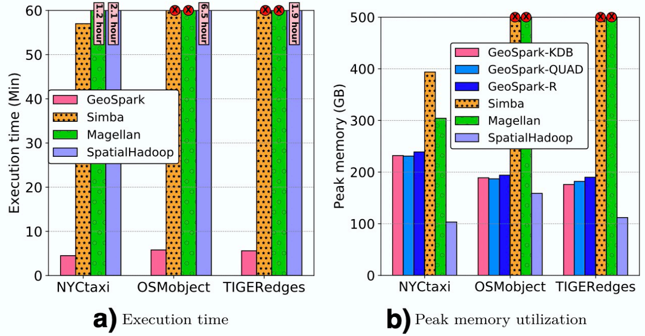



**Fig. 22** Range join query performance on datasets ⋈ OSMpostal


[1] Dong Xie, Feifei Li, and Jeff M. Phillips. 2017. Distributed Trajectory Similarity Search. VLDB 10, 11 (2017), 1478–1489.

[2] Jia Yu, Zongsi Zhang, and Mohamed Sarwat. Spatial data management in apache spark: the GeoSpark perspective and beyond. GeoInformatica 23, 1 (2019), 37–78.

# Bar Chart

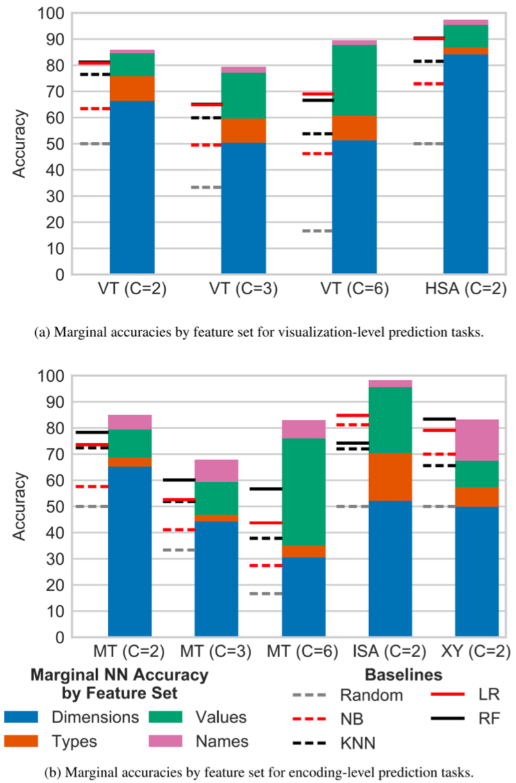
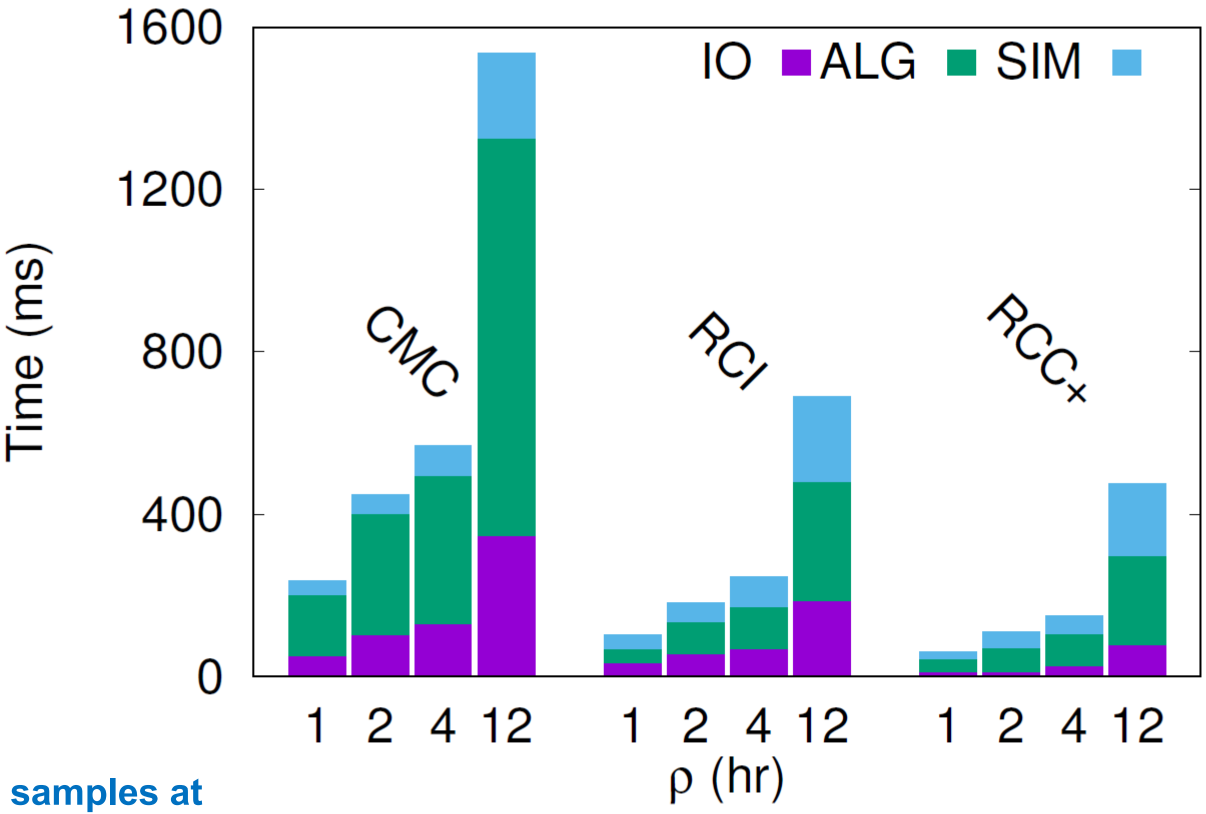



Fig. 9: Marginal contribution to NN accuracy by feature set, for each task. Baseline accuracies are shown as solid and dashed lines.

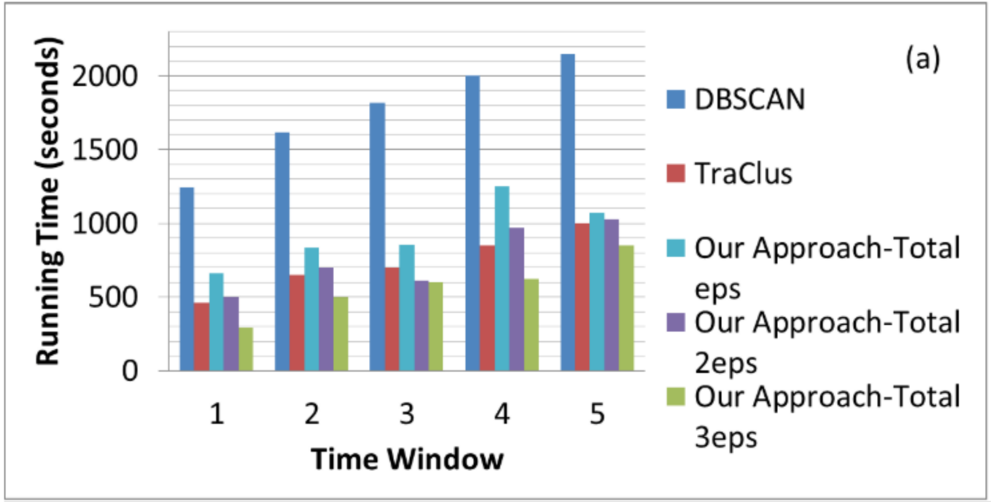
Hu K, Bakker M A, Li S, et al. Vizml: A machine learning approach to visualization recommendation. CHI Conference on Human Factors in Computing Systems. 2019: 1-12.

# Bar Chart




**See samples at BarChart/1.7**

# Bar Chart



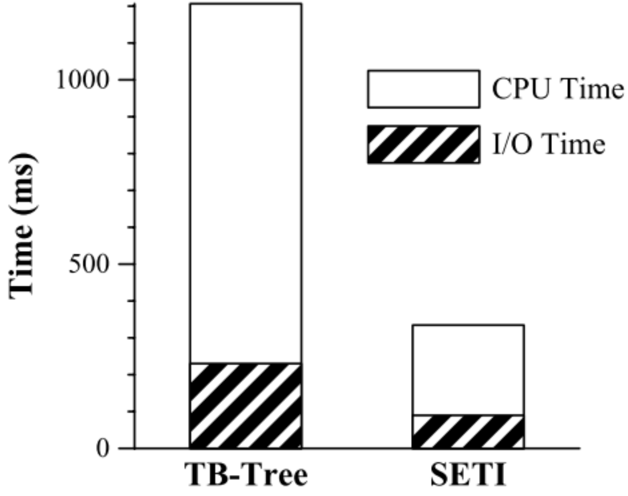
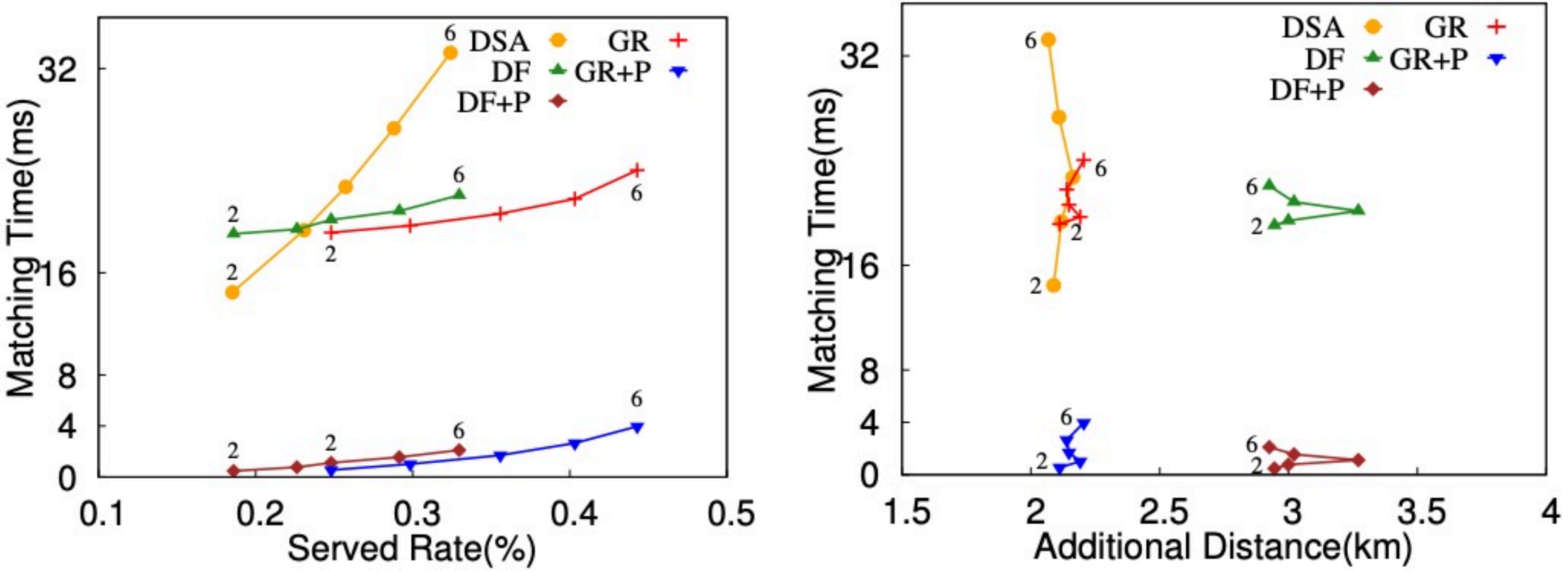




Figure 11: CPU and I/O components, GSTD(1K, 4M), 0.1% Time-interval Query

# 2. Line Charts

# Line Chart (Tradeoff Graph)




(a) Served rate

(b) Additional distance

Hui Luo, Zhifeng Bao, Farhana Choudhury, and Shane Culpepper. Dynamic ridesharing in peak travel periods. TKDE (2019).

**See samples at LineChart/2.1**

# Line Chart (Tradeoff Graph)

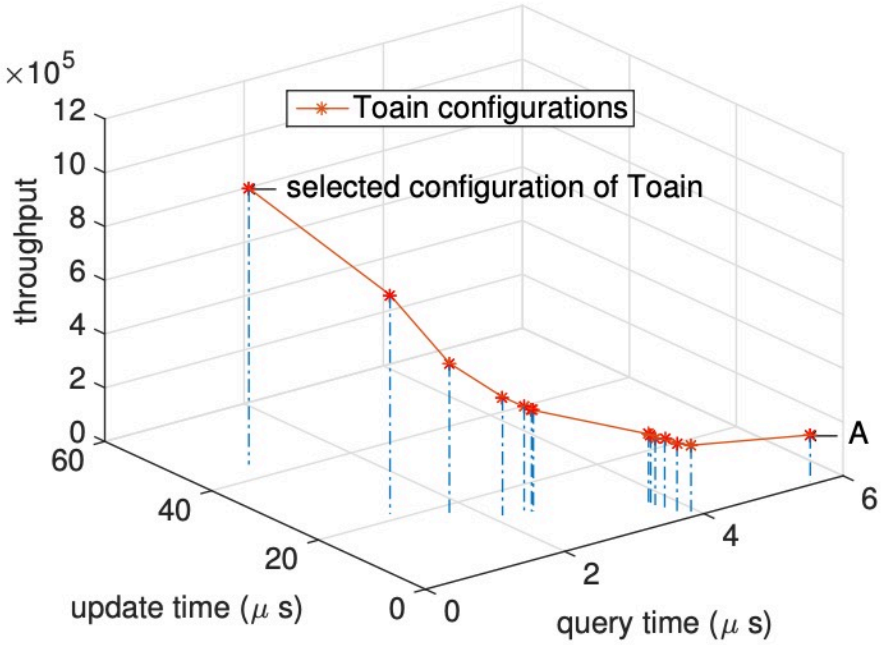
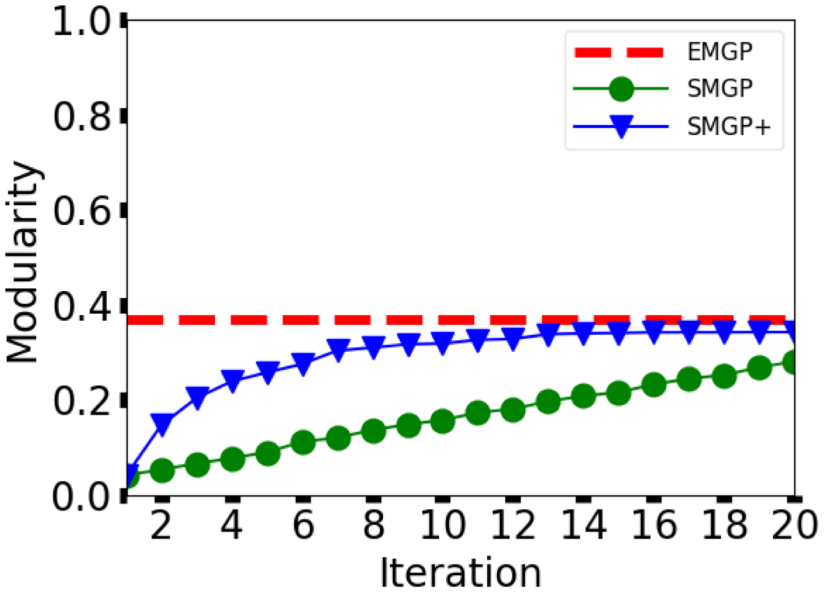
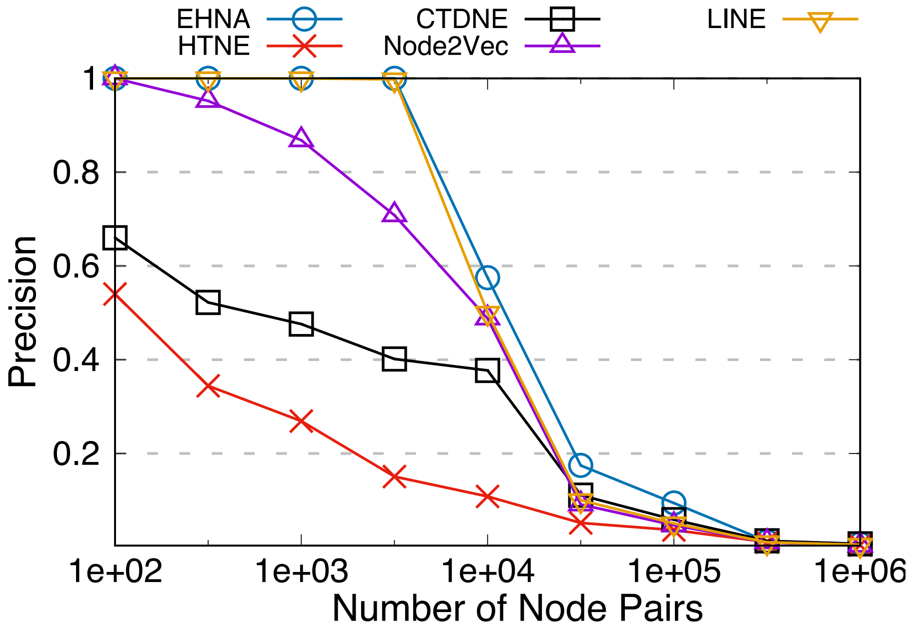



**Figure 7: SCOB configurations (Case Study 2).**

Siqiang Luo, Ben Kao, Guoliang Li. TOAIN: A Throughput Optimizing Adaptive Index for Answering Dynamic kNN Queries on Road Networks. PVLDB 2018: 594-606.

# Line Chart




**See samples at LineChart/2.2**

# Line Chart

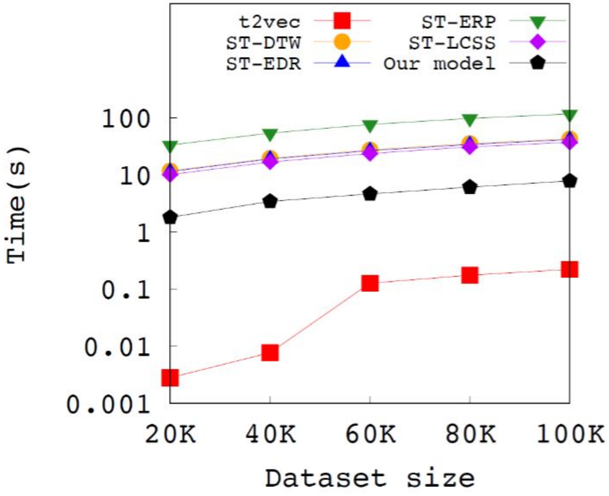



Fig. 9. Efficiency Evaluation (Porto)

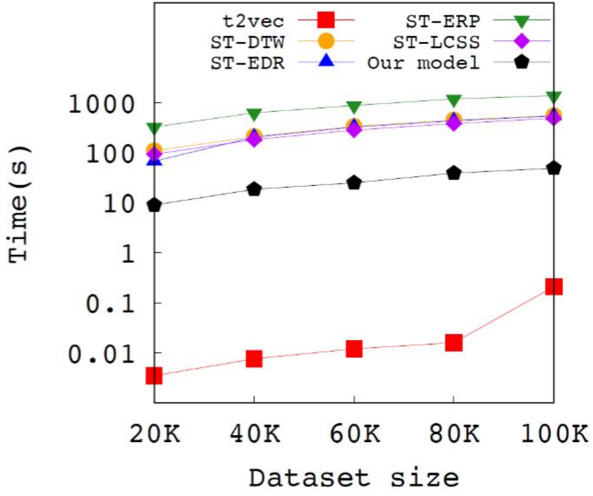
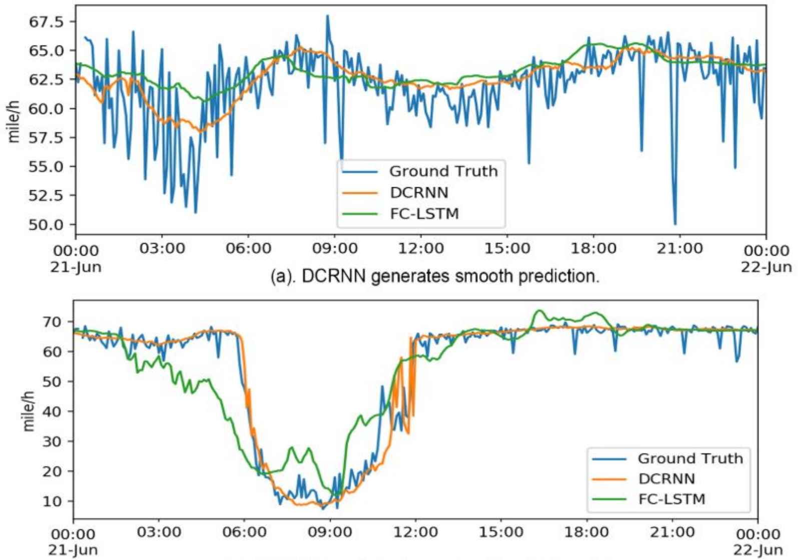



Fig. 10. Efficiency Evaluation (Chengdu)

- I used gnuplot to draw these graphs
  - http://www.gnuplot.info/
- Source code available at: https://github.com/david-tedjopurnomo/dtedjopurnom_graphs/tree/main/lineplot

**See samples at LineChart/2.3**

# Line Chart




(a). DCRNN generates smooth prediction.

(b). DCRNN predicts the start and end of peak hours.

- Shows desired property of a deep neural network prediction

# Line Chart

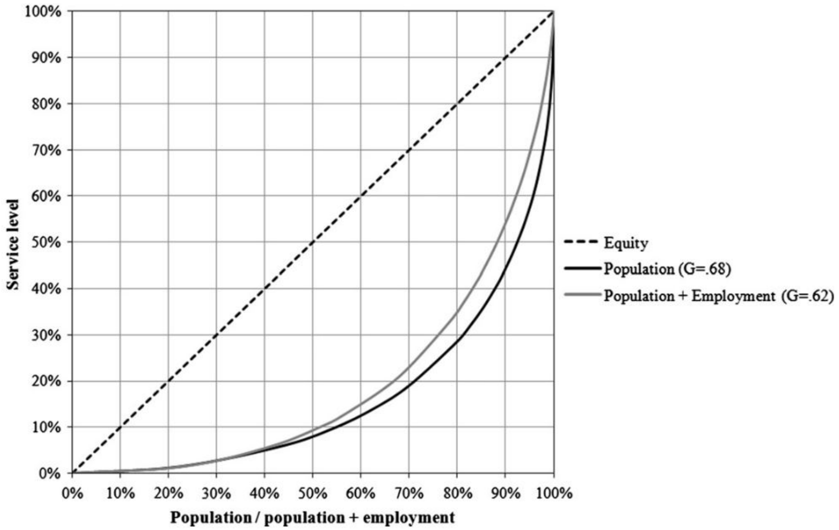
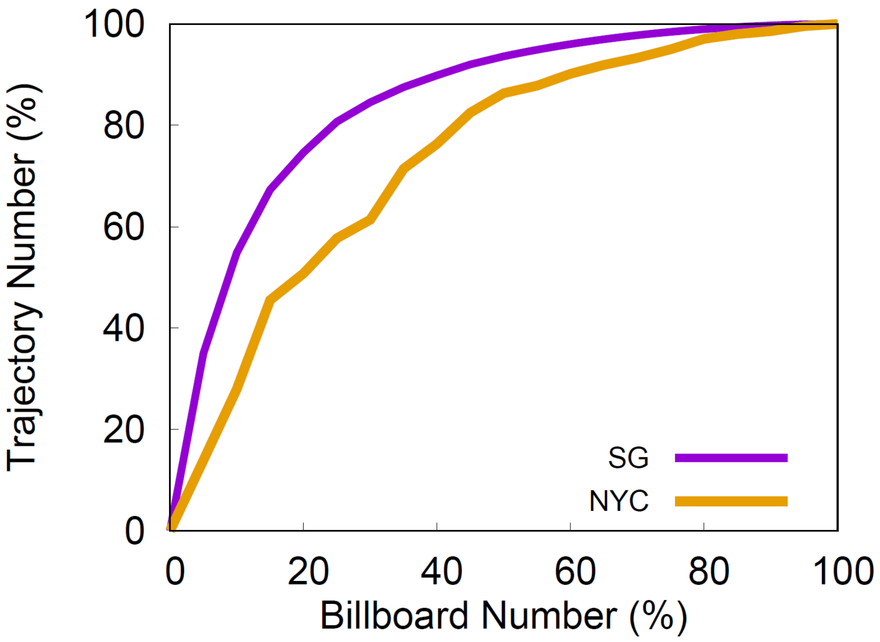
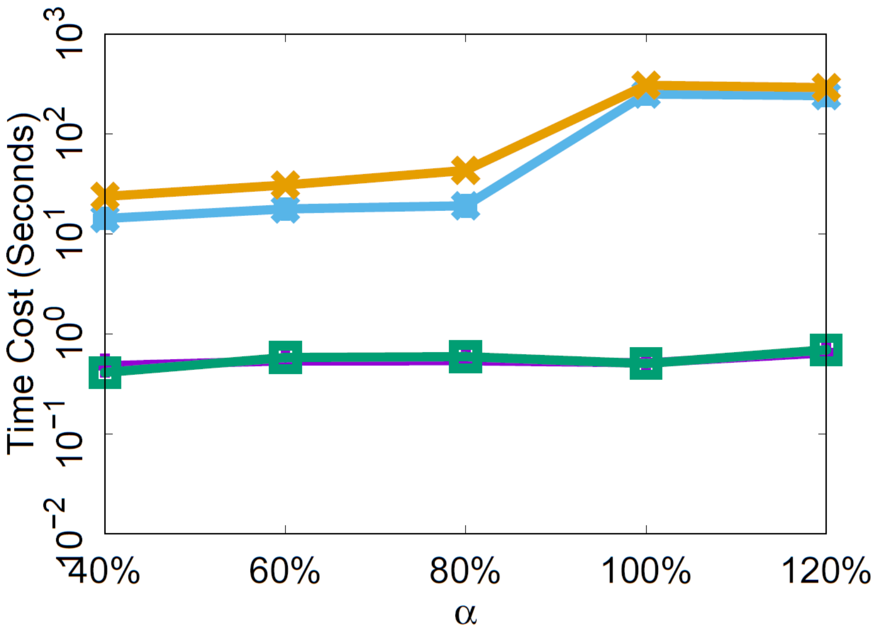



**Fig. 4.** Lorenz curves of population and employment.

- Measures inequality
- Diagonal line = ideal
- Area between straight and curved line = Gini coefficient

# Line Chart




**See samples at LineChart/2.4**

# Line Chart

✓

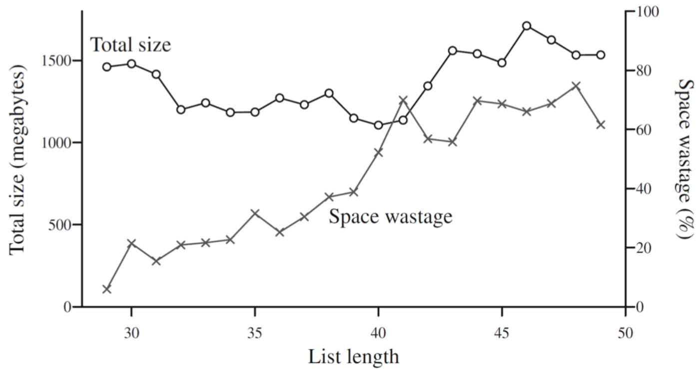
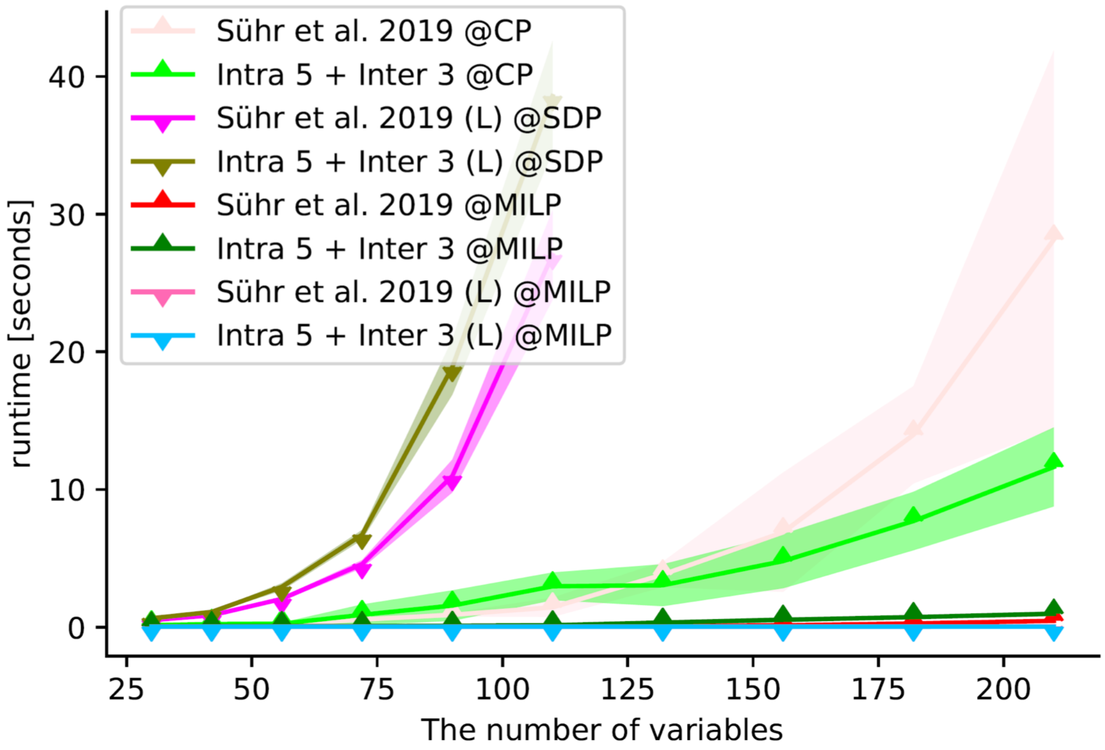



FIGURE 2. *Size and space wastage as a function of average list length.*

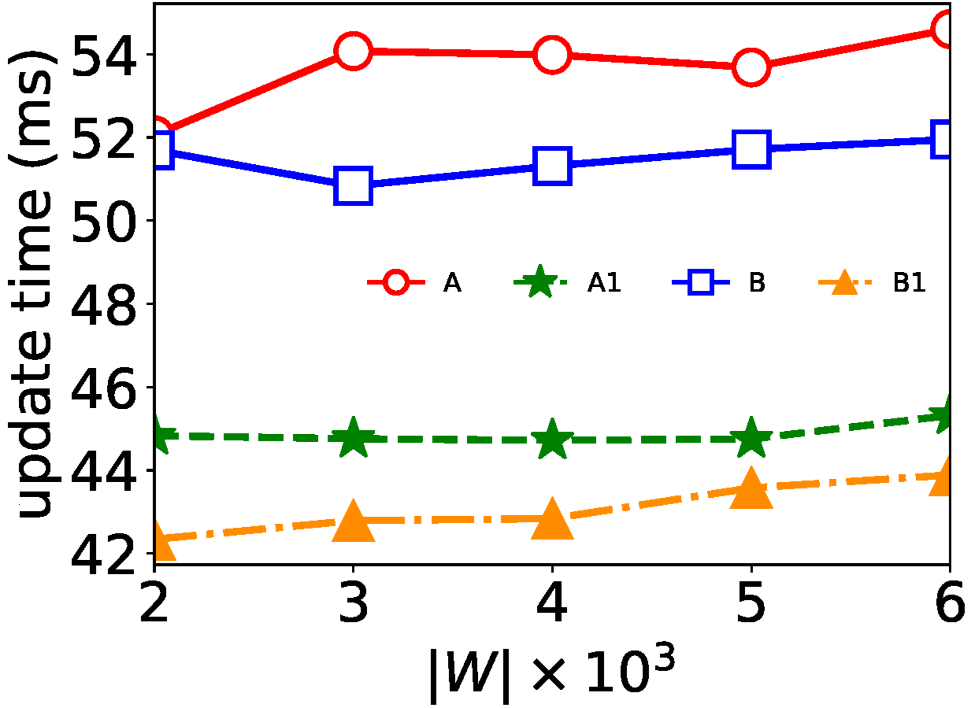
Zobel, Justin. *Writing for computer science*. Vol. 8. New York NY: Springer, 2004

# Line Chart



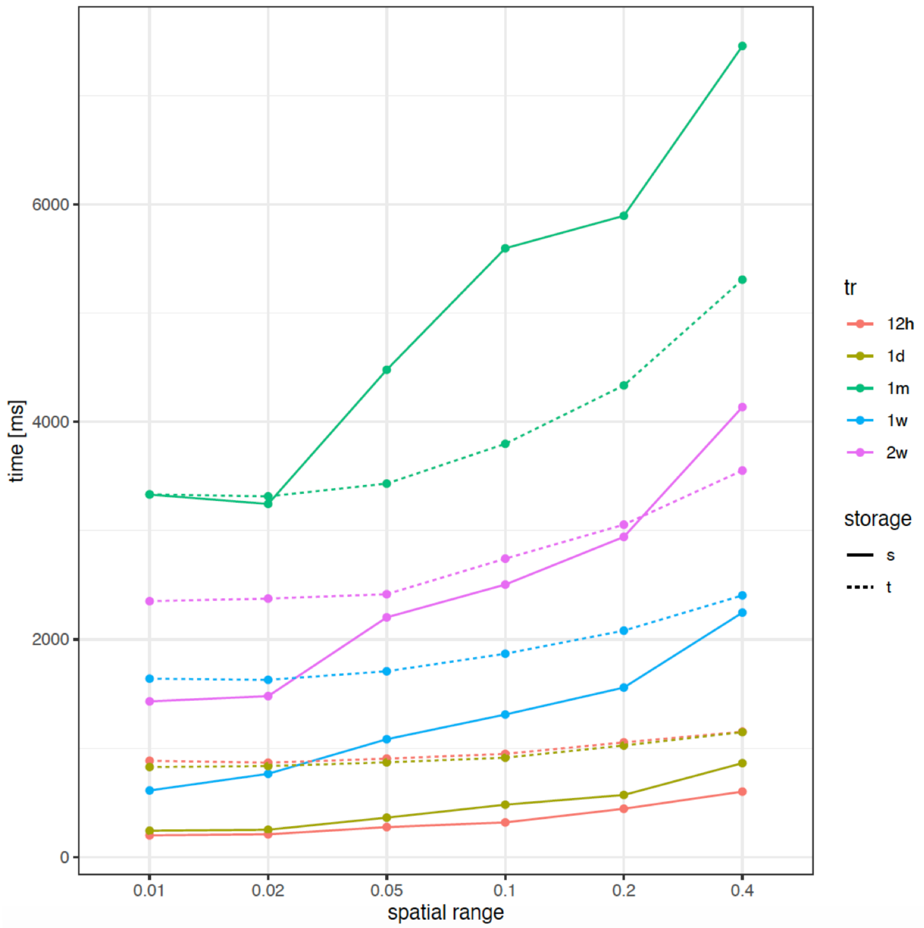

**See samples at LineChart/2.5**

# Line Chart




See samples at LineChart/2.6

# Line Chart




**See samples at LineChart/2.7**

# Line Chart

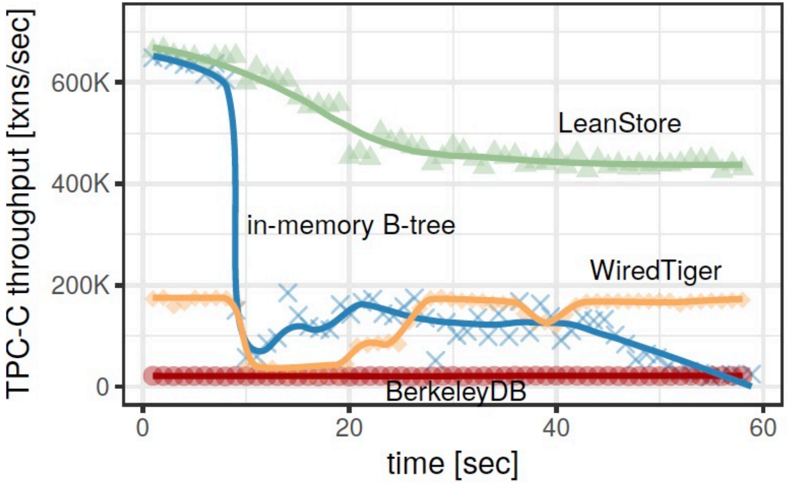



Fig. 9. TPC-C with 20 GB buffer pool (100 warehouses, 20 threads). The data grows from 10 GB to 50 GB—exceeding the buffer pool.

Viktor Leis, Michael Haubenschild, Alfons Kemper, Thomas Neumann: LeanStore: In-Memory Data Management beyond Main Memory. ICDE 2018: 185-196

# Line Chart

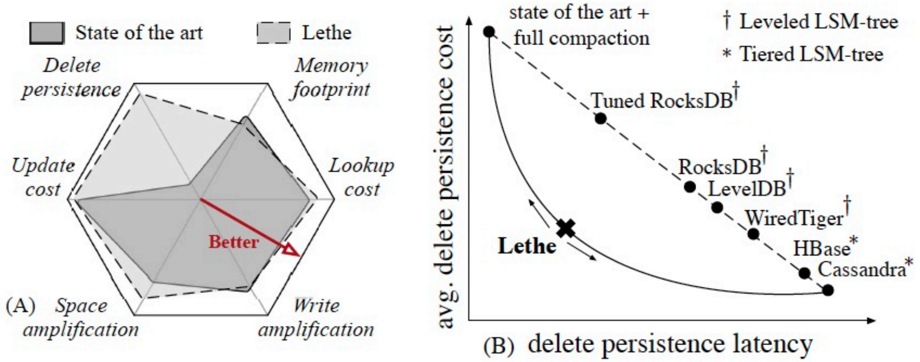
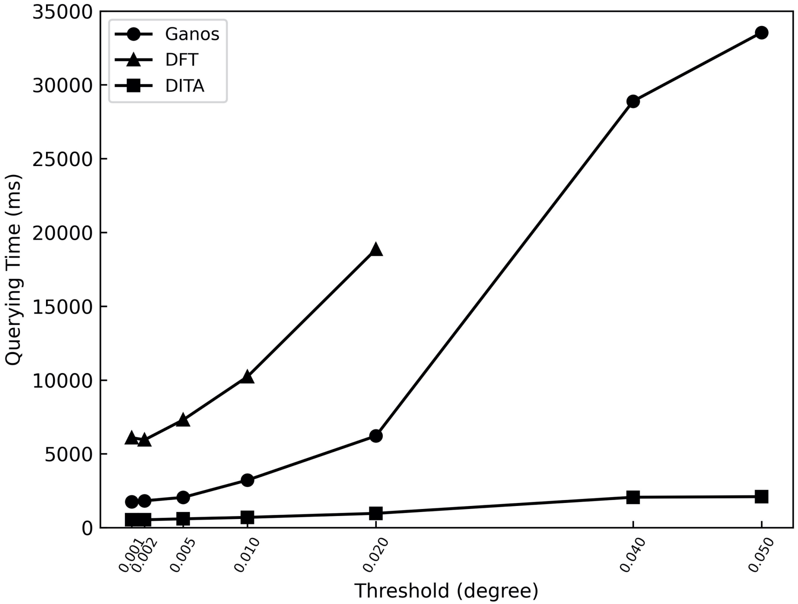



**Figure 1: (A) *Lethe* strikes an optimal balance between the latency/performance for timely delete persistence in LSM-trees, and (B) supports timely delete persistence by navigating the latency/cost trade-off.**

Subhadeep Sarkar, Tarikul Islam Papon, Dimitris Staratzis, Manos Athanassoulis: Lethe: A Tunable Delete-Aware LSM Engine. SIGMOD Conference 2020: 893-908

# Line Chart




See samples at LineChart/2.8

[1]

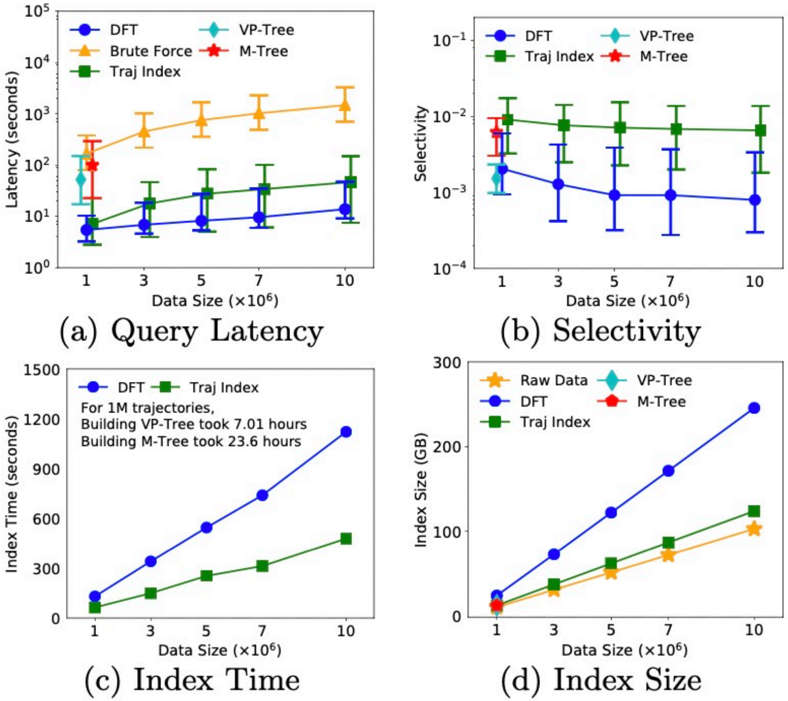
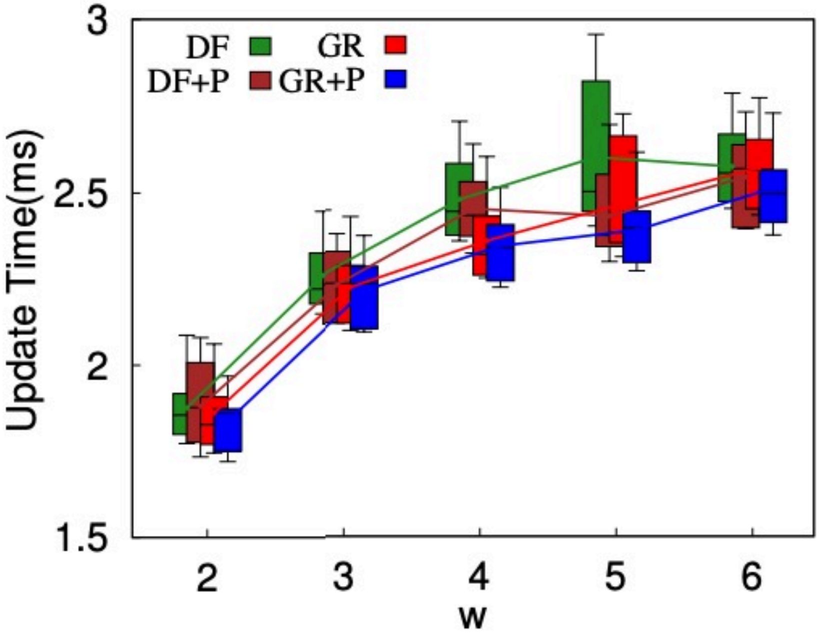



Figure 9: Scalability with respect to data size.

[1] Dong Xie, Feifei Li, and Jeff M. Phillips. 2017. Distributed Trajectory Similarity Search. VLDB 10, 11 (2017), 1478–1489.

# 3. Box Plots

# Box Plot




(c) Update time (The update time of $DSA$ is over 10s, we omit it in the plot for better visualization)

Hui Luo, Zhifeng Bao, Farhana Choudhury, and Shane Culpepper. Dynamic ridesharing in peak travel periods. TKDE (2019).

**See samples at BoxPlot/3.1**

# Box Plot

✓

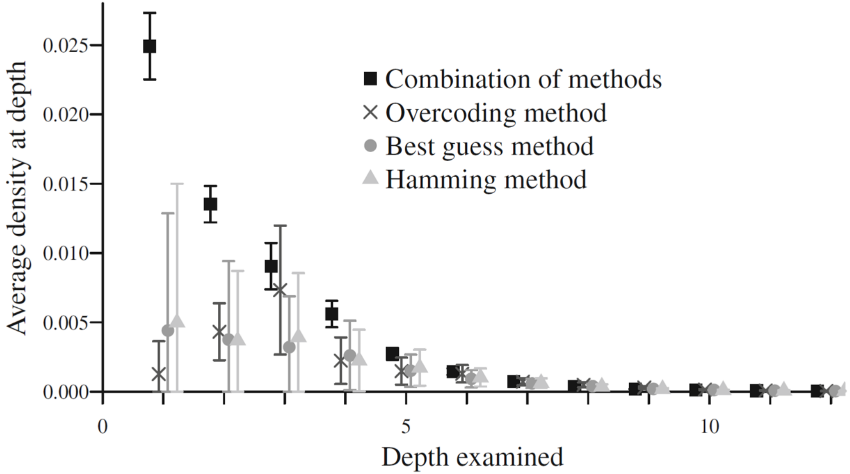
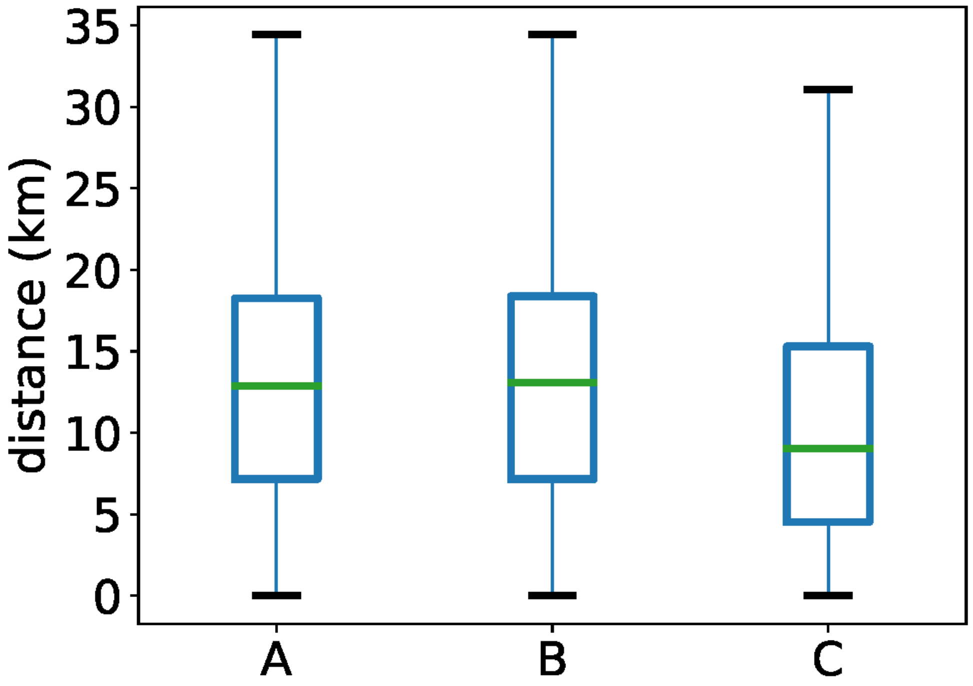



FIGURE 9. *Range of scores with each method, at each depth. The principal mark in each range is the average score. As can be seen, each method returns results within a reasonably narrow band, but they are surprisingly different from each other. Combination is highly effective in this case.*

Zobel, Justin. *Writing for computer science*. Vol. 8. New York NY: Springer, 2004.

# Box Plot




**See samples at BoxPlot/3.2**

# Box Plot

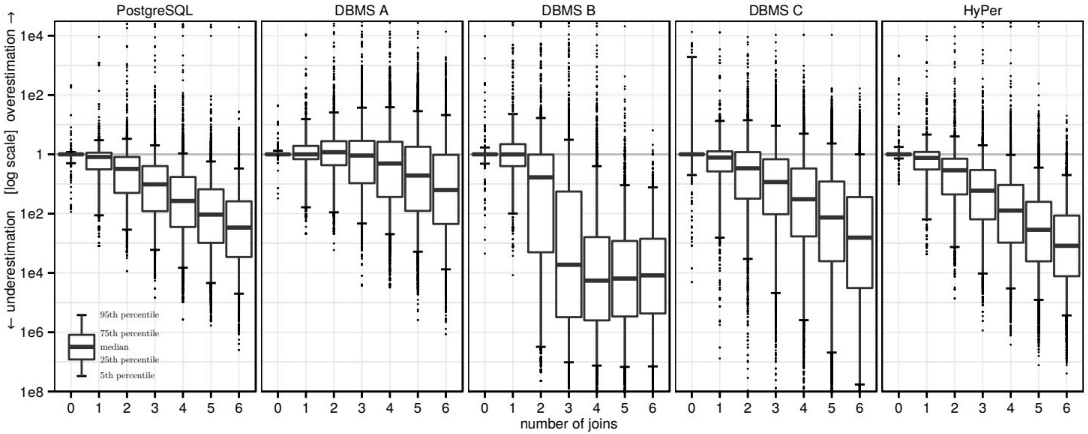



**Figure 3: Quality of cardinality estimates for multi-join queries in comparison with the true cardinalities. Each boxplot summarizes the error distribution of all subexpressions with a particular size (over all queries in the workload)**

Viktor Leis, Andrey Gubichev, Atanas Mirchev, Peter A. Boncz, Alfons Kemper, Thomas Neumann: How Good Are Query Optimizers, Really? Proc. VLDB Endow. 9(3): 204-215 (2015)

# Box Plot

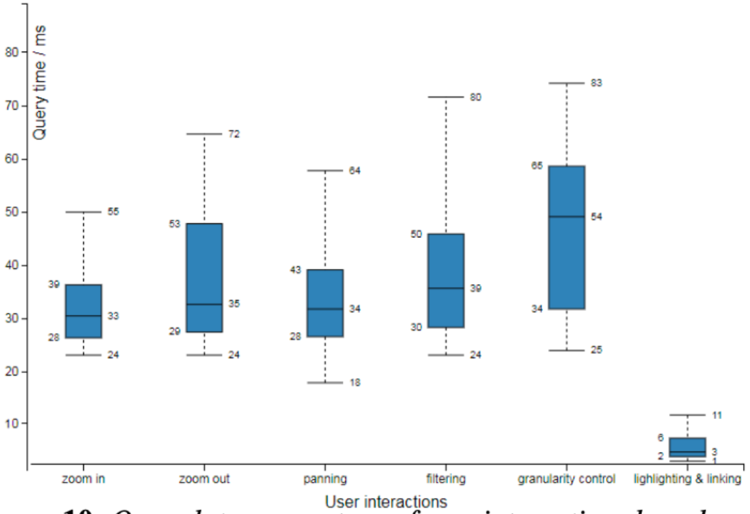
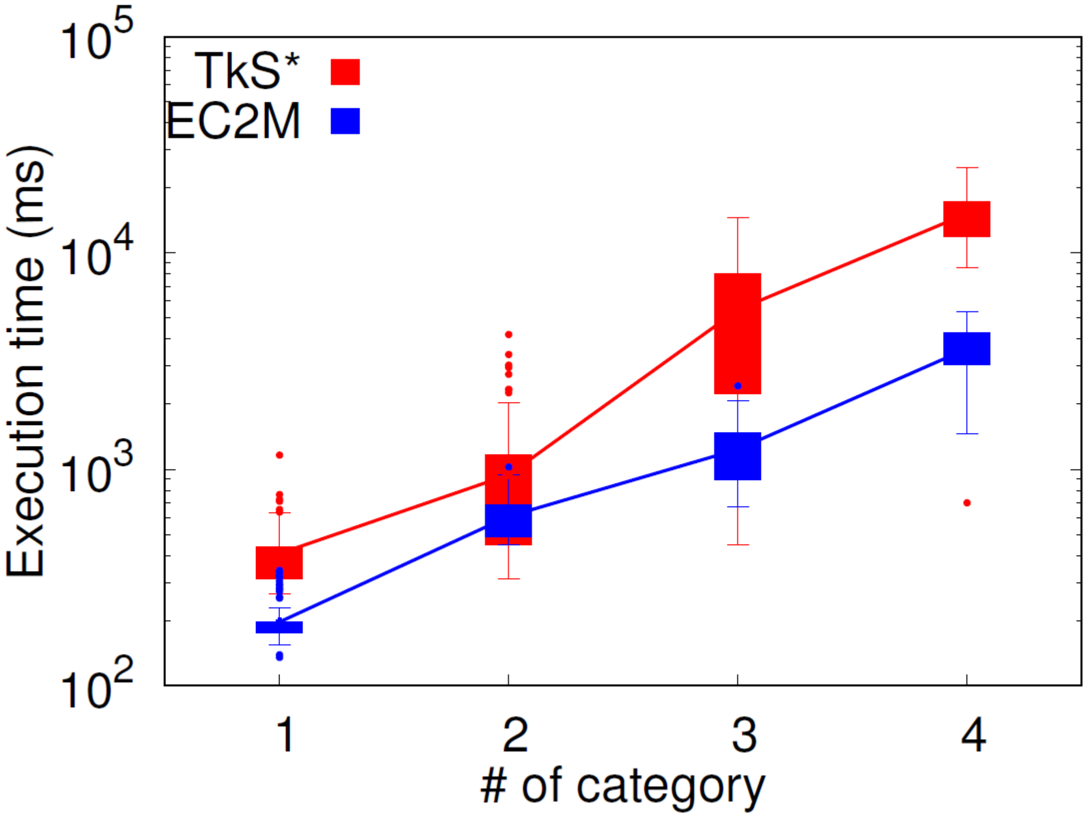



**Figure 10:** *Query latency per type of user interactions based on the real estate dataset.*

] Li, M., Choudhury, F., Bao, Z., et al. ConcaveCubes: Supporting Cluster-based Geographical Visualization in Large Data Scale. Computer Graphics Forum, 37(3), 217–228, 2018.

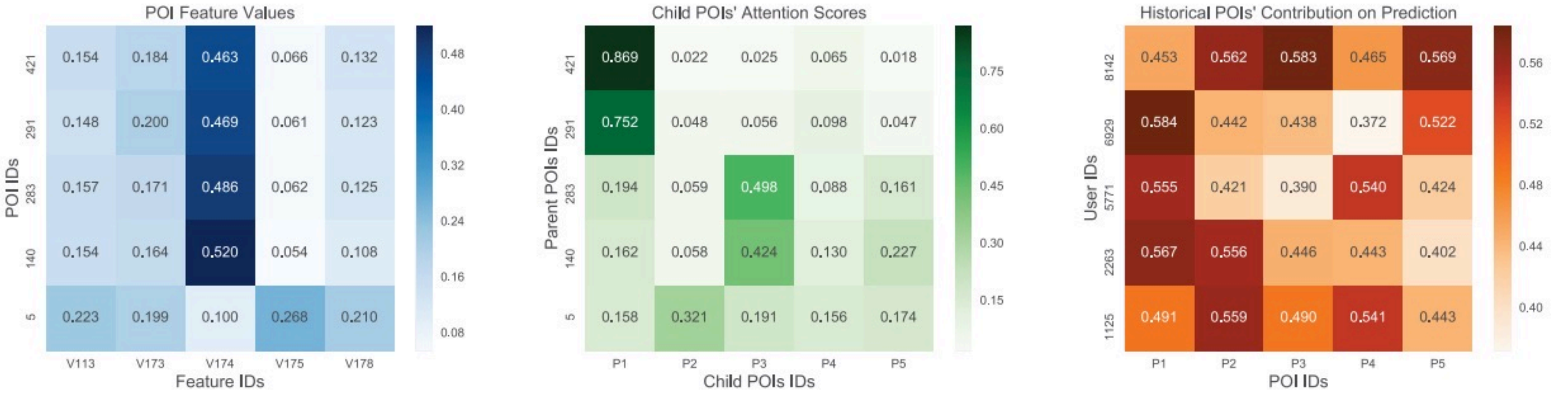
# Box Plot




**See samples at BoxPlot/3.3**

# 4. Heat Maps

# Heat Map



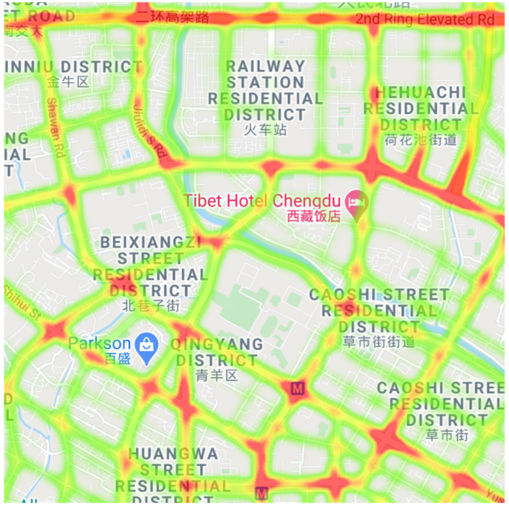

(a) User-apsect hint (b) POI-aspect hint (c) Interaction-aspect hint

**Figure 4: Visualization heat maps of three recommendation hints on the *Beijing* dataset. The larger a value is, the darker color its corresponding cell has.**

See samples at LineChart/4.1

Hui Luo, Jingbo Zhou, Zhifeng Bao, Shuangli Li, J. Shane Culpepper, Haochao Ying, Hao Liu, and Hui Xiong. Spatial object recommendation with hints: When spatial granularity matters. In *SIGIR*, pp. 781-790. 2020.

# Heat Map Geographic Maps


Draw with Google Maps API. Free for 90 days.

# Heat Map

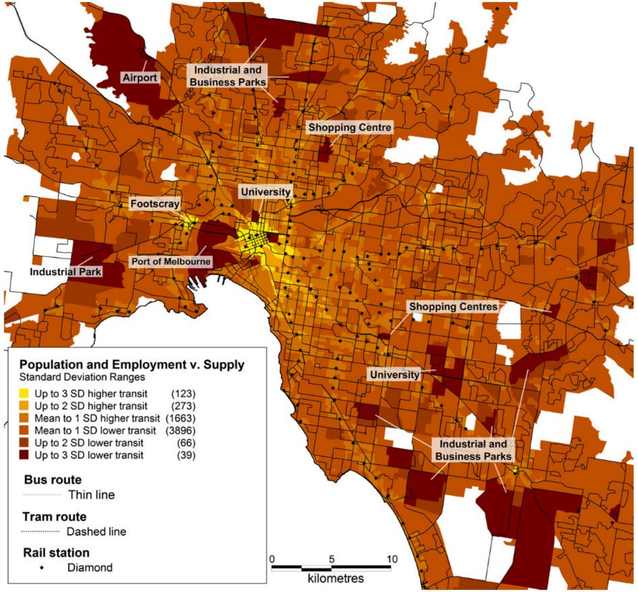



**Fig. 6.** Close-up of supply gaps in Melbourne.

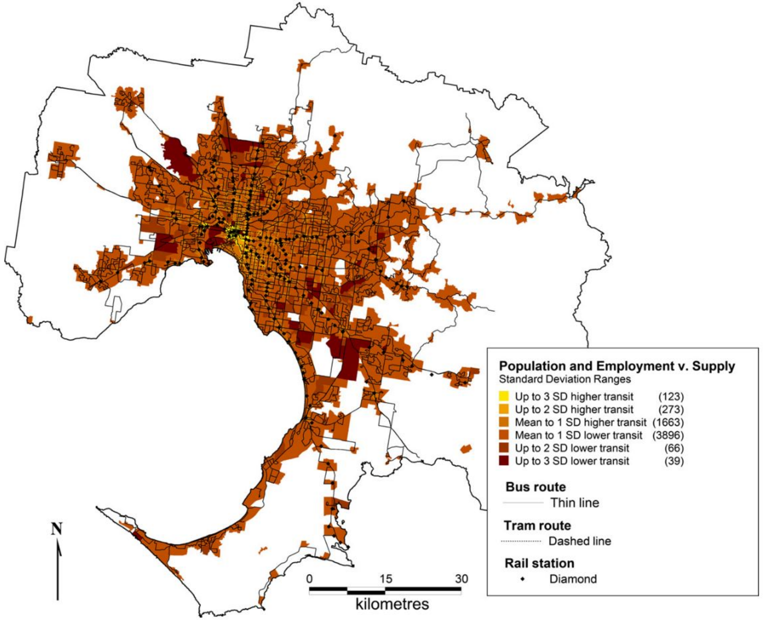



**Fig. 5.** Difference between population + employment and supply.

# Heat Map

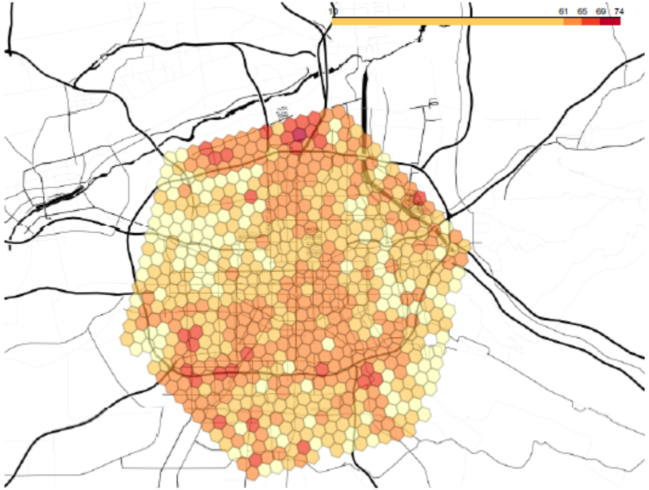



Figure 3: A visualization of CVNet output on a single layer of the hexagon grid system.

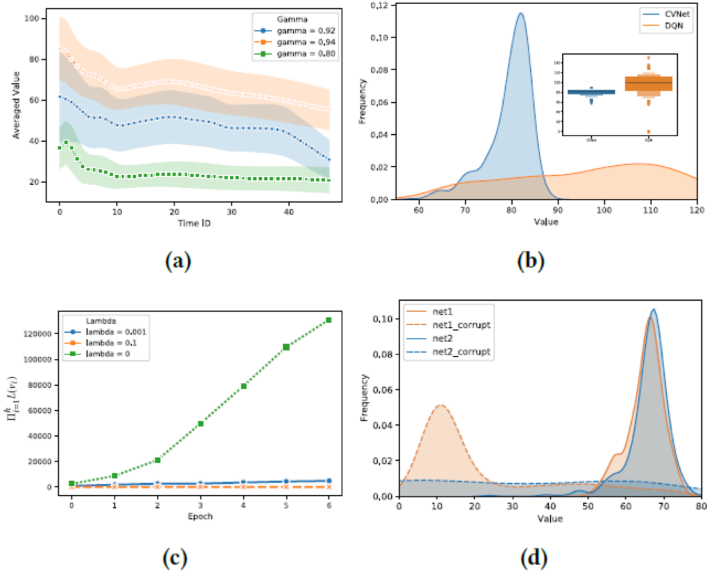



Figure 4: (a). Temporal patterns in the learned value network and how it reacts against the time discount factor $\gamma$; (b). Comparison of value distributions at a given time between DQN [17] and CVNet; (c). The change of global Lipschitz during training under different regularization $\lambda$; (d). Comparison of robustness of CVNet w/o Lipschitz Regularization (net1 is trained with $\lambda = 0.1$ and net2 is trained with $\lambda = 0$).

Xiaocheng Tang, Zhiwei (Tony) Qin, Fan Zhang, Zhaodong Wang, Zhe Xu, Yintai Ma, Hongtu Zhu, Jieping Ye. A Deep Value-network Based Approach for Multi-Driver Order Dispatching. In KDD 2019: 1780-1790

# Heat Map

Table 2: Comparison of lookup times (nanoseconds per lookup) with the SOSD benchmark. The red box indicates the base model (IM) and the enhanced versions.

| | | Algorithmic indexes | | | | On-the-fly search | | | | Learned indexes | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dataset | ART | FAST | RBS | B+tree | BS | TIP | IS | IM | IM + Shift-Table | RMI | RS | RS + Shift-Table |
| Synthetic | logn32 | N/A | 230 | 385 | 375 | 624 | 551 | N/A | 1384 | 166 | **73.9** | 83.9 | 143.5 |
| | norm32 | 173 | 197 | 267 | 390 | 655 | 671 | N/A | 1479 | 88.2 | **51.5** | 60.3 | 96.4 |
| | uden32 | 99.4 | 196 | 235 | 389 | 654 | 126 | **32.3** | 38.6 | 67.5 | 38.1 | 47.8 | 72.3 |
| | uspr32 | N/A | 198 | 230 | 390 | 654 | 298 | 321 | 425 | **89.7** | 141 | 166 | 153.5 |
| | logn64 | 238 | N/A | 622 | 427 | 674 | 377 | N/A | 1075 | 376 | 132 | **109** | 151.0 |
| | norm64 | 214 | N/A | 317 | 427 | 672 | 705 | N/A | 1615 | 88.6 | **51.7** | 61.8 | 93.2 |
| | uden64 | 104 | N/A | 255 | 428 | 670 | 142 | **34.8** | 40.4 | 67.4 | 39.8 | 47.9 | 71.8 |
| | uspr64 | 216 | N/A | 244 | 427 | 673 | 329 | 338 | **472** | **92.8** | 145 | 182 | 154.6 |
| Real-world | amzn32 | N/A | 208 | 243 | 393 | 658 | 569 | 3228 | 1524 | **99.5** | 185 | 236 | 110.8 |
| | face32 | 179 | 203 | 238 | 388 | 654 | 717 | 792 | 861 | **103** | 213 | 310 | 142.8 |
| | amzn64 | N/A | N/A | 284 | 428 | 676 | 578 | 3510 | 1575 | **105** | 189 | 238 | 119.3 |
| | face64 | 290 | N/A | 257 | 427 | 671 | 925 | 1257 | 918 | **149** | 247 | 344 | 204.1 |
| | osmc64 | N/A | N/A | 410 | 428 | 675 | 4617 | N/A | 1462 | 194 | 297 | 339 | **177.2** |
| | wiki64 | N/A | N/A | 271 | 437 | 686 | 767 | 5867 | 1687 | **94.2** | 172 | 191 | 124.1 |

1. Ali Hadian, Thomas Heinis: Shift-Table: A Low-latency Learned Index for Range Queries using Model Correction. CoRR abs/2101.10457 (2021) /EDBT 2021 accepted paper

# 5. Tables

# Table

See samples at Table/5.1

**Table 2:** Model performance comparisons on the *Beijing* and *Chengdu* dataset. Entries marked $^{\triangle}$ and $^{\blacktriangle}$ correspond to statistical significance using a paired t-test with Bonferroni correction at 95% and 99.9% confidence intervals respectively. Comparisons are relative to PACE.

| Level | Model | *Beijing* | | | | | | *Chengdu* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P@5 | NDCG@5 | P@10 | NDCG@10 | P@20 | NDCG@20 | P@5 | NDCG@5 | P@10 | NDCG@10 | P@20 | NDCG@20 |
| $H_1$ | WRMF | 0.056$^{\blacktriangledown}$ | 0.096$^{\blacktriangledown}$ | 0.047$^{\blacktriangledown}$ | 0.121$^{\blacktriangledown}$ | 0.037$^{\blacktriangledown}$ | 0.151$^{\blacktriangledown}$ | 0.063$^{\blacktriangledown}$ | 0.079$^{\blacktriangledown}$ | 0.051$^{\blacktriangledown}$ | 0.098$^{\blacktriangledown}$ | 0.041$^{\blacktriangledown}$ | 0.127$^{\blacktriangledown}$ |
| | BPRMF | 0.079$^{\blacktriangle}$ | 0.123$^{\blacktriangle}$ | 0.064$^{\blacktriangle}$ | 0.150$^{\blacktriangle}$ | 0.050$^{\blacktriangle}$ | 0.187$^{\blacktriangle}$ | 0.110$^{\blacktriangle}$ | 0.142$^{\blacktriangle}$ | 0.086$^{\blacktriangle}$ | 0.170$^{\blacktriangle}$ | 0.061$^{\blacktriangle}$ | 0.202$^{\blacktriangle}$ |
| | PACE | 0.067 | 0.104 | 0.053 | 0.124 | 0.043 | 0.156 | 0.087 | 0.117 | 0.074 | 0.152 | 0.054 | 0.181 |
| | SAE-NAD | 0.078$^{\blacktriangle}$ | 0.125$^{\blacktriangle}$ | 0.064$^{\blacktriangle}$ | 0.155$^{\blacktriangle}$ | 0.051$^{\blacktriangle}$ | 0.194$^{\blacktriangle}$ | 0.100$^{\blacktriangle}$ | 0.128$^{\blacktriangle}$ | 0.081$^{\blacktriangle}$ | 0.155$^{\blacktriangle}$ | 0.057$^{\blacktriangle}$ | 0.185$^{\blacktriangle}$ |
| | MPR | **0.084**$^{\blacktriangle}$ | **0.133**$^{\blacktriangle}$ | **0.067**$^{\blacktriangle}$ | **0.162**$^{\blacktriangle}$ | **0.053**$^{\blacktriangle}$ | **0.203**$^{\blacktriangle}$ | **0.119**$^{\triangle}$ | **0.159**$^{\triangle}$ | **0.094**$^{\blacktriangle}$ | **0.190**$^{\blacktriangle}$ | **0.064**$^{\blacktriangle}$ | **0.222**$^{\blacktriangle}$ |
| $H_2$ | WRMF | 0.009 | 0.017 | 0.007 | 0.022 | 0.005 | 0.026$^{\blacktriangle}$ | 0.022 | 0.027 | 0.018$^{\triangledown}$ | 0.034 | 0.013 | 0.040 |
| | BPRMF | 0.007 | 0.014$^{\blacktriangle}$ | 0.007 | 0.020$^{\blacktriangle}$ | 0.005 | 0.026$^{\blacktriangle}$ | 0.027 | 0.037 | 0.022 | 0.047 | 0.017 | 0.058 |
| | PACE | 0.007 | 0.013 | 0.007 | 0.019 | 0.005 | 0.024 | 0.022 | 0.031 | 0.022 | 0.039 | 0.013 | 0.046 |
| | SAE-NAD | 0.007 | 0.014$^{\blacktriangle}$ | 0.006$^{\blacktriangledown}$ | 0.018$^{\blacktriangledown}$ | 0.005 | 0.024 | **0.033** | 0.043 | 0.019 | 0.049 | 0.017 | 0.059 |
| | MPR | **0.010**$^{\blacktriangle}$ | **0.018**$^{\blacktriangle}$ | **0.008**$^{\blacktriangle}$ | **0.023**$^{\blacktriangle}$ | **0.007**$^{\blacktriangle}$ | **0.030**$^{\blacktriangle}$ | **0.033**$^{\blacktriangle}$ | **0.044**$^{\blacktriangle}$ | **0.026**$^{\triangle}$ | **0.054**$^{\blacktriangle}$ | **0.020**$^{\blacktriangle}$ | **0.067**$^{\blacktriangle}$ |
| $H_3$ | WRMF | 0.008$^{\blacktriangle}$ | 0.015$^{\blacktriangle}$ | 0.006$^{\blacktriangle}$ | 0.018$^{\blacktriangle}$ | 0.004 | 0.022$^{\blacktriangle}$ | 0.021$^{\triangle}$ | 0.027$^{\triangle}$ | 0.017 | 0.033 | 0.013 | 0.041 |
| | BPRMF | 0.006$^{\blacktriangledown}$ | 0.012$^{\blacktriangle}$ | 0.005 | 0.015$^{\blacktriangle}$ | 0.004 | 0.019$^{\blacktriangle}$ | 0.021$^{\triangle}$ | 0.029 | 0.017 | 0.036 | 0.013$^{\triangle}$ | 0.043$^{\blacktriangle}$ |
| | PACE | 0.007 | 0.008 | 0.005 | 0.009 | 0.004 | 0.010 | 0.016 | 0.023 | 0.016 | 0.032 | 0.009 | 0.035 |
| | SAE-NAD | 0.008$^{\blacktriangle}$ | 0.015$^{\blacktriangle}$ | 0.007$^{\blacktriangle}$ | 0.020$^{\blacktriangle}$ | 0.005$^{\blacktriangle}$ | **0.026**$^{\blacktriangle}$ | 0.020$^{\blacktriangle}$ | 0.027$^{\blacktriangle}$ | 0.020$^{\triangle}$ | 0.038$^{\blacktriangle}$ | 0.016$^{\blacktriangle}$ | 0.047$^{\blacktriangle}$ |
| | MPR | **0.009**$^{\blacktriangle}$ | **0.015**$^{\blacktriangle}$ | **0.007**$^{\blacktriangle}$ | **0.021**$^{\blacktriangle}$ | **0.006**$^{\blacktriangle}$ | **0.026**$^{\blacktriangle}$ | **0.032**$^{\blacktriangle}$ | **0.042**$^{\blacktriangle}$ | **0.021**$^{\blacktriangle}$ | **0.046**$^{\blacktriangle}$ | **0.016**$^{\blacktriangle}$ | **0.056**$^{\blacktriangle}$ |

Hui Luo, Jingbo Zhou, Zhifeng Bao, Shuangli Li, J. Shane Culpepper, Haochao Ying, Hao Liu, and Hui Xiong. Spatial object recommendation with hints: When spatial granularity matters. In *SIGIR*, pp. 781-790. 2020.

# Table

| Dataset | Statistics | Query Motif $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_5$ |
|---|---|---|---|---|---|---|
| Epinion | Time (s) | 4.5E+02 | 3.3E+02 | 9.9E+03 | 2.6E+04 | 4.1E+04 |
| | Count | 7.8E+07 | 1.7E+07 | 1.1E+09 | 1.7E+09 | 5.7E+09 |
| Gowalla | Time (s) | 5.0E+02 | 3.0E+02 | 1.6E+04 | 3.2E+04 | 7.2E+04 |
| | Count | 8.6E+07 | 1.5E+07 | 1.1E+09 | 5.4E+09 | 7.3E+09 |
| Flixster | Time (s) | 1.6E+03 | 1.8E+03 | 5.5E+04 | 1.1E+05 | 1.8E+05 |
| | Count | 2.3E+08 | 9.6E+07 | 2.5E+09 | 5.2E+09 | 8.6E+09 |
| Digg | Time (s) | 1.2E+04 | 2.3E+04 | \ | \ | \ |
| | Count | 2.3E+09 | 1.7E+09 | \ | \ | \ |
| Dogster | Time (s) | 1.1E+05 | 2.9E+05 | \ | \ | \ |
| | Count | 3.5E+10 | 1.2E+10 | \ | \ | \ |
| Catster | Time (s) | 5.4E+05 | \ | \ | \ | \ |
| | Count | 2.0E+11 | \ | \ | \ | \ |
| Orkut | Time (s) | 2.8E+05 | \ | \ | \ | \ |
| | Count | 4.8E+10 | \ | \ | \ | \ |

| Dataset | Statistics | Query Motif $Q_0$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_5$ | $Q_6$ |
|---|---|---|---|---|---|---|---|---|
| Epinion | Time(s) | 8.0E+00 | 4.5E+02 | 3.3E+02 | 9.9E+03 | 2.6E+04 | 4.1E+04 | 6.3E+06 |
| | Count | 1.6E+06 | 7.8E+07 | 1.7E+07 | 1.1E+09 | 1.7E+09 | 5.7E+09 | 2.8E+11 |
| Gowalla | Time(s) | 1.2E+01 | 5.0E+02 | 3.0E+02 | 1.6E+04 | 3.2E+04 | 7.2E+04 | 3.9E+06 |
| | Count | 2.3E+06 | 8.6E+07 | 1.5E+07 | 1.1E+09 | 5.4E+09 | 7.3E+09 | 2.4E+11 |
| Flixster | Time(s) | 6.0E+01 | 1.6E+03 | 1.8E+03 | 5.5E+04 | 1.1E+05 | 1.8E+05 | \ |
| | Count | 7.9E+06 | 2.3E+08 | 9.6E+07 | 2.5E+09 | 5.2E+09 | 8.6E+09 | \ |
| Digg | Time(s) | 5.3E+01 | 1.2E+04 | 2.3E+04 | \ | \ | \ | \ |
| | Count | 1.4E+07 | 2.3E+09 | 1.7E+09 | \ | \ | \ | \ |
| Dogster | Time(s) | 2.7E+02 | 1.1E+05 | 2.9E+05 | \ | \ | \ | \ |
| | Count | 8.4E+07 | 3.5E+10 | 1.2E+10 | \ | \ | \ | \ |
| Catster | Time(s) | 5.2E+02 | 5.4E+05 | \ | \ | \ | \ | \ |
| | Count | 1.9E+08 | 2.0E+11 | \ | \ | \ | \ | \ |
| Orkut | Time(s) | 2.7E+03 | 2.8E+05 | \ | \ | \ | \ | \ |
| | Count | 6.3E+08 | 4.8E+10 | \ | \ | \ | \ | \ |
| Sina | Time(s) | 2.1E+04 | \ | \ | \ | \ | \ | \ |
| | Count | 4.7E+09 | \ | \ | \ | \ | \ | \ |

| Category | Method | Diffusion Model IC | LT | TR | CT | Efficiency | Scalability | Effectiveness | Bottleneck or Breakthrough |
|---|---|---|---|---|---|---|---|---|---|
| Simulation | SimuGreedy [113] | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | Bottleneck: Costly Monte Carlo Simulations |
| | CELF [127] | ✓ | ✓ | ✓ | ✓ | | | | |
| | UBLF [218] | ✓ | ✓ | ✗ | ✗ | | | | |
| | CGA [201] | ✓ | ✗ | ✗ | ✗ | | | | |
| Sketch | NewGeneric [32] | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | Bottleneck: High memory cost of storing sketches |
| | SG [41] | ✓ | ✓ | ✓ | ✓ | | | | |
| | SGDU [42] | ✓ | ✓ | ✓ | ✓ | | | | |
| | PMC [155] | ✓ | ✓ | ✓ | ✓ | | | | |
| | RIS [27] | ✓ | ✓ | ✓ | ✓ | | | | |
| | TIM [190] | ✓ | ✓ | ✓ | ✓ | | | | |
| | IMM [191] | ✓ | ✓ | ✓ | ✓ | | | | |
| | Chapter 3 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | Breakthrough: Sketch Compression |
| Heuristic | DegDis [32] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | Bottleneck: No theoretical guarantees |
| | PageRank [159] | ✓ | ✓ | ✓ | ✓ | | | | |
| | PMIA [34] | ✓ | ✗ | ✗ | ✗ | | | | |
| | SimPath [76] | ✗ | ✓ | ✗ | ✗ | | | | |
| | IRIE [105] | ✓ | ✗ | ✗ | ✗ | | | | |
| | IMRank [43] | ✓ | ✗ | ✗ | ✗ | | | | |
| | EasyIM [19] | ✓ | ✓ | ✗ | ✗ | | | | |

**See samples at Table/5.2**

# Table

| Dataset | dim | Methods Use: **A** Eigen Maps | SVD | DNGR | **D** n2v $C=2$ | n2v $C=5$ | Asym Proj | $\mathbb{E}[\mathbf{D}]$ Graph Attention *(ours)* | Error Reduction |
|---|---|---|---|---|---|---|---|---|---|
| wiki-vote | 64 | 61.3 | 86.0 | 59.8 | 64.4 | 63.6 | 91.7 | $\mathbf{93.8 \pm 0.13}$ | 25.2% |
| | 128 | 62.2 | 80.8 | 55.4 | 63.7 | 64.6 | 91.7 | $\mathbf{93.8 \pm 0.05}$ | 25.2% |
| ego-Facebook | 64 | 96.4 | 96.7 | 98.1 | 99.1 | 99.0 | 97.4 | $\mathbf{99.4 \pm 0.10}$ | 33.3% |
| | 128 | 95.4 | 94.5 | 98.4 | 99.3 | 99.2 | 97.3 | $\mathbf{99.5 \pm 0.03}$ | 28.6% |
| ca-AstroPh | 64 | 82.4 | 91.1 | 93.9 | 97.4 | 96.9 | 95.7 | $\mathbf{97.9 \pm 0.21}$ | 19.2% |
| | 128 | 82.9 | 92.4 | 96.8 | 97.7 | 97.5 | 95.7 | $\mathbf{98.1 \pm 0.49}$ | 24.0% |
| ca-HepTh | 64 | 80.2 | 79.3 | 86.8 | 90.6 | 91.8 | 90.3 | $\mathbf{93.6 \pm 0.06}$ | 22.0% |
| | 128 | 81.2 | 78.0 | 89.7 | 90.1 | 92.0 | 90.3 | $\mathbf{93.9 \pm 0.05}$ | 23.8% |
| PPI | 64 | 70.7 | 75.4 | 76.7 | 79.7 | 70.6 | 82.4 | $\mathbf{89.8 \pm 1.05}$ | 43.5% |
| | 128 | 73.7 | 71.2 | 76.9 | 81.8 | 74.4 | 83.9 | $\mathbf{91.0 \pm 0.28}$ | 44.2% |

Abu-El-Haija, S., Perozzi, B., Al-Rfou, R., & Alemi, A. (2017). Watch your step: Learning node embeddings via graph attention. NeurIPS: 9198-9208.

# Table

Table 13. Experiment results of all baselines on different choices of $\alpha$

| $\alpha$ | ST-DTW | ST-EDR | ST-ERP | ST-LCSS |
|---|---|---|---|---|
| 0.5 | 1.5696 | 23.2293 | 29.7937 | 4.0918 |
| 0.6 | 1.6254 | 22.9506 | 20.6814 | 4.1011 |
| 0.7 | 1.6830 | 23.8497 | 16.1187 | 4.1194 |
| 0.8 | 1.7300 | 26.4362 | 14.1944 | 4.1997 |
| 0.9 | 1.7947 | 32.2287 | 14.7551 | 4.6906 |
| 1 | 2.6350 | 41.5255 | 19.1225 | 6.4034 |
| Mean | 1.8396 | 28.3700 | 19.1110 | 4.6010 |

- I used Tables generator:
  - https://www.tablesgenerator.com
  - Excel-esque GUI. Very easy to use
  - Use “Booktabs table style” for better presentation

# Table

| Reference | Authors | Year | Primary Datatype | Primary Dataset | Data Time Range | Data Granularity | Secondary Dataset | Input Sequence Length | Prediction Horizon |
|---|---|---|---|---|---|---|---|---|---|
| [5] | Huang et al. | 2014 | Point | 1) Caltrans PEMS<br>2) China highways | 2011 | 15 minutes | None | 60 minutes | 15 minutes |
| [50] | Yisheng et al. | 2015 | Point | Caltrans PEMS | Jan - Mar 2013 (weekdays) | 5 minutes | None | a) 15 minutes<br>b) 15 minutes<br>c) 20 minutes<br>d) 15 minutes | a) 15 minutes<br>b) 30 minutes<br>c) 45 minutes<br>d) 60 minutes |
| [45] | Ma et al. | 2015 | Point | Beijing Ring Road | June 2013 | 2 minutes | None | 2 minutes | 2 minutes |
| [51] | Tian and Pan | 2015 | Point | Caltrans PEMS | 2014 (workdays) | 5 minutes | None | a) 180 minutes<br>b) 180 minutes<br>c) 180 minutes<br>d) 180 minutes | a) 15 minutes<br>b) 30 minutes<br>c) 45 minutes<br>d) 60 minutes |
| [46] | Jia et al. | 2016 | Point | Beijing traffic | Jun - Aug 2013 | 2 minutes | None | a) 16 minutes<br>b) 24 minutes<br>c) 50 minutes | a) 2 minutes<br>b) 10 minutes<br>c) 30 minutes |
| [52] | Soua et al. | 2016 | Point | Caltrans PEMS | 1 Aug 2013 - 25 Nov 2013 | 15 minutes | - Weather data<br>- Tweets data | Unknown | Unknown |

- More advanced table
- Select some good variables for categorization
- Gives good view of the literature for the domain
- May require landscape view for the pages

# Table

Table 4. Sensitivity of OSPMiner's effectiveness w.r.t. the *aps* threshold $\alpha$

| Dataset | $\alpha$ | Outlierness rank | | | # of outlying patterns | | | Length of outlying patterns | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Min. | Max. | Avg. | Min. | Max. | Avg. | Min. | Max. | Avg. |
| Twitter | 0.025 | 1 | 163 | 3.02 | 1 | 12 | 2.33 | 1 | 71 | 4.14 |
| | 0.05 | 1 | 239 | 5.80 | 1 | 11 | 1.72 | 1 | 37 | 4.53 |
| | 0.075 | 1 | 239 | 9.57 | 1 | 6 | 1.44 | 1 | 85 | 4.94 |
| | 0.1 | 1 | 440 | 14.19 | 1 | 5 | 1.31 | 1 | 86 | 5.08 |
| Foursquare | 0.025 | 1 | 1435 | 121.88 | 1 | 24 | 1.72 | 1 | 75 | 5.48 |
| | 0.05 | 1 | 1435 | 121.07 | 1 | 12 | 1.48 | 1 | 75 | 5.04 |
| | 0.075 | 1 | 1711 | 125.38 | 1 | 12 | 1.35 | 1 | 75 | 4.70 |
| | 0.1 | 1 | 2070 | 136.98 | 1 | 8 | 1.28 | 1 | 75 | 4.34 |

**See samples at Table/5.3**

# Table

**Table 1: Comparing Our Method With Other Distributed Systems (in C1 and C2)**

| Work | Basic Query | | | Advanced Query | | | | Scalability[4] | | Trajectory Properties | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *IDTQ*[1] | *SRQ* | *STRQ* | *Sim* | *SubSim* | *kNN* | *Join* | *Computation* | *Storage* | *NoP* | *STS* | *DoT* |
| Summit [3] | × | ✓ | ✓ | × | × | × | × | ✓ | ✓ | - | - | - |
| DFT [28] | × | × | × | F/H[2] | × | F/H | × | ✓ | ×[5] | × | × | × |
| DITA [24] | × | ✓ | × | F/D/L/E[3] | × | F/D/L/E | ✓ | ✓ | × | × | × | ✓ |
| MobilityDB [5] | × | ✓ | ✓ | × | × | × | × | ✓ | ✓ | ✓ | ✓ | × |
| TrajMesa [17] | ✓ | ✓ | ✓ | F/H | × | F/H | × | × | ✓ | × | × | × |
| Our Method | ✓ | ✓ | ✓ | F/H/D/L/E | ✓ | F/H/D/L/E | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

[1] *IDTQ* refers to id-temporal query. Other queries can be found in Section 2.2.
[2] F, H, D, L, and E refer to Fréchet, Hausdorff, DTW, LCSS, and EDR.
[3] LCSS' definition in DITA is not equivalent to the original paper [25].
[4] Scalability means scaling out here.
[5] Spark is a main-memory processing system, which needs lots of memory for the massive trajectories.

See samples at Table/5.4

# 6. 3D Plots

# 3D Plot

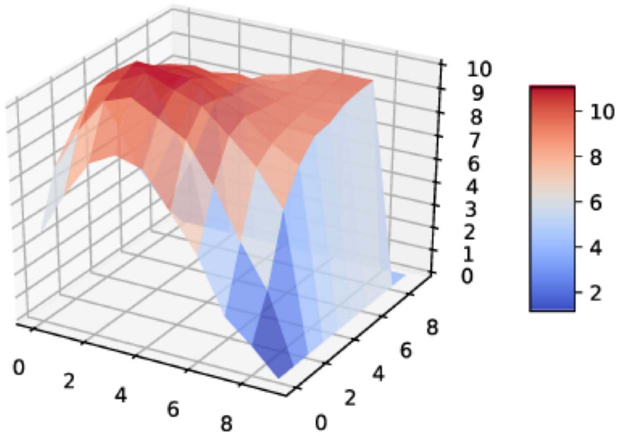
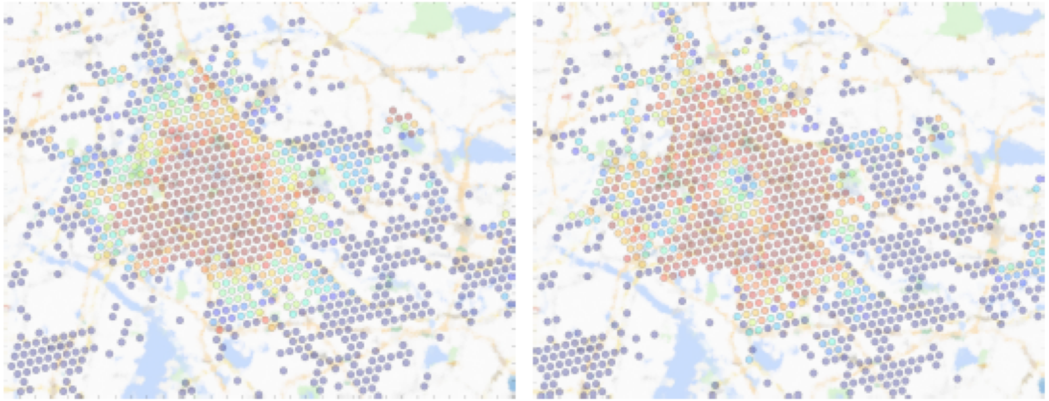



Figure 7: Value function map at time $t = 3$ generated in the toy example environment.



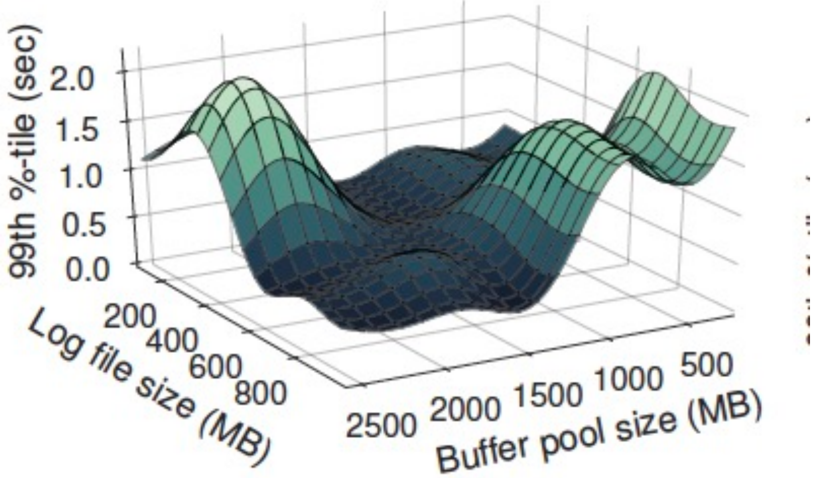
(a) 18:00-18:10, evening peak (b) 09:00-09:10, after morning peak

Figure 8: Sampled value function for the same city at different times. Red indicates higher values, blue for lower ones. Better viewed in color.

**See samples at 3DPlot/6.1**

Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, Jieping Ye. Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach. In KDD 2018: 905-913.

# 3D Plot

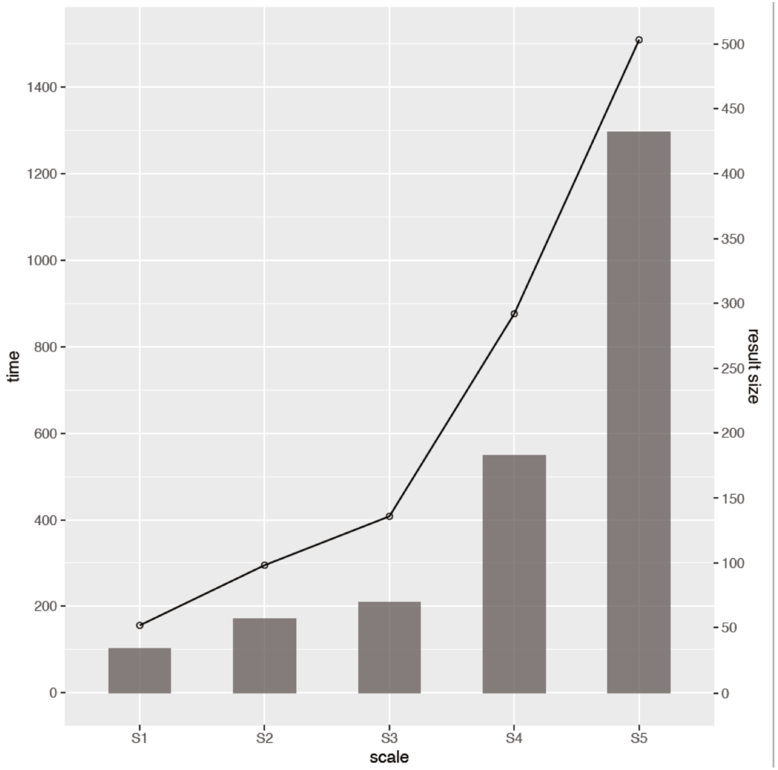



(a) Dependencies

Dana Van Aken, Andrew Pavlo, Geoffrey J. Gordon, Bohan Zhang: Automatic Database Management System Tuning Through Large-scale Machine Learning. SIGMOD Conference 2017: 1009-1024

# 7. Hybrid Graphs

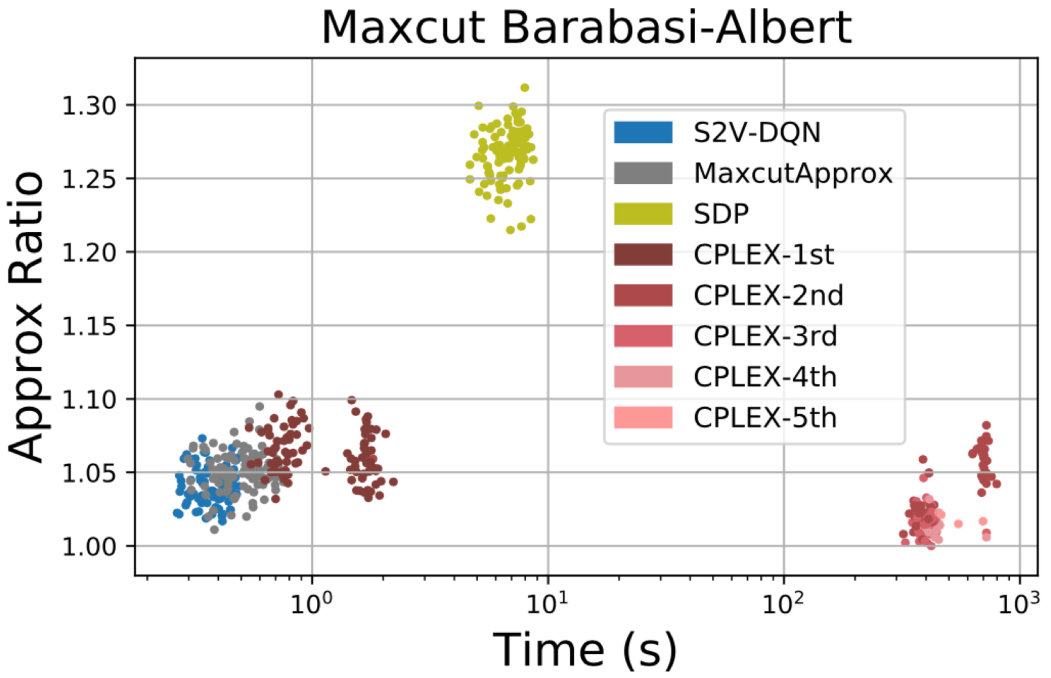
# Hybrid Graph




**See samples at HybridGraph/7.1**

# 8. Scatter Plots

# Scatter Plot




Dai, H., Khalil, E. B., Zhang, Y., Dilkina, B., & Song, L. (2017). Learning combinatorial optimization algorithms over graphs. NeurIPS: *6348-6358*

# Scatter Plot

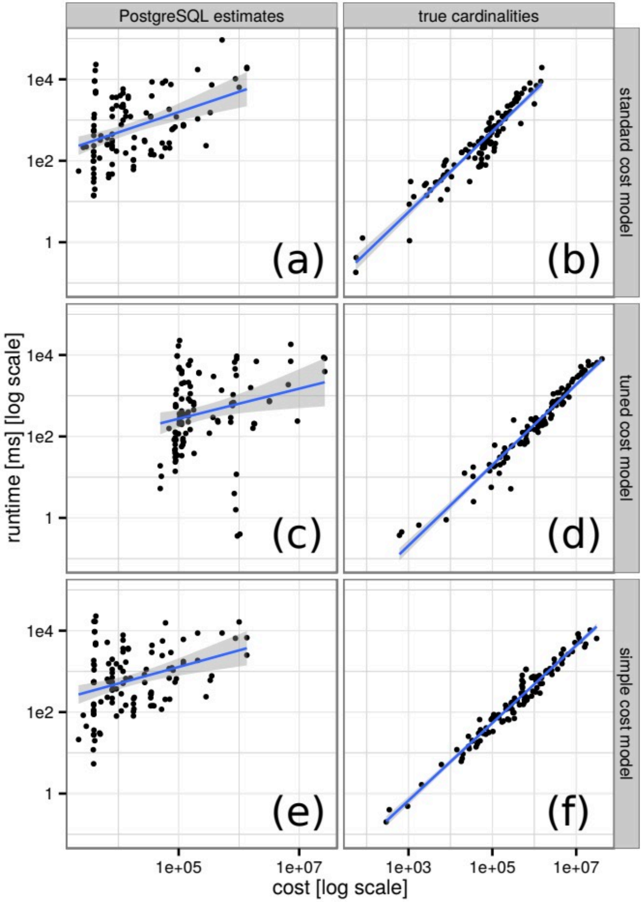
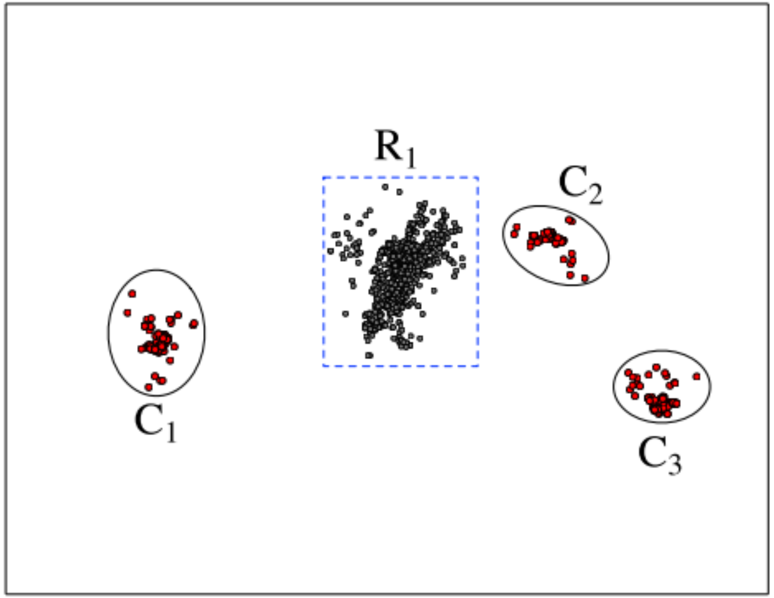



**Figure 8: Predicted cost vs. runtime for different cost models**

Viktor Leis, Andrey Gubichev, Atanas Mirchev, Peter A. Boncz, Alfons Kemper, Thomas Neumann: How Good Are Query Optimizers, Really? Proc. VLDB Endow. 9(3): 204-215 (2015)

# Scatter Plot




(b) Split with $h = 0.02$.

# 9. Others

# Others

[1]

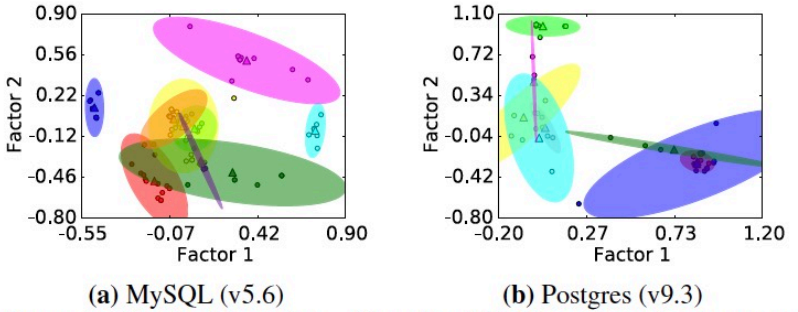



**Figure 4: Metric Clustering** – Grouping DBMS metrics using $k$-means based on how similar they are to each other as identified by Factor Analysis and plotted by their (f1, f2) coordinates. The color of each metric shows its cluster membership. The triangles represent the cluster centers.

[2]

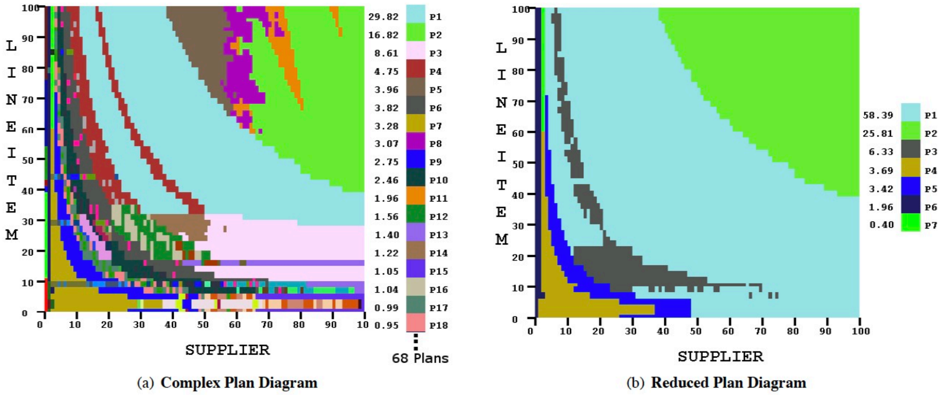
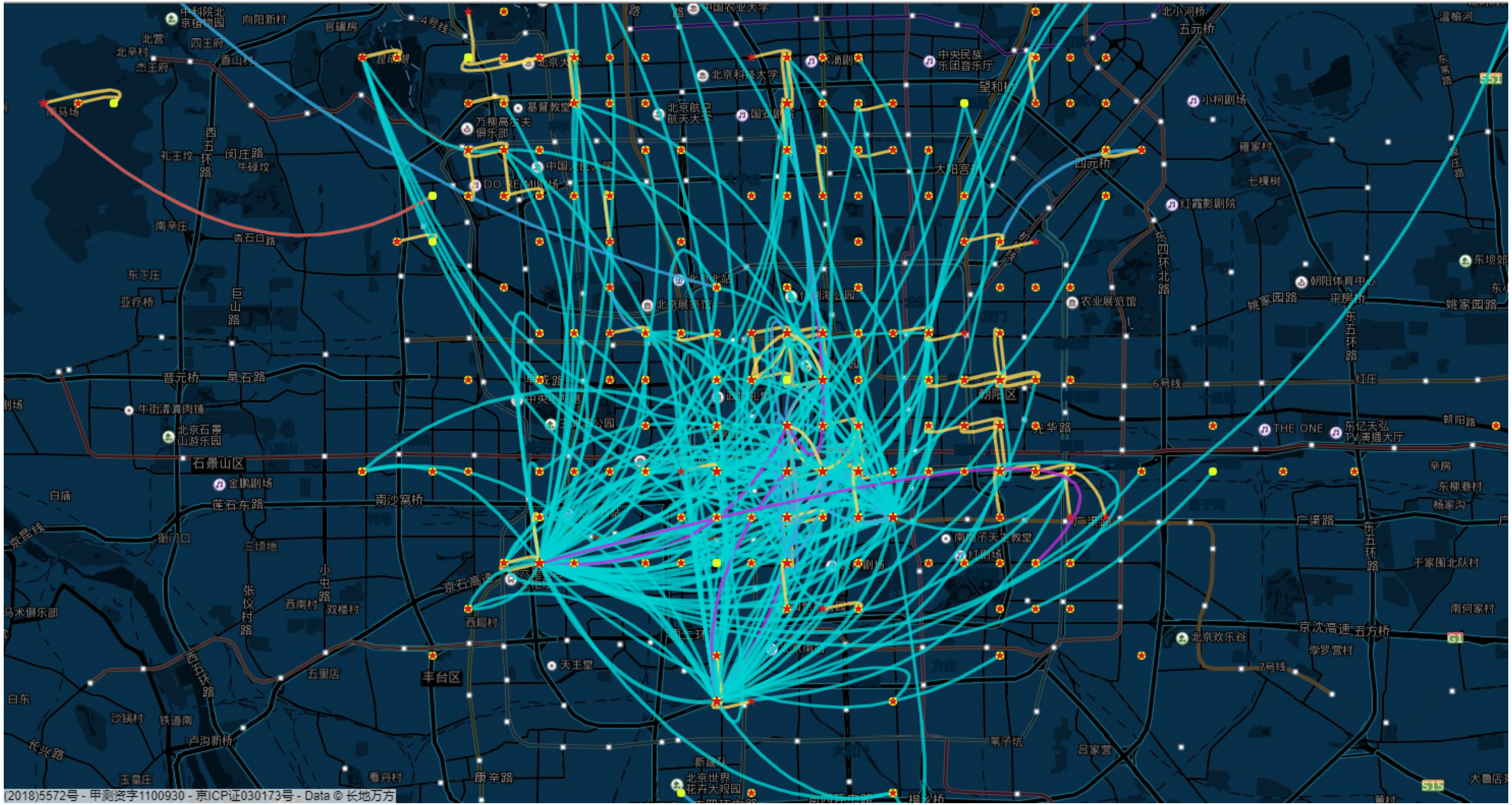



Figure 2: **Complex Plan and Reduced Plan Diagram (Query 8, OptA)**


[1] Dana Van Aken, Andrew Pavlo, Geoffrey J. Gordon, Bohan Zhang: Automatic Database Management System Tuning Through Large-scale Machine Learning. SIGMOD Conference 2017: 1009-1024

[2] Naveen Reddy, Jayant R. Haritsa: Analyzing Plan Diagrams of Database Query Optimizers. VLDB 2005: 1228-1240

# Others



石景山区
丰台区
朝阳区
北京科技大学
农业展览馆
北京欢乐谷
THE ONE
五元桥
四元桥
五方桥
红庄
广渠路
京沈高速
6号线
7号线
(2018)5572号 - 甲测资字1100930 - 京ICP证030173号 - Data © 长地万方

# Thoughts in Figure/Table Plotting

# Before Drawing…

- Make all experimental results be prepared and organized.
- Thinking about:
  - what phenomena you want to show (e.g. data distribution, desirable properties of your model)
  - do your data and findings support the phenomena?
  - what is the clearest way to visualize this phenomena?
- Think about your audience
  - E.g. when describing a deep neural network model to deep neural network audience, use the common visualization of the model and incorporate your novel improvements.
- Use your familiar tools
  - Gnuplot, Python (matplotlib), R …
  - Sometimes, Gnuplot produces the wrong fig. Tying to reboot it.

# While Drawing…

- No needs to be fancy, but make it clear. Do not mislead the reader.
- Pay attention to:
  - Value elegance and simplicity.
  - Captions are not optional.
- Highlight the difference/observation in an elegant way,
  - Color, pattern, bold …
- You may do more follow-up work after the first drawing
  - Adjust scale/color/font size/font weight/font type …
  - Remove the white margin
- Guarantee the image quality when export
  - e.g. Pdf (cut), eps …