The V*-Diagram: A Query-Dependent Approach to Moving KNN Queries

Sarana Nutanong, Rui Zhang, Egemen Tanin, Lars Kulik

Dept. of Computer Science and Software Engineering
University of Melbourne

Motivation

Consider two scenarios:

- a driver in a GPS-equipped car finding the nearest gas station along the route of a trip;
- a tourist walking in the city looking for the nearest ATM.

These scenarios are examples of *moving k nearest neighbor queries (MkNN).*

Simple Approach
The Voronoi Diagram

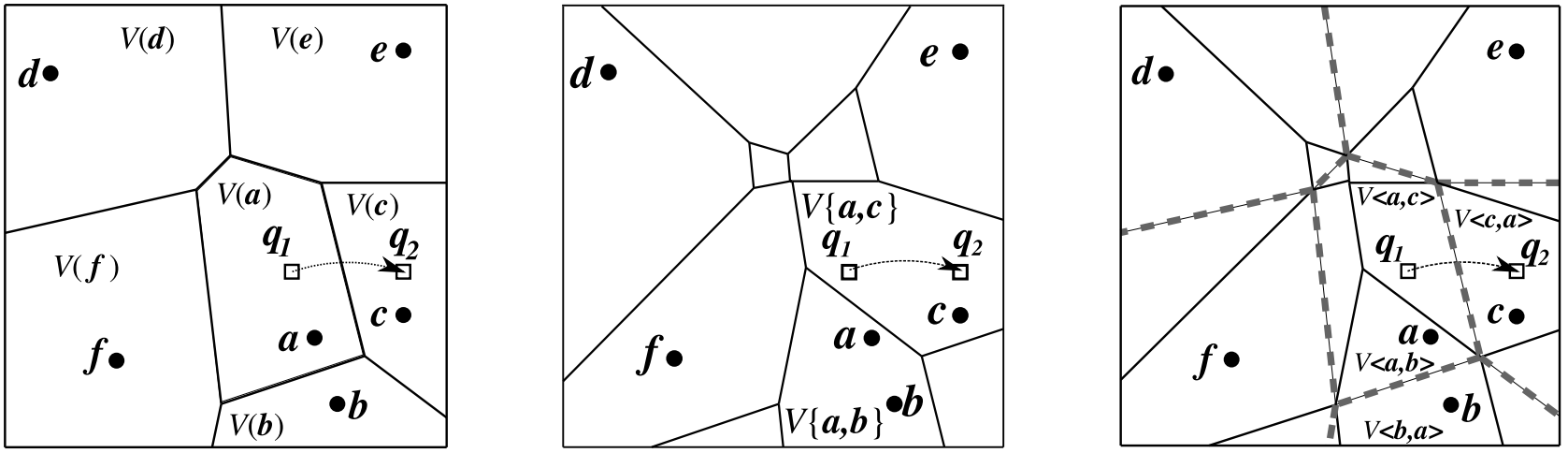


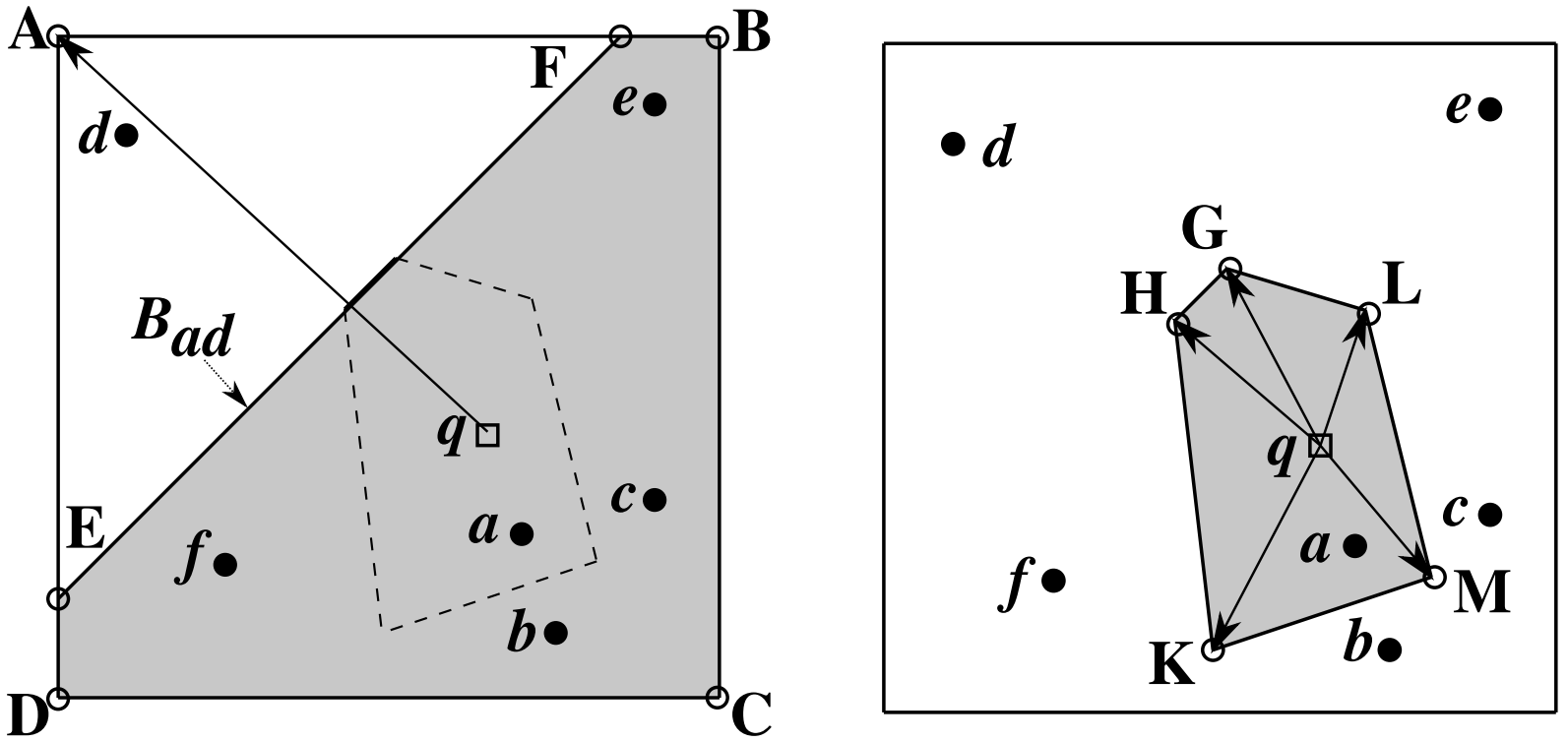
Figure 1: Voronoi diagrams

Drawbacks:

1. Expensive precomputations
2. Inefficient update operations
3. No support for dynamically changing k values

Best Existing Approach

Influence-set Retrieval [Zhang et al., 2003]



(a) Bisector B_{ad} is discovered as a boundary.

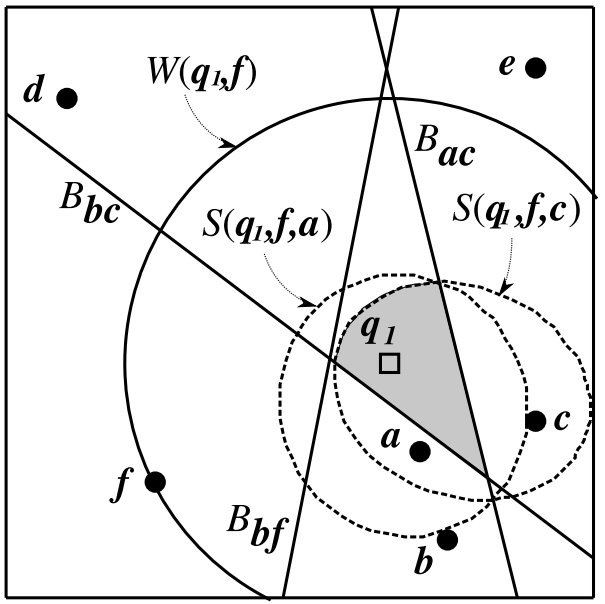
(b) All boundaries are discovered

Figure 2: Computing a Voronoi cell locally

Our Approach: V*-Diagram

Objectives:

1. Requires *no precomputation*
2. Supports dynamic *insertions*/*deletions* of objects
3. Handles *dynamically changing* k

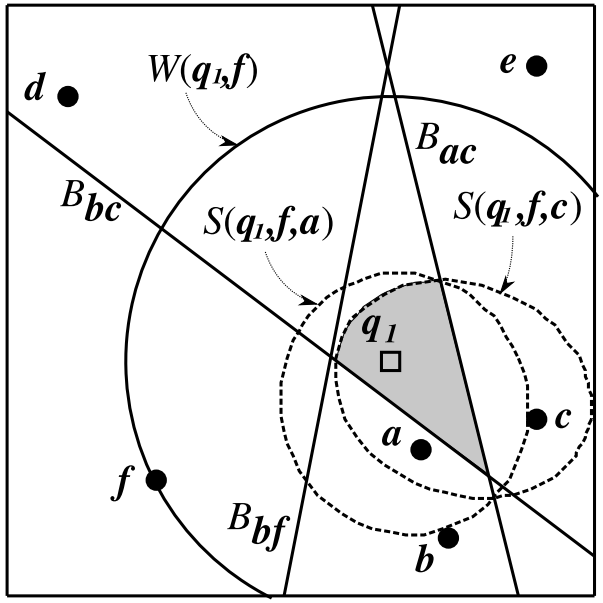


Our Approach: V*-Diagram

Objectives:

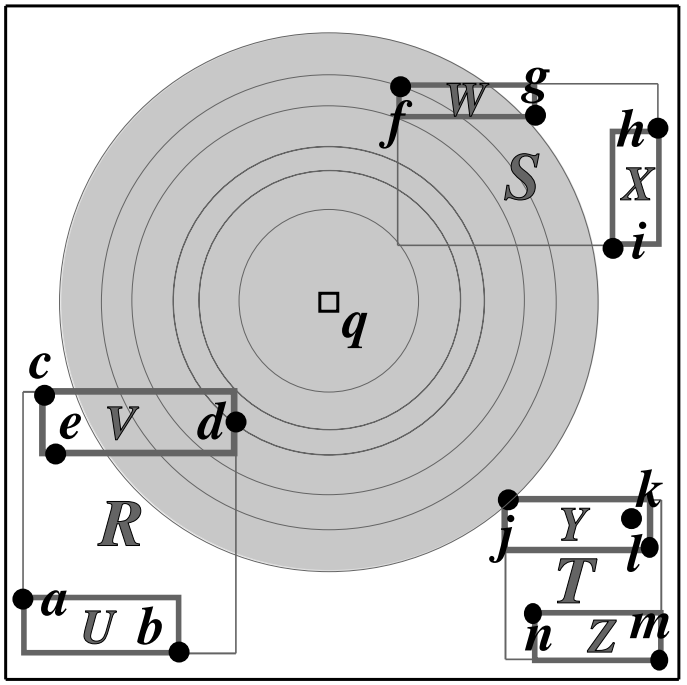
1. Requires *no precomputation*
2. Supports dynamic *insertions*/*deletions* of objects
3. Handles *dynamically changing* k

Result: Outperforms the best practice [Zhang et al.] by *2 orders of magnitude*



The V*-Diagram

Known Region



If the known NNs to q are

$$\{d, f, j\},$$

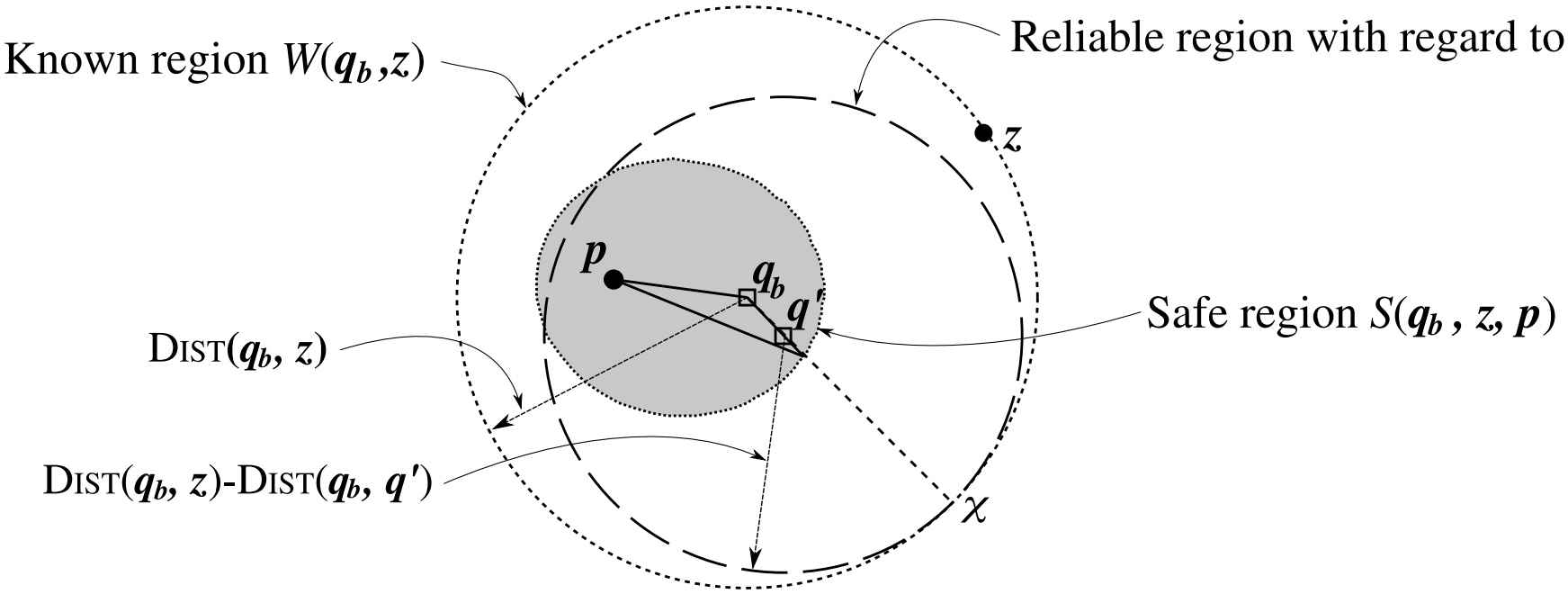
the know region $W(q, j)$ is

$$\{v : dist(q, v) \leq dist(q, j)\}.$$

The V*-Diagram

Safe region wrt a data point

We retrieve $(k + x)$ objects. In this example, k and x are 1, so we retrieve $\boldsymbol{p}$ and $\boldsymbol{z}$.



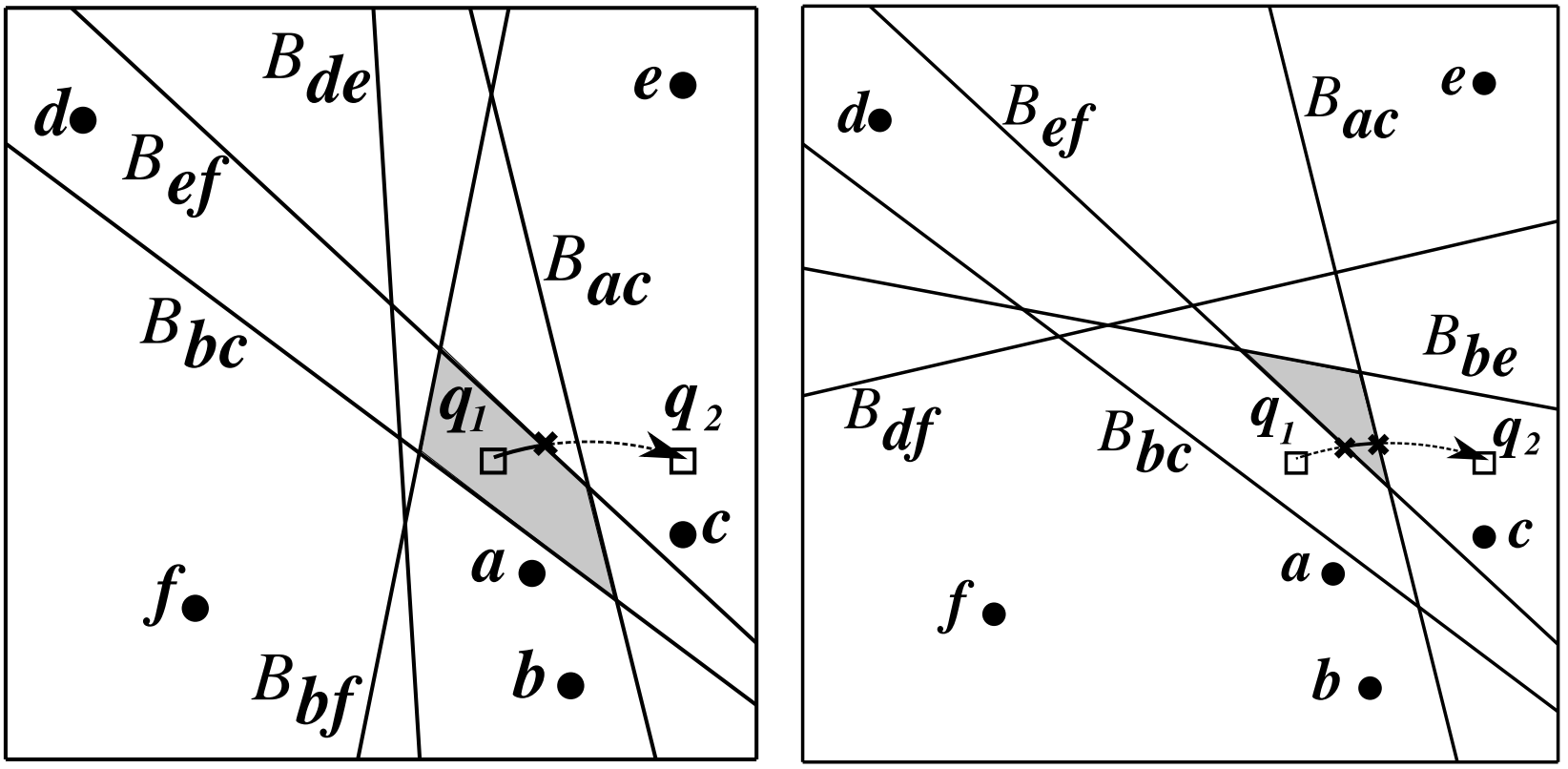
If $\boldsymbol{q'} \in S(\boldsymbol{q_b}, \boldsymbol{z}, \boldsymbol{p})$ then,

$$\forall \boldsymbol{p'} \notin W(\boldsymbol{q_b}, \boldsymbol{z}), dist(\boldsymbol{q'}, \boldsymbol{p}) < dist(\boldsymbol{q'}, \boldsymbol{p'}).$$

$$S(\boldsymbol{q_b}, \boldsymbol{z}, \boldsymbol{p}) = \{\boldsymbol{q'} : dist(\boldsymbol{p}, \boldsymbol{q'}) \leq dist(\boldsymbol{q_b}, \boldsymbol{z}) - dist(\boldsymbol{q_b}, \boldsymbol{q'})\}$$

The V*-Diagram

The Fixed-rank Region (FRR) [Kulik and Tanin, 2006]



(a) $\langle a, c, b, f, e, d \rangle$ (b) $\langle a, c, b, e, f, d \rangle$

Figure 3: Incremental rank update

The V*-Diagram

Integrated Safe Region (ISR) and V*-kNN

ISR is an intersection of

1. the safe region wrt k^{th} NN, $S(\boldsymbol{q_b}, \boldsymbol{z}, \boldsymbol{p_k})$;
2. the FRR of the $(k+x)$ NNs of $\boldsymbol{q_b}$.

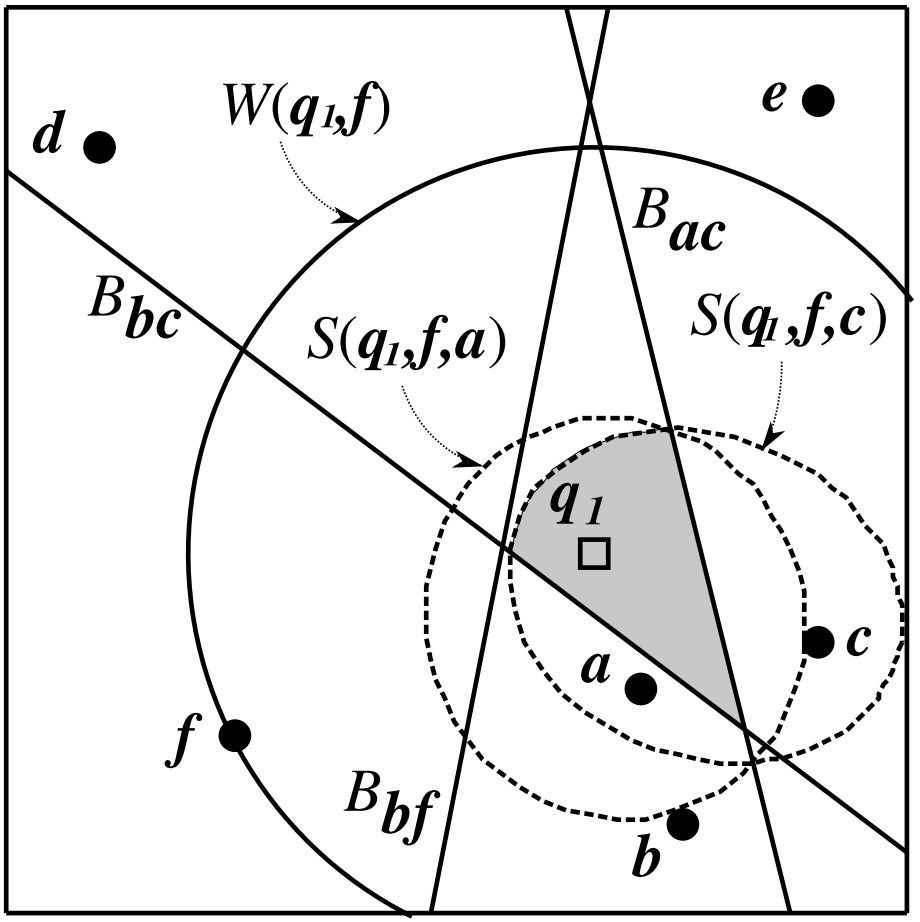
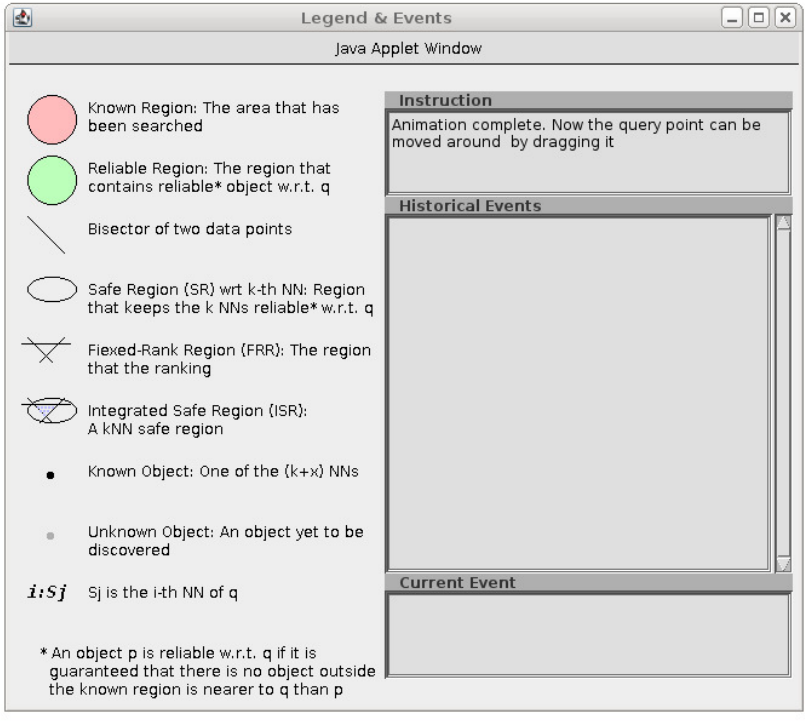
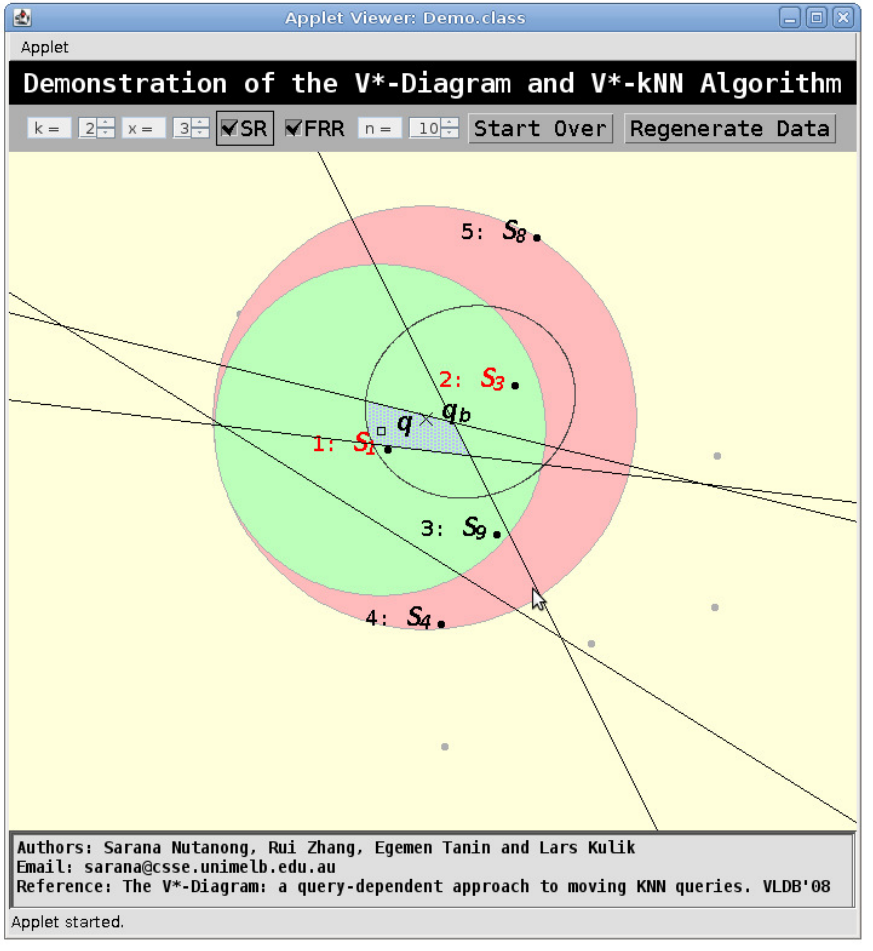


Figure 4: V*-kNN Example ($k = 2$, $x = 2$)

V*-kNN Algorithm

http://www.csse.unimelb.edu.au/~sarana/demo.html

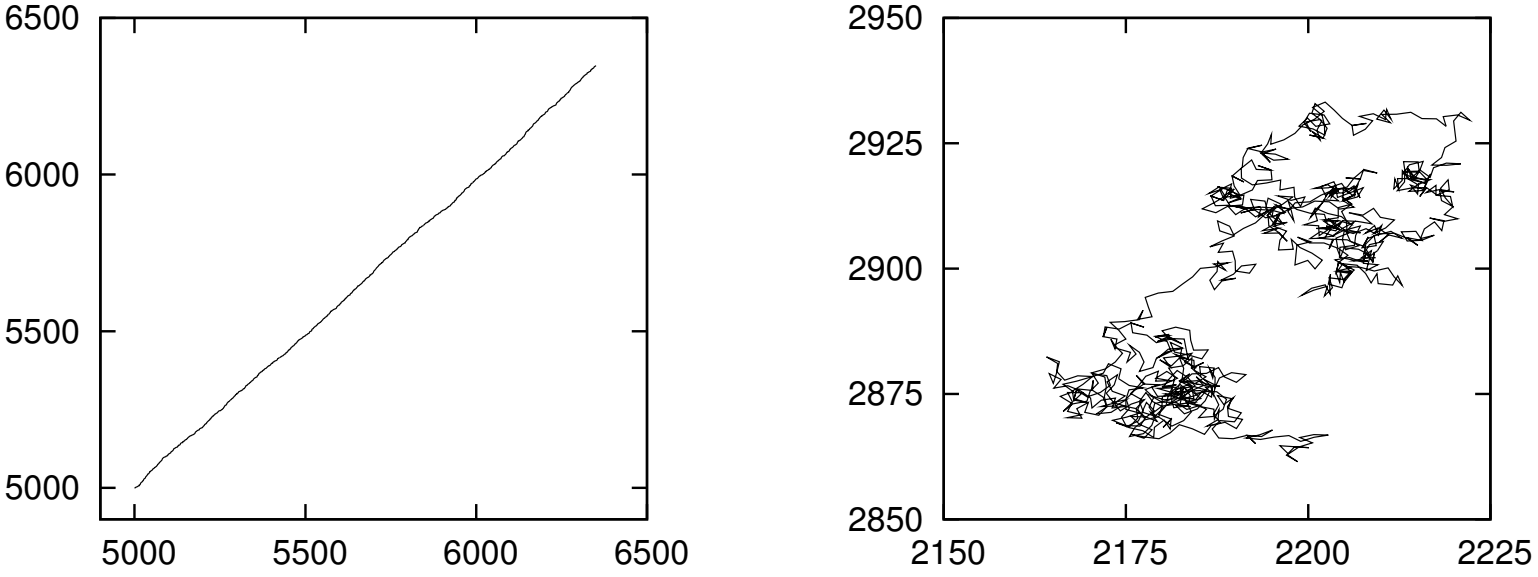




Experiments

- Data Structure: R*-trees (1-kB block size).
- Comparative Method: RIS-kNN [Zhang et al.]
- Datasets:
  - (U) 25,000 of data points in uniform distribution
  - (Z) 25,000 of data points in Zipfian distribution
  - (C) 65,743 postal addresses from California
  - (N) 119,897 postal addresses from North-Eastern USA

Experiments
Trajectories

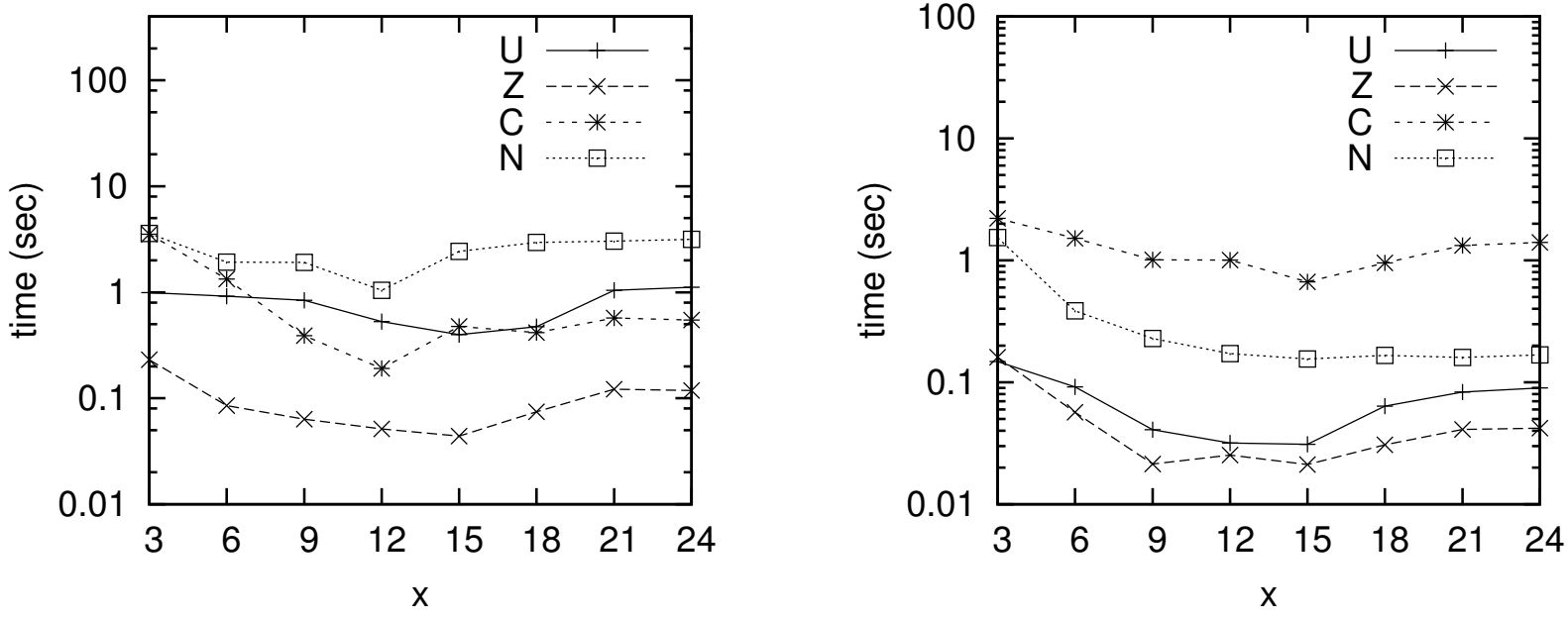


(a) Directional (D) (b) Random (R)

Figure 5: Trajectory types

Experiments

total cost wrt x



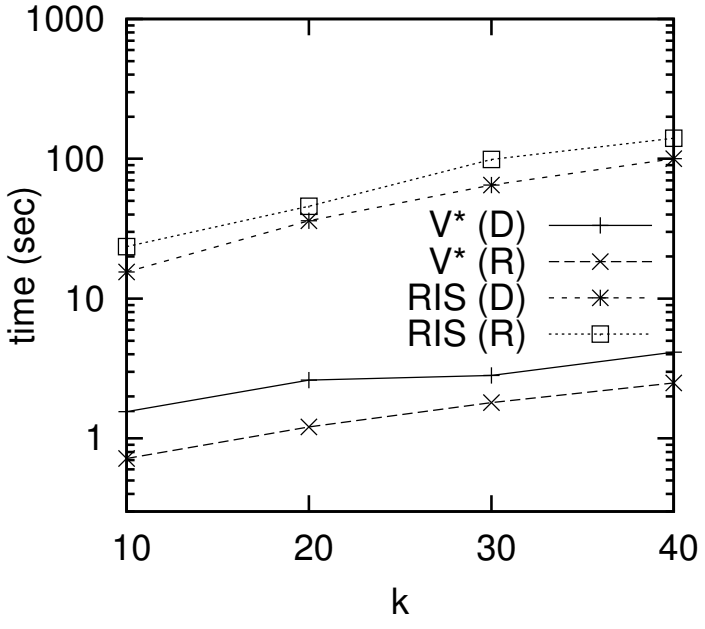
(a) Total cost (D)

(b) Page access (D)

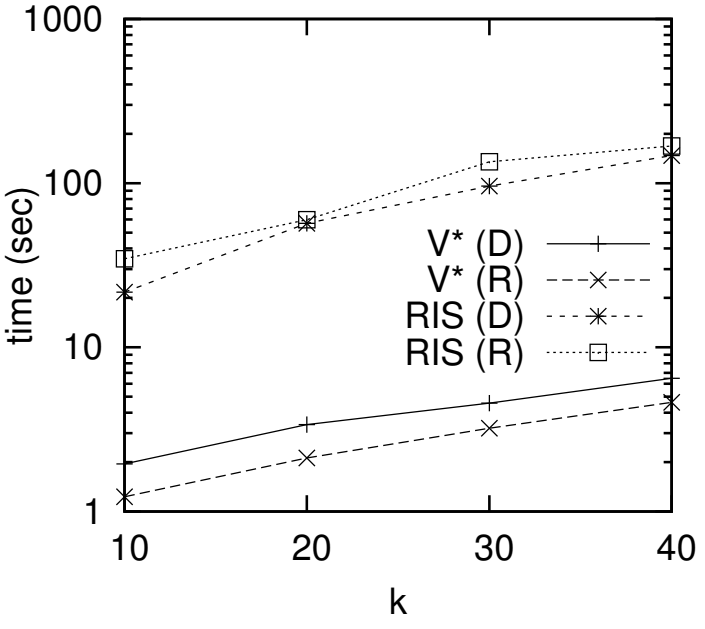
Figure 6: Effect of x

Experiments

total cost wrt k



(a) Total Cost (California)

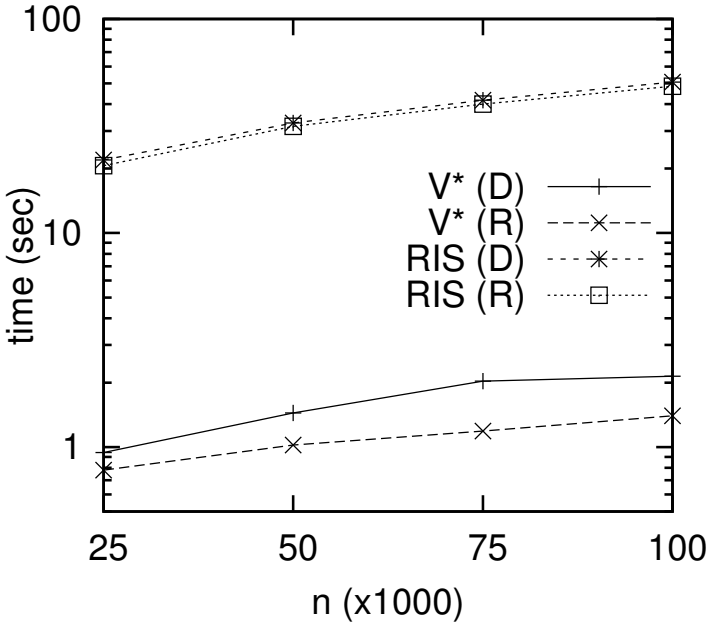


(b) Total Cost (North-Eastern USA)

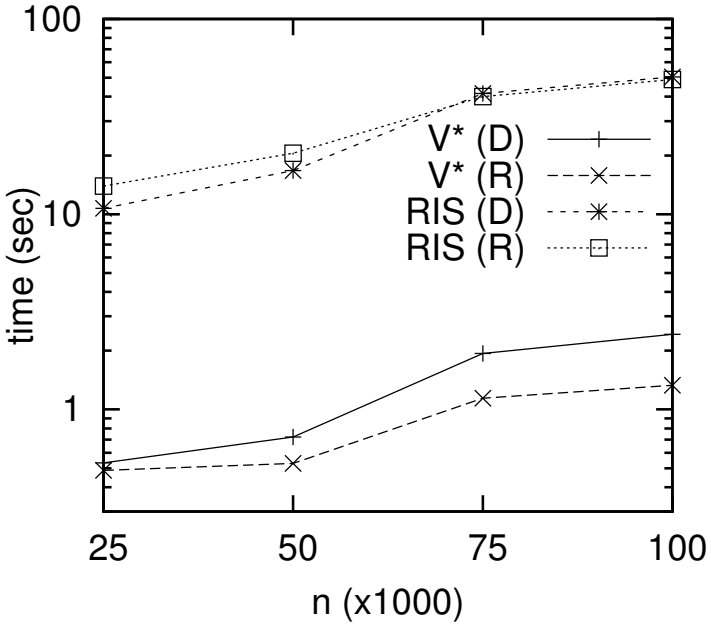
Figure 7: Effect of k

Experiments

total cost wrt n



(a) Total Cost (Uniform)



(b) Total Cost (Zipfian)

Figure 8: Effect of dataset size

Cost model

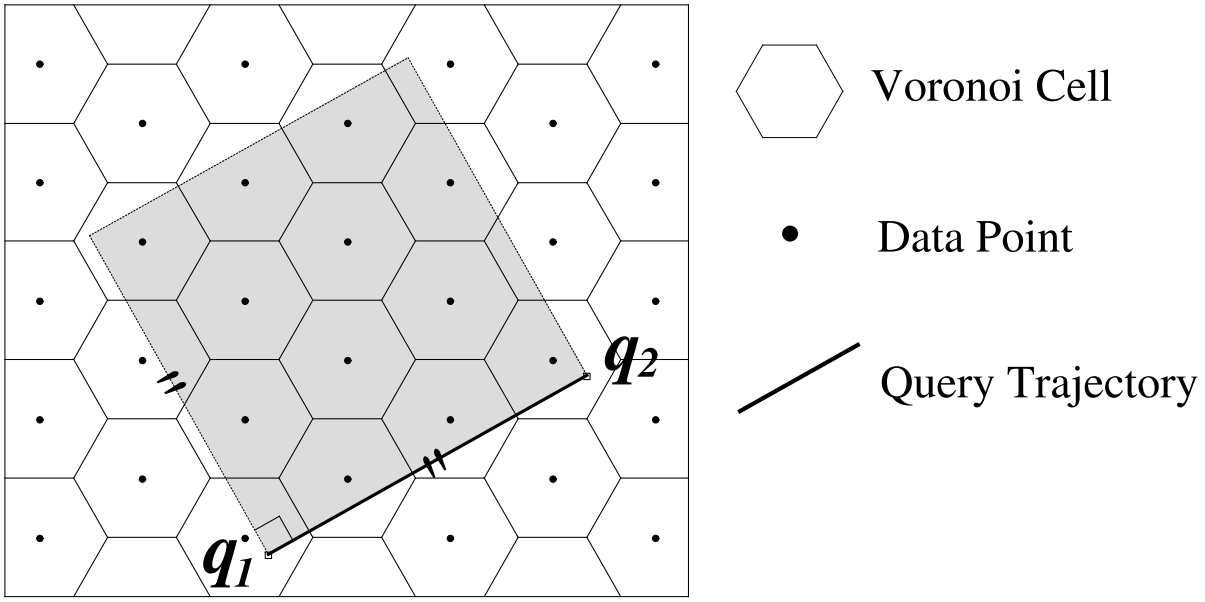
RIS-kNN

The number of the kVD cells in 2D space is approximated as

$2kn$ [Okabe et al., 1992].

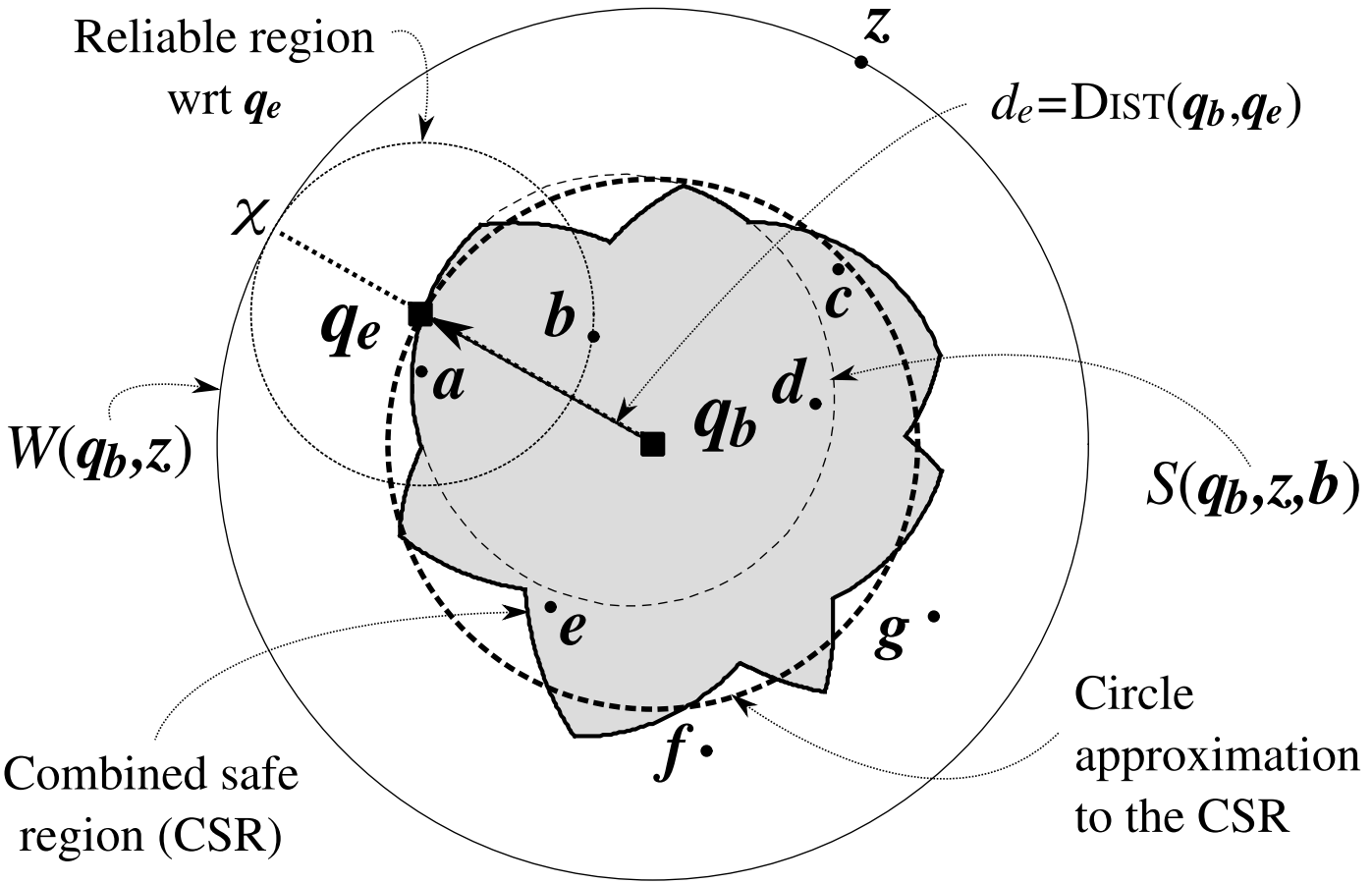
For a given trajectory length l, the number n_v of kVD cells crossed by the trajectory is given by

$$n_v = l\sqrt{2kn}.$$



Cost model

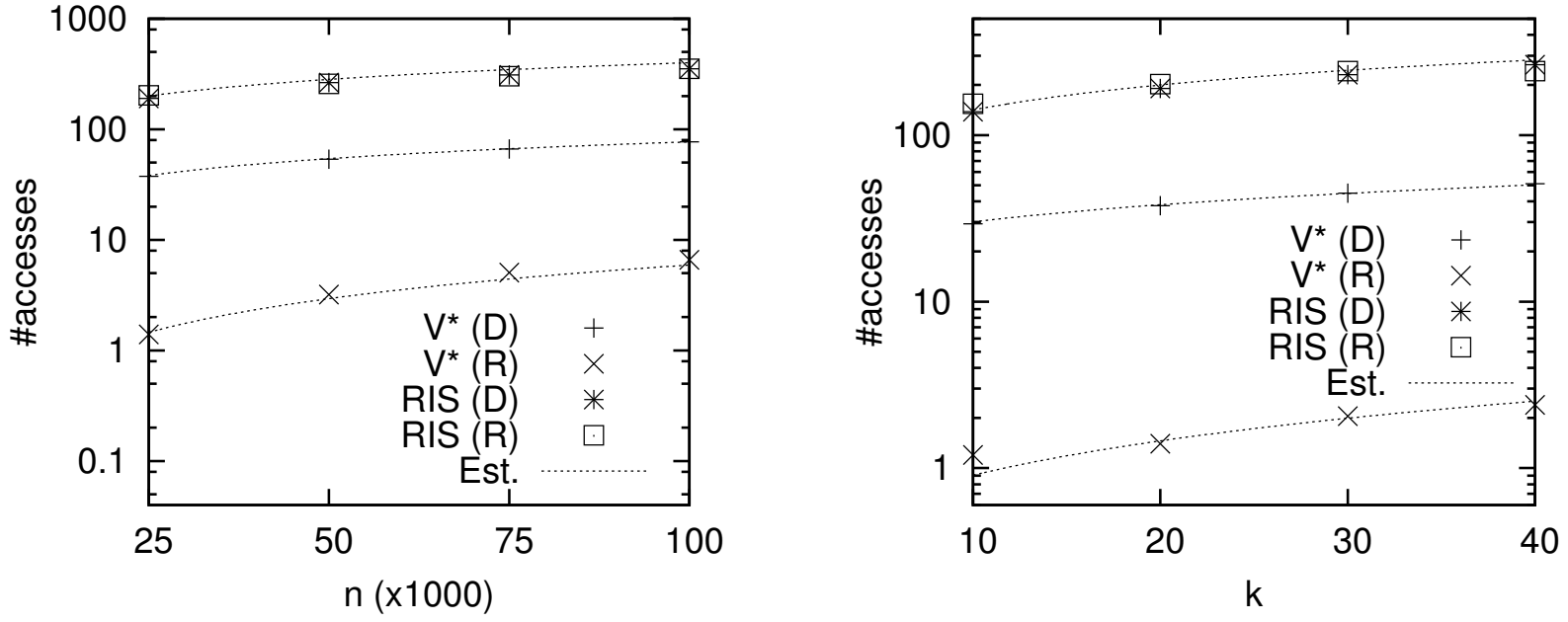
V*-kNN



Directional: $n_b = l/d_e$.

Random: $n_b = ls/d_e^2$, where s is the step size.

Experiments
Cost Model



(a) Effect of n

(b) Effect of k

Figure 9: Cost model validation

The V*-Diagram in a spatial network

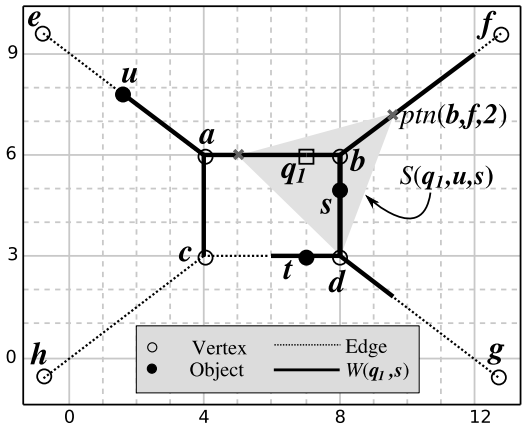


Figure 10: Safe region

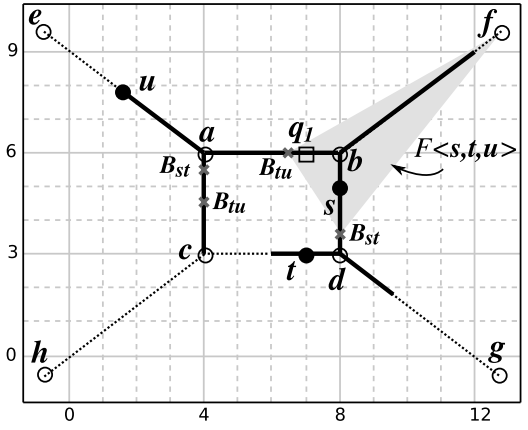


Figure 11: Fixed-rank region

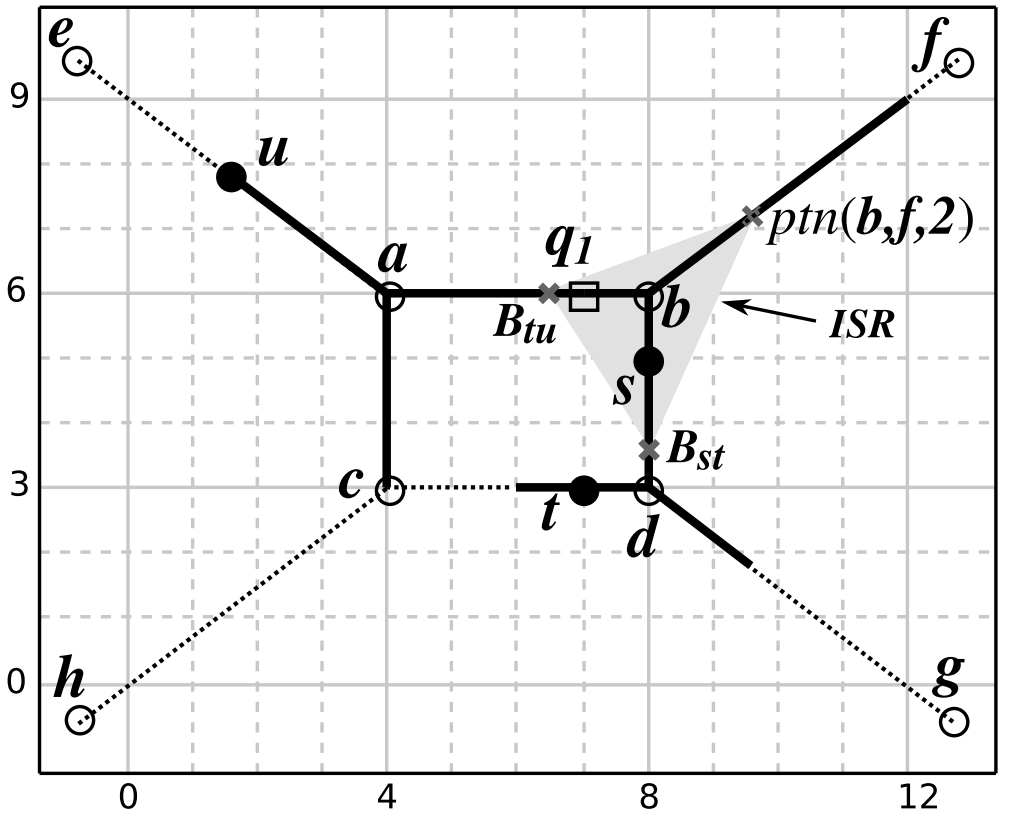


Figure 12: ISR is $S(\boldsymbol{q_1}, \boldsymbol{u}, \boldsymbol{s}) \cap F\langle \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{u} \rangle$

Experiments

The V*-Diagram in a spatial network

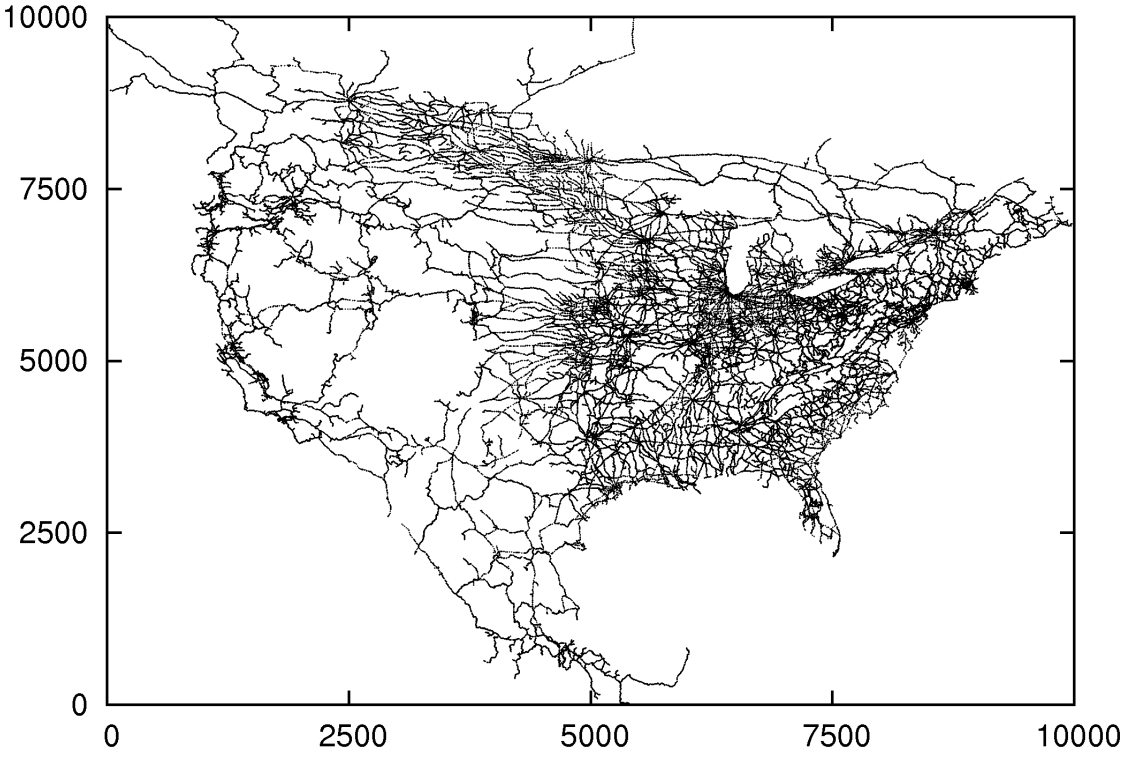
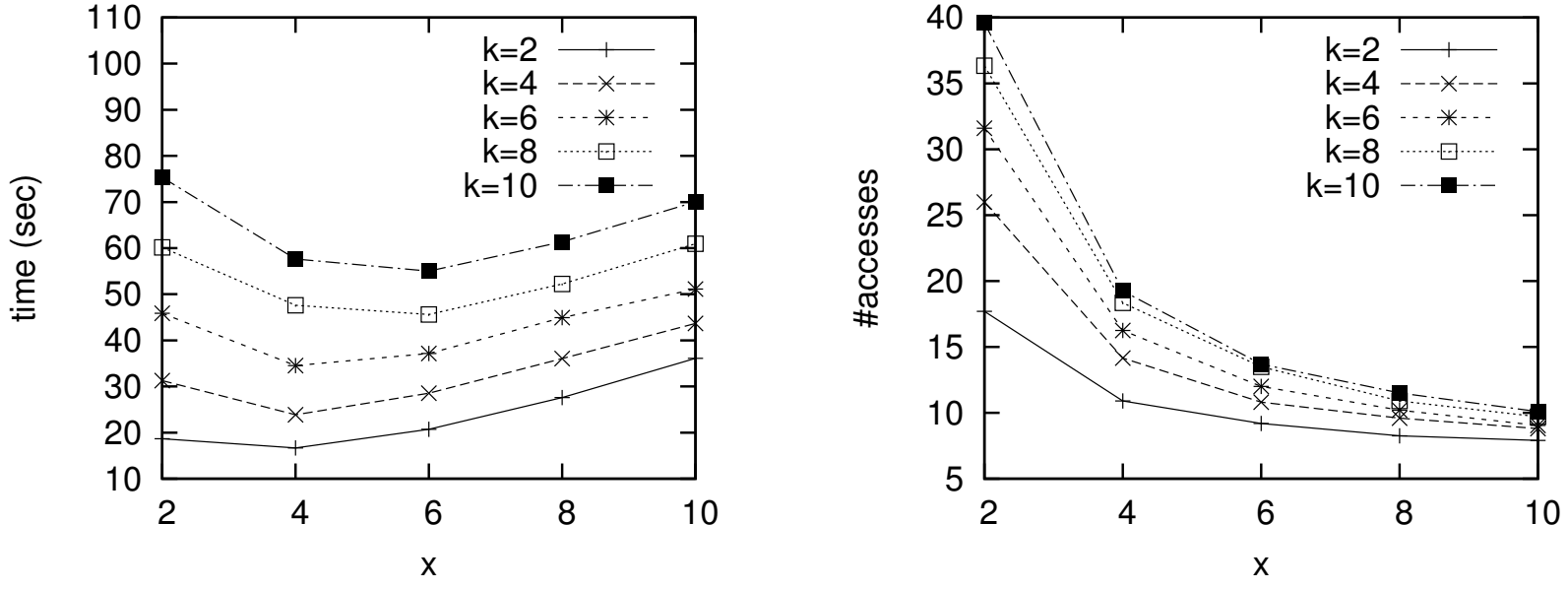


Figure 13: Road network in north America (175,813 nodes and 179,179 edges)

Experiments

The V*-Diagram in a spatial network



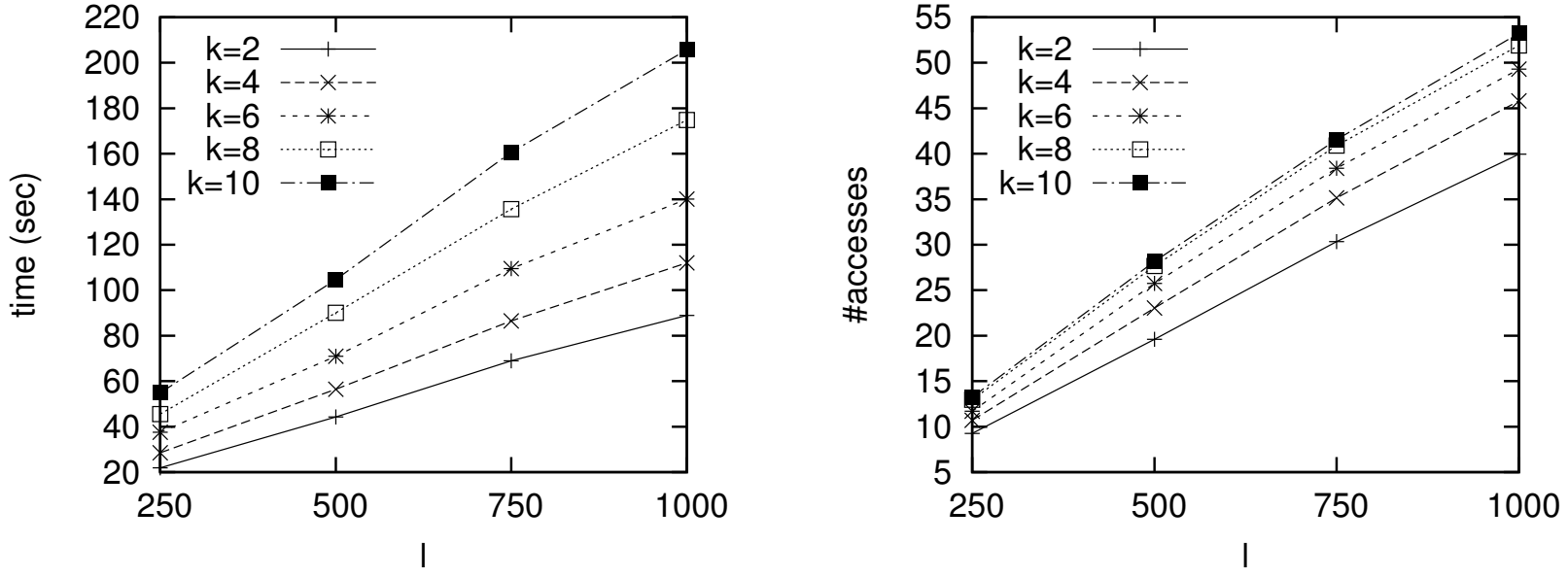
(a) Total Response Time

(b) Access Cost

Figure 14: Spatial network: effect of x

Experiments

The V*-Diagram in a spatial network



(a) Total Response Time

(b) Access Cost

Figure 15: Spatial network: effect of l

Conclusions

- The V*-Diagram constructs a safe region using:
  1. the location of the query point,
  2. kNN-search coverage (known region),
  3. known data points.
- V*-kNN is *local*, *incremental* and *dynamic*.
- V*-kNN outperforms the best existing technique by two orders of magnitude.
- The V*-diagram is a general philosophy, which can be applied to most safe region based techniques.

Related Publications

- S. Nutanong, R. Zhang, E. Tanin, L. Kulik: Analysis and Evaluation of V*-kNN: An Efficient Algorithm for Moving k Nearest Neighbor Queries. To appear in VLDB Journal.
- S. Nutanong, R. Zhang, E. Tanin, L. Kulik: V*-kNN: An Efficient Algorithm for Moving k Nearest Neighbor Queries (Demo). ICDE 2009: 1519-1522.
- S. Nutanong, R. Zhang, E. Tanin, L. Kulik: The V*-Diagram: a query-dependent approach to moving KNN queries. PVLDB 1(1): 1095-1106 (2008).

Key References

- Lars Kulik, Egemen Tanin: Incremental Rank Updates for Moving Query Points. GIScience 2006:251-268.
- Atsuyuki Okabe, Berry Boots, Kokichi Sugihara, Sung Nok Chiu: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley & Sons, Inc., 1992.
- Jun Zhang, Manli Zhu, Dimitris Papadias, Yufei Tao, Dik Lun Lee: Location-based Spatial Queries. SIGMOD 2003:443-454.