Hierarchical Graph Embedding for Efficient Query Processing in Very Large Traffic Networks

Matthias Renz

Hans-Peter Kriegel, Peer Kröger, Tim Schmidt

Ludwig-Maximilians-Universität München

Munich, Germany

www.dbs.ifi.lmu.de

Outline

- **Introduction**

- Embedding of Large Road Networks
  - flat embedding
  - multi-level embedding
  - distance approximations

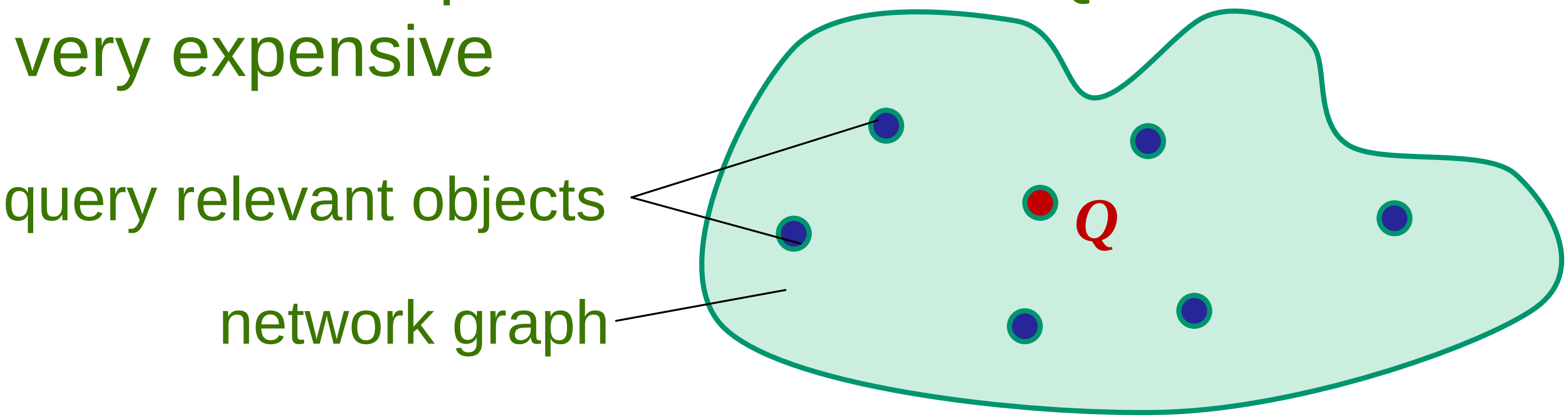
- Multi-Step Distance Query Processing
- Experimental Evaluation

- Summary

Introduction

- Distance Queries in Road-Networks
  - Query Types:
    - **distance range query**
    - **k-nearest-neighbor query**
  - Applications
    - **Location-Based Service Applications**
    - **new applications in the car / navigation system industry**
    - **…**

- traditional approaches:
  - mainly focus on small search space
- problem addressed here:
  - very dense road-network graph
  - query relevant objects are sparsely distributed
  - distance computations between Q and O∈DB very expensive



- we propose:
  - efficient filter-refinement query processor
    - **efficient computation of lower/upper-bounding distance approximations (filter step)**
    - **use distance approximations to accelerate the exact distance computations (refinement step)**
  - hierarchical graph embedding
    - **support queries covering large network graphs**

Outline

- Introduction
- **Embedding of Large Road Networks**
  - flat embedding
  - multi-level embedding
  - distance approximations
- Multi-Step Distance Query Processing
- Experimental Evaluation
- Summary

Embedding of Large Road Networks

- Flat Embedding:
  - Lipschitz embedding based on *k* reference nodes (landmarks) $N' = \{n_{r_1}, n_{r_2}, \ldots, n_{r_k}\} \subseteq N$
  - *reference node embedding* of *G*(*N*,*E*) based on *N*‘ defines the function:

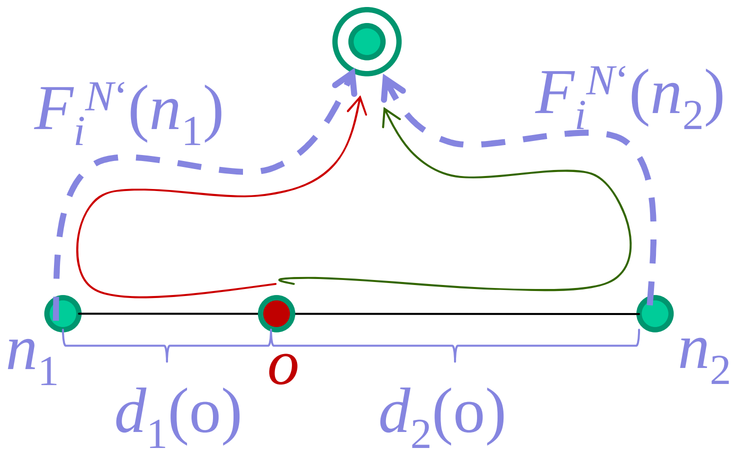
$$F^{N'}(n) = (F_1^{N'}(n), \ldots, F_k^{N'}(n)) \qquad (k = |N'|)$$

where

$$F_i^{N'}(n) = d_{net}(n, n_{r_i})$$

Embedding of Large Road Networks

– embedding of objects located on an edge between two nodes:



$$F_i^{N'}(o) = \min\{d_1(o) + F_i^{N'}(n_1), d_2(o) + F_i^{N'}(n_2)\}$$

- Distance Approximations:
  - lower bounding distance approximation

$$D(F^{N'}(x), F^{N'}(y)) = \max_{i=1..k}\left\{F_i^{N'}(x) - F_i^{N'}(y)\right\}$$

  - upper bounding distance approximation

$$D^*(F^{N'}(x), F^{N'}(y)) = \min_{i=1..k}\left\{F_i^{N'}(x) + F_i^{N'}(y)\right\}$$

Embedding of Large Road Networks

- problem with flat embedding:
  - few reference nodes
    - → **bad distance approximation**
    - → **low filter selectivity**
  - many reference nodes
    - → **large reference node vectors**
    - → **high storage costs**

  → bad query performance
- approximations for long distances do not need to be as accurate as short distances

Embedding of Large Road Networks

- Hierarchical Embedding:
  - idea: introducing further embedding levels
  - lower level embedding for local distance approximations

    → **only a small set of (k<<|N'|) local reference nodes are required (1st-level)**

  - higher level embedding for global distance approximations

    → **distances between sets of local reference nodes are materialized in an additional graph (2nd-level)**

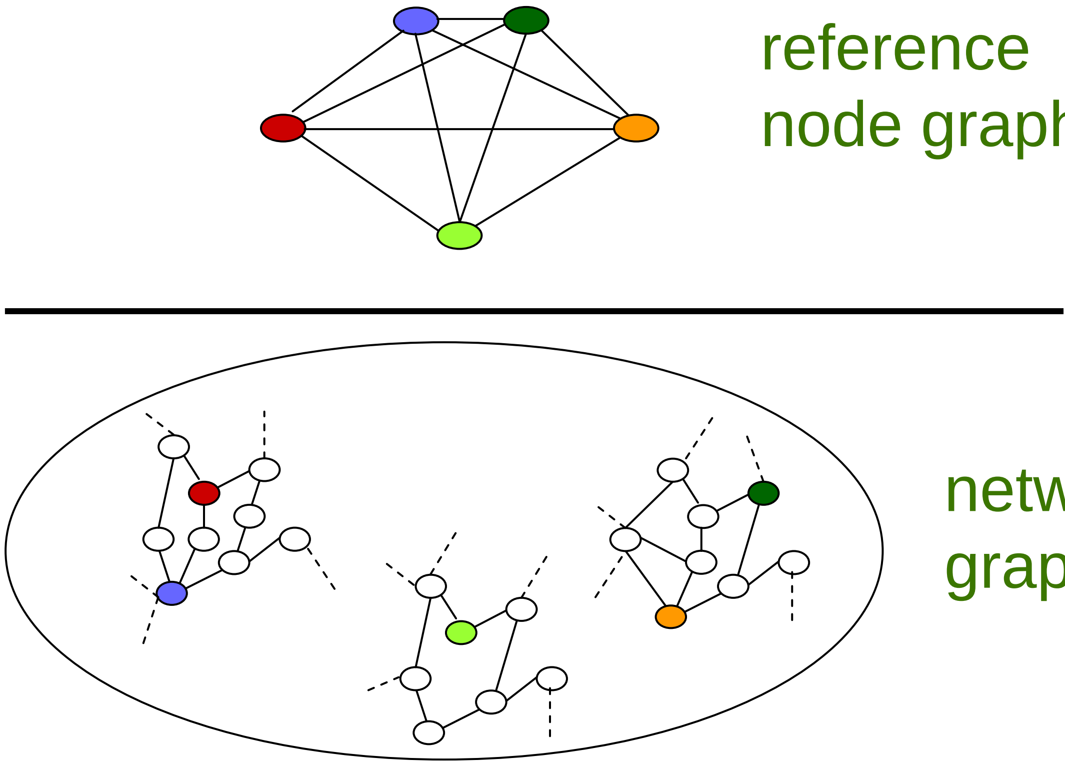
  - distance approximations are composed of 1st-level distances and 2nd-level distances

Embedding of Large Road Networks

- Schema of the 2-Level Embedding
  - Matrix *M'* stores pair-wise distances between 1st-level reference nodes

2nd level

reference node graph *M'*



1st level

network graph

Embedding of Large Road Networks

- Distance Approximations:
  - lower bounding distance approximation

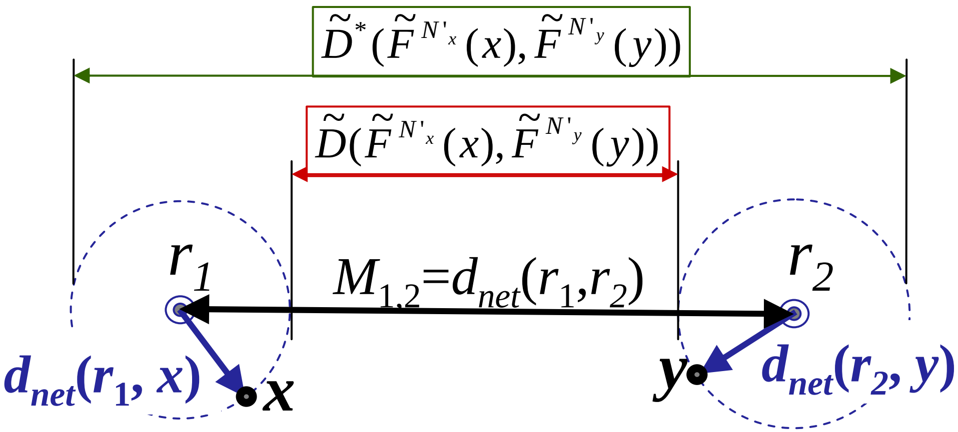
$$\tilde{D}(\tilde{F}^{N_x}(x), \tilde{F}^{N_y}(y)) = \max_{k \in N_x, l \in N_y} \left\{ \begin{array}{l} M_{i_k, i_l} - \tilde{F}_k^{N_x}(x) - \tilde{F}_l^{N_y}(y), \\ \tilde{F}_k^{N_x}(x) - M_{i_k, i_l} - \tilde{F}_l^{N_y}(y), \\ \tilde{F}_l^{N_y}(y) - \tilde{F}_k^{N_x}(x) - M_{i_k, i_l}, \\ 0 \end{array} \right\}$$

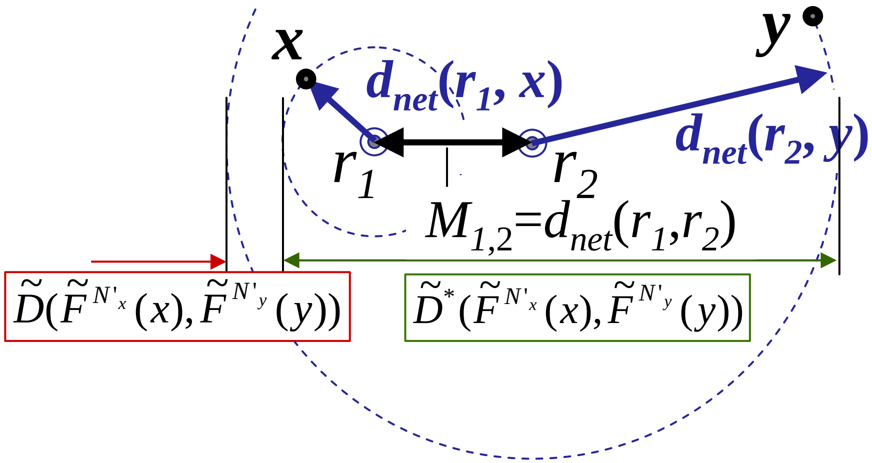
  - upper bounding distance approximation

$$\tilde{D}^*(\tilde{F}^{N_x}(x), \tilde{F}^{N_y}(y)) = \min_{k \in N_x, l \in N_y} \left\{ M_{i_k, i_l} + \tilde{F}_k^{N_x}(x) + \tilde{F}_l^{N_y}(y) \right\}$$

Embedding of Large Road Networks

- Examples:





$d_{net}(r_1,r_2)>d_{net}(x,r_1)+d_{net}(y,r_2)$

$d_{net}(y,r_2)>d_{net}(x,r_1)+d_{net}(r_1,r_2)$

Outline

- Introduction
- Embedding of Large Road Networks
  - flat embedding
  - multi-level embedding
  - distance approximations
- **Multi-Step Distance Query Processing**
- Experimental Evaluation
- Summary

Multi-Step Distance Query Processing

- Multi-Step Query Processing
  - Distance Range Queries
    - **filter step: scan over all objects and filter out true hits and true drops according to D and D***
    - **refinement: shortest path computation for all remaining candidates**
  - *k*-Nearest Neighbor Queries
    - **refinement optimal multi-step k-NN query as proposed in [1]**

[1] Kriegel H.-P., Kröger P, Kunath P., Renz M.:
Generalizing the Optimality of Multi-Step k-Nearest Neighbor Query Processing.
In Proc. 10th International Symposium on Spatial and Temporal Databases (SSTD'07), Boston, U.S.A., 2007, pp. 75-92.

- Refinement Step:

  Accelerated Shortest Path Computation
  - computation of the shortest path between query object *q* and database object *o*:
    - **A*-search method.**
    - **use distance approximation *D* for the forward estimation**
    - **use distance approximation *D** to further prune the candidate list**

Outline

- Introduction
- Embedding of Large Road Networks
  - flat embedding
  - multi-level embedding
  - distance approximations
- Multi-Step Distance Query Processing
- **Experimental Evaluation**
- Summary

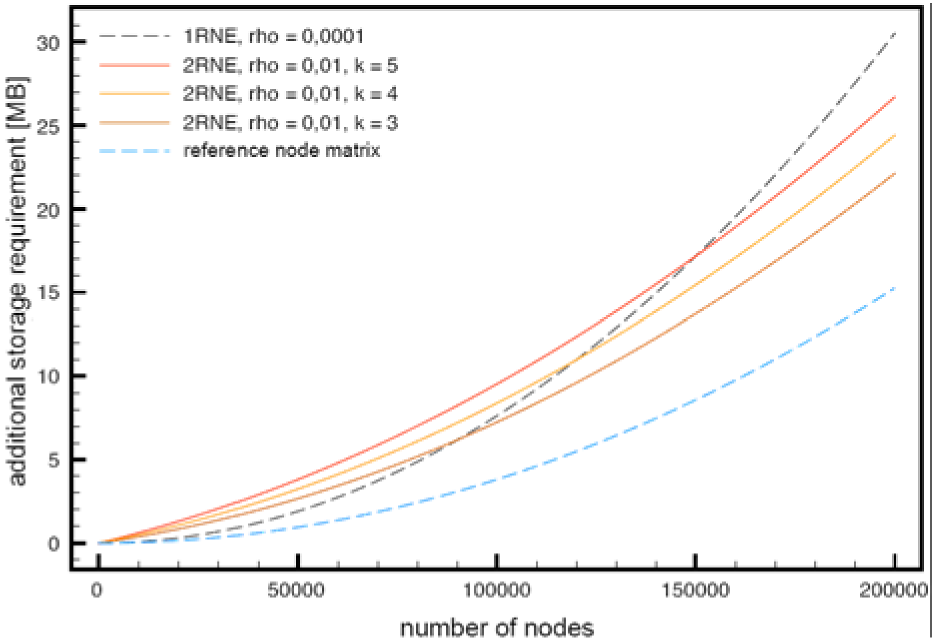
Experimental Evaluation

- Datasets:
  - SA: San Francisco, 175 000 nodes
  - TG: San Joaquin County, 18 300 nodes

- Approaches:
  - REF: no embedding
  - 1RNE: flat embedding
  - 2RNE: 2-level embedding

Experimental Evaluation

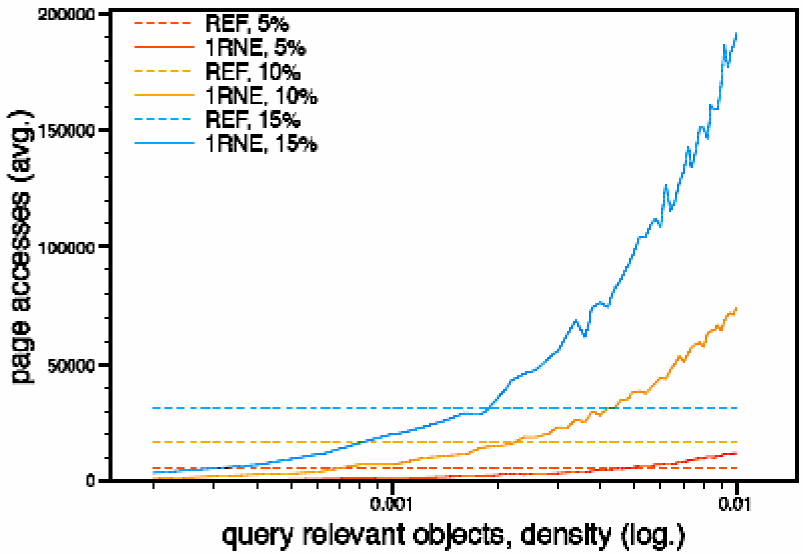
- Storage Requirements
  - size of the embedding, w.r.t. size of the network graph



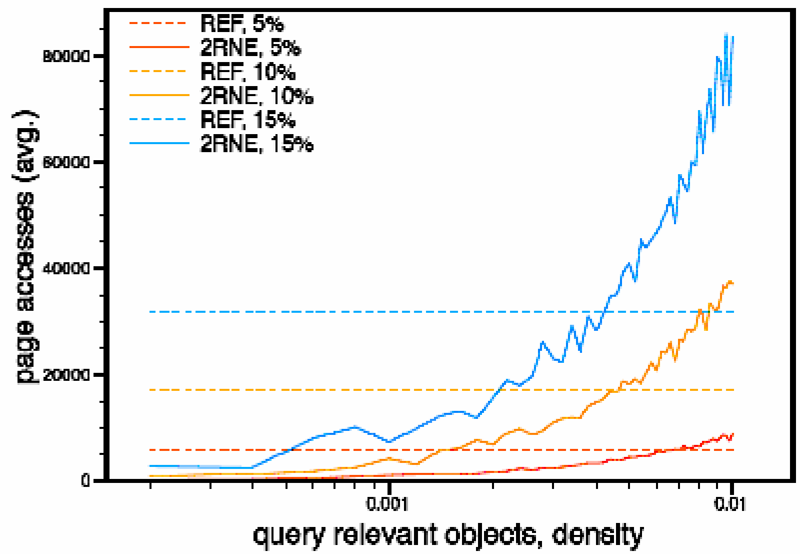
- **2-level embedding allows 2 orders of magnitude more reference nodes**

Experimental Evaluation

- Multi-Step Query Performance
  - Distance Range Queries DRQ
    - **performance measured in page accesses:**

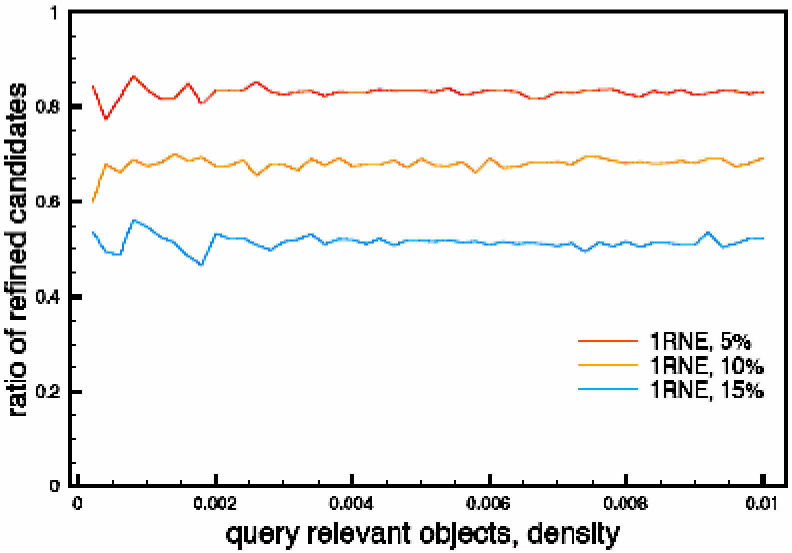


flat embedding (SA)

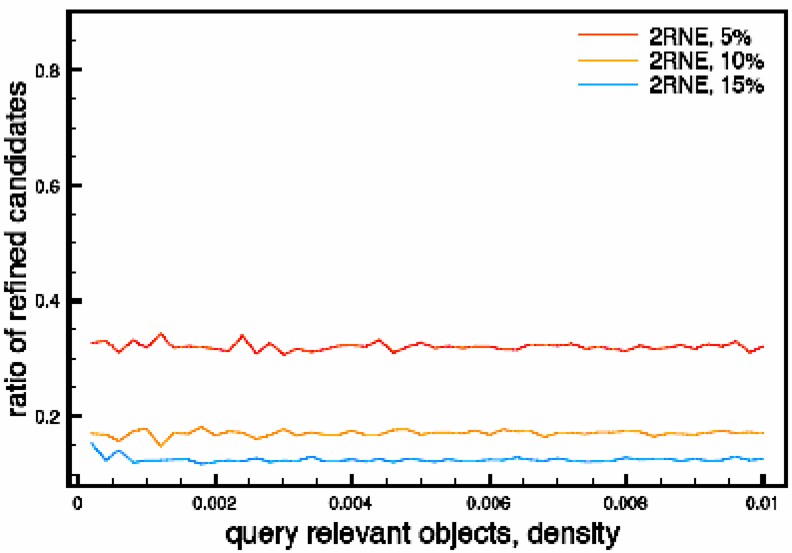


2-level embedding (SA)

- Multi-Step Query Performance
  - Distance Range Queries DRQ
    - **filter selectivity:**



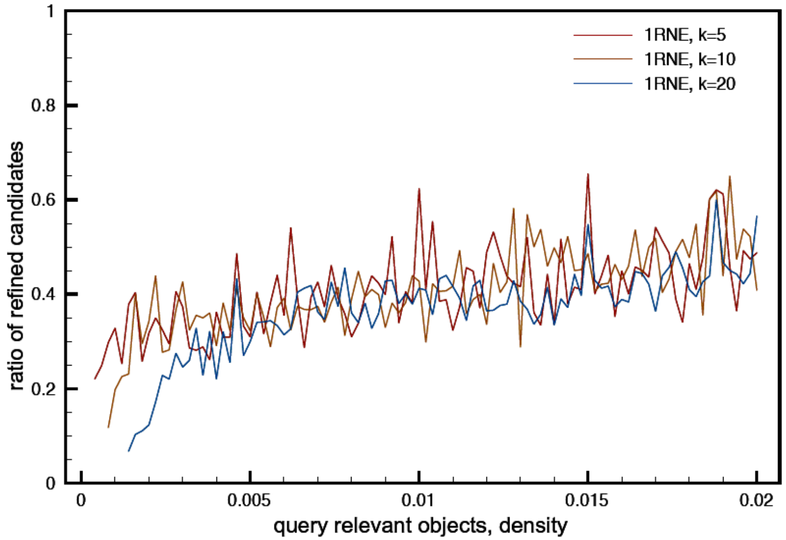
flat embedding (SA)



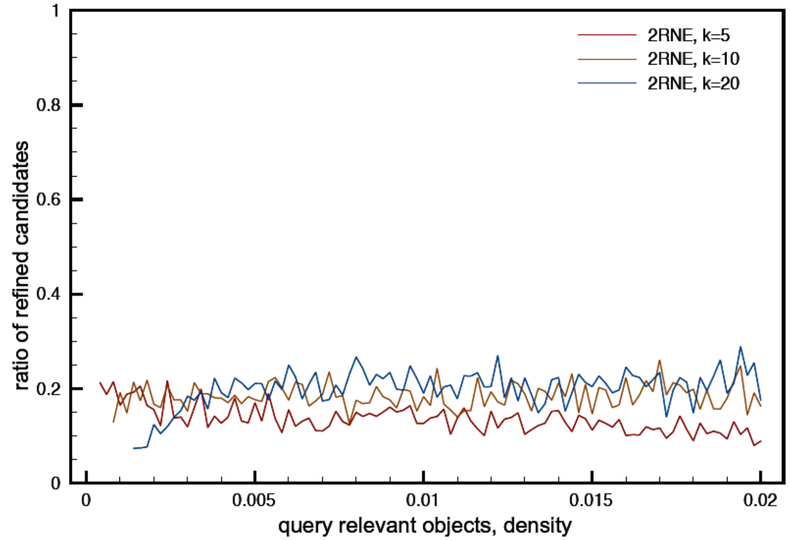
2-level embedding (SA)

Experimental Evaluation

- Multi-Step Query Performance
  - *k*-Nearest-Neighbor Queries *k*NNQ
    - **filter selectivity:**



flat embedding (TG)



2-level embedding (TG)

Experimental Evaluation

- Refinement: Exact Distance Computation
  - search spaces for one shortest path computation

Dijkstra	Euclidean (A*-Search)	flat embedding (A*-Search)	2-level embedding (A*-Search)

Outline

- Introduction
- Embedding of Large Road Networks
  - flat embedding
  - multi-level embedding
  - distance approximations
- Multi-Step Distance Query Processing
- Experimental Evaluation
- **Summary**

Summary

- Summary
  - we proposed
    - **hierarchical graph embedding**
    - **multi-step query processing based on lower and upper bounding distance estimations**
    - **accelerated A*-search based refinement**
  - hierachical embedding
    - **is appropriate for large graphs**
    - **outperforms flat embedding and other existing competitors in terms of pruning power and overall query performance**

Summary

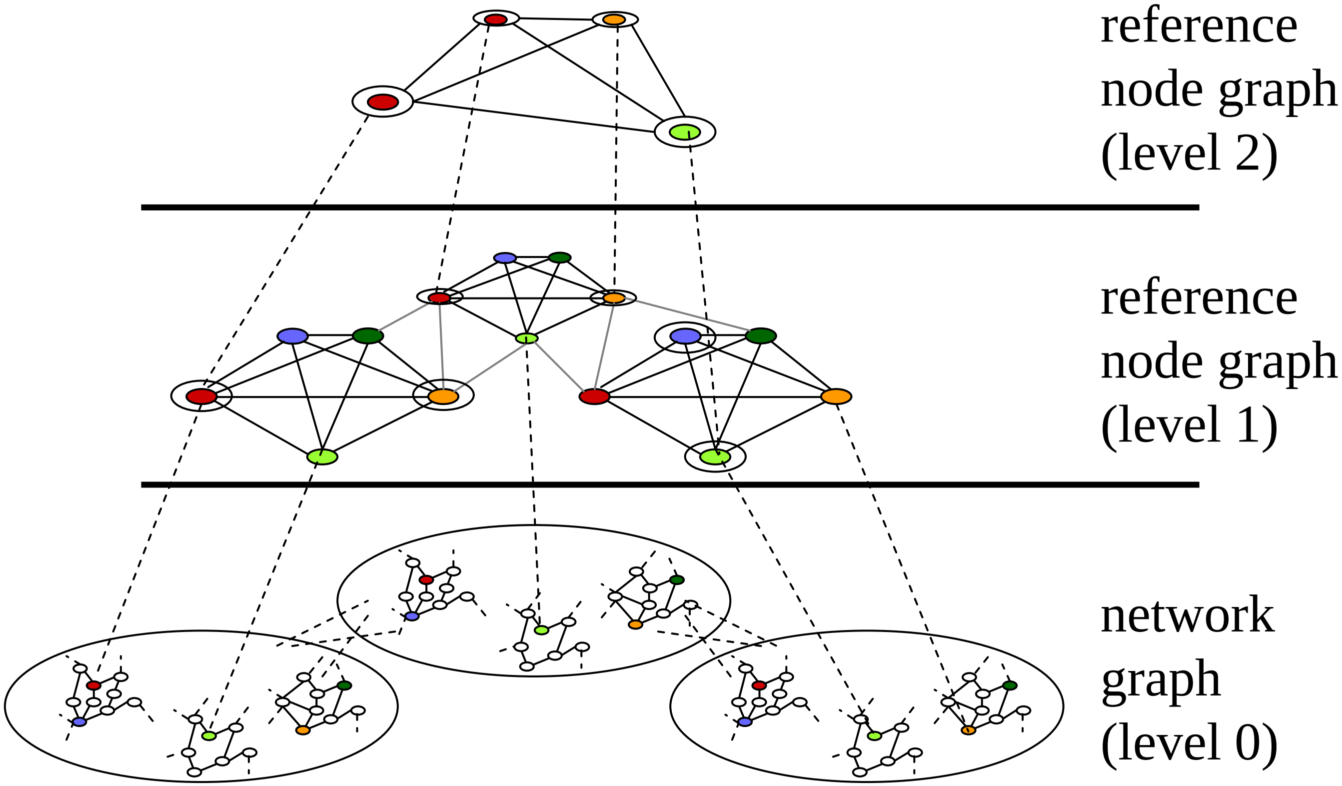
- Future Work
  - detailed evaluation of multi-level embeddings (more than 2 levels)
  - development of methods for efficient handling of updates

Thank you,

for your attention,

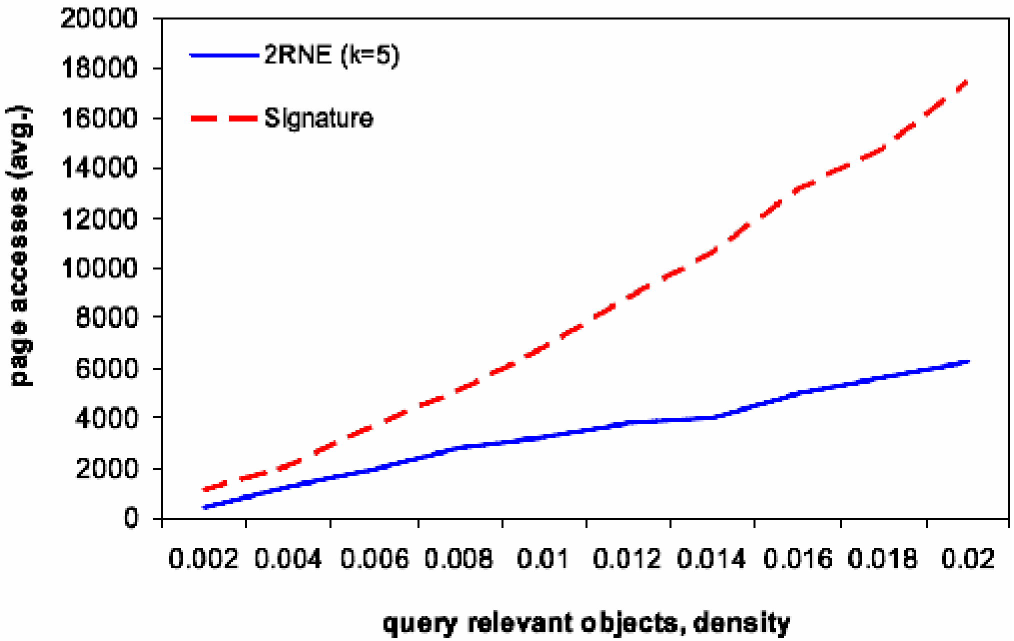
any questions?

• Multi-Level Graph Embedding



Experimental Evaluation

- Comparison to Competing Approach
  - 2-level embedding approach (DRQ) compared to Signature based approach [1]



[1] *Hu, H., Lee, D.L., Lee, V.C.S.: "Distance Indexing on Road Networks". In: Proc. Int. Conf. on Very Large Databases (VLDB'06). 2006*