CMU SCS

15-826: Multimedia Databases and Data Mining

Lecture #12: Fractals - case studies Part III
(quadtrees, knn queries)
C. Faloutsos

CMU SCS

Must-read Material

- Alberto Belussi and Christos Faloutsos, Estimating the Selectivity of Spatial Queries Using the `Correlation' Fractal Dimension Proc. of VLDB, p. 299-310, 1995

CMU SCS

Optional Material

Optional, but very useful: Manfred Schroeder *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* W.H. Freeman and Company, 1991

CMU SCS

Optional

Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing - similarity search
- Data Mining

Optional = NOT in exam
(but useful as mental drill!)

CMU SCS

Optional

Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
  - z-ordering
  - R-trees
  - misc
- fractals
  - intro
  - applications
- text

CMU SCS

Optional

Indexing - Detailed outline

- fractals
  - intro
  - applications
    - ✔ disk accesses for R-trees (range queries)
    - ✔ dimensionality reduction
    - ✔ dim. curse revisited
    - quad-tree analysis [Gaede+]
    - nn queries [Belussi+]

CMU SCS

Optional

Fractals and Quadtrees

- Problem: how many quadtree nodes will we need, to store a region in some level of approximation? [Gaede+96]

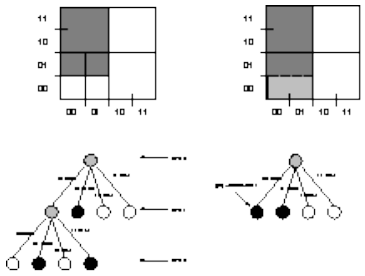
Faloutsos, C. and V. Gaede (Sept. 1996). Analysis of the z-ordering Method Using the Hausdorff Fractal Dimension. VLDB, Bombay, India.

CMU SCS

Optional

Fractals and Quadtrees

- I.e.:

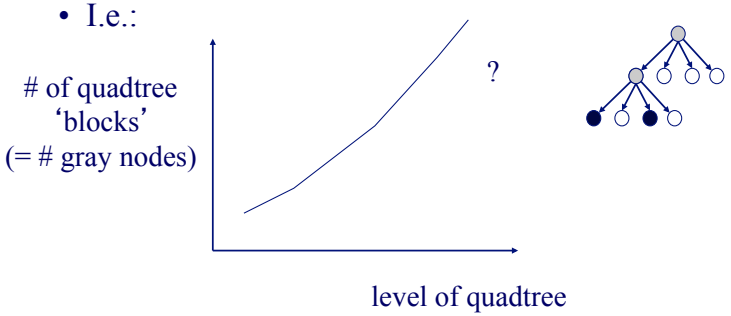


CMU SCS

Optional

Fractals and Quadtrees

- I.e.:



CMU SCS

Optional

Fractals and Quadtrees

- Datasets:

Franconia

Brain Atlas

CMU SCS

Optional

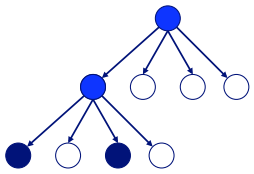
Fractals and Quadtrees

- Hint:
  - assume that the boundary is self-similar, with a given fd
  - how will the quad-tree (oct-tree) look like?

CMU SCS

Optional

Fractals and Quadtrees



white

gray

black

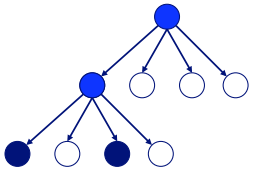
CMU SCS

Optional

Fractals and Quadtrees

Let $p_g(i)$ the prob. to find a gray node at level i.

If self-similar, what can we say for $p_g(i)$?



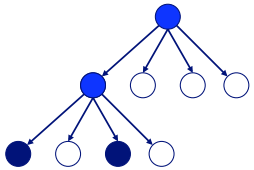
CMU SCS

Optional

Fractals and Quadtrees

Let $p_g(i)$ the prob. to find a gray node at level i.

If self-similar, what can we say for $p_g(i)$?



A: $p_g(i) = p_g$= constant

CMU SCS

Optional

Fractals and Quadtrees

Assume only 'gray' and 'white' nodes (ie., no volume')

Assume that p_g is given - how many gray nodes at level i?

CMU SCS

Optional

Fractals and Quadtrees

Assume only 'gray' and 'white' nodes (ie., no volume')

Assume that p_g is given - how many gray nodes at level i?

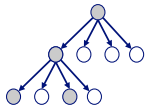
A: 1 at level 0;

$4*p_g$

$(4*p_g)* (4*p_g)$

...

$(4*p_g)^i$



...

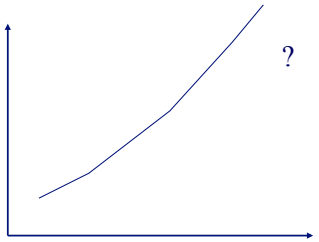
CMU SCS

Optional

Fractals and Quadtrees

- I.e.:

# of quadtree 'blocks'



? $(4*p_g)^i$

level of quadtree ('i')

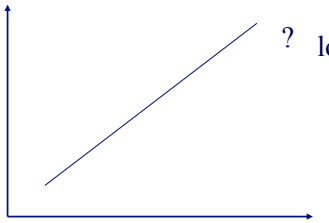
CMU SCS

Optional

Fractals and Quadtrees

- I.e.:

log(# of quadtree 'blocks')



? $\log[(4*p_g)^i]$

level of quadtree

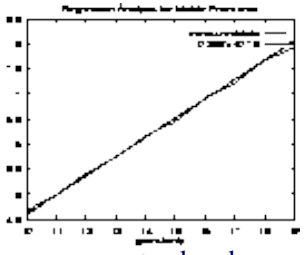
CMU SCS

Optional

Fractals and Quadtrees

- Conclusion: Self-similarity leads to easy and accurate estimation

$\log_2$(#blocks)



level

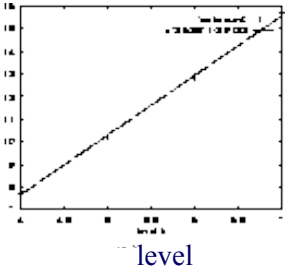
CMU SCS

Optional

Fractals and Quadtrees

- Conclusion: Self-similarity leads to easy and accurate estimation

$\log_2$(#blocks)

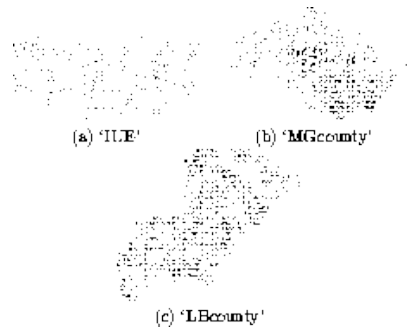


level

CMU SCS

Optional

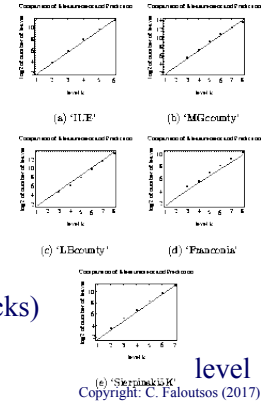
Fractals and Quadtrees



CMU SCS

Optional

Fractals and Quadtrees



log(#blocks)

level

CMU SCS

Optional

Fractals and Quadtrees

- Final observation: relationship between p_g and fractal dimension?

CMU SCS

Optional

Fractals and Quadtrees

- Final observation: relationship between p_g and fractal dimension?
- A: very close:

  $(4*p_g)^i$ = # of gray nodes at level i =

  # of Hausdorff grid-cells of side $(1/2)^i = r$

  Eventually: $D_H = 2 + \log_2(p_g)$

  and, for E-d spaces: $D_H = E + \log_2(p_g)$

CMU SCS

Optional

Fractals and Quadtrees

for E-d spaces: $D_H = E + \log_2(p_g)$

Sanity check:

- point in 2-d: $D_H = 0$ $p_g = ??$
- line in 2-d: $D_H = 1$ $p_g = ??$
- plane in 2-d: $D_H=2$ $p_g = ??$
- point in 3-d: $D_H = 0$ $p_g = ??$

CMU SCS

Optional

Fractals and Quadtrees

for E-d spaces: $D_H = E + \log_2(p_g)$

Sanity check:

- point in 2-d: $D_H = 0$ $p_g = 1/4$
- line in 2-d: $D_H = 1$ $p_g = 1/2$
- plane in 2-d: $D_H=2$ $p_g= 1$
- point in 3-d: $D_H = 0$ $p_g = 1/8$

CMU SCS

Optional

Fractals and Quadtrees

Final conclusions:

- self-similarity leads to estimates for # of z-values = # of quadtree/oct-tree blocks
- close dependence on the Hausdorff fractal dimension of the boundary

CMU SCS

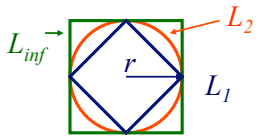
Indexing - Detailed outline

- fractals
  - intro
  - applications
    - ✔ disk accesses for R-trees (range queries)
    - ✔ dimensionality reduction
    - ✔ dim. curse revisited
    - ✔ quad-tree analysis [Gaede+]
    - nn queries [Belussi+]

CMU SCS

NN queries

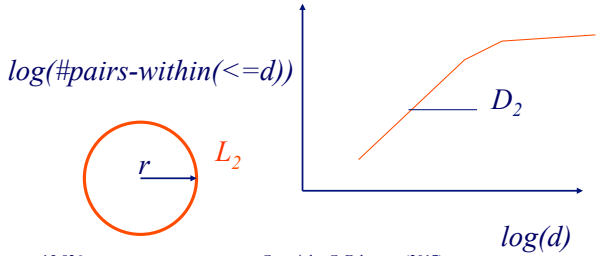
- Q: in NN queries, what is the effect of the shape of the query region? [Belussi+95]



CMU SCS

NN queries

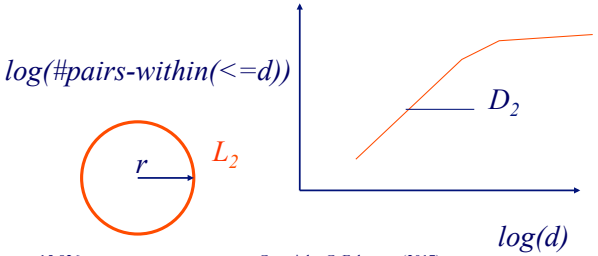
- Q: in NN queries, what is the effect of the shape of the query region?
- that is, for L2, and self-similar data:



CMU SCS

NN queries

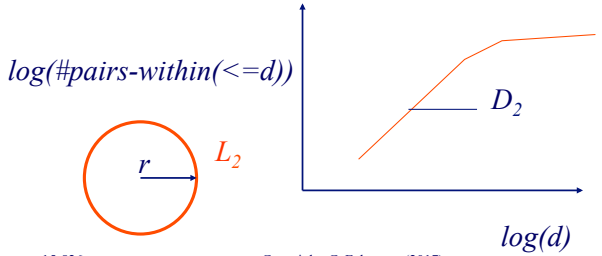
- Q: What about L_1, L_{inf}?



CMU SCS

NN queries

- Q: What about L_1, L_{inf}?
- A: **Same slope**, different intercept

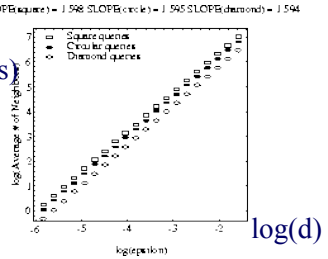


CMU SCS

NN queries

- Q: What about L_1, L_{inf}?
- A: **Same slope**, different intercept

log(#neighbors)



log(d)

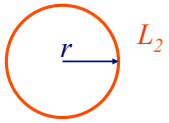
CMU SCS

Optional

NN queries

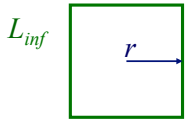
- Q: what about the intercept? Ie., what can we say about N_2 and N_{inf}

N_2 neighbors



volume: V_2

N_{inf} neighbors



volume: V_{inf}

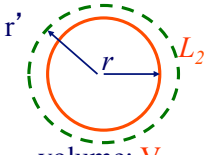
CMU SCS

Optional

NN queries

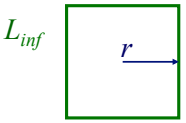
- Consider sphere with volume V_{inf} and r' radius

N_2 neighbors



volume: V_2

N_{inf} neighbors



volume: V_{inf}

CMU SCS

Optional

NN queries

- Consider sphere with volume V_{inf} and r' radius
- $(r/r')^E = V_2 / V_{inf}$
- $(r/r')^{D_2} = N_2 / N_2'$
- $N_2' = N_{inf}$ (since shape does not matter)
- and finally:

CMU SCS

Optional

NN queries

$(N_2 / N_{inf}) ^ 1/D_2 = (V_2 / V_{inf}) ^ 1/E$

CMU SCS

NN queries

Conclusions: for self-similar datasets

- Avg # neighbors: grows like *(distance)^D_2*, regardless of query shape (circle, diamond, square, e.t.c.)

CMU SCS

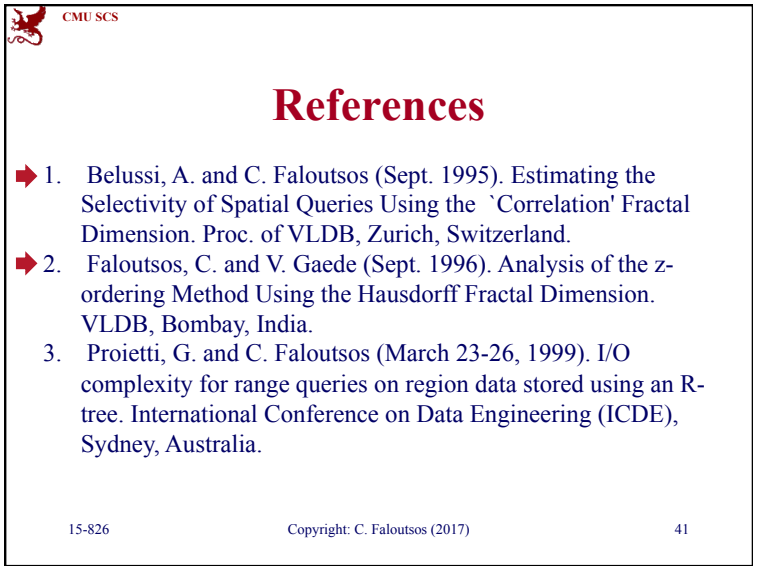
Indexing - Detailed outline

- fractals
  - intro
  - applications
    - disk accesses for R-trees (range queries)
    - dimensionality reduction
    - dim. curse revisited
    - quad-tree analysis [Gaede+]
    - nn queries [Belussi+]
  - Conclusions

CMU SCS

Fractals - overall conclusions

- self-similar datasets: appear often
- powerful tools: correlation integral, NCDF, rank-frequency plot
- intrinsic/fractal dimension helps in
  - estimations (selectivities, quadtrees, etc)
  - dim. reduction / dim. curse
- (later: can help in image compression...)


CMU SCS
References
1. Belussi, A. and C. Faloutsos (Sept. 1995). Estimating the Selectivity of Spatial Queries Using the `Correlation' Fractal Dimension. Proc. of VLDB, Zurich, Switzerland.
2. Faloutsos, C. and V. Gaede (Sept. 1996). Analysis of the z-ordering Method Using the Hausdorff Fractal Dimension. VLDB, Bombay, India.
3. Proietti, G. and C. Faloutsos (March 23-26, 1999). I/O complexity for range queries on region data stored using an R-tree. International Conference on Data Engineering (ICDE), Sydney, Australia.
15-826
Copyright: C. Faloutsos (2017)
41