


**15-826: Multimedia Databases
and Data Mining**


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



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


15-826 Copyright: C. Faloutsos (2011) 2



Optional Material

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

15-826 Copyright: C. Faloutsos (2011) 3



CMU SCS


Optional

Outline

Goal: 'Find **similar** / **interesting** things'

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

15-826 Copyright: C. Faloutsos (2011) 4




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

15-826 Copyright: C. Faloutsos (2011) 5




CMU SCS

Optional

Indexing - Detailed outline

- fractals
 - intro
 - applications
 - disk accesses for R-trees (range queries)
 - dimensionality reduction
 - selectivity in M-trees
 - dim. curse revisited
- ➔ • "fat fractals"
- quad-tree analysis [Gaede+]
- nn queries [Belussi+]

15-826 Copyright: C. Faloutsos (2011) 6




CMU SCS

'Fat' fractals & R-tree performance on region data

- Problem [Proietti+, '99]
- Given
 - N (# of data regions)
- estimate how many of them will qualify for the average range query ($q_1 \times q_2 \times \dots \times q_E$)

Of course, we need more info
Q: what?

15-826 Copyright: C. Faloutsos (2011) 7




CMU SCS

R-tree performance on region data

A: the distributions of their sizes

Q: do we also need some info about the locations?

15-826 Copyright: C. Faloutsos (2011) 8



CMU SCS

R-tree performance on region data

A: the distributions of their sizes

Q: do we also need some info about the locations?

A: no (not for range queries)

15-826 Copyright: C. Faloutsos (2011) 9

CMU SCS

R-tree performance on region data Optional

A: the distributions of their sizes

Q: what exactly would we need?

15-826 Copyright: C. Faloutsos (2011) 10

CMU SCS

R-tree performance on region data Optional

A: the distributions of their sizes


Q: what exactly would we need?

A: for self-similar regions (\sim 'fat' fractals),
we just need the slope of the Korcak law!
(and the total area) [Proietti+]

15-826 Copyright: C. Faloutsos (2011) 11


CMU SCS

More power laws: areas – Korcak's law Optional



Scandinavian lakes
Any pattern?

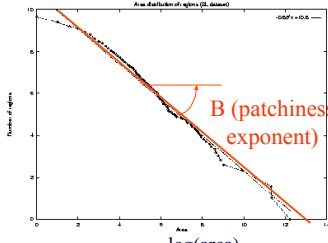
15-826 Copyright: C. Faloutsos (2011) 12

CMU SCS

Optional

More power laws: areas – Korcak’s law

log(count(\geq area))




Scandinavian lakes
area vs
complementary
cumulative count
(log-log axes)

15-826

Copyright: C. Faloutsos (2011)

13

CMU SCS

Optional


R-tree performance on regions

- Once we know ‘B’ (and the total area)
- we can second-guess the individual sizes
- and then apply the [Pagel+93] formula
- Bottom line:

15-826

Copyright: C. Faloutsos (2011)

14

CMU SCS

Optional


R-tree performance on regions

Dataset	N	A	B
LAKES	816	75,910	0.85
ISLANDS	470	136,893	0.60
REGIONS	757	190,526	0.70

15-826

Copyright: C. Faloutsos (2011)

15

CMU SCS

Optional

R-tree performance on regions

LAKES Dataset

ISLANDS Dataset

REGIONS Dataset


sel. error

query side

15-826

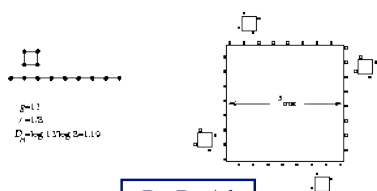
Copyright: C. Faloutsos (2011)

16

CMU SCS

Optional

'Fat' fractals - observation



$B = D_H / d$


B: patchiness exp; d: embedding dim

D_H : Hausdorff of periphery

15-826

Copyright: C. Faloutsos (2011)

17

CMU SCS

Optional

'Fat' fractals - observation

Dataset	D_H	B	$D_H - 2B$
LAKES	1.78	0.85	0.08
ISLANDS	1.23	0.60	0.03
REGIONS	1.48	0.70	0.08
Aegean Island	1.08	0.52	0.04
Japan archipelago	1.19	0.59	0.01
Italy plains	1.32	0.63	0.06
Whole Earth	1.2	0.6	0
Cyprens vegetation	0.62	1.23	0.01

15-826

Copyright: C. Faloutsos (2011)

18

CMU SCS

‘Fat’ fractals

Optional

- intuition behind $B = D_H / d$?


15-826 Copyright: C. Faloutsos (2011) 19

CMU SCS

‘Fat’ fractals

Optional

- intuition behind $B = D_H / d$?
- A: consider ‘flooding’:



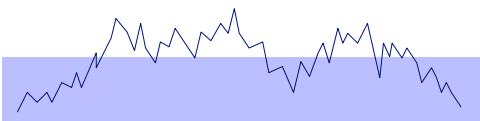
15-826 Copyright: C. Faloutsos (2011) 20

CMU SCS


‘Fat’ fractals

Optional

- intuition behind $B = D_H / d$?
- A: consider ‘flooding’:



15-826 Copyright: C. Faloutsos (2011) 21




CMU SCS

Conclusions

Optional

- ‘Fat’ fractals model regions well
- patchiness exp.: $B = D_H / d$
- can help us estimate selectivities

15-826 Copyright: C. Faloutsos (2011) 22




CMU SCS

Indexing - Detailed outline

Optional

- fractals
 - intro
 - applications
 - disk accesses for R-trees (range queries)
 - dimensionality reduction
 - selectivity in M-trees
 - dim. curse revisited
 - “fat fractals”
- ➡ • quad-tree analysis [Gaede+]
- nn queries [Belussi+]

15-826 Copyright: C. Faloutsos (2011) 23



CMU SCS

Fractals and Quadtrees

Optional

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

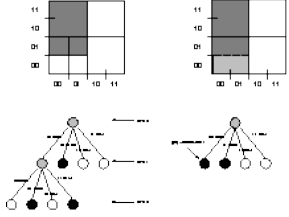
15-826 Copyright: C. Faloutsos (2011) 24

CMU SCS

Optional

Fractals and Quadrees

- I.e.:



15-826 Copyright: C. Faloutsos (2011) 25

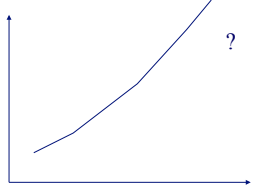
CMU SCS

Optional

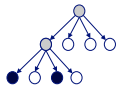
Fractals and Quadrees

- I.e.:

of quadtree 'blocks' (= # gray nodes)



level of quadtree




15-826 Copyright: C. Faloutsos (2011) 26

CMU SCS

Optional

Fractals and Quadrees

- Datasets:



Franconia Brain Atlas


15-826 Copyright: C. Faloutsos (2011) 27

CMU SCS

Fractals and Quadrees

Optional

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

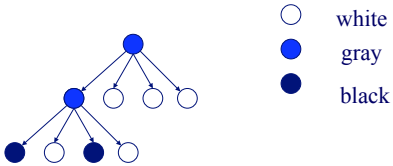


15-826 Copyright: C. Faloutsos (2011) 28

CMU SCS

Fractals and Quadrees

Optional



○ white
● gray
● black

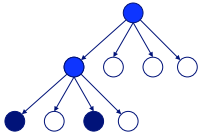
15-826 Copyright: C. Faloutsos (2011) 29

CMU SCS


Fractals and Quadrees

Optional

Let $p_g(i)$ the prob. to find a gray node at level i .
If self-similar, what can we say for $p_g(i)$?



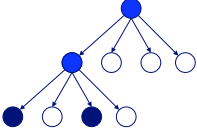
15-826 Copyright: C. Faloutsos (2011) 30

CMU SCS

Optional

Fractals and Quadrees

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: $p_g(i) = p_g = \text{constant}$

15-826

Copyright: C. Faloutsos (2011)


31

CMU SCS

Optional

Fractals and Quadrees


Assume only ‘gray’ and ‘white’ nodes (ie., no volume’)
Assume that p_g is given - how many gray nodes at level i ?



15-826

Copyright: C. Faloutsos (2011)


32

CMU SCS

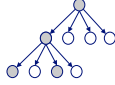
Optional

Fractals and Quadrees

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;
 $4 * p_g$
 $(4 * p_g) * (4 * p_g)$
...
 $(4 * p_g)^i$



15-826

Copyright: C. Faloutsos (2011)

33

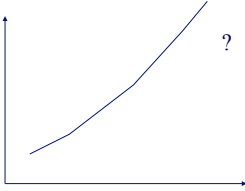
CMU SCS

Optional

Fractals and Quadrees

- I.e.:

of quadtree
'blocks'




level of quadtree ('i')

15-826

Copyright: C. Faloutsos (2011)

34

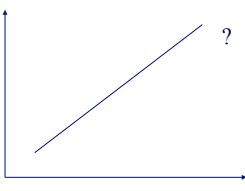
CMU SCS

Optional

Fractals and Quadrees

- I.e.:

log(# of quadtree
'blocks')




level of quadtree

15-826

Copyright: C. Faloutsos (2011)

35

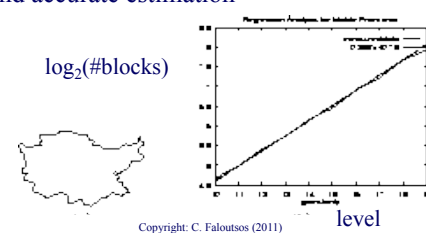
CMU SCS

Optional

Fractals and Quadrees

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

log₂(#blocks)



level

15-826

Copyright: C. Faloutsos (2011)

36


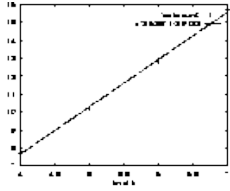
CMU SCS

Fractals and Quadrees

Optional

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

$\log_2(\# \text{blocks})$

level

15-826 Copyright: C. Faloutsos (2011) 37

CMU SCS

Fractals and Quadrees

Optional

(a) 'ILF'

(b) 'MGCommunity'

(c) 'LBCommunity'

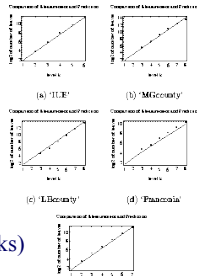
15-826 Copyright: C. Faloutsos (2011) 38

CMU SCS

Fractals and Quadrees


Optional

log(#blocks)



level

15-826 Copyright: C. Faloutsos (2011) 39


CMU SCS

Optional

Fractals and Quadrees

- Final observation: relationship between p_g and fractal dimension?

15-826Copyright: C. Faloutsos (2011)40


CMU SCS

Optional

Fractals and Quadrees

- Final observation: relationship between p_g and fractal dimension?
- A: very close:
 $(4 \cdot p_g)^i = \#$ 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: $D_H = E + \log_2(p_g)$

15-826Copyright: C. Faloutsos (2011)41

CMU SCS

Optional


Fractals and Quadrees

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: $D_H = 0$ $p_g = ??$

15-826Copyright: C. Faloutsos (2011)42



CMU SCS

Fractals and Quadtrees


Optional

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: $D_H = 0$ $p_g = 1/8$

15-826 Copyright: C. Faloutsos (2011) 43



CMU SCS


Fractals and Quadtrees

Optional

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

15-826 Copyright: C. Faloutsos (2011) 44



CMU SCS

Indexing - Detailed outline

Optional

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

15-826 Copyright: C. Faloutsos (2011) 45

CMU SCS

NN queries

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

15-826 Copyright: C. Faloutsos (2011) 46

CMU SCS

NN queries

• Q: in NN queries, what is the effect of the shape of the query region?

• that is, for L_2 , and self-similar data:

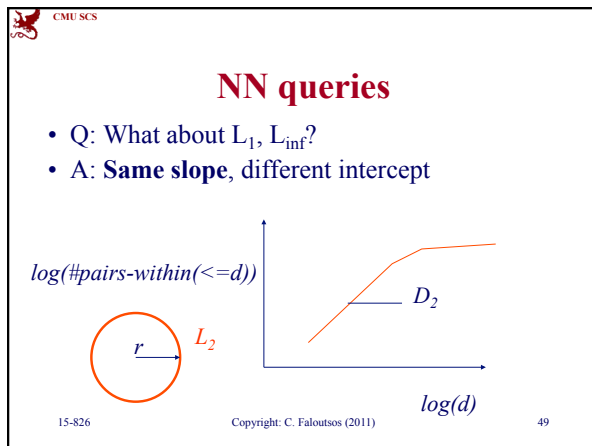
15-826 Copyright: C. Faloutsos (2011) 47

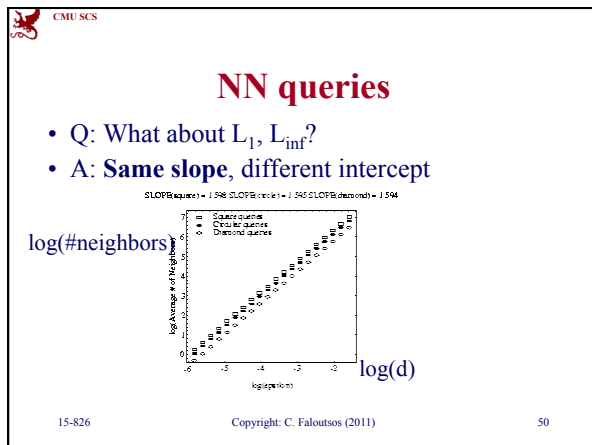
CMU SCS

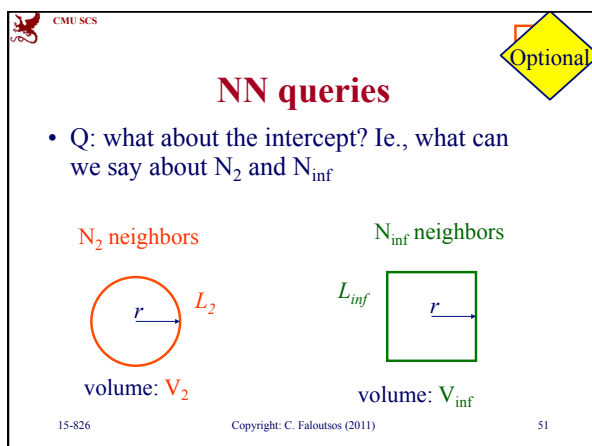
NN queries

• Q: What about L_1 , L_{inf} ?

15-826 Copyright: C. Faloutsos (2011) 48







CMU SCS

NN queries

Optional

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

N_2 neighbors

volume: V_2

N_{inf} neighbors

volume: V_{inf}

15-826 Copyright: C. Faloutsos (2011) 52

CMU SCS

NN queries

Optional

- 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:

15-826 Copyright: C. Faloutsos (2011) 53


CMU SCS

NN queries

Optional

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

15-826 Copyright: C. Faloutsos (2011) 54



CMU SCS


Optional

NN queries

Conclusions: for self-similar datasets

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

15-826 Copyright: C. Faloutsos (2011) 55




CMU SCS

Optional

Indexing - Detailed outline

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

15-826 Copyright: C. Faloutsos (2011) 56




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...)

15-826 Copyright: C. Faloutsos (2011) 57

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 (2011)

58
