

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

Lecture #30: Conclusions
C. Faloutsos



Outline

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


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

15-826 (c) 2010, C. Faloutsos 2



Indexing - similarity search


15-826 (c) 2010, C. Faloutsos 3

 CMU SCS

Indexing - similarity search


- R-trees
- z-ordering / hilbert curves
- M-trees
- beware of high intrinsic dimensionality

15-826 (c) 2010, C. Faloutsos 4

 CMU SCS

Multimedia indexing

15-826 (c) 2010, C. Faloutsos 5

 CMU SCS

Multimedia indexing

- Feature extraction for indexing
- MDS/FastMap

15-826 (c) 2010, C. Faloutsos 6

CMU SCS

Time series + forecasting

15-826 (c) 2010, C. Faloutsos 7

CMU SCS

Time series + forecasting

- Fourier; Wavelets
- Box/Jenkins and AutoRegression
- non-linear/chaotic forecasting (fractals again)

15-826 (c) 2010, C. Faloutsos 8

CMU SCS

Data Mining - DB

15-826 (c) 2010, C. Faloutsos 13

CMU SCS

Data Mining - DB

- Association Rules
- OLAP
- Classifiers [NOT in the final exam]

15-826 (c) 2010, C. Faloutsos 14

CMU SCS

Taking a step back:

We saw some fundamental, recurring concepts and tools:

15-826 (c) 2010, C. Faloutsos 15

CMU SCS

Powerful, recurring tools

- SVD (optimal L2 approx)
 - LSI, KL, PCA

15-826 (c) 2010, C. Faloutsos 16

CMU SCS

Powerful, recurring tools

- Discrete Fourier Transform
- Wavelets

15-826 (c) 2010, C. Faloutsos 17

CMU SCS

Powerful, recurring tools

- Fractals/ self similarity
 - Zipf, Korcak, Pareto’s laws
 - intrinsic dimension (Sierpinski triangle)
 - correlation integral
 - Barnsley’s IFS compression
 - Kronecker graphs


15-826 (c) 2010, C. Faloutsos 18

CMU SCS

Powerful, recurring tools

- Matrix inversion lemma
 - Recursive Least Squares
 - Sherman-Morrison(-Woodbury)


15-826 (c) 2010, C. Faloutsos 19



Among them

- **fractals / power laws** probably lead to the most startling discoveries ('the mean may be meaningless')
- **SVD**: behind PageRank/HITS/tensors
- **Wavelets**: Nature seems to prefer them
- **RLS**: seems to achieve the impossible

15-826 (c) 2010, C. Faloutsos 20



Thank you!

- feel free to contact me:
 - `christos@cs` GHC 8019
- Reminder: faculty course eval's:
 - www.cmu.edu/fce/index.shtml
- Final: **Tue. May 11 1:00p.m.-4:00p.m. @GHC 4307** - check
 - <http://www.cmu.edu/hub/current-finals.pdf>
 - Demos: Fri April 30 '10, 12-5pm, @GHC6115
- Have a great summer!

15-826 (c) 2010, C. Faloutsos 21
