CMU SCS

15-826: Multimedia Databases and Data Mining

Lecture #30: Conclusions

C. Faloutsos

CMU SCS

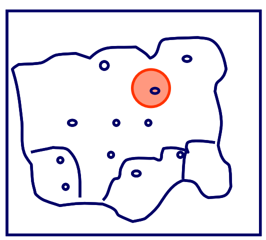
Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing - similarity search
  - Points
  - Text
  - Time sequences; images etc
  - Graphs

CMU SCS

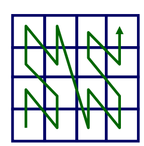
Indexing - similarity search

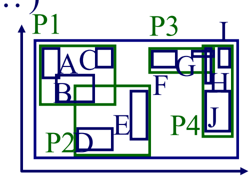


CMU SCS

Indexing - similarity search

- R-trees
- z-ordering / hilbert curves
- M-trees
- (DON'T FORGET ...)

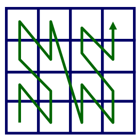


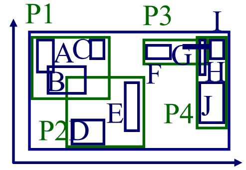


CMU SCS

Indexing - similarity search

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





15-826 (c) 2017, C. Faloutsos 5

CMU SCS

Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing - similarity search
  - Points
  - Text
  - Time sequences; images etc
  - Graphs

15-826 (c) 2017, C. Faloutsos 6

CMU SCS

Text searching

- 'find all documents with word *bla*'

15-826 (c) 2017, C. Faloutsos 7

CMU SCS

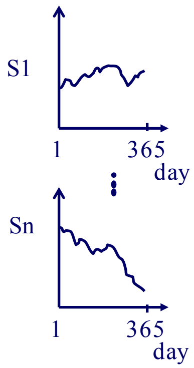
Text searching

- Full text scanning ('grep')
- Inversion (B-tree or hash index)
- signature files – Bloom filters
- Vector space model
  - Ranked output
  - Relevance feedback
- String editing distance (-> dynamic prog.)

15-826 (c) 2017, C. Faloutsos 8

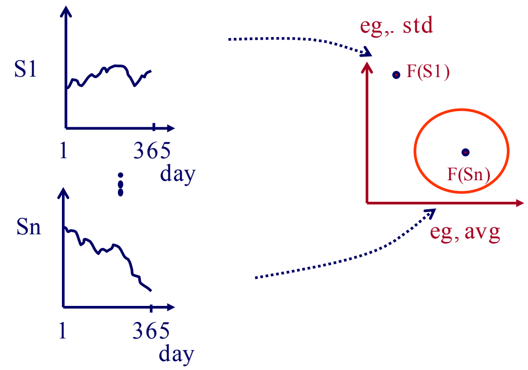
CMU SCS

Multimedia indexing



CMU SCS

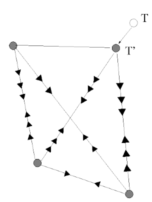
'GEMINI' - Pictorially



CMU SCS

Multimedia indexing

- Feature extraction for indexing (GEMINI)
  - Lower-bounding lemma, to guarantee no false dismissals
- MDS/FastMap



CMU SCS

Outline

Goal: 'Find similar / interesting things'

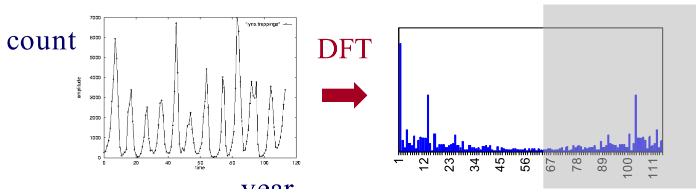
- Intro to DB
- Indexing - similarity search
  - Points
  - Text
  - Time sequences; images etc
  - Graphs

CMU SCS

Time series & forecasting

Goal: given a signal (eg., sales over time and/or space)

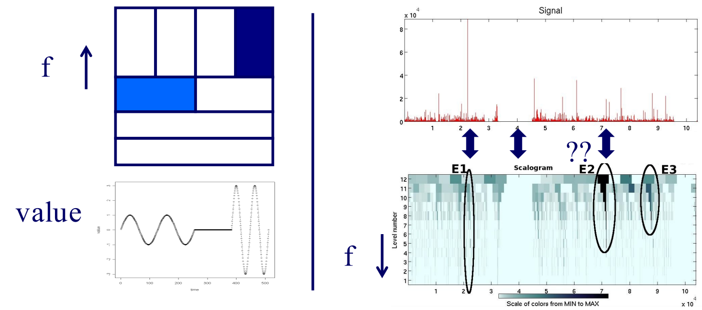
Find: patterns and/or compress



CMU SCS

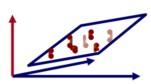
Wavelets

- Q: baritone/silence/soprano - DWT?

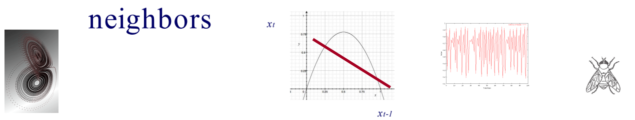


CMU SCS

Time series + forecasting



- Fourier; Wavelets
- Box/Jenkins and AutoRegression
- non-linear/chaotic forecasting (fractals again)
  - 'Delayed Coordinate Embedding' ~ nearest neighbors



CMU SCS

Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing - similarity search
  - Points
  - Text
  - Time sequences; images etc
  - Graphs

CMU SCS

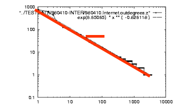
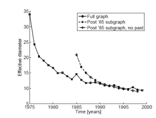
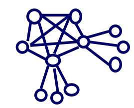
Graphs

- Real graphs: surprising patterns
  - ??

CMU SCS

Graphs

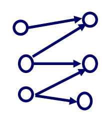
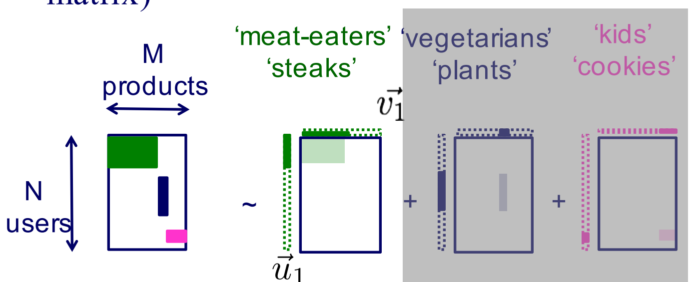
- Real graphs: surprising patterns
  - '**six degrees**'
  - **Skewed** degree distribution ('rich get richer')
  - Super-linearities (2x nodes -> 3x edges)
  - Diameter: **shrinks** (!)
  - Might have **no** good cuts

CMU SCS

Graphs - SVD

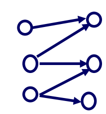
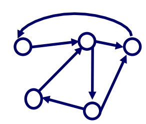
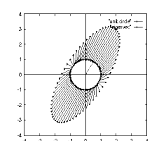
- Hubs/Authorities (SVD on adjacency matrix)

CMU SCS

Graphs - PageRank

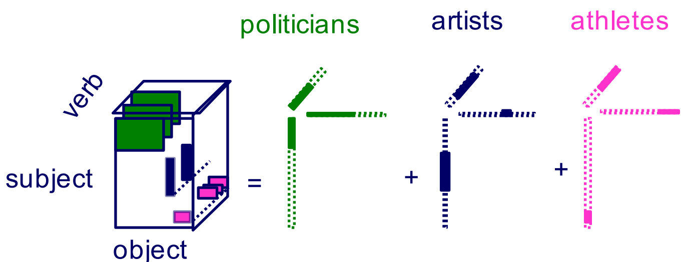
- Hubs/Authorities (SVD on adjacency matrix)
- PageRank (fixed point -> eigenvector)

CMU SCS

Tensors

- Eg., time evolving graphs; Subject-verb-object triplets; etc



15-826 (c) 2017, C.Faloutsos 21

CMU SCS

Taking a step back:

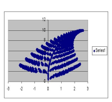
We saw some fundamental, recurring concepts and tools:

15-826 (c) 2017, C.Faloutsos 25

CMU SCS

T1: Powerful, recurring tools

- Fractals/ self similarity

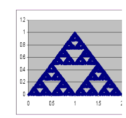


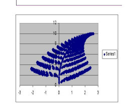
15-826 (c) 2017, C.Faloutsos 26

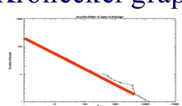
CMU SCS

T1: Powerful, recurring tools

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







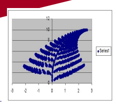
15-826 (c) 2017, C.Faloutsos 27

CMU SCS

T1: Powerful, recurring tools

- Fractals/ self similarity
  - Zipf, K...
  - ...ession
  - ...er graphs)

- 'Take logarithms'
- mean-> meaningless
- Gaussian trap

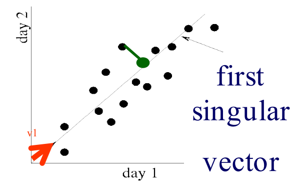


15-826 (c) 2017, C.Faloutsos 28

CMU SCS

T2: Powerful, recurring tools

- SVD (optimal L2 approx)

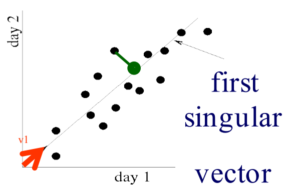


15-826 (c) 2017, C.Faloutsos 29

CMU SCS

T2: Powerful, recurring tools

- SVD (optimal L2 approx)
  - LSI, KL, PCA, 'eigenSpokes', (& in ICA)
  - HITS (PageRank)

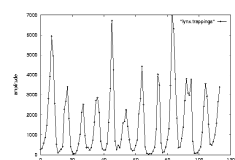


15-826 (c) 2017, C.Faloutsos 30

CMU SCS

T3: Powerful, recurring tools

- Discrete Fourier Transform
- Wavelets

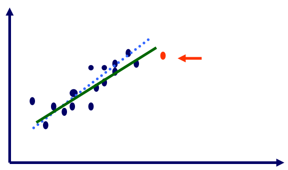


15-826 (c) 2017, C.Faloutsos 31

CMU SCS

T4: Powerful, recurring tools

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



15-826 (c) 2017, C. Faloutsos 32

CMU SCS

Summary

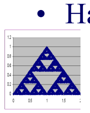
- **T1: fractals / power laws** lead to startling discoveries
  - 'the mean may be meaningless'
  - Don't assume Gaussian (average, k-means, etc)
- **T2: SVD**: behind PageRank/HITS/tensors/…
- **T3: Wavelets**: Nature seems to prefer them
- **T4: RLS**: matrix inversion, without inverting

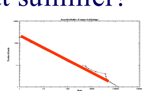
15-826 (c) 2017, C. Faloutsos 33

CMU SCS

Thank you!

- Feel free to contact me:
  - Cell#; christos@cs; GHC 8019
- Reminder: faculty course eval's:
  - www.cmu.edu/hub/fce/
- Final: as announced in Hub
  - Mo 5/8/2017, 8:30-11:30am, DH 2315
  - (**double-check** with www.cmu.edu/hub/docs/final-exams.pdf)
- Have a great summer!





- 'Take logarithms'
- mean -> meaningless
- Gaussian trap

15-826 (c) 2017, C. Faloutsos 35