

CarnegieMellon

# 15-826: Multimedia Databases and Data Mining

Lecture #22: DSP tools –  
Wavelets

*C. Faloutsos*


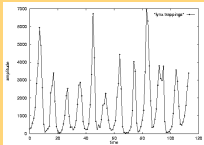
1

CarnegieMellon

## Problem


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

Q: Find patterns and/or compress

count


year



15-826 Copyright (c) 2019 C. Faloutsos 2

2


CarnegieMellon



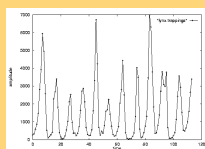
## Solutions:

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

Q: Find patterns and/or compress

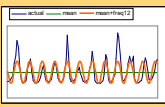


count

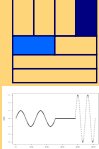


year

A1: Fourier (DFT)



A2: Wavelets (DWT)



15-826 Copyright (c) 2019 C. Faloutsos

3

CarnegieMellon

## Must-read Material

- Wavelets: In [PTVF](#) ch. 13.10; in [MM Textbook](#) Appendix C

15-826 Copyright (c) 2019 C. Faloutsos

4

4

CarnegieMellon

## Outline

Goal: 'Find similar / interesting things'

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

15-826 Copyright (c) 2019 C. Faloutsos 5

5

CarnegieMellon

## Indexing - Detailed outline

- primary key indexing
- ..
- Multimedia –
  - Digital Signal Processing (DSP) tools
    - Discrete Fourier Transform (DFT)
    - ➡ • Discrete Wavelet Transform (DWT)

15-826 Copyright (c) 2019 C. Faloutsos 6


6

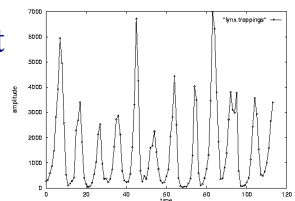
CarnegieMellon

## Reminder: Problem:

Goal: given a signal (eg., #packets over time)  
 Find: patterns, periodicities, and/or compress

count






lynx caught per year  
 (packets per day;  
 virus infections per month)

year

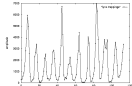
15-826 Copyright (c) 2019 C. Faloutsos 7

7

CarnegieMellon



## Important observations




Patterns, rules, forecasting and similarity indexing are closely related:

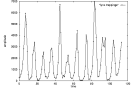
15-826 Copyright (c) 2019 C. Faloutsos 8

8

CarnegieMellon


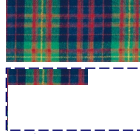



## Important observations



Patterns, rules, forecasting and similarity indexing are closely related:


- To do **forecasting**, we need
  - to find **patterns**/rules
  - compress
  - to find **similar** settings in the past
- to find outliers, we need to have forecasts
  - (outlier = too far away from our forecast)

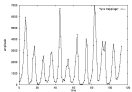
15-826 Copyright (c) 2019 C. Faloutsos 9

9


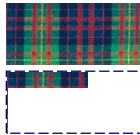

CarnegieMellon



## Important observations



Patterns, rules, forecasting and similarity indexing are closely related:

15-826 Copyright (c) 2019 C. Faloutsos 10

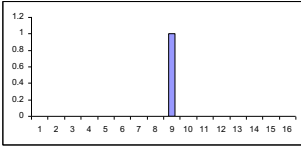
10

CarnegieMellon

## Wavelets - DWT

- DFT is great - but, how about compressing a spike?

value



time

15-826 Copyright (c) 2019 C. Faloutsos 11

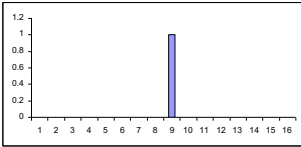
11

CarnegieMellon

## Wavelets - DWT

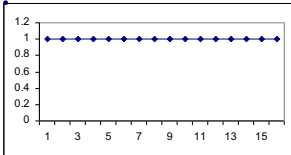
- DFT is great - but, how about compressing a spike?
- A: Terrible - all DFT coefficients needed!

value



time

Ampl.



Freq.

15-826 Copyright (c) 2019 C. Faloutsos 12

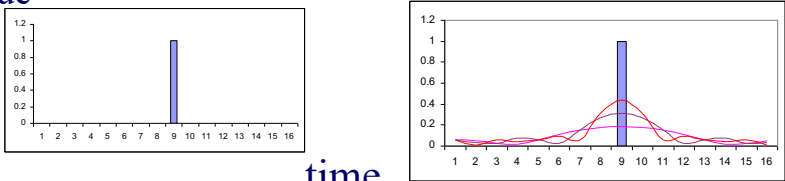
12

CarnegieMellon

## Wavelets - DWT

- DFT is great - but, how about compressing a spike?
- A: Terrible - all DFT coefficients needed!

value



time

15-826 Copyright (c) 2019 C. Faloutsos 13

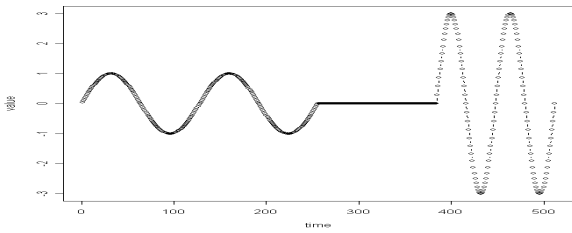
13

CarnegieMellon

## Wavelets - DWT

- Similarly, DFT suffers on short-duration waves (eg., baritone, silence, soprano)

value



time

15-826 Copyright (c) 2019 C. Faloutsos 14

14

CarnegieMellon

## Wavelets - DWT

- Solution#1: Short window Fourier transform (SWFT)
- But: how short should be the window?

15-826 Copyright (c) 2019 C. Faloutsos

15

CarnegieMellon

## Wavelets - DWT

- Answer: **multiple** window sizes! -> DWT

Time domain

freq

15-826 Copyright (c) 2019 C. Faloutsos 16

16



CarnegieMellon

## Haar Wavelets

- subtract sum of left half from right half
- repeat recursively for quarters, eight-ths, ...

15-826
Copyright (c) 2019 C. Faloutsos
17

17

CarnegieMellon

## Wavelets - construction

$x_0$   $x_1$   $x_2$   $x_3$   $x_4$   $x_5$   $x_6$   $x_7$

15-826
Copyright (c) 2019 C. Faloutsos
18

18

CarnegieMellon

## Wavelets - construction

level 1  $d_{1,0}$   $s_{1,0}$   $d_{1,1}$   $s_{1,1}$  .....

$x_0$   $x_1$   $x_2$   $x_3$   $x_4$   $x_5$   $x_6$   $x_7$

15-826 Copyright (c) 2019 C. Faloutsos 19

19

CarnegieMellon

## Wavelets - construction

level 2  $d_{2,0}$   $s_{2,0}$   $d_{1,0}$   $s_{1,0}$   $d_{1,1}$   $s_{1,1}$  .....

$x_0$   $x_1$   $x_2$   $x_3$   $x_4$   $x_5$   $x_6$   $x_7$

15-826 Copyright (c) 2019 C. Faloutsos 20

20

CarnegieMellon

## Wavelets - construction

etc ...

15-826 Copyright (c) 2019 C. Faloutsos 21

21

CarnegieMellon

## Wavelets - construction

Q: map each coefficient  
on the time-freq. plane

15-826 Copyright (c) 2019 C. Faloutsos 22

22

CarnegieMellon

## Wavelets - construction

Q: map each coefficient  
on the time-freq. plane

15-826 Copyright (c) 2019 C. Faloutsos 23

23

CarnegieMellon

## Wavelets - construction

Q: map each coefficient  
on the time-freq. plane

15-826 Copyright (c) 2019 C. Faloutsos 24

24

CarnegieMellon

## Haar wavelets - code

```
#!/usr/bin/perl5
# expects a file with numbers
# and prints the dwt transform
# The number of time-ticks should be a power of 2
# USAGE
# haar.pl <fname>

my @vals=();
my @smooth; # the smooth component of the signal
my @diff; # the high-freq. component

# collect the values into the array @val
while(<>){
    @vals = ( @vals , split );
}

my $len = scalar(@vals);
my $half = int($len/2);
while($half >= 1 ){
    for(my $i=0; $i<$half; $i++){
        $diff [$i] = ($vals[2*$i] - $vals[2*$i + 1]) / sqrt(2);
        print "\t", $diff[$i];
        $smooth [$i] = ($vals[2*$i] + $vals[2*$i + 1]) / sqrt(2);
    }
    print "\n";
    @vals = @smooth;
    $half = int($half/2);
}
print "\t", $vals[0], "\n"; # the final, smooth component
```

Also at: [www.cs.cmu.edu/~christos/SRC/DWT-Haar-all.tar](http://www.cs.cmu.edu/~christos/SRC/DWT-Haar-all.tar)

15-826 Copyright (c) 2019 C. Faloutsos 25

25

CarnegieMellon

## Wavelets - construction

Observation1:

- ‘+’ can be some weighted addition
- ‘-’ is the corresponding weighted difference  
( ‘Quadrature mirror filters’ )

Observation2: unlike DFT/DCT,

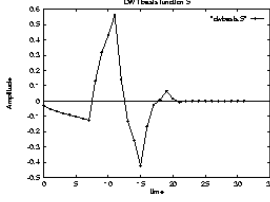
there are *many* wavelet bases: Haar, Daubechies-4, Daubechies-6, Coifman, Morlet, Gabor, ...

15-826 Copyright (c) 2019 C. Faloutsos 26

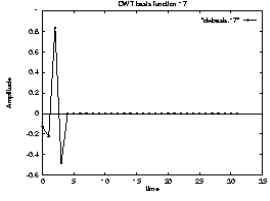
26

**CarnegieMellon**

## Wavelets - how do they look like?



DMT basis function 5  
"daubech\_5"



DMT basis function 7  
"daubech\_7"

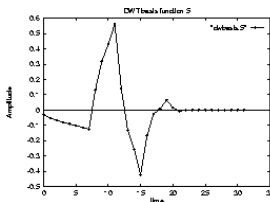
- E.g., Daubechies-4

15-826 Copyright (c) 2019 C. Faloutsos 27

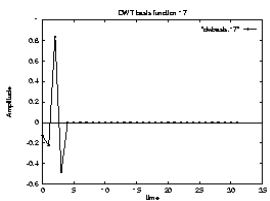
27

**CarnegieMellon**

## Wavelets - how do they look like?

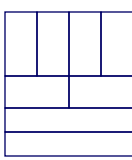


DMT basis function 5  
"daubech\_5"



DMT basis function 7  
"daubech\_7"

- E.g., Daubechies-4

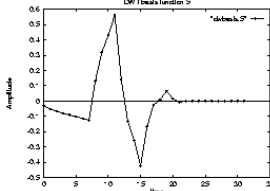


15-826 Copyright (c) 2019 C. Faloutsos 28

28

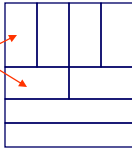
**CarnegieMellon**

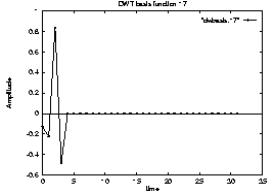
## Wavelets - how do they look like?



DWT basis function 5  
Amplitude vs. time

- E.g., Daubechies-4





DWT basis function 7  
Amplitude vs. time

15-826
Copyright (c) 2019 C. Faloutsos
29

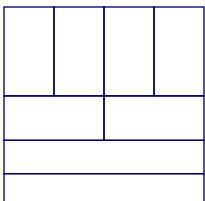
29

**CarnegieMellon**

## Wavelets - Drill#1:

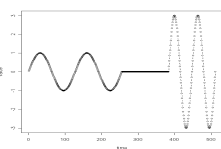
- Q: baritone/silence/soprano - DWT?

f



t

value



time

15-826
Copyright (c) 2019 C. Faloutsos
30

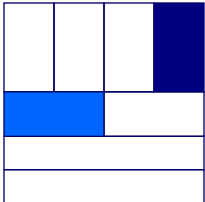
30

CarnegieMellon

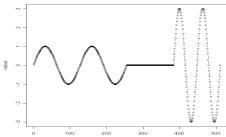
## Wavelets - Drill#1:

- Q: baritone/silence/soprano - DWT?

f



value



15-826 Copyright (c) 2019 C. Faloutsos 31

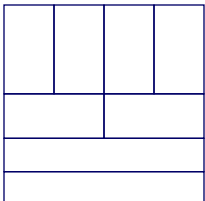
31

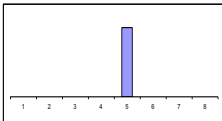
CarnegieMellon

## Wavelets - Drill#2:

- Q: spike - DWT?

f





15-826 Copyright (c) 2019 C. Faloutsos 32

32



CarnegieMellon

## Wavelets - Drill#2:

- Q: spike - DWT?

f

0.00 0.00 **0.71** 0.00

0.00 **0.50**

**-0.35**

t **0.35**

15-826
Copyright (c) 2019 C. Faloutsos
33

33

CarnegieMellon

## Wavelets - Drill#3:

- Q: weekly + daily periodicity, + spike - DWT?

f

0.00 0.00 **0.71** 0.00

0.00 **0.50**

**-0.35**

t **0.35**

15-826
Copyright (c) 2019 C. Faloutsos
34

34

CarnegieMellon

## Wavelets - Drill#3:

- Q: **weekly** + daily periodicity, + spike - DWT?

15-826 Copyright (c) 2019 C. Faloutsos 35

35

CarnegieMellon

## Wavelets - Drill#3:

- Q: weekly + **daily** periodicity, + spike - DWT?

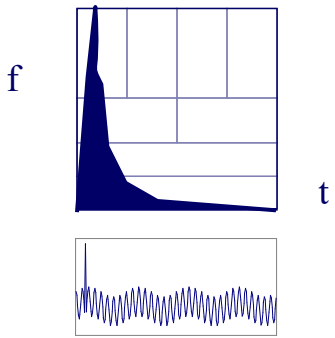
15-826 Copyright (c) 2019 C. Faloutsos 36

36

CarnegieMellon

## Wavelets - Drill#3:

- Q: weekly + daily periodicity, + spike - DWT?



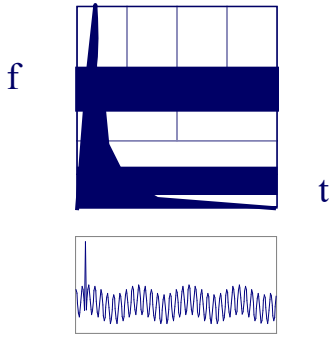
15-826 Copyright (c) 2019 C. Faloutsos 37

37

CarnegieMellon

## Wavelets - Drill#3:

- Q: weekly + daily periodicity, + spike - DWT?



15-826 Copyright (c) 2019 C. Faloutsos 38

38

CarnegieMellon

## Wavelets - Drill#3:

- Q: DFT?

DWT

DFT

15-826 Copyright (c) 2019 C. Faloutsos 39

39

CarnegieMellon

## Wavelets - Drill:

Let's see it live:

<http://dsp.rice.edu/software/dsp-teaching-tools>

delta; cosine; cosine2; chirp

- Haar vs Daubechies-4, -6, etc

15-826 Copyright (c) 2019 C. Faloutsos 40

40

CarnegieMellon

## Delta?

$x(0)=1; x(t)=0$  elsewhere

15-826 Copyright (c) 2019 C. Faloutsos 41

41

CarnegieMellon

## Delta?

$x(0)=1; x(t)=0$  elsewhere

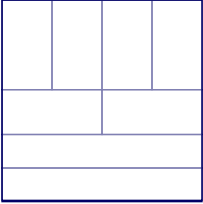
15-826 Copyright (c) 2019 C. Faloutsos 42

42

CarnegieMellon

## 2 cosines?

$$x(t) = \cos(2 * \pi * 4 * t / 1024) + 5 * \cos(2 * \pi * 8 * t / 1024)$$

f

t

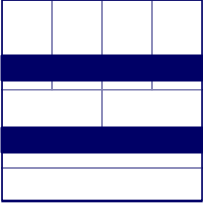
15-826
Copyright (c) 2019 C. Faloutsos
43

43

CarnegieMellon

## 2 cosines?

$$x(t) = \cos(2 * \pi * 4 * t / 1024) + 5 * \cos(2 * \pi * 8 * t / 1024)$$

f

t

Which one is for freq.=4?

15-826
Copyright (c) 2019 C. Faloutsos
44

44

CarnegieMellon

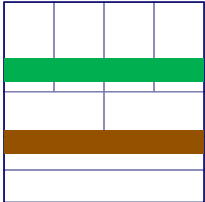
## 2 cosines?

$$x(t) = \cos(2 * \pi * 4 * t / 1024) + 5 * \cos(2 * \pi * 8 * t / 1024)$$

$f \sim 8$  →

$f \sim 4$  →

f



Which one is for freq.=4?

15-826 Copyright (c) 2019 C. Faloutsos 45

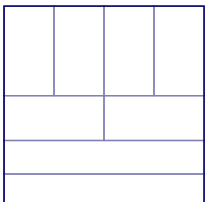
45

CarnegieMellon

## Chirp?

$$x(t) = \cos(2 * \pi * t * t / 1024)$$

f



t

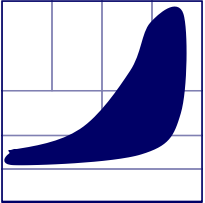
15-826 Copyright (c) 2019 C. Faloutsos 46

46

CarnegieMellon

## Chirp?

$$x(t) = \cos(2 * \pi * t * t / 1024)$$

f

t

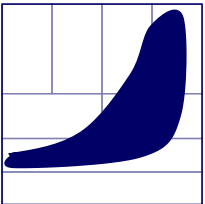
15-826 Copyright (c) 2019 C. Faloutsos 47

47

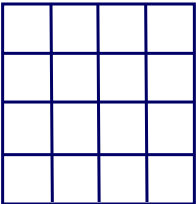
CarnegieMellon

## Chirp?

$$x(t) = \cos(2 * \pi * t * t / 1024)$$

f

t

SWFT?



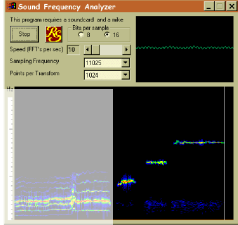
15-826 Copyright (c) 2019 C. Faloutsos 48

48



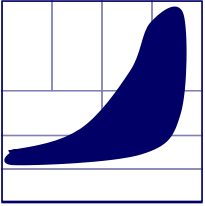
CarnegieMellon

# Chirp?

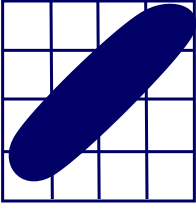
$$x(t) = \cos(2 * \pi * t * t / 1024)$$


SWFT

f



t




15-826 Copyright (c) 2019 C. Faloutsos 49

49

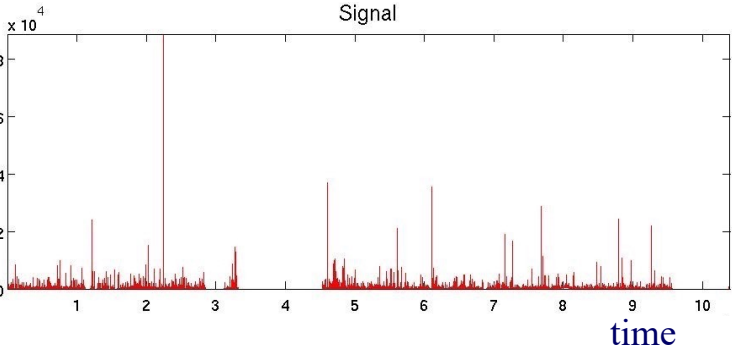
CarnegieMellon

# More examples (BGP updates)

#updates

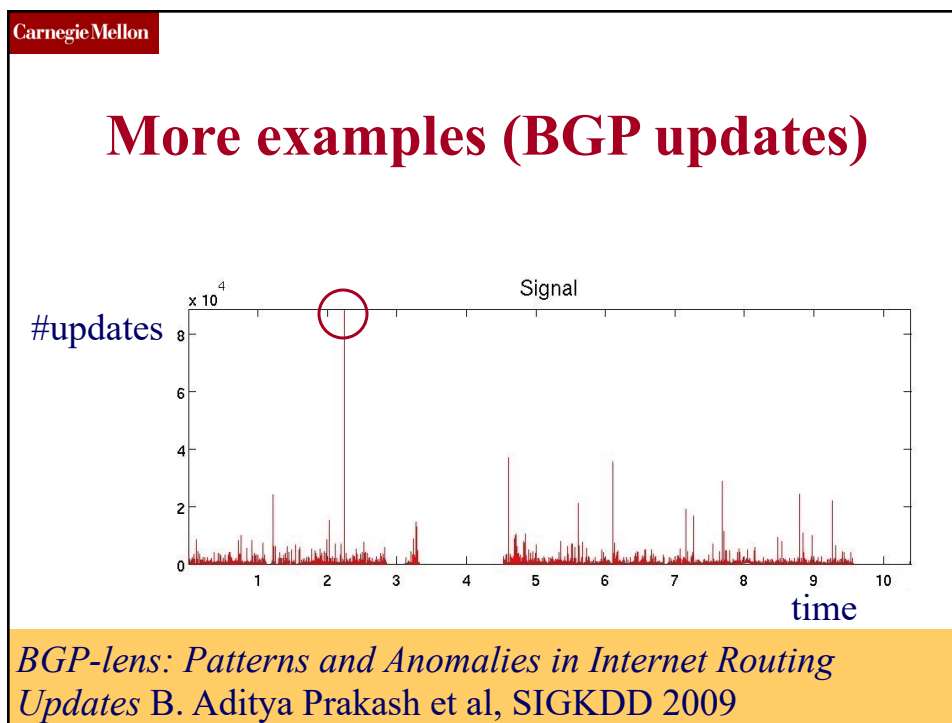


Signal

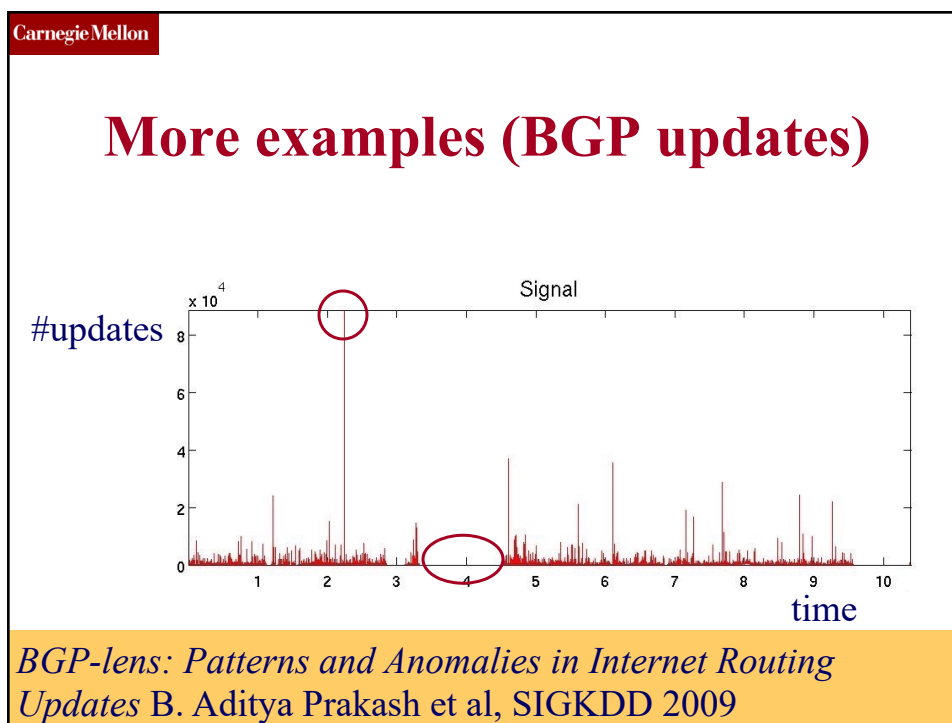


*BGP-lens: Patterns and Anomalies in Internet Routing Updates* B. Aditya Prakash et al, SIGKDD 2009

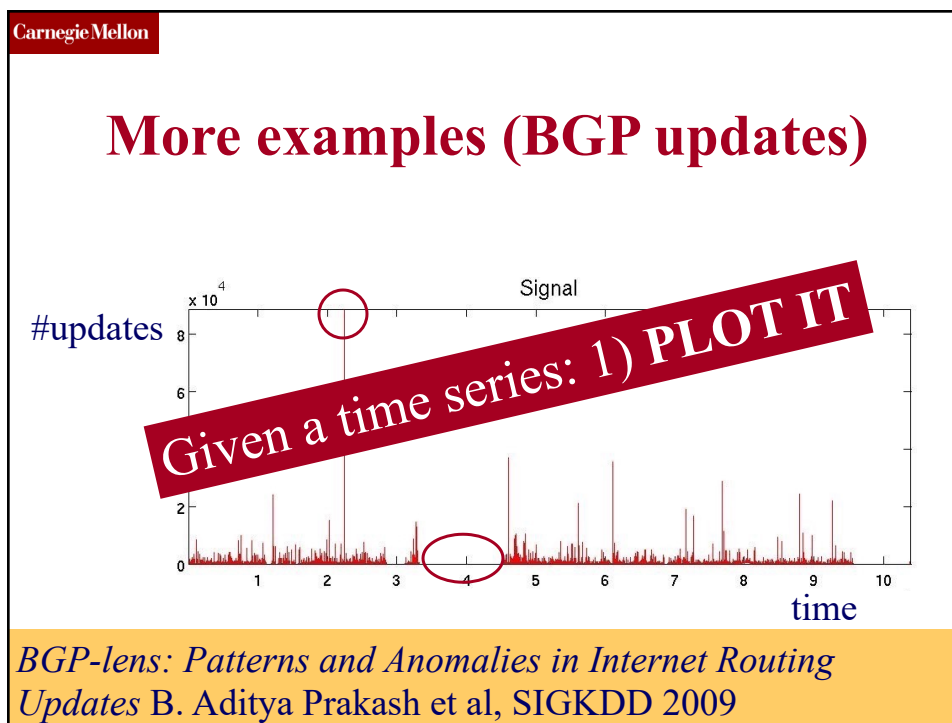
50



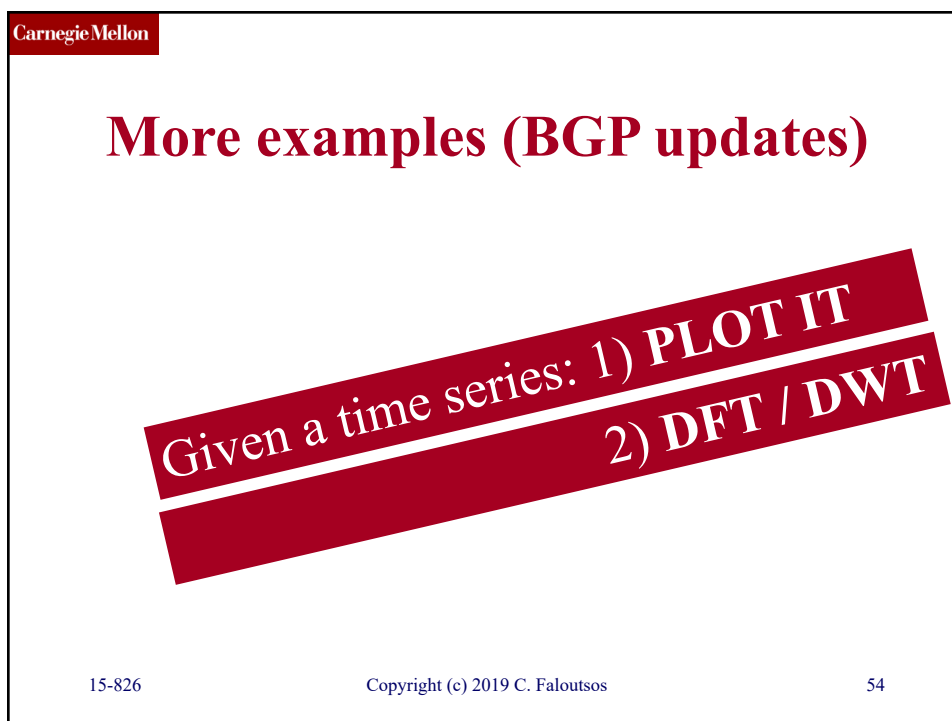
51



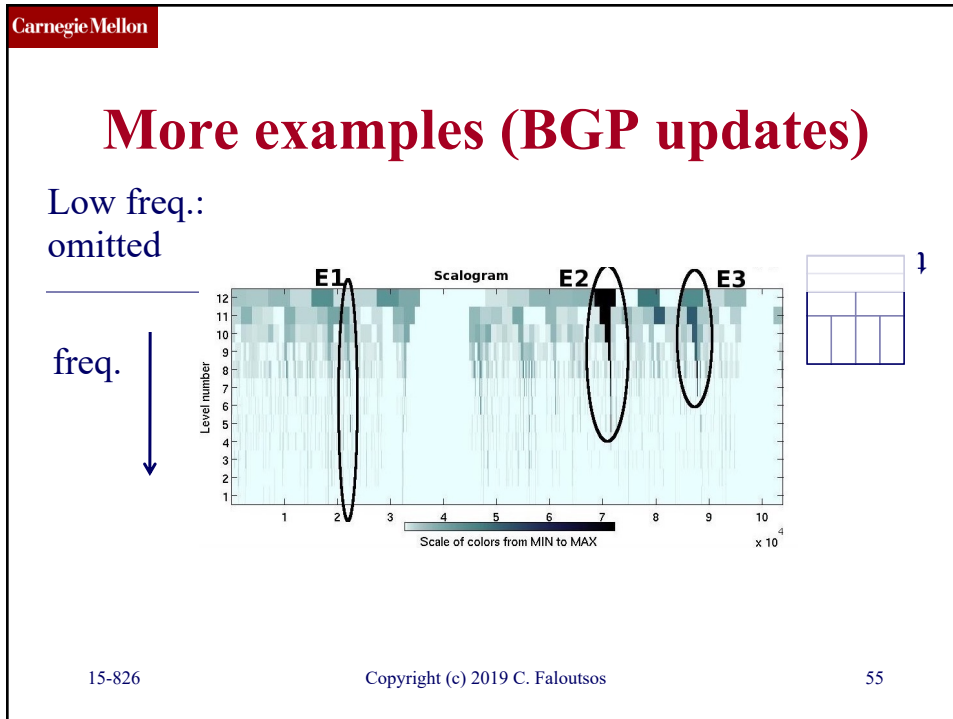
52



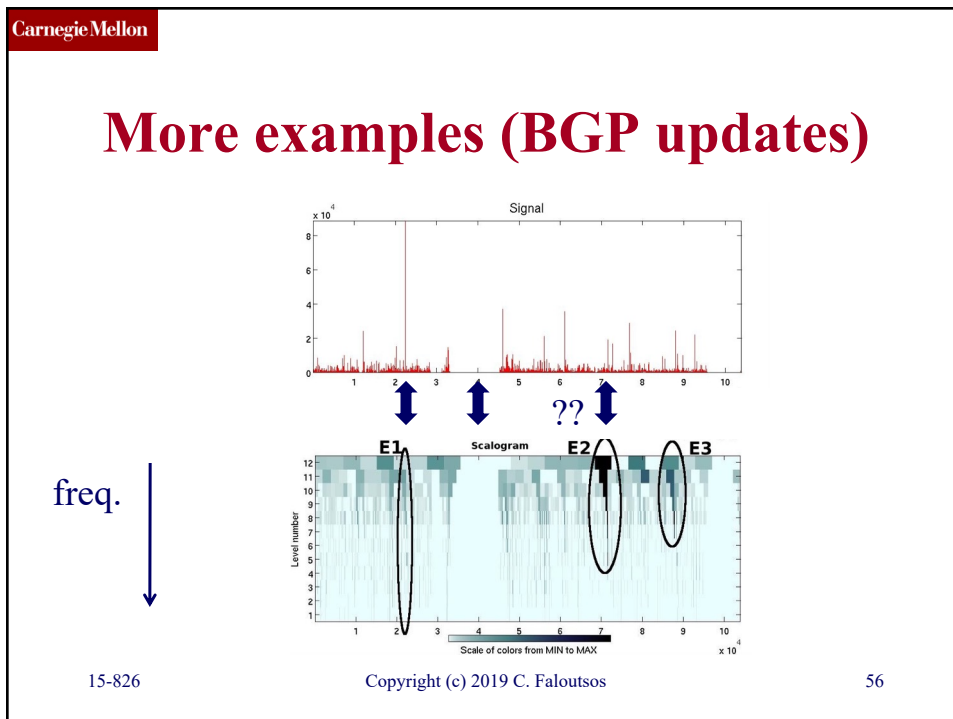
53



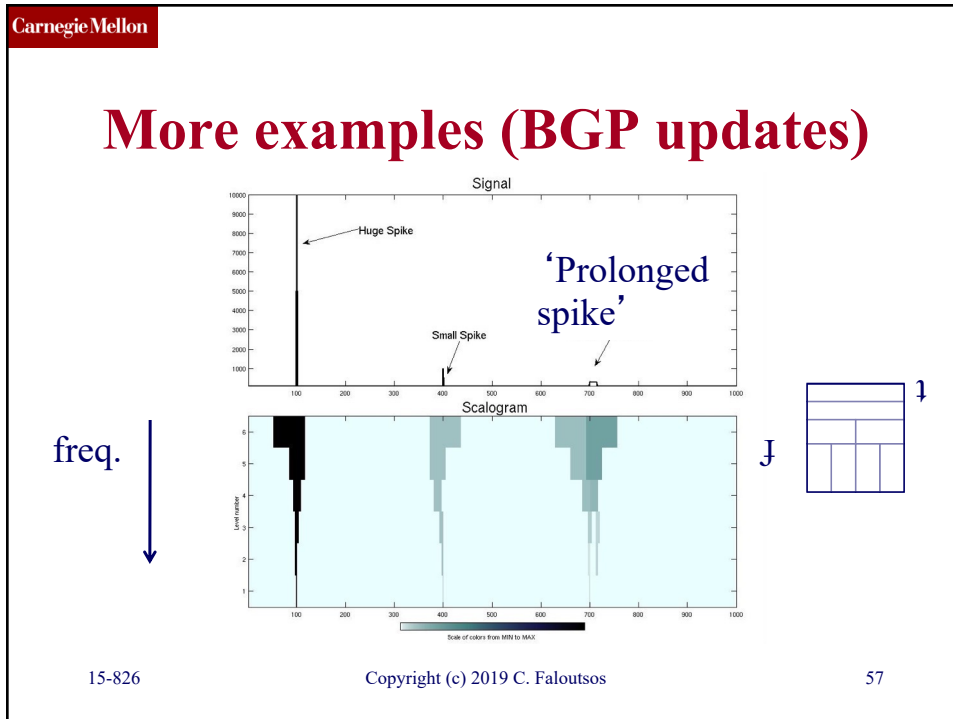
54



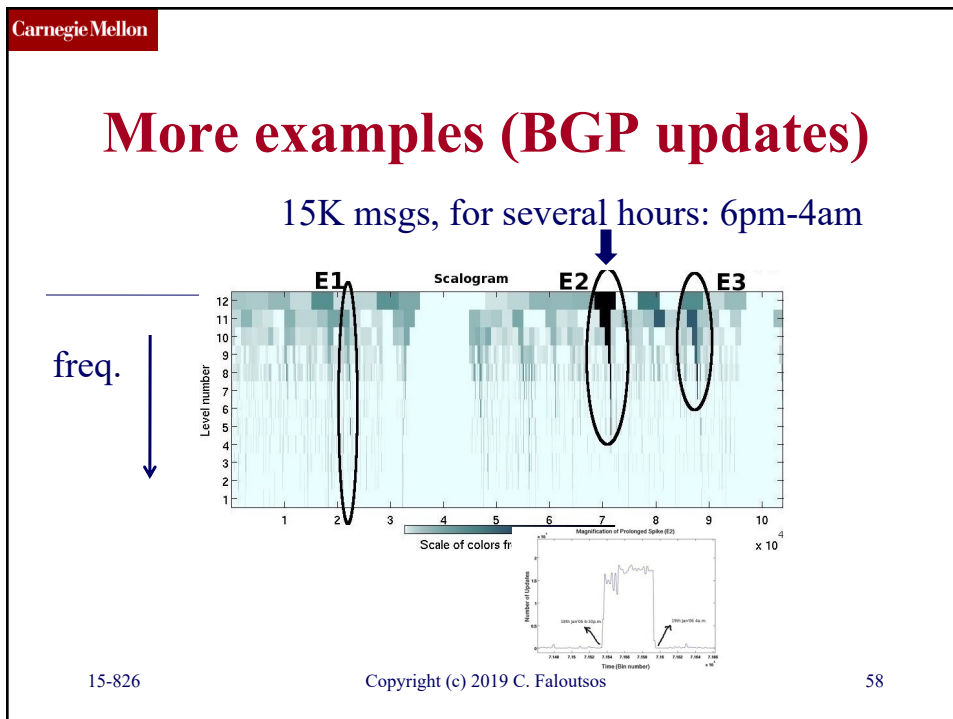
55



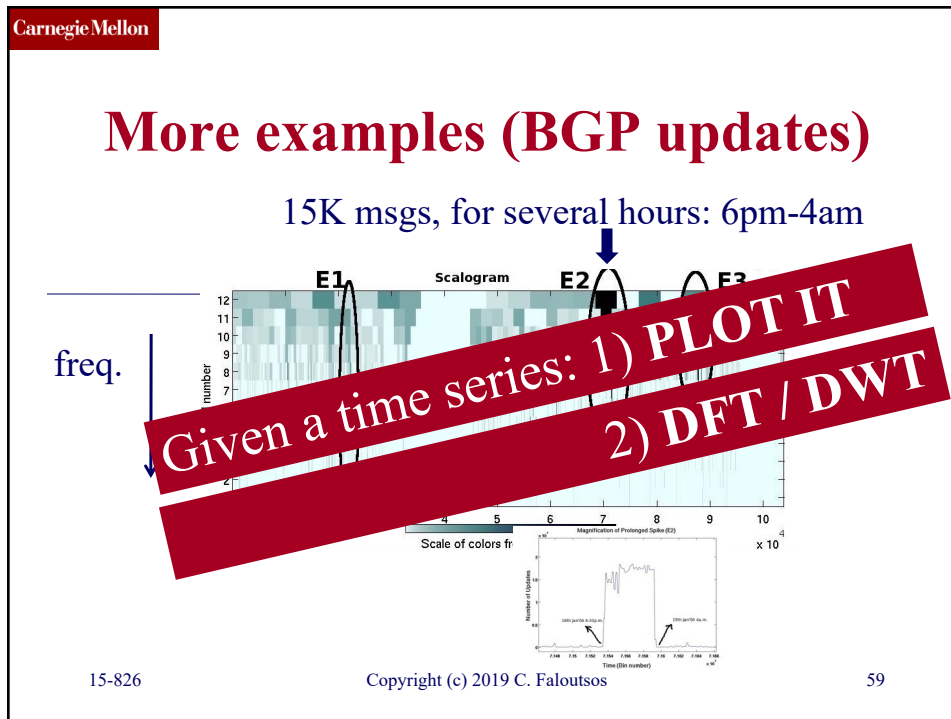
56



57



58



59

CarnegieMellon

## Wavelets - Drill

- Or use 'R', 'octave' or 'matlab' – R:

```
install.packages("wavelets")
library("wavelets")
X1<-c(1,2,3,4,5,6,7,8)
dwt(X1, n.levels=3, filter="d4")
mra(X1, n.levels=3, filter="d4")
```

15-826 Copyright (c) 2019 C. Faloutsos 60

60

CarnegieMellon

## Wavelets - k-dimensions?

- easily defined for any dimensionality (like DFT, DCT)

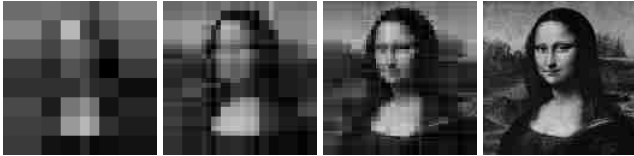
15-826 Copyright (c) 2019 C. Faloutsos 61

61

CarnegieMellon

## Wavelets - example

<http://grail.cs.washington.edu/projects/query/>  
Wavelets achieve \*great\* compression:



20      100      400      16,000  
# coefficients

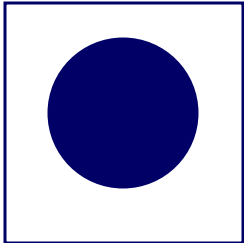
15-826 Copyright (c) 2019 C. Faloutsos 62

62

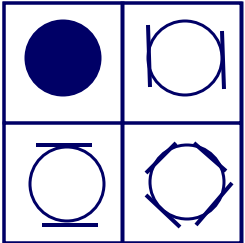
CarnegieMellon

## Wavelets - intuition

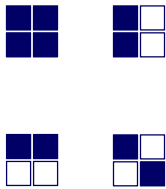
- Edges (horizontal; vertical; diagonal)



15-826



Copyright (c) 2019 C. Faloutsos



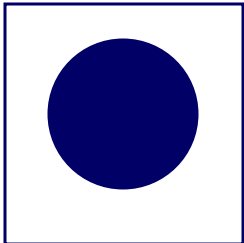
63

63

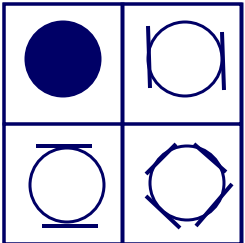
CarnegieMellon

## Wavelets - intuition

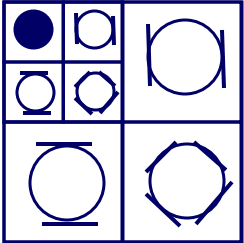
- Edges (horizontal; vertical; diagonal)
- recurse



15-826



Copyright (c) 2019 C. Faloutsos



64

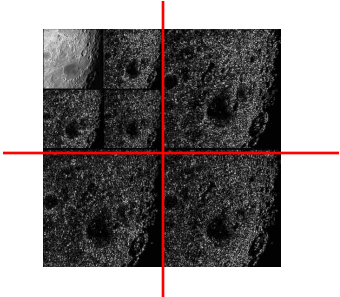
64



CarnegieMellon

## Wavelets - intuition

- Edges (horizontal; vertical; diagonal)
- <http://www331.jpl.nasa.gov/public/wave.html>





15-826 Copyright (c) 2019 C. Faloutsos 65

65

CarnegieMellon

## Advantages of Wavelets


- Better compression (better RMSE with same number of coefficients)
- closely related to the processing of the mammalian eye and ear
- Good for progressive transmission
- handle spikes well
- usually, fast to compute ( $O(n)$ !)

15-826 Copyright (c) 2019 C. Faloutsos 66

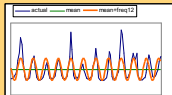

66

CarnegieMellon



## Overall Conclusions


- DFT spots periodicities
- **DWT** : multi-resolution - matches processing of mammalian ear/eye better
- Both: powerful tools for **compression**, **pattern detection** in real signals
- Both: included in math packages (matlab, R, python)

15-826 Copyright (c) 2019 C. Faloutsos 67

67


CarnegieMellon



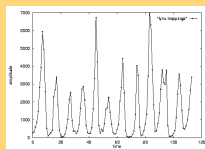
## Solutions:

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

Q: Find patterns and/or compress

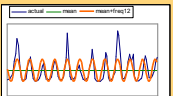


count





year

✓ A1: Fourier (DFT)




✓ A2: Wavelets (DWT)

15-826 Copyright (c) 2019 C. Faloutsos

68


CarnegieMellon



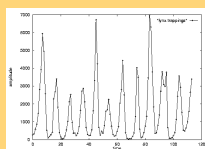
## Solutions:

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

Q: Find patterns and/or compress



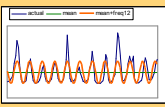
count



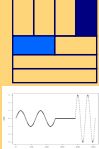
year

**A0: PLOT IT**

✓ A1: Fourier (DFT)



✓ A2: Wavelets (DWT)



15-826 Copyright (c) 2019 C. Faloutsos

69

CarnegieMellon

## Resources

- Numerical Recipes in C: great description, intuition and code for all three tools
- *xwpl*: open source wavelet package from Yale, with excellent GUI.

15-826 Copyright (c) 2019 C. Faloutsos 70

70

CarnegieMellon

## Resources (cont' d)

- [www-dsp.rice.edu/software/EDU/mra.shtml](http://www-dsp.rice.edu/software/EDU/mra.shtml)  
(wavelets and other demos)
- R ( 'install.packages("wavelets") )

15-826

Copyright (c) 2019 C. Faloutsos

71

71