

Carnegie Mellon


# Mining graphs and time series: patterns, anomalies, and fraud detection

Part 1: Graphs  
Intro & patterns

*Christos Faloutsos*

CMU SCS

<https://www.cs.cmu.edu/~christos/TALKS/19-Gol>



1

---

---

---

---

---


---

---

---

Carnegie Mellon

## Roadmap



- Introduction
- Part#1: Graphs
- Part#2: Time series
- Part#3: extras (visualization, etc)
- Conclusions

Gov. of India

Copyright (C) 2019 C. Faloutsos

2

2

---

---

---

---

---


---

---

---

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - P1.1: properties/patterns in graphs
  - P1.2: node importance
  - P1.3: community detection
  - P1.4: fraud/anomaly detection
  - P1.5: belief propagation

Gov. of India

Copyright (C) 2019 C. Faloutsos

3

3

---

---

---

---

---

---

---

---

Carnegie Mellon

Why study graphs?

Gov. of India

Copyright (C) 2019 C. Faloutsos

4

4

---

---

---

---

---

---

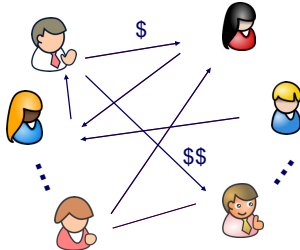
---

---

Carnegie Mellon

Why study graphs?

- Fraud – money laundering etc



Gov. of India

Copyright (C) 2019 C. Faloutsos

5

5

---

---

---

---

---

---

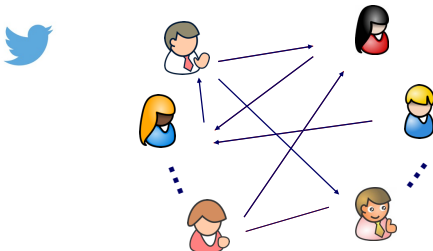
---

---

Carnegie Mellon

Why study graphs?

- Fake followers/trends; botnets/trolls



Gov. of India

Copyright (C) 2019 C. Faloutsos

6

6

---

---

---

---

---

---

---

---

Carnegie Mellon University

## Why study graphs?

- Fake friends

Gov. of India Copyright (C) 2019 C. Faloutsos 7

7

---

---

---

---

---

---

---

---

Carnegie Mellon University

## e-commerce examples

- Recommendation systems
- ....

Gov. of India Copyright (C) 2019 C. Faloutsos 8

8

---

---

---

---

---

---

---

---

Carnegie Mellon University

## e-commerce examples

Who-buys-what

Gov. of India Copyright (C) 2019 C. Faloutsos 9

9

---

---

---

---

---

---

---

---

Carnegie Mellon

### e-commerce examples

Who-buys-what ←
Who-sells-what →

Gov. of India
Copyright (C) 2019 C. Faloutsos
10

---

---

---

---

---

---

---

---

10

Carnegie Mellon

### e-commerce examples

Who-buys-what ←
Who-sells-what →
Who-reviews-what ~

Gov. of India
Copyright (C) 2019 C. Faloutsos
11

---

---

---

---

---

---

---

---

11

Carnegie Mellon

### Cyber-security

Who-buys-what ←
Who-sells-what →
Who-reviews-what ~

Which\_machine - connects\_to - what ~
...
<subject> related-to <object> : graph

Gov. of India
Copyright (C) 2019 C. Faloutsos
12

---

---

---

---

---

---


---

---


12

Carnegie Mellon

## Why study graphs?



fb > \$10B; ~1B users



Gov. of India Copyright (C) 2019 C. Faloutsos 13

13

---

---

---

---

---


---

---

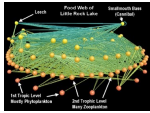
---

Carnegie Mellon

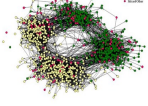
## Why study graphs?



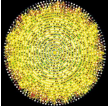
Internet Map  
[lumeta.com]



Food Web  
[Martinez '91]



Friendship Network  
[Moody '01]



Protein Interactions  
[genomebiology.com]

Gov. of India Copyright (C) 2019 C. Faloutsos 14

14

---

---

---

---

---

---


---

---

Carnegie Mellon

## Graphs - why should we care?

- IR: bi-partite graphs (doc-terms)
- web: hyper-text graph
- ... and more:



Gov. of India Copyright (C) 2019 C. Faloutsos 15

15

---

---

---

---

---

---

---

---

Carnegie Mellon

## Graphs - why should we care?

- ‘viral’ marketing
- web-log ( ‘blog’ ) news propagation
- computer network security: email/IP traffic and anomaly detection
- ....

Gov. of India

Copyright (C) 2019 C. Faloutsos

16

16

---

---

---

---

---


---

---




---

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - ➔ – P1.1: properties/patterns in graphs
  - P1.2: node importance
  - P1.3: community detection
  - P1.4: fraud/anomaly detection
  - P1.5: belief propagation

Gov. of India

Copyright (C) 2019 C. Faloutsos

17

17

---

---

---

---

---


---

---



---

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - ➔ – P1.1: properties/patterns in graphs
  - P1.2: node importance
  - P1.3: community detection
  - P1.4: fraud/anomaly detection
  - P1.5: belief propagation

Gov. of India

Copyright (C) 2019 C. Faloutsos

18

18

---

---

---

---

---


---

---



---

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - P1.1: properties/patterns in graphs
  - P1.2: node importance
  - ➔ – P1.3: community detection
  - P1.4: fraud/anomaly detection
  - P1.5: belief propagation

Gov. of India

Copyright (C) 2019 C. Faloutsos

19

---

---

---

---

---

---


---

---



19

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - P1.1: properties/patterns in graphs
  - P1.2: node importance
  - ➔ – P1.3: community detection
  - P1.4: fraud/anomaly detection
  - P1.5: belief propagation

Gov. of India

Copyright (C) 2019 C. Faloutsos

20

---

---

---

---

---

---


---

---


20

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - P1.1: properties/patterns in graphs
  - P1.2: node importance
  - P1.3: community detection
  - ➔ – P1.4: fraud/anomaly detection
  - P1.5: belief propagation



Gov. of India

Copyright (C) 2019 C. Faloutsos

21

---

---

---

---

---

---


---

---


21

Carnegie Mellon

## Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - P1.1: properties/patterns in graphs
  - P1.2: node importance
  - P1.3: community detection
  - P1.4: fraud/anomaly detection
  - ➔ – P1.5: belief propagation



Gov. of India
Copyright (C) 2019 C. Faloutsos
22

---

---

---

---

---

---

---





---

22

Carnegie Mellon

## ‘Recipe’ Structure:

- Problem definition
- Short answer/solution
- LONG answer – details
- Conclusion/short-answer

Gov. of India
Copyright (C) 2019 C. Faloutsos
23

---

---

---

---

---

---

---





---

23

Carnegie Mellon

## ‘Recipe’ Structure:

- Problem definition
- Short answer/solution
- LONG answer – details
- Conclusion/short-answer

Gov. of India
Copyright (C) 2019 C. Faloutsos
24

---

---

---

---

---

---

---


---




24



Carnegie Mellon

# Roadmap



- Introduction – Motivation
- Part#1: Graphs
  - ➔ – P1.1: properties/patterns in graphs 
  - P1.2: node importance 
  - P1.3: community detection 
  - P1.4: fraud/anomaly detection
  - P1.5: belief propagation

Gov. of India
Copyright (C) 2019 C. Faloutsos
25

---

---

---

---

---

---

---

---

25

Carnegie Mellon

# Why care about patterns?

Gov. of India
Copyright (C) 2019 C. Faloutsos
26

---

---

---

---

---

---


---


---

26

Carnegie Mellon

# Why care about patterns?



1. Anomalies 
2. Faster algorithms
3. Graph generators (‘what if’ scenarios)

Gov. of India
Copyright (C) 2019 C. Faloutsos
27

---

---

---

---

---

---

---


---



27

Carnegie Mellon

## Why care about patterns?

1. Anomalies



Patterns

anomalies

2. Faster algorithms
3. Graph generators ('what if' scenarios)

Gov. of India

Copyright (C) 2019 C. Faloutsos

28

---

---

---

---

---

---

---


---



28



Carnegie Mellon

## Why care about patterns?

1. Anomalies



Patterns

anomalies

2. Faster algorithms



3. Graph generators ('what if' scenarios)

Gov. of India

Copyright (C) 2019 C. Faloutsos

29

---

---

---

---

---

---

---


---



29



Carnegie Mellon

## Why care about patterns?

1. Anomalies



Patterns

anomalies

2. Faster algorithms



3. Graph generators ('what if' scenarios)

– [Graph500.org](http://Graph500.org)

Gov. of India

Copyright (C) 2019 C. Faloutsos

30

---

---

---

---

---


---

---


---

30

Carnegie Mellon






## 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)

Gov. of India Copyright (C) 2019 C. Faloutsos 31

31

---

---

---

---


---

---


---

---

Carnegie Mellon



## Problem definition



- Are real graphs random?
  - S\*: what do **static** graphs look like?
  - T\*: how do graphs evolve over **time**?

Gov. of India Copyright (C) 2019 C. Faloutsos 32

32

---

---

---

---


---

---


---

---


Carnegie Mellon



## Short answer(s)



- Are real graphs random?
  - S\*: what do **static** graphs look like?
    - S.0: 'six degrees'
    - S.1: skewed degree distribution
    - S.2: skewed eigenvalues
    - S.3: triangle power-laws
    - S.4: GCC; and skewed distr. of conn. comp.
  - D\*: how do graphs evolve over **time**?
    - D.1: diameters
    - D.2: densification



Gov. of India Copyright (C) 2019 C. Faloutsos 33

33

---

---

---

---

---

---

---

---

Carnegie Mellon

Short answer(?)

- Are real graphs random?
  - S\*: what do we see?

**Power laws:  $y \sim x^a$**

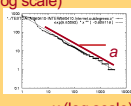
**NOT Gaussians**

**Take logarithms**

Power laws

and skewed distr. of conn. comp.

- T\*: how do graphs evolve over time?
  - T.1: diameters
  - T.2: densification



Gov. of India Copyright (C) 2019 C. Faloutsos 34

34

---

---

---

---

---

---


---

---

Carnegie Mellon

**Graph mining**

- Are real graphs random?



Gov. of India Copyright (C) 2019 C. Faloutsos 35

35

---

---

---

---

---

---


---

---

Carnegie Mellon

**Laws and patterns**

- Are real graphs random?
  - A: NO!!
    - Diameter ( '6 degrees' , 'Kevin Bacon' )
    - in- and out- degree distributions
    - other (surprising) patterns
- So, let's look at the data



Gov. of India Copyright (C) 2019 C. Faloutsos 36

36

---

---

---

---

---

---

---

---

Carnegie Mellon

Solution# S.1

- Power law in the degree distribution [SIGCOMM99]

internet domains

log(degree)

att.com

ibm.com

log(rank)

Gov. of India

Copyright (C) 2019 C. Faloutsos

37

37

---

---

---

---

---

---

---

---

Carnegie Mellon

Solution# S.1

- Power law in the degree distribution [SIGCOMM99]

internet domains

log(degree)

att.com

ibm.com

-0.82

log(rank)

Gov. of India

Copyright (C) 2019 C. Faloutsos

38

38

---

---

---

---

---

---

---

---

Carnegie Mellon

Solution# S.1

- Q: So what?

internet domains

log(degree)

att.com

ibm.com

-0.82

log(rank)

Gov. of India

Copyright (C) 2019 C. Faloutsos

39

39

---

---

---

---

---

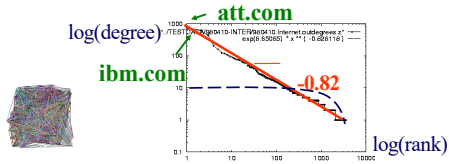
---

---

---

**Solution# S.1**

- Q: So what? = friends of friends (F.O.F.)
- A1: # of two-step-away pairs: internet domains



Gov. of India Copyright (C) 2019 C. Faloutsos 40

---

---

---

---

---

---

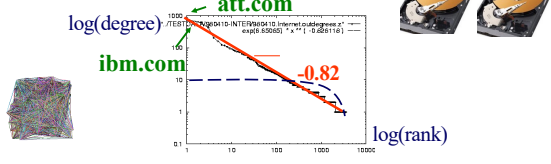
---

---

40

**Solution# S.1**

- Q: So what? = friends of friends (F.O.F.)
- A1: # of two-step-away pairs:  $100^2 * N = 10$  Trillion internet domains



Gov. of India Copyright (C) 2019 C. Faloutsos 41

---

---

---

---

---

---

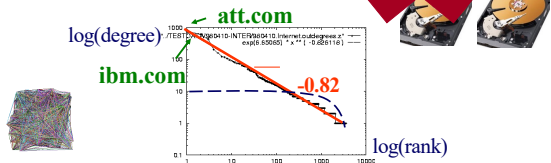
---

---

41

**Solution# S.1**

- Q: So what? = friends of friends (F.O.F.)
- A1: # of two-step-away pairs:  ~~$100^2 * N = 10$~~  Trillion internet domains



Gov. of India Copyright (C) 2019 C. Faloutsos 42

---

---

---

---

---

---

---

---

42

Carnegie Mellon

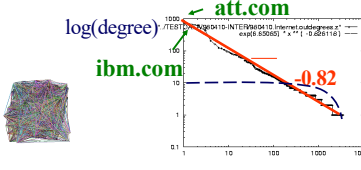
### Gaussian trap

#### Solution# S.1

- Q: So what? = friends of friends (F.O.F.)
- A1: # of two-step-away pairs:  $O(d_{\max}^2) \sim 10M^2$  internet domains

$\downarrow$

$\sim 0.8PB \rightarrow$   
a data center(!)



Gov. of India Copyright (C) 2019 C. Faloutsos DCO @ CMU 43

---

---

---

---

---

---

---

---

43

Carnegie Mellon

### Gaussian trap

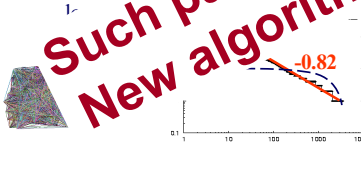
#### Solution# S.1

- Q: So what?
- A1: # of two-step-away pairs:  $? \sim 10M^2$

$\downarrow$

$\sim 0.8PB \rightarrow$   
a data center(!)

**Such patterns  $\rightarrow$   
New algorithms**



Gov. of India Copyright (C) 2019 C. Faloutsos 44

---

---

---

---

---

---

---


---

44

Carnegie Mellon

### Observation – big-data:

- $O(N^2)$  algorithms are ~intractable -  $N=1B$
- $N^2$  seconds = 31B years (>2x age of universe)



Gov. of India Copyright (C) 2019 C. Faloutsos 45

---

---

---

---

---

---

---


---

45

**Observation – big-data:**

- $O(N^2)$  algorithms are ~intractable -  $N=1B$
- $N^2$  seconds = ~~31B~~<sup>31M</sup> years
- 1,000 machines

1B



Gov. of India Copyright (C) 2019 C. Faloutsos 46

46

---

---

---

---

---

---

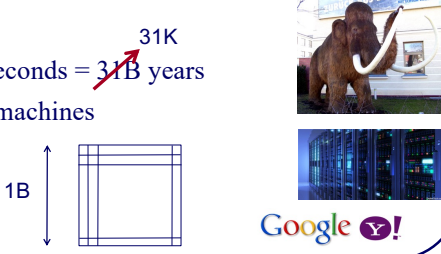
---

---

**Observation – big-data:**

- $O(N^2)$  algorithms are ~intractable -  $N=1B$
- $N^2$  seconds = ~~31B~~<sup>31K</sup> years
- 1M machines

1B



Gov. of India Copyright (C) 2019 C. Faloutsos 47

47

---

---

---

---

---

---

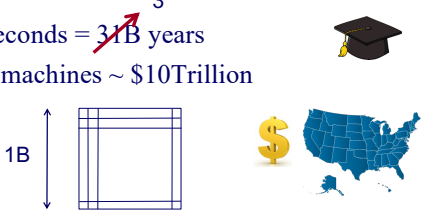
---

---

**Observation – big-data:**

- $O(N^2)$  algorithms are ~intractable -  $N=1B$
- $N^2$  seconds = ~~31B~~<sup>3</sup> years
- 10B machines ~ \$10Trillion

1B



Gov. of India Copyright (C) 2019 C. Faloutsos 48

48

---

---

---

---

---

---

---

---



Carnegie Mellon

### Observation – big-data:

- $O(N^2)$  algorithms are ~intractable -  $N=1B$

**And parallelism might not help**

- $N^2$  seconds =  $3 \times 10^8$  years
- 10B machines ~ \$10Trillion

Gov. of India Copyright (C) 2019 C. Faloutsos 49

49

---

---

---

---

---

---

---

---

Carnegie Mellon

### Solution# S.2: Eigen Exponent $E$

Eigenvalue

Exponent = slope  
 $E = -0.48$   
May 2001

$Ax = \lambda x$

- A2: power law in the eigenvalues of the adjacency matrix

Gov. of India Copyright (C) 2019 C. Faloutsos 50

50

---

---

---

---

---

---

---

---

Carnegie Mellon

### Solution# S.2: Eigen Exponent $E$

Eigenvalue

Exponent = slope  
 $E = -0.48$   
May 2001

Rank of decreasing eigenvalue

- [Mihail, Papadimitriou '02]: slope is  $\frac{1}{2}$  of rank exponent

Gov. of India Copyright (C) 2019 C. Faloutsos 51

51

---

---

---

---

---

---

---

---

Carnegie Mellon

**But:**

How about graphs from other domains?

Gov. of India

Copyright (C) 2019 C. Faloutsos

52

---

---

---

---

---

---

---

---

52

Carnegie Mellon

**More power laws:**

- web hit counts [w/ A. Montgomery]

Gov. of India

Copyright (C) 2019 C. Faloutsos

53

---

---

---

---

---

---

---

---

53

Carnegie Mellon

**epinions.com**

- who-trusts-whom [Richardson + Domingos, KDD 2001]

Gov. of India

Copyright (C) 2019 C. Faloutsos

54

---

---

---

---

---

---

---


---

54

Carnegie Mellon

### And numerous more

- # of sexual contacts
- Income [Pareto] – ‘80-20 distribution’
- Duration of downloads [Bestavros+]
- Duration of UNIX jobs ( ‘mice and elephants’ )
- Size of files of a user
- ...
- ‘Black swans’



Gov. of India Copyright (C) 2019 C. Faloutsos 55

55

---

---

---

---

---


---

---

---

Carnegie Mellon

### List of Static Patterns



- ✓ • S.1 degree
- ✓ • S.2 eigenvalues
- S.3 small diameter
- S.4/5 Triangle laws
- (S.6) NLCC non-largest conn. components
- (S.7) eigen plots
- (S.8) radius plot

Gov. of India Copyright (C) 2019 C. Faloutsos 56

56

---

---

---

---

---

---


---

---

Carnegie Mellon

### S.3 small diameters

- Small diameter (~ constant!) –
  - six degrees of separation / ‘Kevin Bacon’
  - small worlds [Watts and Strogatz]



Gov. of India Copyright (C) 2019 C. Faloutsos 57

57

---

---

---


---

---

---

---

---

Carnegie Mellon


## List of Static Patterns

- ✓ • S.1 degree
- ✓ • S.2 eigenvalues
- ✓ • S.3 small diameter
  - S.4/5 Triangle laws
  - (S.6) NLCC non-largest conn. components
  - (S.7) eigen plots
  - (S.8) radius plot

Gov. of India
Copyright (C) 2019 C. Faloutsos
58

58

---

---

---

---

---


---

---

---

Carnegie Mellon

## Solution# S.4: Triangle ‘Laws’



- Real social networks have a lot of triangles

Gov. of India
Copyright (C) 2019 C. Faloutsos
59

59

---

---

---

---

---


---

---

---

Carnegie Mellon

## Solution# S.4: Triangle ‘Laws’



- Real social networks have a lot of triangles
  - Friends of friends are friends
- Any patterns?

Gov. of India
Copyright (C) 2019 C. Faloutsos
60

60

---

---

---

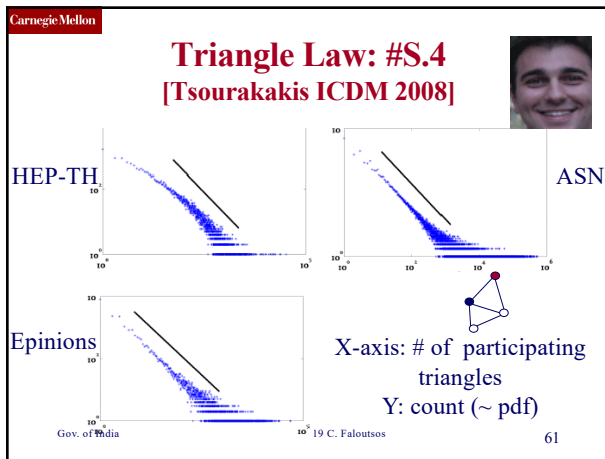
---

---

---

---

---



61

---

---

---

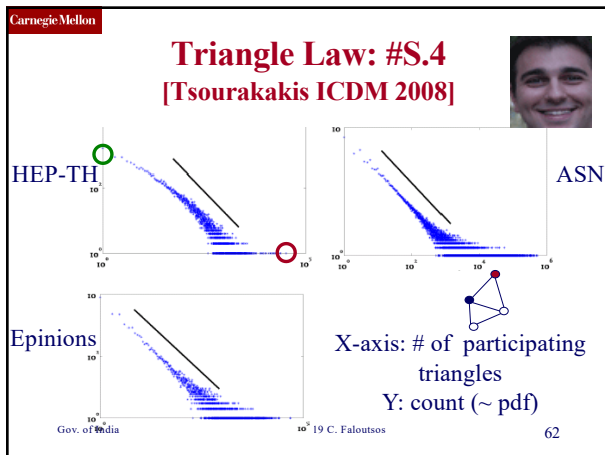
---

---

---

---

---



62

---

---

---

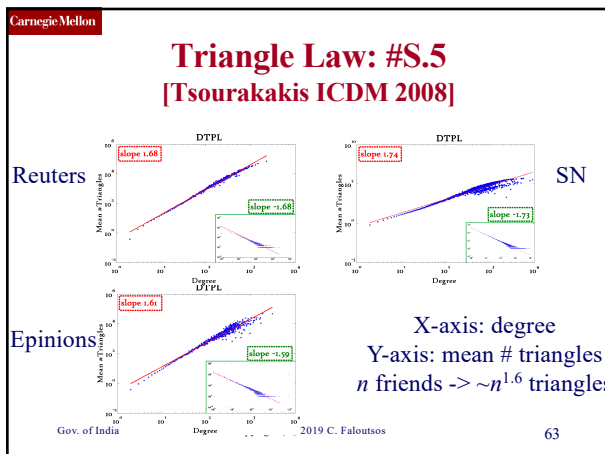
---

---

---

---

---



63

---

---

---

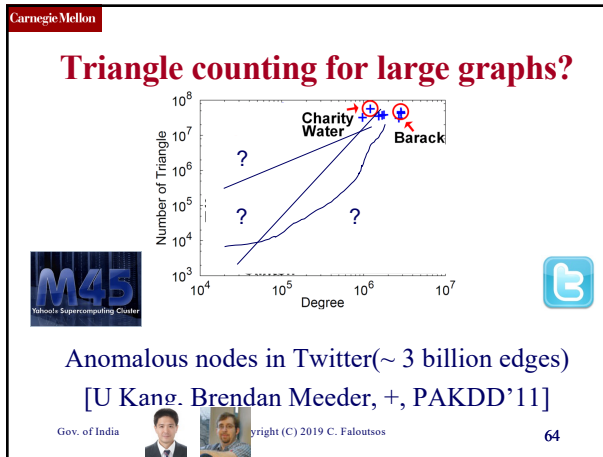
---

---

---

---

---



64

---

---

---

---

---

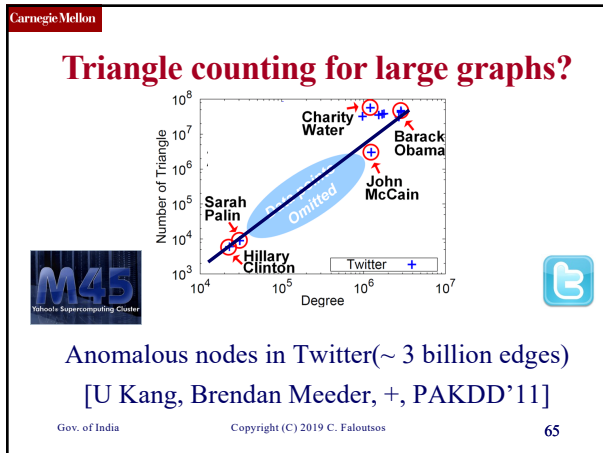
---

---

---

---

---



65

---

---

---

---

---

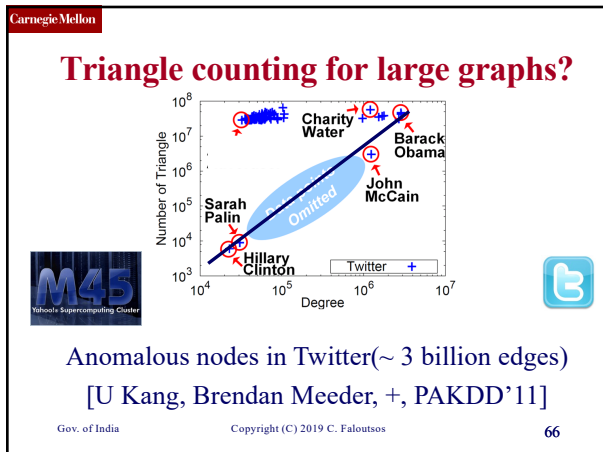
---

---

---

---

---



66

---

---

---

---

---

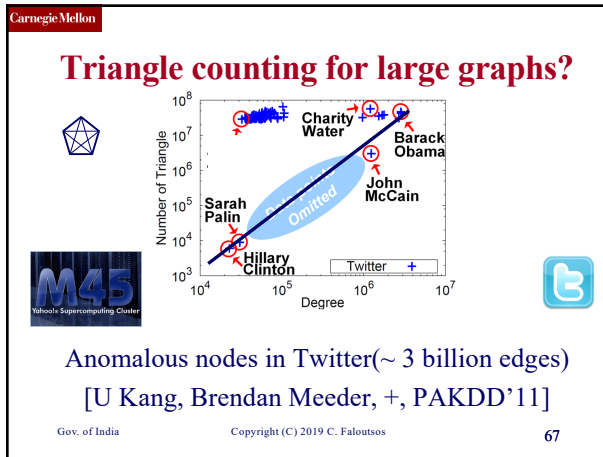
---

---

---

---

---



67

---

---

---

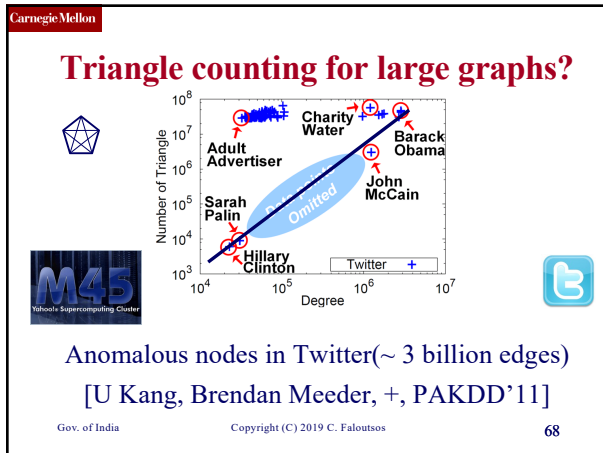
---

---

---

---

---



68

---

---

---

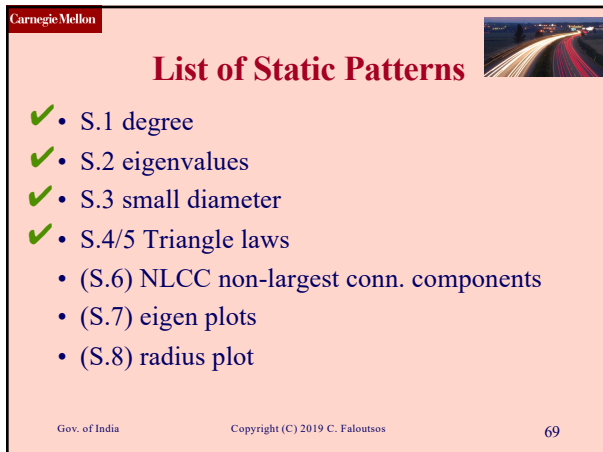
---

---

---

---

---



69

---

---

---

---

---

---

---

---

Carnegie Mellon

## Generalized Iterated Matrix Vector Multiplication (GIMV)

[PEGASUS: A Peta-Scale Graph Mining System - Implementation and Observations.](#)  
 U Kang, Charalampos E. Tsourakakis,  
 and Christos Faloutsos.  
 (ICDM) 2009, Miami, Florida, USA.  
 Best Application Paper (runner-up).

Gov. of India

Copyright (C) 2019 C. Faloutsos

70

70

---

---

---

---

---

---

---

---

Carnegie Mellon

## S.6: NLCC

- Connected Components – 4 observations:

Count

YahooWeb

Size

Giant Connected Component

Gov. of India

Copyright (C) 2019 C. Faloutsos

71

71

---

---

---

---

---

---

---

---

Carnegie Mellon

## S.6: NLCC

- Connected Components

Count

YahooWeb

Size

Giant Connected Component

1) 10K x larger than next

Gov. of India

Copyright (C) 2019 C. Faloutsos

72

72

---

---

---

---

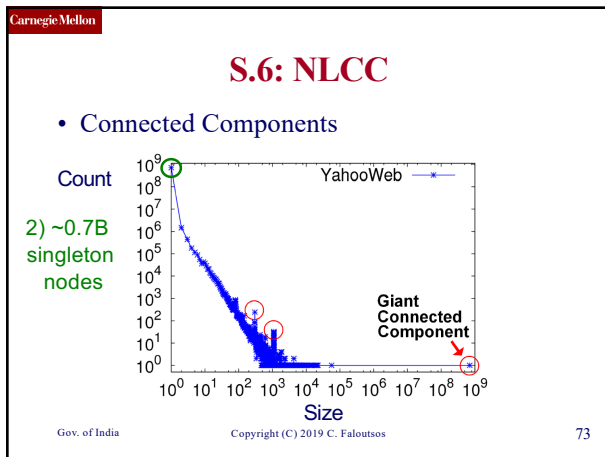
---

---

---

---






---

---

---

---

---

---

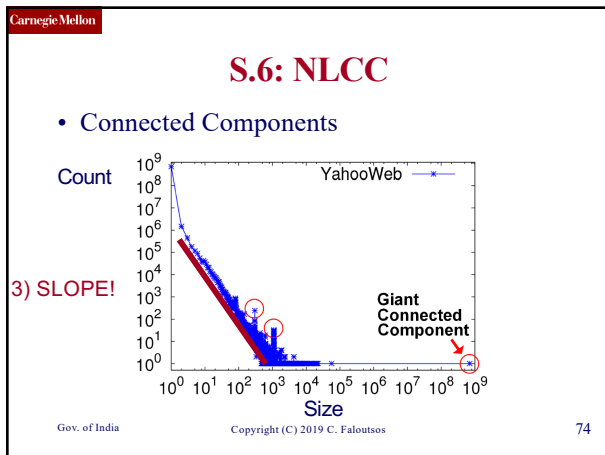
---

---

---

---

73




---

---

---

---

---

---

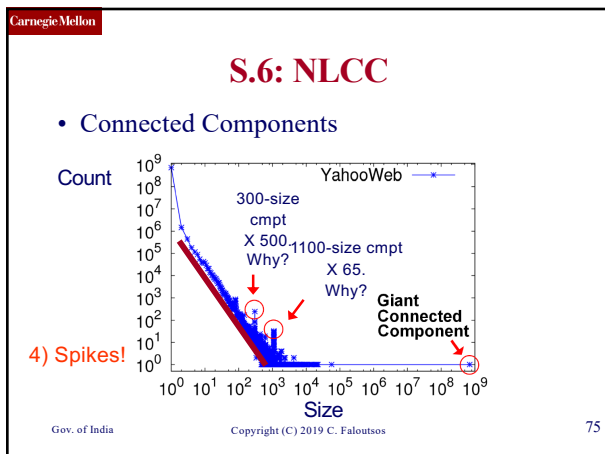
---

---

---

---

74




---

---

---

---

---

---

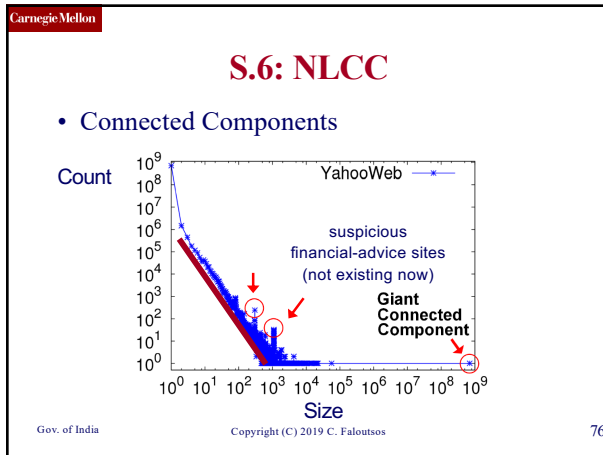
---

---

---

---

75



76

---

---

---

---

---

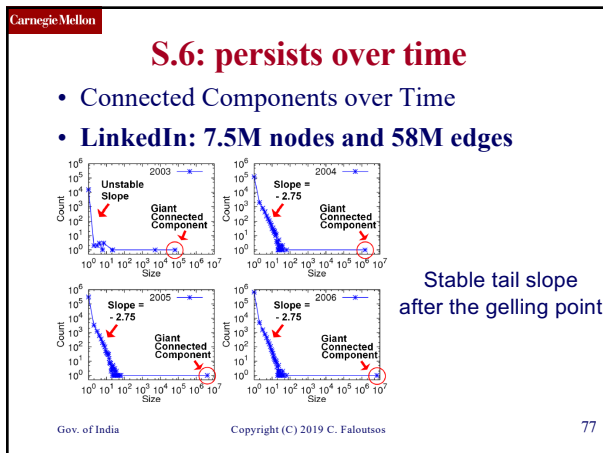
---

---

---

---

---



77

---

---

---

---

---

---

---

---

---

---



78

---

---

---

---

---

---

---


---

---

---

Carnegie Mellon

## EigenSpokes



B. Aditya Prakash, Mukund Seshadri, Ashwin Sridharan, Sridhar Machiraju and Christos Faloutsos: *EigenSpokes: Surprising Patterns and Scalable Community Chipping in Large Graphs*, PAKDD 2010, Hyderabad, India, 21-24 June 2010.

**Useful for fraud detection!**

Gov. of India

Copyright (C) 2019 C. Faloutsos

79

---

---

---

---

---

---

---

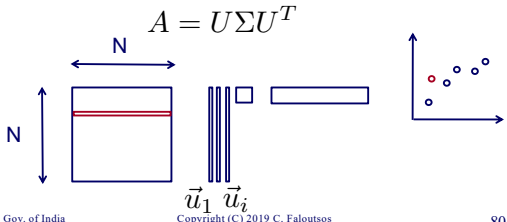
---

79

Carnegie Mellon

## EigenSpokes

- Eigenvectors of adjacency matrix
  - equivalent to singular vectors (symmetric, undirected graph)

$$A = U \Sigma U^T$$


Gov. of India

Copyright (C) 2019 C. Faloutsos

80

---

---

---

---

---

---

---

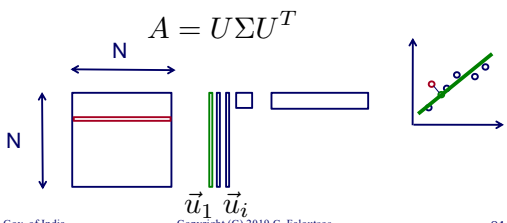
---

80

Carnegie Mellon

## EigenSpokes

- Eigenvectors of adjacency matrix
  - equivalent to singular vectors (symmetric, undirected graph)

$$A = U \Sigma U^T$$


Gov. of India

Copyright (C) 2019 C. Faloutsos

81

---

---

---

---

---

---

---

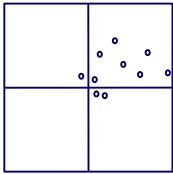
---

81

Carnegie Mellon

## EigenSpokes

- EE plot:
- Scatter plot of scores of  $u_1$  vs  $u_2$
- One would expect
  - Many points @ origin
  - A few scattered ~randomly



$u_2$   
2<sup>nd</sup> Principal component

$u_1$   
1<sup>st</sup> Principal component

Gov. of India Copyright (C) 2019 C. Faloutsos 82

82

---

---

---

---

---

---

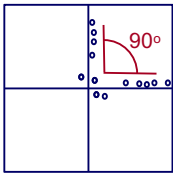
---

---

Carnegie Mellon

## EigenSpokes

- EE plot:
- Scatter plot of scores of  $u_1$  vs  $u_2$
- One would expect
  - Many points @ origin
  - A few scattered ~randomly



$u_2$

$u_1$

Gov. of India Copyright (C) 2019 C. Faloutsos 83

83

---

---

---

---

---

---

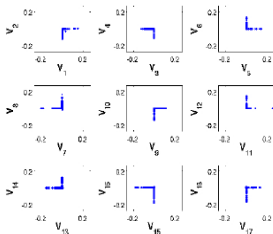
---

---

Carnegie Mellon

## EigenSpokes - pervasiveness

- Present in mobile social graph
  - across time and space
- Patent citation graph



Gov. of India Copyright (C) 2019 C. Faloutsos 84

84

---

---

---

---

---

---

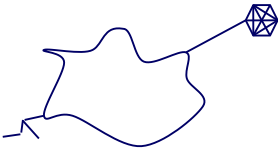
---

---

Carnegie Mellon

### EigenSpokes - explanation

Near-cliques, or near-bipartite-cores, loosely connected



Gov. of India

Copyright (C) 2019 C. Faloutsos

85

---

---

---

---

---

---

---

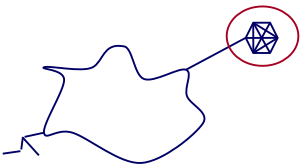
---

85

Carnegie Mellon

### EigenSpokes - explanation

Near-cliques, or near-bipartite-cores, loosely connected



Gov. of India

Copyright (C) 2019 C. Faloutsos

86

---

---

---

---

---

---

---

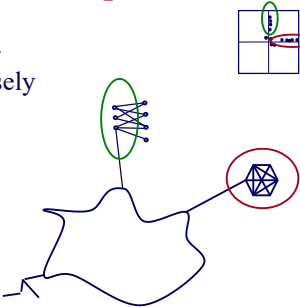
---

86

Carnegie Mellon

### EigenSpokes - explanation

Near-cliques, or near-bipartite-cores, loosely connected



Gov. of India

Copyright (C) 2019 C. Faloutsos

87

---

---

---

---

---

---

---

---

87

Carnegie Mellon

## EigenSpokes - explanation

Near-cliques, or near-bipartite-cores, loosely connected

So what?

- Extract nodes with high *scores*
- high connectivity
- Good “communities”

Gov. of India Copyright (C) 2019 C. Faloutsos

88

88

---

---

---

---

---

---

---

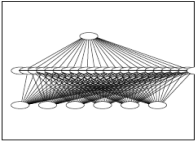
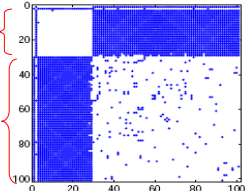
---

Carnegie Mellon

## Bipartite Communities!

patents from same inventor(s)  
‘cut-and-paste’ bibliography!

magnified bipartite community

Useful for fraud detection!

Gov. of India Copyright (C) 2019 C. Faloutsos

89

89

---

---

---

---

---

---

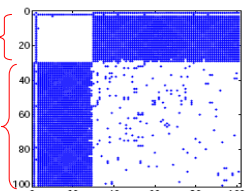
---

---

Carnegie Mellon

## Bipartite Communities!

IP – port scanners  
victims



Useful for fraud detection!

Gov. of India Copyright (C) 2019 C. Faloutsos

90

90

---

---

---

---

---


---

---

---

Carnegie Mellon

## List of Static Patterns



- ✓ • S.1 degree
- ✓ • S.2 eigenvalues
- ✓ • S.3 small diameter
- ✓ • S.4/5 Triangle laws
- ✓ • S.6 NLCC non-largest conn. components
- ✓ • S.7 eigen plots
  - S.8 radius plot

Gov. of India

Copyright (C) 2019 C. Faloutsos

91

91

---

---

---

---

---


---

---

---

Carnegie Mellon

## HADI for diameter estimation



- *Radius Plots for Mining Tera-byte Scale Graphs* **U Kang**, Charalampos Tsourakakis, Ana Paula Appel, Christos Faloutsos, Jure Leskovec, SDM'10
- Naively: diameter needs  $O(N^{**2})$  space and up to  $O(N^{**3})$  time – **prohibitive** ( $N \sim 1B$ )
- Our HADI: linear on  $E$  ( $\sim 10B$ )
  - Near-linear scalability wrt # machines
  - Several optimizations  $\rightarrow$  5x faster

Gov. of India

Copyright (C) 2019 C. Faloutsos

92

92

---

---

---

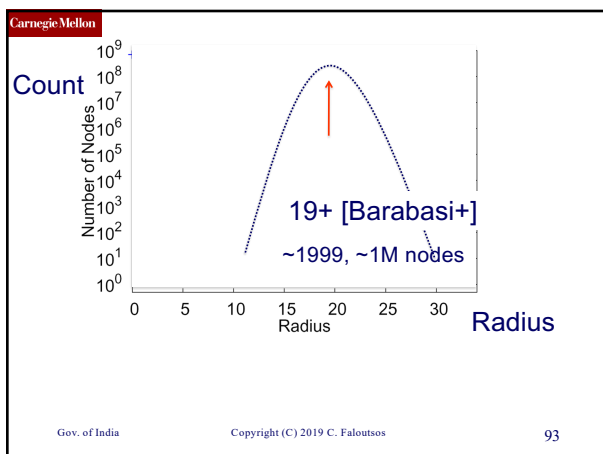
---

---

---

---

---



93

---

---

---

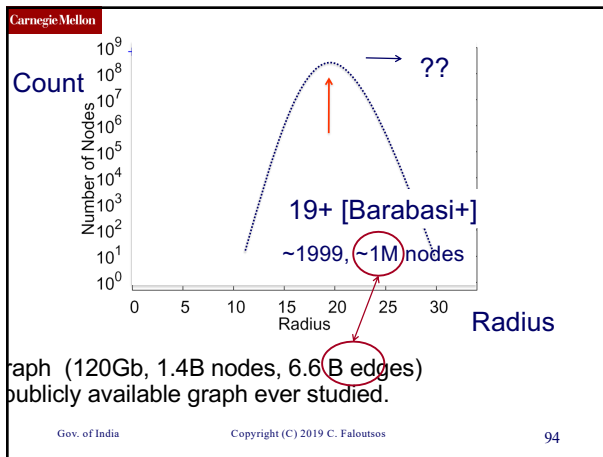
---

---

---

---

---



94

---

---

---

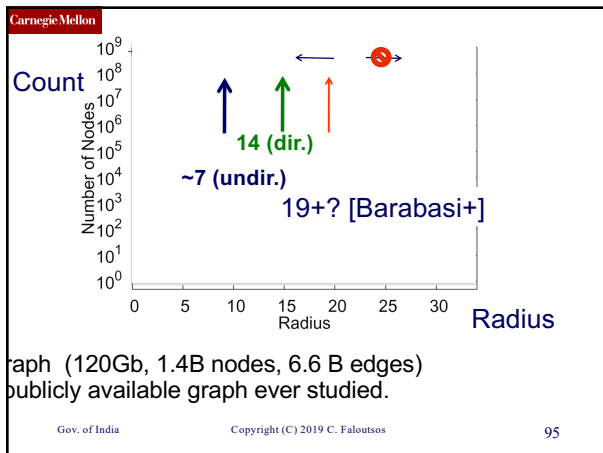
---

---

---

---

---



95

---

---

---

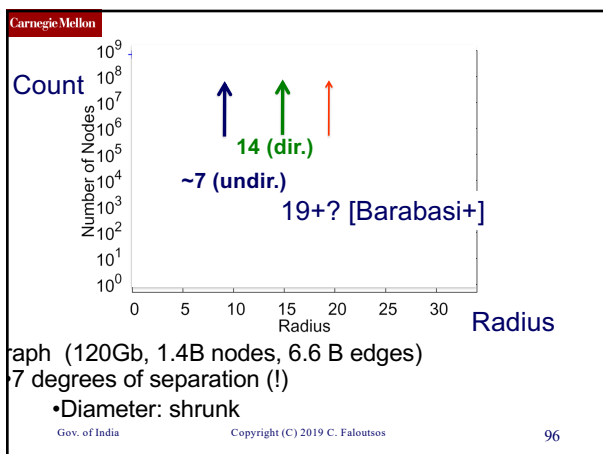
---

---

---

---

---



96

---

---

---

---

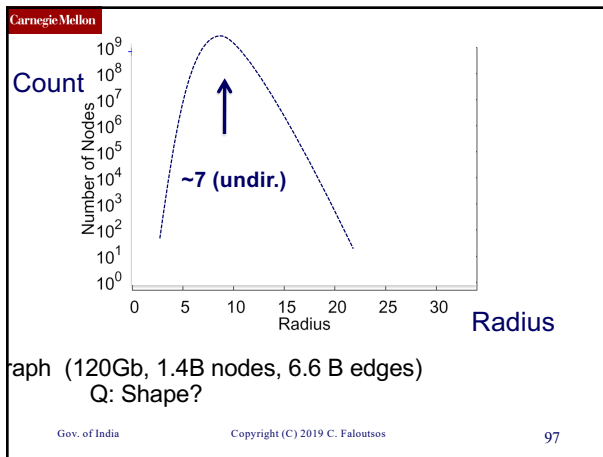
---

---

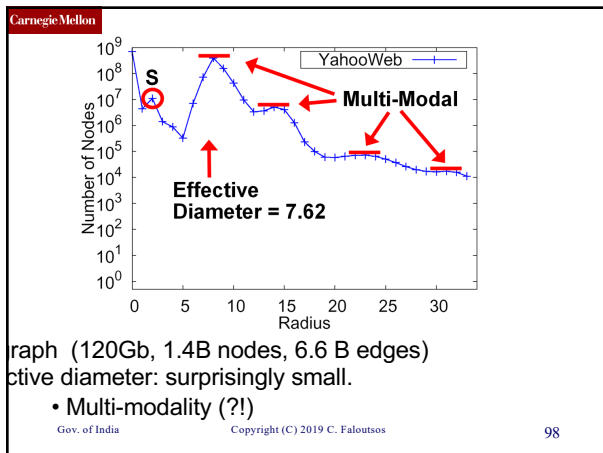
---

---

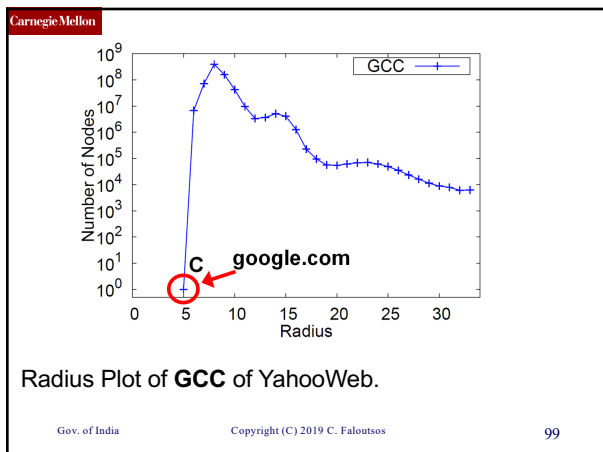




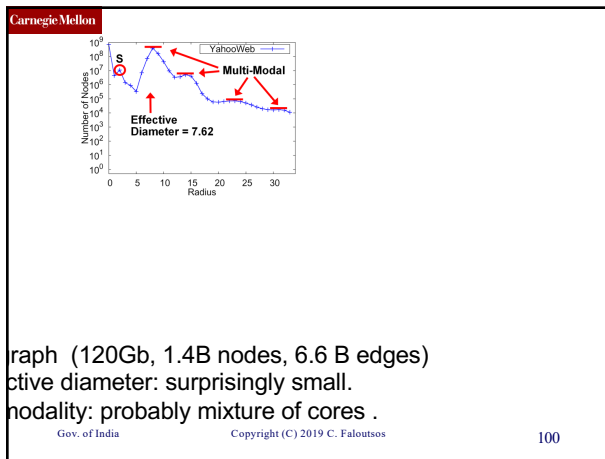
97



98



99



100

---

---

---

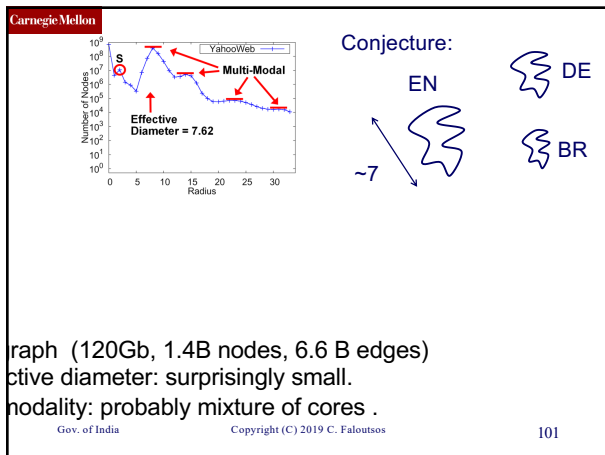
---

---

---

---

---



101

---

---

---

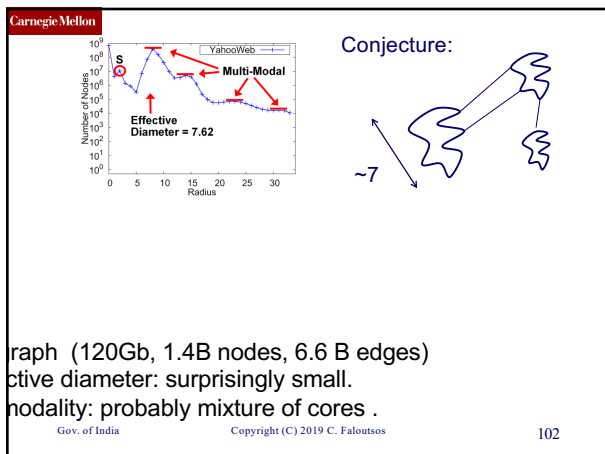
---

---

---

---

---



102

---

---

---

---

---

---


---

---

Carnegie Mellon

Outline

- Introduction – Motivation
- Problem: Patterns in graphs
  - Static graphs
    - degree, diameter, eigen,
    - Triangles
  - ➡ – Weighted graphs
  - Time evolving graphs



Gov. of India

Copyright (C) 2019 C. Faloutsos

103

---

---

---

---

---

---

---


---


103

Carnegie Mellon

Observations on weighted graphs?

- A: yes - even more ‘laws’ !





M. McGlohon, L. Akoglu, and C. Faloutsos

*Weighted Graphs and Disconnected Components: Patterns and a Generator.*

SIG-KDD 2008

Gov. of India

Copyright (C) 2019 C. Faloutsos

104

---

---

---

---

---

---

---

---

104

Carnegie Mellon

Observation W.1: Fortification

*Q: How do the weights of nodes relate to degree?*

Gov. of India

Copyright (C) 2019 C. Faloutsos

105

---

---

---

---

---

---

---

---

105

Carnegie Mellon

## Observation W.1: Fortification

**More donors, more \$ ?**

Gov. of India

Copyright (C) 2019 C. Faloutsos

106

106

---

---

---

---

---

---

---

---

Carnegie Mellon

## Observation W.1: fortification: Snapshot Power Law

- Weight: super-linear on in-degree
- exponent 'iw':  $1.01 < iw < 1.26$

**More donors, even more \$**

Gov. of India

Copyright (C) 2019 C. Faloutsos

107

107

---

---

---

---

---

---

---

---

Carnegie Mellon

## Outline

- Introduction – Motivation
- Problem: Patterns in graphs
  - Static graphs
  - Weighted graphs
  - Time evolving graphs

Gov. of India

Copyright (C) 2019 C. Faloutsos

108

108

---

---

---

---

---

---



---

---

Carnegie Mellon

## Problem: Time evolution

- with Jure Leskovec (CMU -> Stanford)
- and Jon Kleinberg (Cornell – sabb. @ CMU)

Gov. of India Copyright (C) 2019 C. Faloutsos 109

109

---

---

---

---

---


---

---

---

Carnegie Mellon

## List of Dynamic Patterns



- D.1 diameter
- D.2 densification
- D.3 gelling point
- D.4 NLCC over time
- D.5 Eigenvalue over time
- D.6 Popularity over time
- D.7 phonecall duration

Gov. of India Copyright (C) 2019 C. Faloutsos 110

110

---

---

---

---

---

---

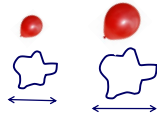
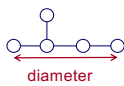
---

---

Carnegie Mellon

## D.1 Evolution of the Diameter

- Prior work on Power Law graphs hints at **slowly growing diameter**:
  - [diameter  $\sim O(N^{1/3})$ ]
  - diameter  $\sim O(\log N)$
  - diameter  $\sim O(\log \log N)$
- What is happening in real data?

Gov. of India Copyright (C) 2019 C. Faloutsos 111

111

---

---

---

---

---

---

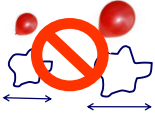
---

---

Carnegie Mellon

## D.1 Evolution of the Diameter

- Prior work on Power Law graphs hints at **slowly growing diameter**:
  - [diameter  $\sim O(N^{1/3})$ ]
  - diameter  $\sim O(\log N)$
  - diameter  $\sim O(\log \log N)$
- What is happening in real data?
- Diameter **shrinks** over time



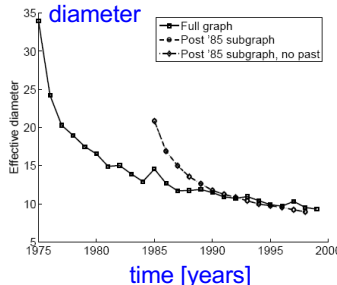
Gov. of India Copyright (C) 2019 C. Faloutsos 112

112

Carnegie Mellon

## D.1 Diameter – “Patents”

- Patent citation network
- 25 years of data
- @1999
  - 2.9 M nodes
  - 16.5 M edges



Gov. of India Copyright (C) 2019 C. Faloutsos 113

113

Carnegie Mellon

## List of Dynamic Patterns

- ✓ D.1 diameter
- D.2 densification
- D.3 gelling point
- D.4 NLCC over time
- D.5 Eigenvalue over time
- D.6 Popularity over time
- D.7 phonecall duration



Gov. of India Copyright (C) 2019 C. Faloutsos 114

114

Carnegie Mellon

## D.2 Temporal Evolution of the Graphs

- $N(t)$  ... nodes at time  $t$
- $E(t)$  ... edges at time  $t$
- Suppose that
 
$$N(t+1) = 2 * N(t)$$
- Q: what is your guess for
 
$$E(t+1) = ? 2 * E(t)$$

Gov. of India

Copyright (C) 2019 C. Faloutsos

115

115

---

---

---

---

---

---

---

---

Carnegie Mellon

## D.2 Temporal Evolution of the Graphs

- $N(t)$  ... nodes at time  $t$
- $E(t)$  ... edges at time  $t$
- Suppose that
 
$$N(t+1) = 2 * N(t)$$
- Q: what is your guess for
 
$$E(t+1) = ? 2 * E(t)$$
- A: over-doubled!
  - But obeying the ``Densification Power Law''

Gov. of India

Copyright (C) 2019 C. Faloutsos

116

116

---

---

---

---

---

---

---

---

Carnegie Mellon

## D.2 Densification – Patent Citations

- Citations among patents granted
- @1999
  - 2.9 M nodes
  - 16.5 M edges
- Each year is a datapoint

Gov. of India

Copyright (C) 2019 C. Faloutsos

117

117

---

---

---

---


---

---

---

---

Carnegie Mellon

List of Dynamic Patterns

- ✓ • D.1 diameter
- ✓ • D.2 densification
  - D.3 gelling point
  - D.4 NLCC over time
  - D.5 Eigenvalue over time
  - D.6 Popularity over time
  - D.7 phonecall duration

Gov. of IndiaCopyright (C) 2019 C. Faloutsos118

118

---

---

---

---

---

---

---

---

Carnegie Mellon

More on Time-evolving graphs

M. McGlohon, L. Akoglu, and C. Faloutsos  
*Weighted Graphs and Disconnected Components: Patterns and a Generator.*  
SIG-KDD 2008

Gov. of IndiaCopyright (C) 2019 C. Faloutsos119

119

---

---

---

---

---

---

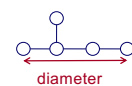
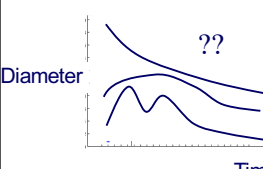
---

---

Carnegie Mellon

D.3 Gelling Point

- Diameter, over time



Gov. of IndiaCopyright (C) 2019 C. Faloutsos120

120

---

---

---

---

---

---

---

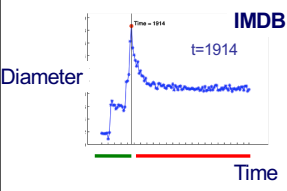
---



Carnegie Mellon

### D.3 Gelling Point

- Most real graphs display a gelling point
- After gelling point, they exhibit typical behavior. This is marked by a spike in diameter.



Gov. of India Copyright (C) 2019 C. Faloutsos 121

121

---

---

---

---

---

---

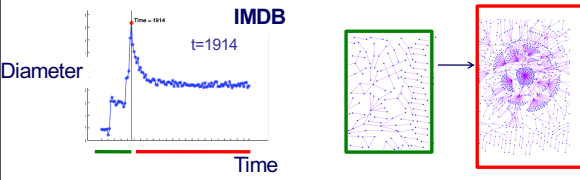
---

---

Carnegie Mellon

### D.3 Gelling Point

- Most real graphs display a gelling point
- After gelling point, they exhibit typical behavior. This is marked by a spike in diameter.



Gov. of India Copyright (C) 2019 C. Faloutsos 122

122

---

---

---

---

---

---

---

---

Carnegie Mellon

### List of Dynamic Patterns

- ✓ • D.1 diameter
- ✓ • D.2 densification
- ✓ • D.3 gelling point
- D.4 NLCC over time
- D.5 Eigenvalue over time
- D.6 Popularity over time
- D.7 phonecall duration



Gov. of India Copyright (C) 2019 C. Faloutsos 123

123

---

---

---

---

---

---

---

---


Carnegie Mellon

### Observation D.4: NLCC behavior

*Q: How do NLCC's emerge and join with the GCC?*

('NLCC' = non-largest conn. components)

- Do they continue to grow in size?
- or do they shrink?
- or stabilize?



Gov. of India Copyright (C) 2019 C. Faloutsos 124

124

---

---

---

---

---

---

---

---


Carnegie Mellon

### Observation D.4: NLCC behavior

*Q: How do NLCC's emerge and join with the GCC?*

('NLCC' = non-largest conn. components)

- Do they continue to grow in size?
- or do they shrink?
- or stabilize?



Gov. of India Copyright (C) 2019 C. Faloutsos 125

125

---

---

---

---

---

---

---

---

Carnegie Mellon

### Observation D.4: NLCC behavior

*Q: How do NLCC's emerge and join with the GCC?*

('NLCC' = non-largest conn. components)

**YES** - Do they continue to grow in size?

**YES** - or do they shrink?

**YES** - or stabilize?

Gov. of India Copyright (C) 2019 C. Faloutsos 126

126

---

---

---

---

---

---

---

---

Carnegie Mellon

### Observation D.4: NLCC behavior

- After the gelling point, the GCC takes off, but NLCC's remain ~constant (actually, **oscillate**).

CC size

IMDB

Gov. of India Copyright (C) 2019 C. Faloutsos 127

127

---

---

---

---

---

---

---

---

Carnegie Mellon

### List of Dynamic Patterns

- ✓ D.1 diameter
- ✓ D.2 densification
- ✓ D.3 gelling point
- ✓ D.4 NLCC over time
- ~~D.5 Eigenvalue over time~~
- D.6 Popularity over time
- D.7 phonecall duration

Gov. of India Copyright (C) 2019 C. Faloutsos 128

128

---

---

---

---

---

---

---

---

Carnegie Mellon

### Timing for Blogs

*Cascading Behavior in Large Blog Graphs: Patterns and a model*

Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, Matthew Hurst

SDM'07

Gov. of India Copyright (C) 2019 C. Faloutsos 129

129

---

---

---

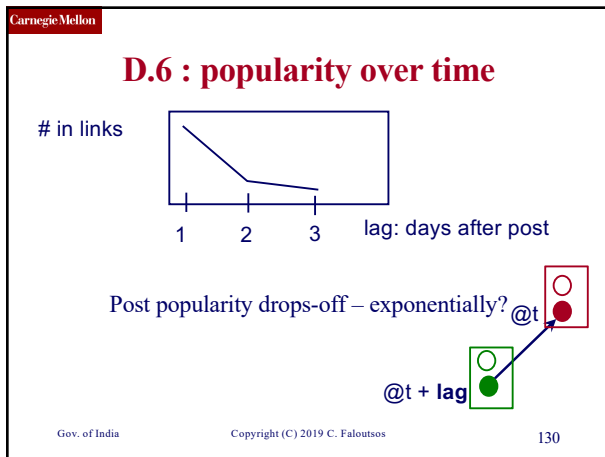
---

---

---

---

---



130

---

---

---

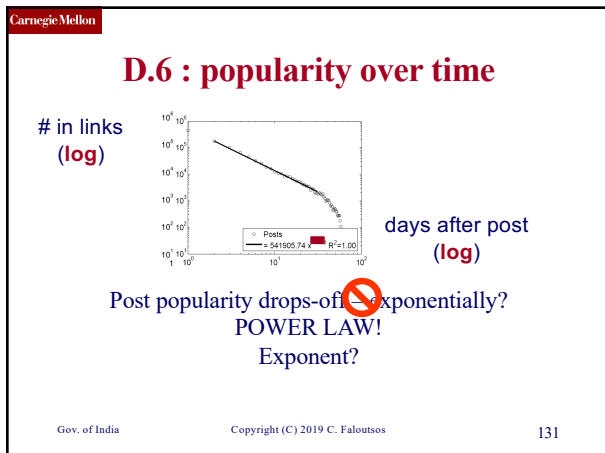
---

---

---

---

---



131

---

---

---

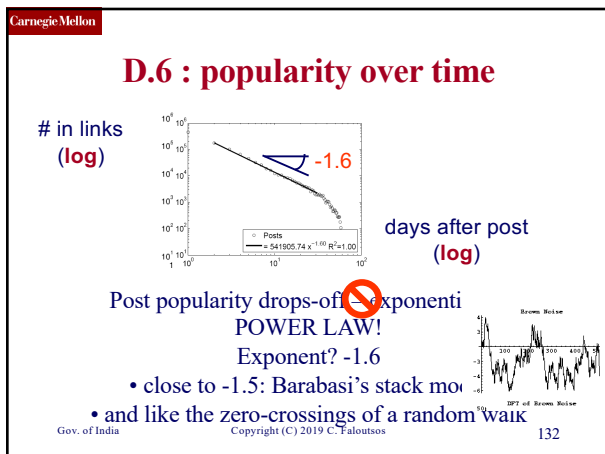
---

---

---

---

---



132

---

---

---

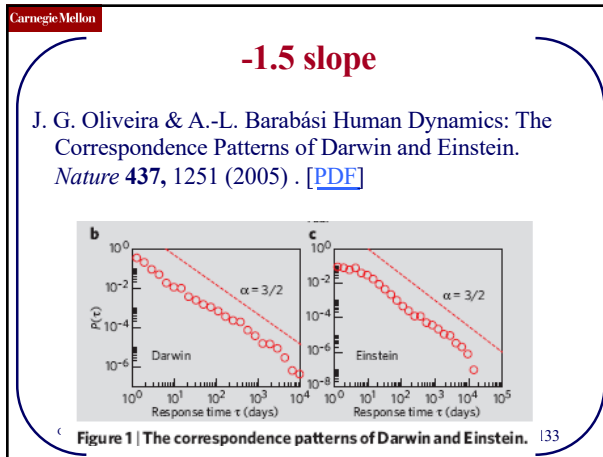
---

---

---

---

---



133

Carnegie Mellon

**List of Dynamic Patterns**

- ✓ • D.1 diameter
- ✓ • D.2 densification
- ✓ • D.3 gelling point
- ✓ • D.4 NLCC over time
- ~~D.5 Eigenvalue over time~~
- ✓ • D.6 Popularity over time
- D.7 phonecall duration

Gov. of India Copyright (C) 2019 C. Faloutsos 134

134

Carnegie Mellon

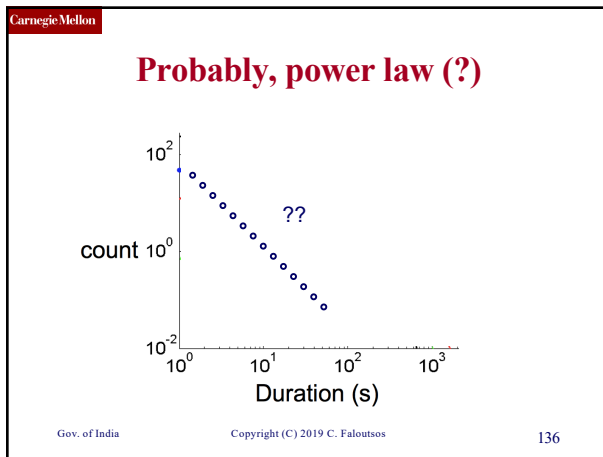
**D.7: duration of phonecalls**

*Surprising Patterns for the Call Duration Distribution of Mobile Phone Users*

Pedro O. S. Vaz de Melo, Leman Akoglu, Christos Faloutsos, Antonio A. F. Loureiro  
 PKDD 2010

Gov. of India Copyright (C) 2019 C. Faloutsos 135

135



136

---

---

---

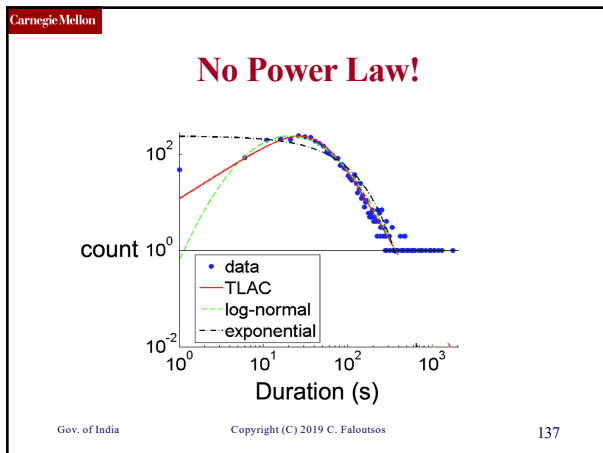
---

---

---

---

---



137

---

---

---

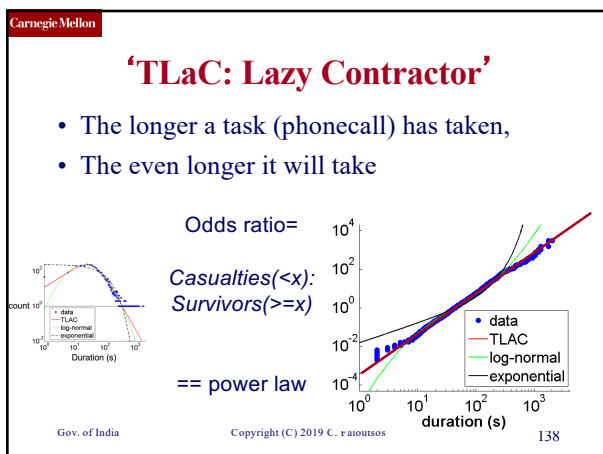
---

---

---

---

---



138

---

---

---

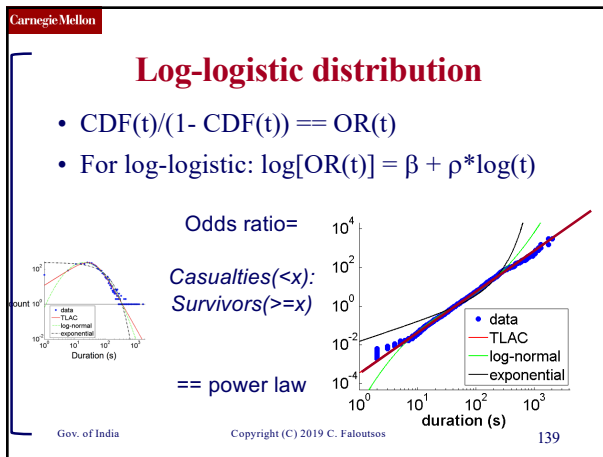
---

---

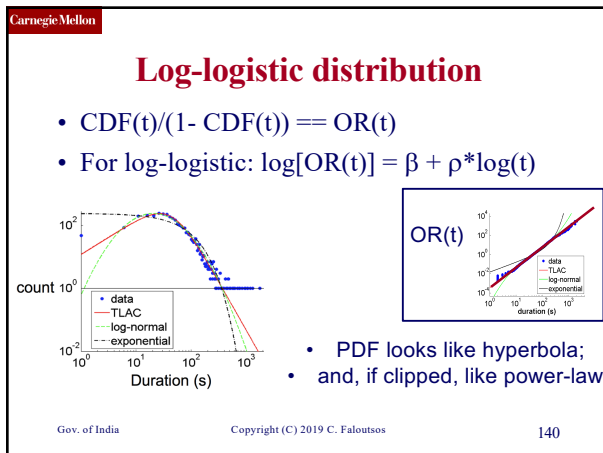
---

---

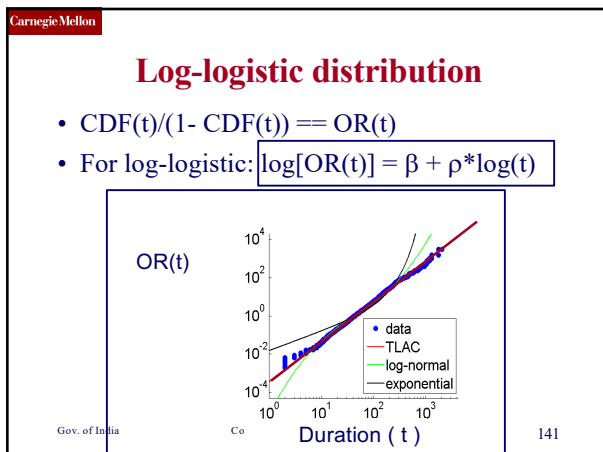
---



139



140



141

Carnegie Mellon

Attention:  
Phase I

## Log-logistic distribution

Nice 1 page description: section II of

Pravallika Devineni, Danai Koutra, Michalis Faloutsos, and Christos Faloutsos.  
[\*If walls could talk: Patterns and anomalies in Facebook wallposts.\*](#)  
*ASONAM 2015, pp 367-374.*

Gov. of India

Copyright (C) 2019 C. Faloutsos

142

---

---

---

---

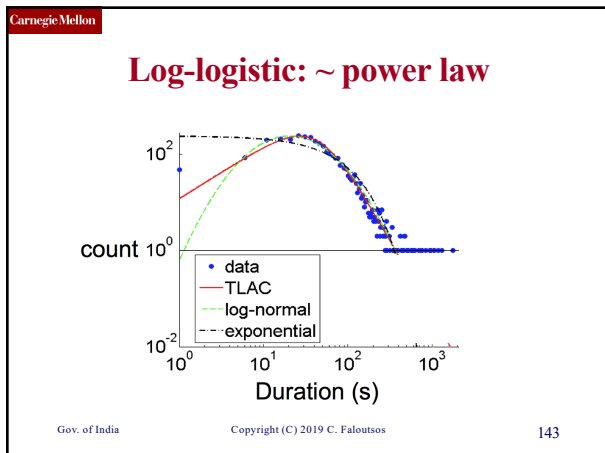
---

---

---

---

142




---

---

---

---

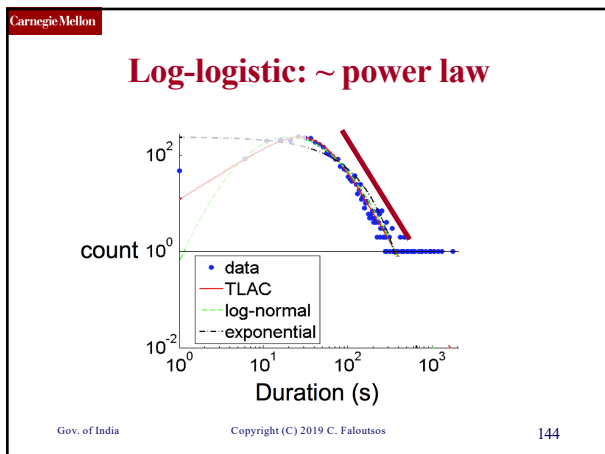
---

---

---

---

143




---

---

---

---

---

---

---

---

144



Carnegie Mellon

## Data Description

- Data from a private mobile operator of a large city
  - 4 months of data
  - 3.1 million users
  - more than 1 billion phone records
- Over 96% of ‘talkative’ users obeyed a TLAC distribution (‘talkative’: >30 calls)

Gov. of India Copyright (C) 2019 C. Faloutsos 145

145

---

---

---

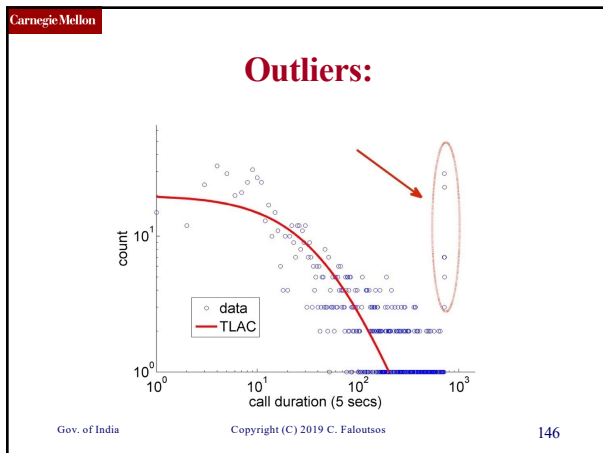
---

---

---

---

---



146

---

---

---

---

---

---

---

---

Carnegie Mellon

## Short answer(s)

- Are real graphs random?
  - S\*: what do **static** graphs look like?
    - S.0: ‘six degrees’
    - S.1: skewed degree distribution
    - S.2: skewed eigenvalues
    - S.3: triangle power-laws
    - S.4: GCC; and skewed distr. of conn. comp.
  - D\*: how do graphs evolve over **time**?
    - D.1: diameters
    - D.2: densification

Gov. of India Copyright (C) 2019 C. Faloutsos 147

147

---

---

---

---



---

---

---


---

Carnegie Mellon

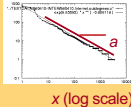
### Short answer

- Are real graphs random?
  - S\*: what do we see?



## Power laws: $y \sim x^a$

- NOT Gaussians
- Take logarithms



(log scale)

x (log scale)

- Power laws
- Power-law; and skewed distr. of conn. comp.
- T\*: how do graphs evolve over time?
  - T.1: diameters
  - T.2: densification

Gov. of India

Copyright (C) 2019 C. Faloutsos

148

---

---

---

---

---

---

---

---

148

Carnegie Mellon

## References

- Deepayan Chakrabarti and Christos Faloutsos [\*Graph Mining: Laws, Tools and Case Studies\*](#), Morgan Claypool, 2012
  - Part I (patterns)

Gov. of India

Copyright (C) 2019 C. Faloutsos

149

---

---

---

---

---

---

---

---

149

Carnegie Mellon

## References

- Mary McGlohon, Leman Akoglu, Christos Faloutsos. *Statistical Properties of Social Networks*. in "Social Network Data Analytics" (Ed.: Charu Aggarwal)
- Deepayan Chakrabarti and Christos Faloutsos, [\*Graph Mining: Laws, Tools, and Case Studies\*](#) Oct. 2012, Morgan Claypool.






Gov. of India

Copyright (C) 2019 C. Faloutsos

150

---

---

---

---

---

---

---

---

150

Carnegie Mellon

References, cont'd

- Michalis Faloutsos, Petros Faloutsos and Christos Faloutsos, On Power-Law Relationships of the Internet Topology, SIGCOMM 1999.
- R. Albert, H. Jeong, and A.-L. Barabasi, Diameter of the World Wide Web Nature, 401, 130-131 (1999).
- Reka Albert and Albert-Laszlo Barabasi Statistical mechanics of complex networks, Reviews of Modern Physics, 74, 47 (2002).
- Jure Leskovec, Jon Kleinberg, Christos Faloutsos Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations, KDD 2005, Chicago, IL, USA

Gov. of India

Copyright (C) 2019 C. Faloutsos

151

151

---

---

---

---

---

---

---

Carnegie Mellon

Project info



[www.cs.cmu.edu/~pegasus](http://www.cs.cmu.edu/~pegasus)



Chau,  
Polo

McGlohon,  
Mary

Tsourakakis,  
Babis



Akoglu, Leman      Kang, U      Prakash, Aditya      Tong, Hanghang

Thanks to: NSF IIS-0705359, IIS-0534205, CTA-INARC; Yahoo (M45), LLNL, IBM, SPRINT, INTEL, HP

152

---

---

---

---

---

---

---