

Carnegie Mellon


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

Part 3: Extras - Visualization,
explanation, influence propagation

Christos Faloutsos

CMU SCS


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



1

Carnegie Mellon

Roadmap



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

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

2

2

Carnegie Mellon

Outline - Extras

- 3.1: Visualization / Explanation
 - ➡ - 3.1.1 SVD
 - 3.1.2 Social network
 - 3.1.3 point processes
- 3.2: Virus/influence propagation

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



3

3

Carnegie Mellon

NBA dataset
~500 players;
~30 attributes

Any patterns?
Clusters?
Outliers?

Gov. of India

Copyright (C) 2019 C. Faloutsos

4

4

Carnegie Mellon

PCA - ‘Ratio Rules’

[Korn+98]

Typically: ‘*Association Rules*’ (eg.,
{bread, milk} -> {butter})

But, can we discover more details? like:
\$-bread : \$-milk : \$-butter ~ \$2 : \$4 : \$3

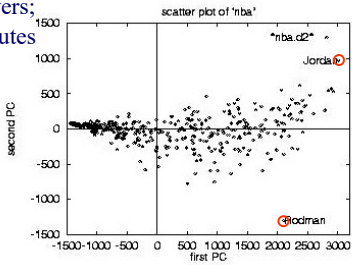
Flip Korn, Alexandros Labrinidis, Yannis Kotidis, and Christos Faloutsos. *Ratio Rules: A New Paradigm for Fast, Quantifiable Data Mining*. (VLDB '98), 582-593.

5

Carnegie Mellon

PCA - Ratio Rules

NBA dataset
~500 players;
~30 attributes



Gov. of India

Copyright (C) 2019 C. Faloutsos



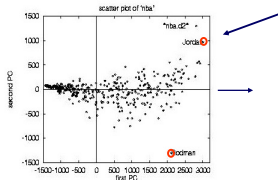
6

6

Carnegie Mellon

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?



Gov. of India

Copyright (C) 2019 C. Faloutsos

7

7

Carnegie Mellon

PCA - Ratio Rules

- PCA: get singular vectors v_1, v_2, \dots
- ignore entries with small abs. value
- try to interpret the rest

Gov. of India

Copyright (C) 2019 C. Faloutsos

8

8

Carnegie Mellon

PCA - Ratio Rules

NBA dataset - V matrix (term to 'concept' similarities)

field	RR_1	RR_2	RR_3
minutes played	.808	-.4	
field goals			
goal attempts			
points	.406	.199	
total rebounds		-.489	.602
assists			-.486
steals			-.07

Gov. of India

Copyright (C) 2019 C. Faloutsos

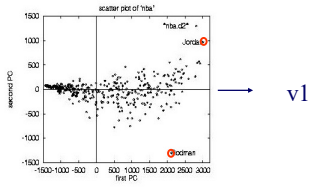
9

9

Carnegie Mellon

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?



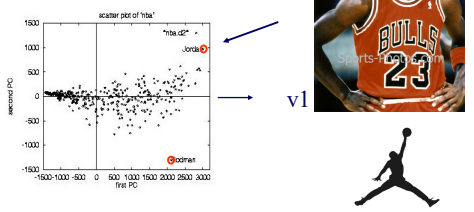
Gov. of India Copyright (C) 2019 C. Faloutsos 10

10

Carnegie Mellon

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?



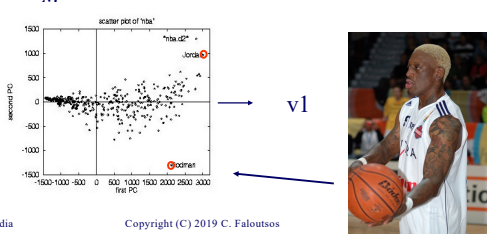
Gov. of India Copyright (C) 2019 C. Faloutsos 11

11

Carnegie Mellon

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?



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

12

Carnegie Mellon

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?
- A: ‘goodness’ of player

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

13

Carnegie Mellon

Ratio Rules - example

- RR2: points: rebounds negatively correlated(!)

field	RR_1	RR_2	RR_3
minutes played	.808	-.4	
field goals			
goal attempts			
points	.406	.199	
total rebounds		-.489	.602
assists			-.486
steals			-.07

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

14

Carnegie Mellon

Ratio Rules - example

- RR2: points: rebounds negatively correlated(!) - concept?

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

15

Carnegie Mellon

Ratio Rules - example

- RR2: points: rebounds negatively correlated(!) - concept?
- A: position: offensive/defensive

Gov. of India Copyright (C) 2019 C. Faloutsos 16

16

Carnegie Mellon

More datasets

- Customer activity over time
- Stock prices over time

Gov. of India Copyright (C) 2019 C. Faloutsos 17

17

Carnegie Mellon

Compression - Visualization

- no Gaussian clusters; Zipf-like distribution

(a) 'phone2000' (b) 'stocks'

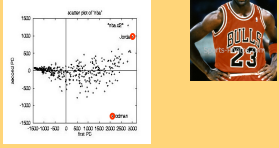
Gov. of India Copyright (C) 2019 C. Faloutsos 18

18

Carnegie Mellon

Solutions

- SVD helps with visualization (2-d - low-dim scatterplots)
- ... and finds rules (minutes : points = 2:1)



Gov. of India Copyright (C) 2019 C. Faloutsos 19

19

Carnegie Mellon

Outline - Extras

- 3.1: Visualization / Explanation
 - 3.1.1 SVD
 - ➔ – 3.1.2 Social network
 - 3.1.3 point processes
- 3.2: Virus/influence propagation

Gov. of India Copyright (C) 2019 C. Faloutsos 20


20

Carnegie Mellon

Catchsync: catch synchronized behavior in large directed graphs

Meng Jiang

Joint work with Peng Cui, Alex Beutel,
Christos Faloutsos and Shiqiang Yang
August 26, 2014 – NYC, USA



Gov. of India Copyright (C) 2019 C. Faloutsos 21

21

Fraud Detection: Graph Analysis Problem

Twitter [www.buyfollowz.org]

5,000 FOLLOWERS	2,000 FOLLOWERS	1,000 FOLLOWERS	10,000 FOLLOWERS	20,000 FOLLOWERS
\$69.99	\$29.99	\$15.99	\$119.99	\$229.99

facebook [buymorelikes.com]

25,000 Facebook Likes	50,000 Facebook Likes	100,000 Facebook Likes	200,000 Facebook Likes
265	525	1,000	1,750

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

22

Fraud Detection: Graph Analysis Problem

YouTube [buycheapllikes.com]

1,000 LIKES	2,000 LIKES	5,000 LIKES	10,000 LIKES
\$1.99	\$3.99	\$9.99	\$19.99

amazon.com [reviewsteria.com]

It's easy to buy Amazon reviews. Just choose the number of reviews you would like to receive. High quality reviews that customers love. 100% unique content by native speaking professional writers.

Choose the number of reviews and click Buy Now button to ramp up your Amazon business NOW!

Choose the number of reviews: 25

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

23

Behavior-based Features

Follower behavior

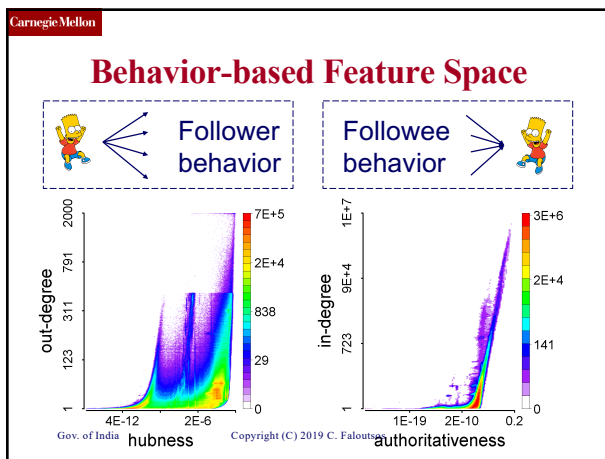
Followee behavior

Out-degree
1st left singular vector (Hubness)
2nd left singular vector

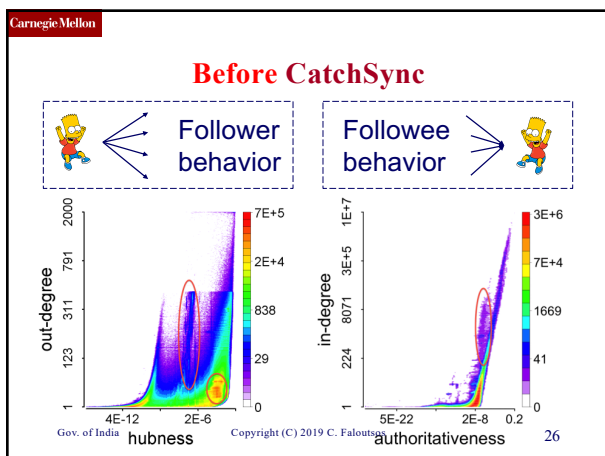
In-degree
1st right singular vector (Authoritativeness)
2nd right singular vector

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

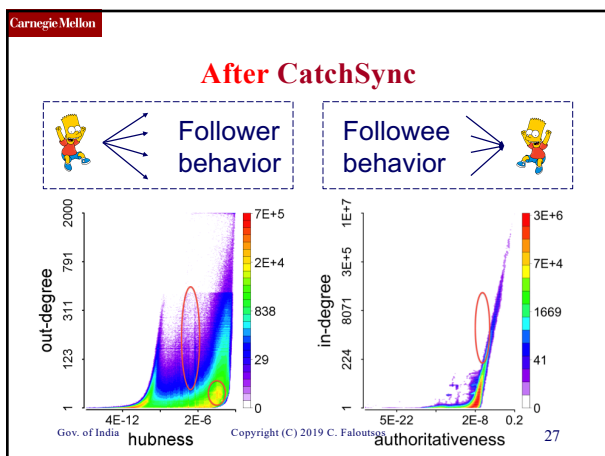
24



25



26



27

Carnegie Mellon


Outline - Extras

- 3.1: Visualization / Explanation
 - 3.1.1 SVD
 - 3.1.2 Social network
 - ➔ 3.1.3 point processes
- 3.2: Virus/influence propagation


Gov. of India
Copyright (C) 2019 C. Faloutsos
28

28


Carnegie Mellon



USP
Universidade de São Paulo




筑波大学
University of Tsukuba



Carnegie Mellon University

KDD 2015 – Sydney, Australia

RSC: Mining and Modeling Temporal Activity in Social Media



Alceu F. Costa*
Yuto Yamaguchi
Agma J. M. Traina
Caetano Traina Jr.
Christos Faloutsos

*alceufc@icmc.usp.br

29

Carnegie Mellon

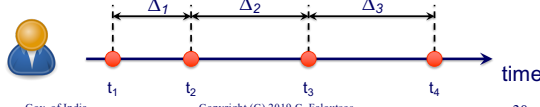
Pattern Mining: Datasets

Reddit Dataset	Twitter Dataset
Time-stamp from comments	Time-stamp from tweets
21,198 users	6,790 users
20 Million time-stamps	16 Million time-stamps

For each user we have:

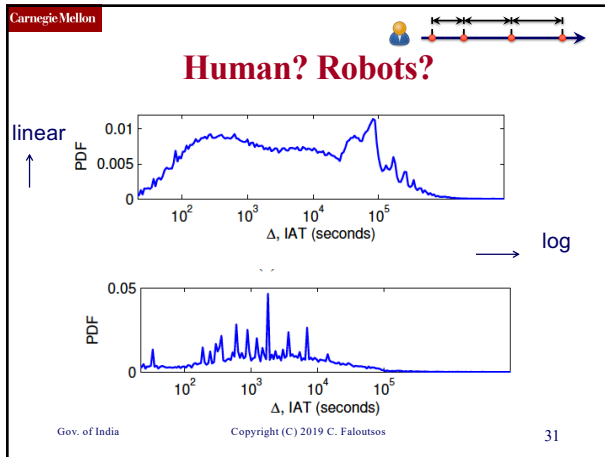
Sequence of postings time-stamps: $T = (t_1, t_2, t_3, \dots)$

Inter-arrival times (IAT) of postings: $(\Delta_1, \Delta_2, \Delta_3, \dots)$

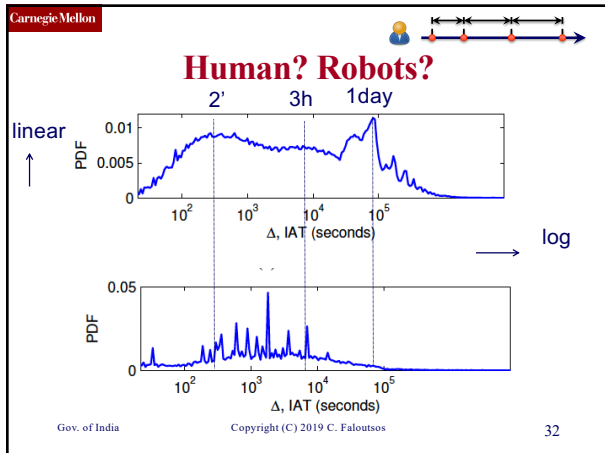


Gov. of India
Copyright (C) 2019 C. Faloutsos
30

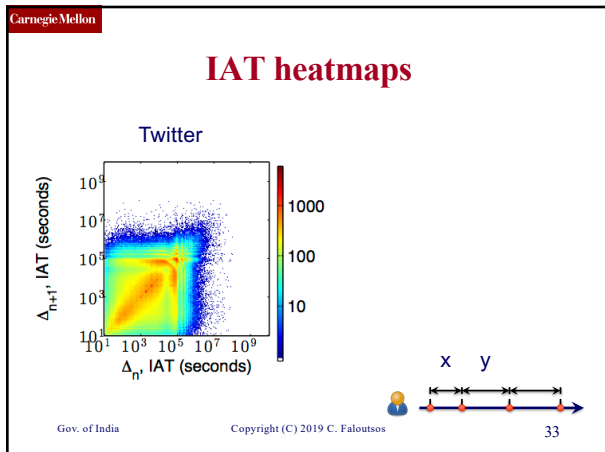
30



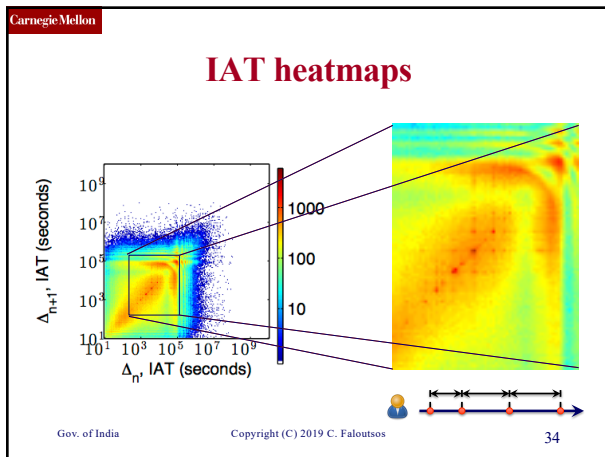
31



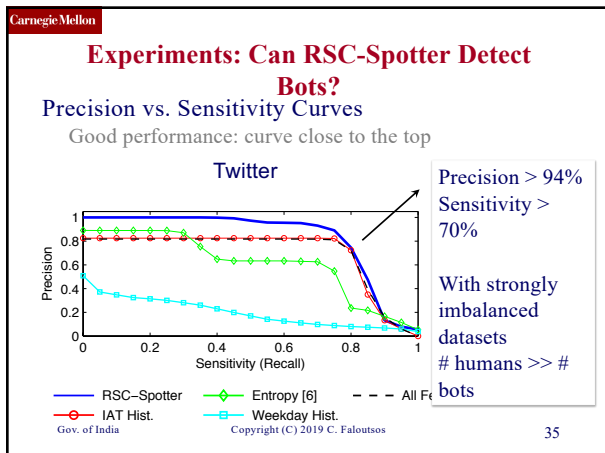
32



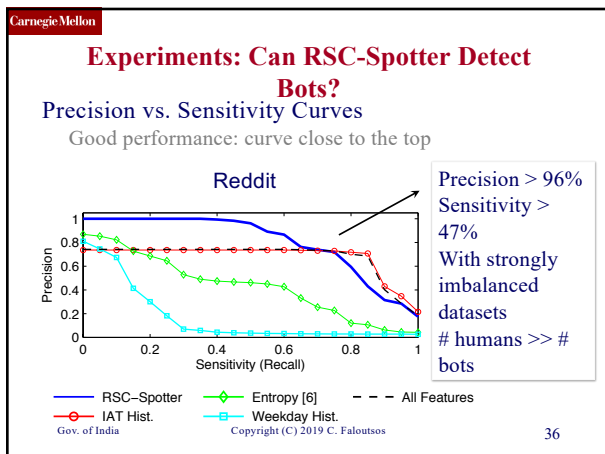
33



34



35



36

Carnegie Mellon

Outline - Extras

- 3.1: Visualization / Explanation
 - 3.1.1 SVD
 - 3.1.2 Social network
 - 3.1.3 point processes
- ➔ • 3.2: Virus/influence propagation

Gov. of India

Copyright (C) 2019 C. Faloutsos

37

37

Carnegie Mellon

Cascades & Immunization

Gov. of India

Copyright (C) 2019 C. Faloutsos

38

38

Carnegie Mellon

Why do we care?

- Information Diffusion
- Viral Marketing
- Epidemiology and Public Health
- Cyber Security
- Human mobility
- Games and Virtual Worlds
- Ecology
-



Gov. of India

Copyright (C) 2019 C. Faloutsos


39

39

Carnegie Mellon

Roadmap

- P3.1: visualization
- P3.2: Cascade analysis
 - ➡ – (Fractional) Immunization
 - Epidemic thresholds
- Conclusions




Gov. of IndiaCopyright (C) 2019 C. Faloutsos40


40

Carnegie Mellon


Fractional Immunization of Networks




B. Aditya Prakash,



Lada Adamic,



Theodore Iwashyna (M.D.),



Hanghang Tong,

Christos Faloutsos

SDM 2013, Austin, TX

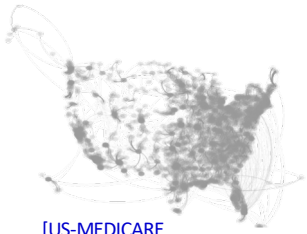
Gov. of IndiaCopyright (C) 2019 C. Faloutsos41

41

Carnegie Mellon

Whom to immunize?

- Dynamical Processes over networks



- Each circle is a hospital
- ~3,000 hospitals
- More than 30,000 patients transferred

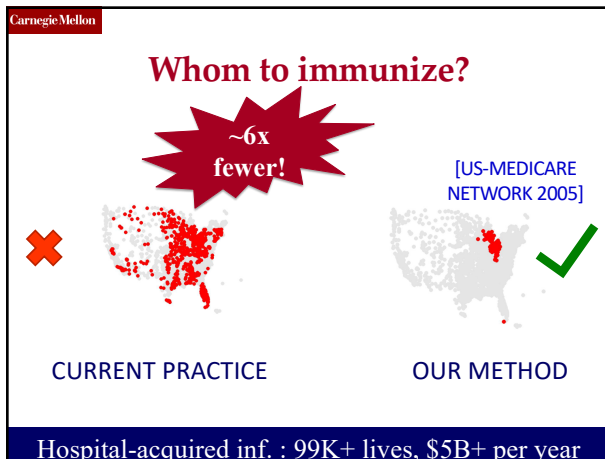
[US-MEDICARE NETWORK 2005]

Problem: Given k units of disinfectant, whom to immunize?

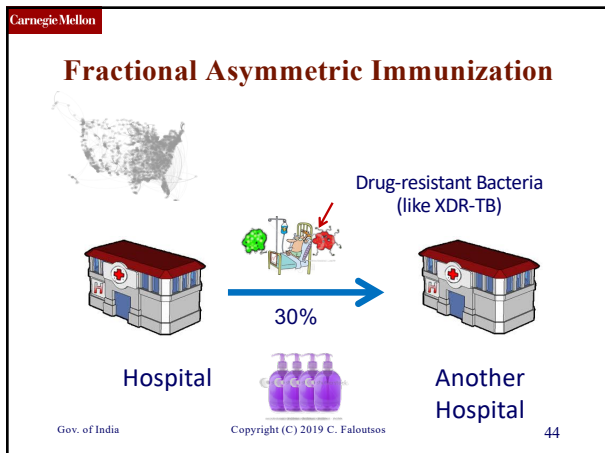
Gov. of IndiaCopyright (C) 2019 C. Faloutsos42

42

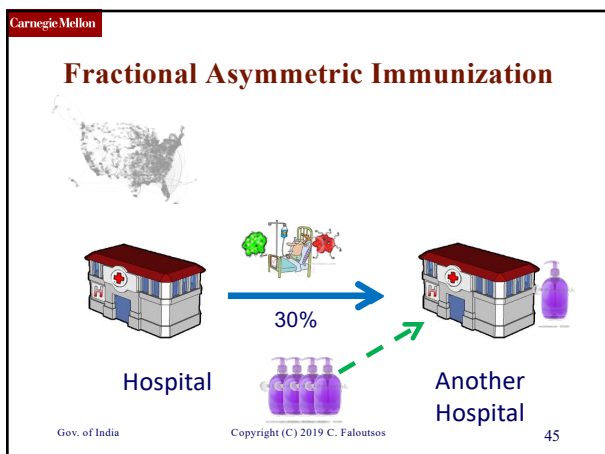
14



43



44



45

Carnegie Mellon

Fractional Asymmetric Immunization

Gov. of India

Copyright (C) 2019 C. Faloutsos

46

46

Carnegie Mellon

Fractional Asymmetric Immunization

Problem:
Given k units of disinfectant,
distribute them
to maximize hospitals saved

Gov. of India

Copyright (C) 2019 C. Faloutsos

47

47

Carnegie Mellon

Fractional Asymmetric Immunization

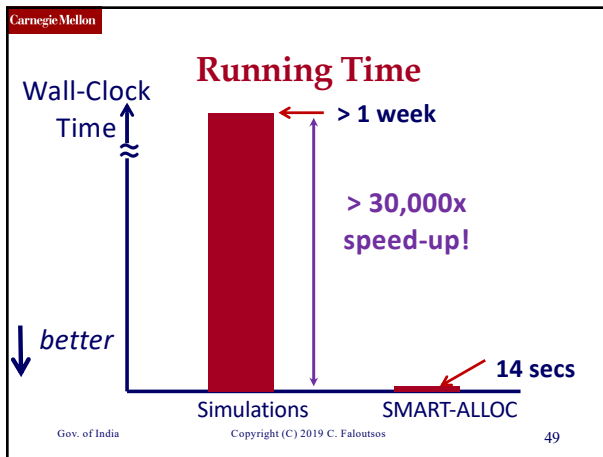
Problem:
Given k units of disinfectant,
distribute them
to maximize hospitals saved @ 365 days

Gov. of India

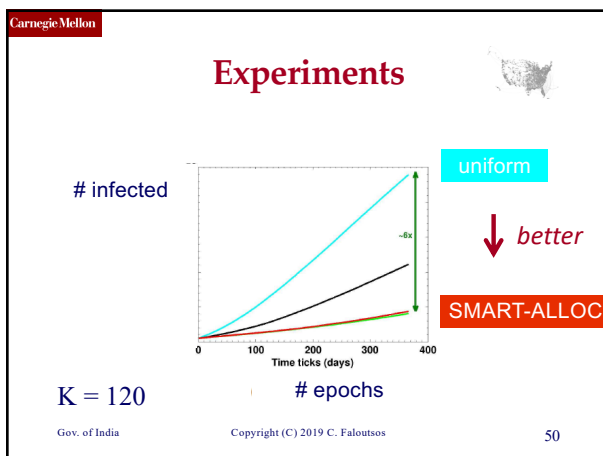
Copyright (C) 2019 C. Faloutsos

48

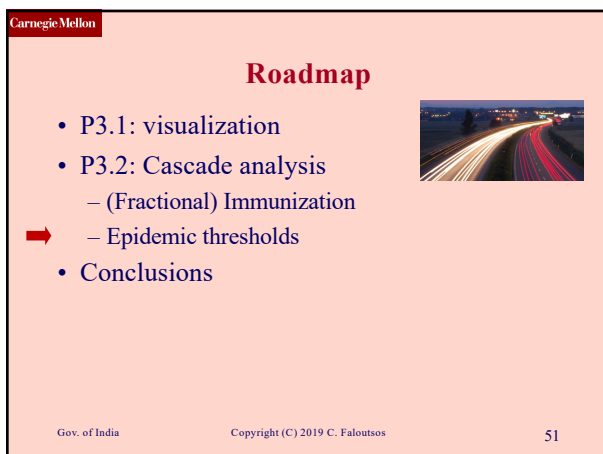
48



49



50



51

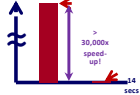
Carnegie Mellon

What is the 'silver bullet'?

A: Try to decrease connectivity of graph

Q: how to measure connectivity?

- Avg degree? Max degree?
- Std degree / avg degree ?
- Diameter?
- Modularity?
- 'Conductance' (~min cut size)?
- Some combination of above?



Gov. of India Copyright (C) 2019 C. Faloutsos 52

52

Carnegie Mellon

What is the 'silver bullet'?

A: Try to decrease connectivity of graph

Q: how to measure connectivity?

A: first **eigenvalue** of adjacency matrix

Q1: why??

(Q2: dfn & intuition of eigenvalue ?)

Avg degree
Max degree
Diameter
Modularity
'Conductance'

Gov. of India Copyright (C) 2019 C. Faloutsos 53

53


Carnegie Mellon

Why eigenvalue?

A1: 'G2' theorem and 'eigen-drop':

- For (almost) **any** type of virus
- For **any** network
- -> no epidemic, if small-enough first eigenvalue (λ_1) of *adjacency* matrix


Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks, B. Aditya Prakash, Deepayan Chakrabarti, Michalis Faloutsos, Nicholas Valler, Christos Faloutsos, ICDM 2011, Vancouver, Canada



Gov. of India Copyright (C) 2019 C. Faloutsos 54

54

Carnegie Mellon



Why eigenvalue?


A1: ‘G2’ theorem and ‘eigen-drop’:

- For (almost) **any** type of virus
- For **any** network
- > no epidemic, if small-enough first eigenvalue (λ_1) of *adjacency* matrix
- Heuristic: for immunization, try to min λ_1
- The smaller λ_1 , the closer to extinction.

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



55

Carnegie Mellon



G2 theorem

Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks

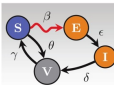



B. Aditya Prakash, Deepayan Chakrabarti,
Michalis Faloutsos, Nicholas Valler,
Christos Faloutsos
IEEE ICDM 2011, Vancouver

extended version, in arxiv
<http://arxiv.org/abs/1004.0060>
~10 pages proof

56

Carnegie Mellon



Our thresholds for some models

- $s = \text{effective strength}$
- $s < 1$: below threshold

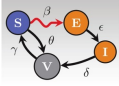
Models	Effective Strength (s)	Threshold (tipping point)
SIS, SIR, SIRS, SEIR	$s = \lambda \cdot \left(\frac{\beta}{\delta} \right)$	$s = 1$
SIV, SEIV	$s = \lambda \cdot \left(\frac{\beta\gamma}{\delta(\gamma + \theta)} \right)$	
SI ₁ I ₂ V (H.I.V.)	$s = \lambda \cdot \left(\frac{\beta_1 v_2 + \beta_2 \varepsilon}{v_2(\varepsilon + v_1)} \right)$	

57

Carnegie Mellon

Our thresholds for some models

- $s = \text{effective strength}$
- $s < 1$: below threshold




No immunity	Temp. immunity	Effective Strength	Threshold (tipping point)
SIS, SIR, SIRS, SEIR		$s = \lambda \left(\frac{\beta}{\delta} \right)$	$s = 1$
SIV, SEIV	w/ incubation	$s = \lambda \left(\frac{\beta\gamma}{\delta(\gamma + \theta)} \right)$	
SIV (H.I.V.)		$s = \lambda \left(\frac{\beta_1 v_2 + \beta_2 \epsilon}{v_2(\epsilon + v_1)} \right)$	

58

Carnegie Mellon

Roadmap

- P3.1: visualization
- P3.2: Cascade analysis
 - (Fractional) Immunization
 - Epidemic thresholds
 - Intuition behind λ_1
- Conclusions



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

59

Carnegie Mellon

Intuition for λ


“Official” definitions:

- Let A be the adjacency matrix. Then λ is the root with the largest magnitude of the characteristic polynomial of A [$\det(A - \lambda I)$].
- Also: $Ax = \lambda x$

Neither gives much intuition!

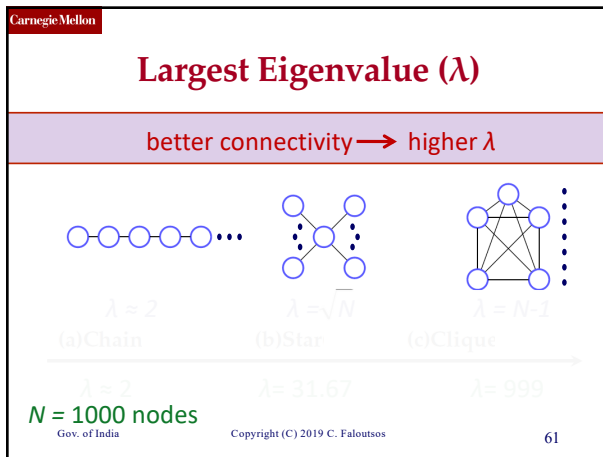
“Un-official” Intuition

- For ‘homogeneous’ graphs, $\lambda \approx \text{degree}$
- $\lambda \sim \text{avg degree}$
 - done right, for skewed degree distributions

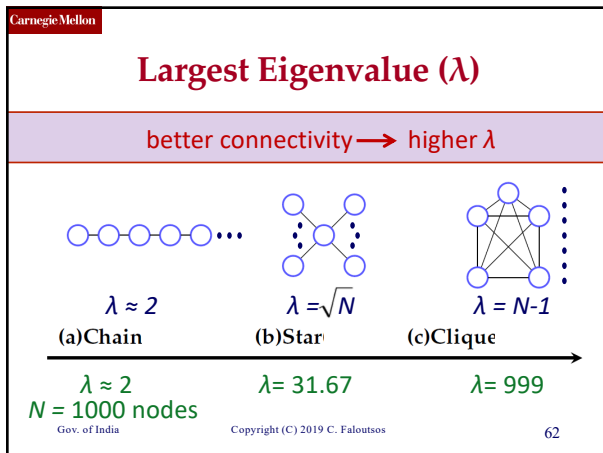


Gov. of India Copyright (C) 2019 C. Faloutsos

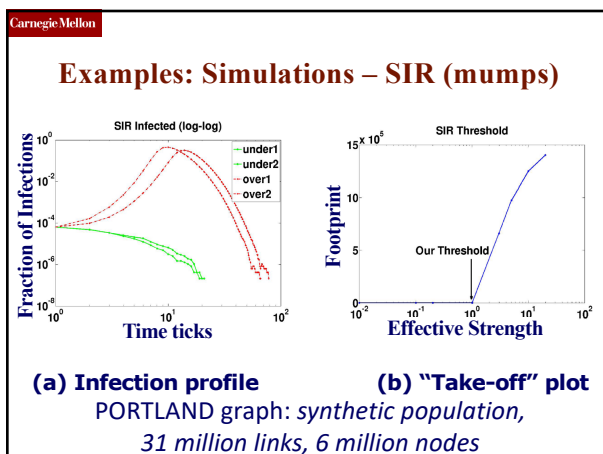
60



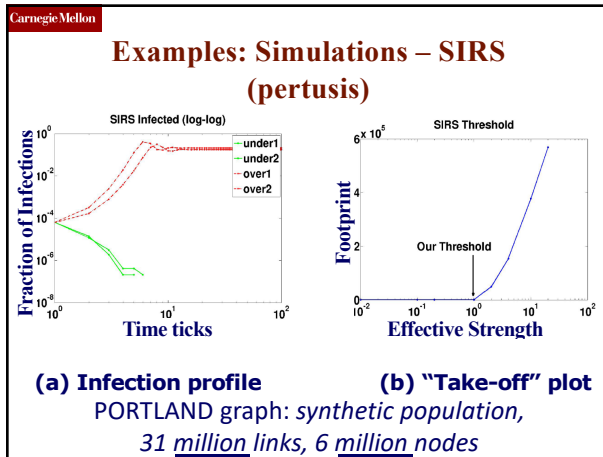
61



62



63



64

Carnegie Mellon

Part3: Immunization - conclusion

In (almost any) immunization setting,

- Allocate resources, to
- **Minimize λ_1**
- (regardless of virus specifics)

• Conversely, in a market penetration setting

- Allocate resources to
- Maximize λ_1

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

65

Carnegie Mellon

Cast

					
Akoglu, Leman	Araujo, Miguel	Beutel, Alex	Chau, Polo	Eswaran, Dhivya	Hooi, Bryan
					
Kang, U	Koutra, Danai	Papalexakis, Vagelis	Shah, Neil	Shin, Kijung	Song, Hyun Ah

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

66

Carnegie Mellon

Overall conclusions

- Graphs
- Time series
- (visualization, immunization)

Gov. of India

Copyright (C) 2019 C. Faloutsos

67

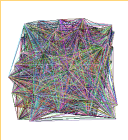
67

Carnegie Mellon

P1: Graphs

Over-arching conclusion

- MANY, time-tested, algorithms for graph mining
- (more, are needed)



Gov. of India

Copyright (C) 2019 C. Faloutsos

68


68

Carnegie Mellon


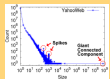


P1: Graphs

Over-arching conclusion

Problems



(some) solutions

Gov. of India


Copyright (C) 2019 C. Faloutsos

69




69

Carnegie Mellon

P2: Time series - Answers



- Similarity search: **Euclidean**/time-warping; **feature extraction** and **SAMs**
- Periodicities: **DFT/DWT**
- Linear Forecasting: **AR** (Box-Jenkins)
- Non-linear forecasting: **lag-plots**
- Gray-box modeling: **Lotka-Volterra**

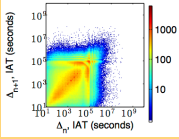

Gov. of India Copyright (C) 2019 C. Faloutsos 70

70

Carnegie Mellon

P3: visualization etc

- Plots help explain (and also they catch errors)
- SVD: super-useful tool
- Immunization: reduce λ_1


Gov. of India Copyright (C) 2019 C. Faloutsos 71

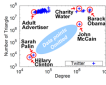

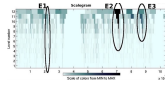
71

Carnegie Mellon

THANK YOU

christos@cs.cmu.edu



Gov. of India Copyright (C) 2019 C. Faloutsos 72

72