Carnegie Mellon

15-826: Multimedia Databases and Data Mining

Lecture #29: Graph mining -
virus propagation & immunization

Christos Faloutsos

1

Carnegie Mellon

Must-read material

- [Graph-Textbook], Ch.18: virus propagation

2

Carnegie Mellon

Main outline

- Introduction
- Indexing
- Mining
  - Graphs – patterns
  - Graphs – generators and tools
  - Association rules
  - …

15-826 Copyright (c) 2019 A. Prakash and C. Faloutsos 3

3

Carnegie Mellon

Detailed outline

- Graphs – generators
- Graphs – tools
  - Community detection / graph partitioning
  - 'Belief Propagation' & fraud detection
  - Influence/virus propagation & immunization
    - Will we have an epidemic?
    - Whom to immunize?
    - (two competing viruses – what will happen?)

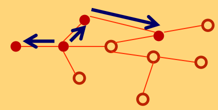
15-826 Copyright (c) 2019 A. Prakash and C. Faloutsos 4

4

Carnegie Mellon

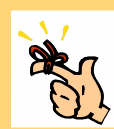
Problem

- Q1: epidemic?



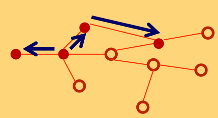
- Q2: whom to immunize
- (Q3: 2 competing viruses – end result?)

Carnegie Mellon



Short answers

- Q1: epidemic?
- A1: tipping point: eigenvalue
- Q2: whom to immunize
- A2: eigen-drop
- (Q3: 2 competing viruses – end result?)



Carnegie Mellon

Influence propagation in large graphs - theorems and algorithms

Prof. B. Aditya Prakash

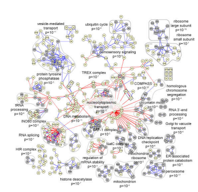
http://people.cs.vt.edu/~badityap/

7

Carnegie Mellon

Networks are everywhere!

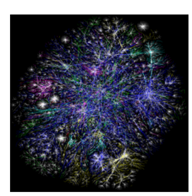
Facebook Network [2010]



Gene Regulatory Network [Decourty 2008]

Human Disease Network [Barabasi 2007]

The Internet [2005]

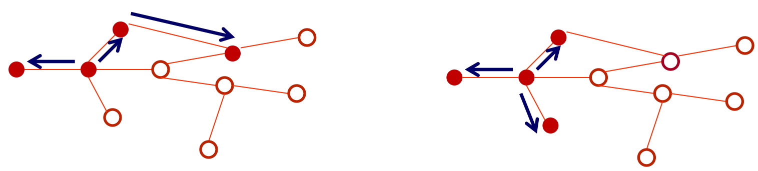


15-826

8

8

Carnegie Mellon



Dynamical Processes ***over*** networks are also everywhere!

Carnegie Mellon

Why do we care?

Carnegie Mellon

Why do we care?

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

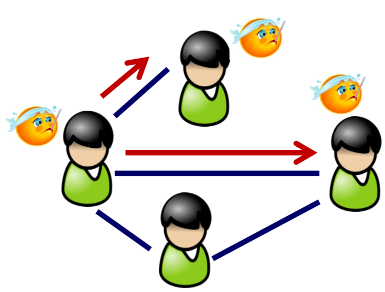
15-826 Copyright (c) 2019 A. Prakash and C. Faloutsos 11

11

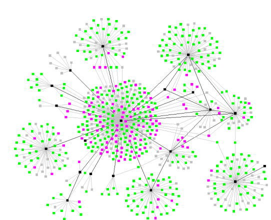
Carnegie Mellon

Why do we care? (1: Epidemiology)

- Dynamical Processes over networks [AJPH 2007]



Diseases over contact networks



CDC data: Visualization of the first 35 tuberculosis (TB) patients and their 1039 contacts

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

12

Carnegie Mellon

Why do we care? (2: Online Diffusion)

> 800m users, ~$1B revenue [WSJ 2010]

facebook

~100m active users

twitter

> 50m users

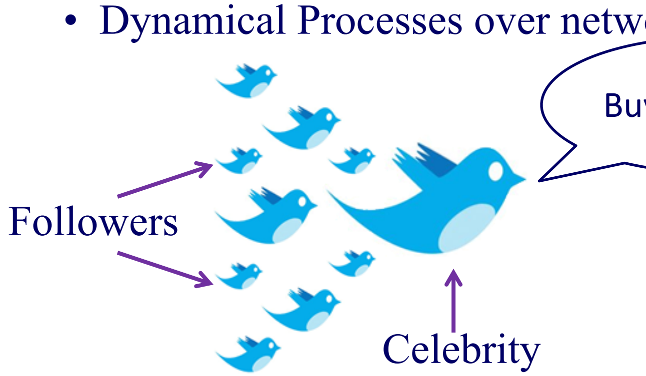
LinkedIn

13

Carnegie Mellon

Why do we care? (2: Online Diffusion)

- Dynamical Processes over networks



Social Media Marketing

14

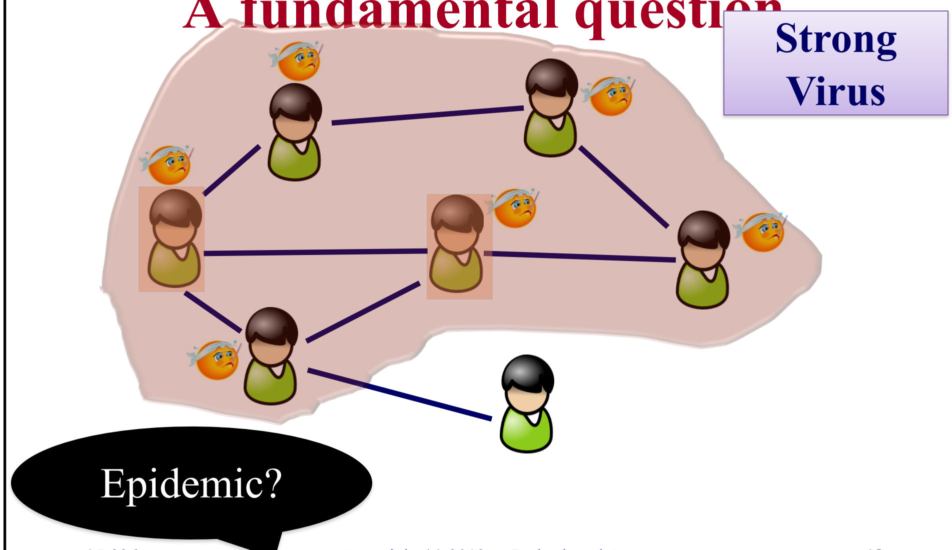
Carnegie Mellon

Outline

- Motivation
- **Q1: Epidemics: what happens? (Theory)**
- Q2: Action: Whom to immunize? (Algorithms)

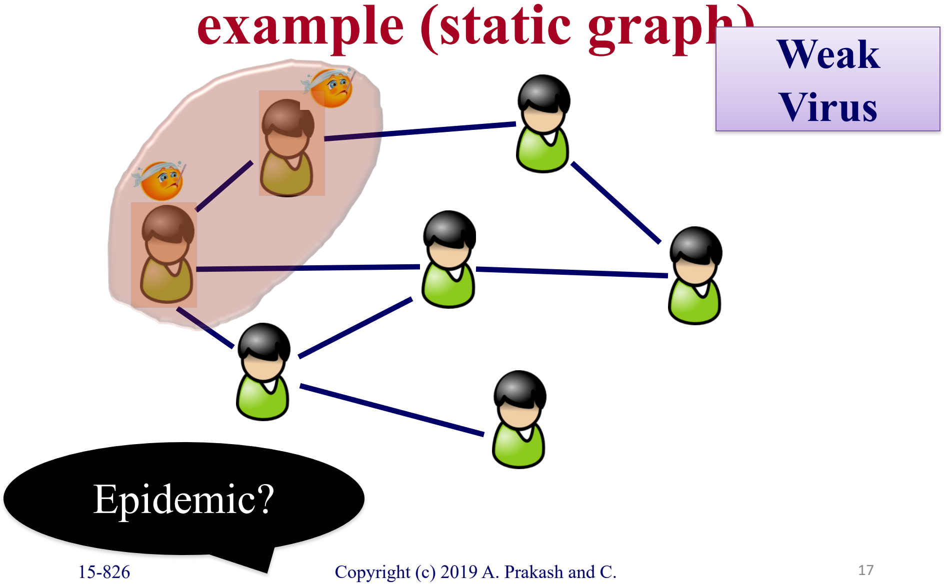
Carnegie Mellon

A fundamental question



Carnegie Mellon

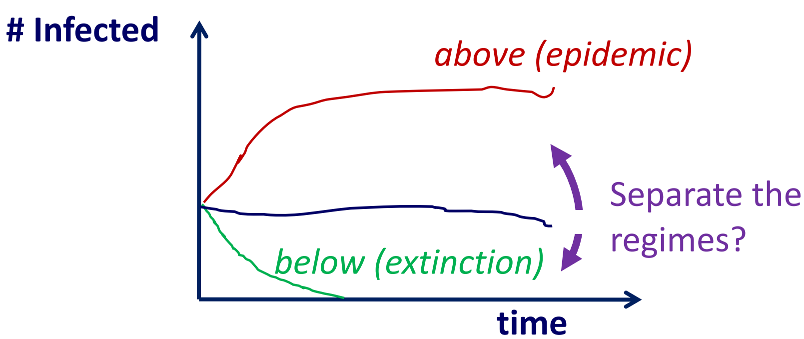
example (static graph)



17

Carnegie Mellon

Problem Statement



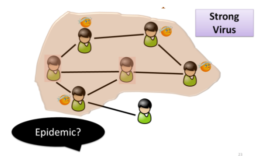
Find, a condition under which

- *virus will die out exponentially quickly*
- *regardless of initial infection condition*

18

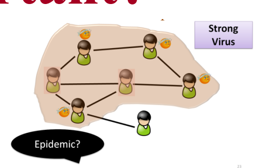
Threshold (static version)

Problem Statement



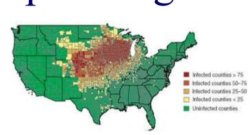
- Given:
  - Graph G, *and*
  - Virus specs (attack prob. etc.)
- Find:
  - A condition for virus extinction/invasion

Threshold: Why important?



- Accelerating simulations
- Forecasting ('What-if' scenarios)
- Design of contagion and/or topology
- A great handle to manipulate the spreading
  - Immunization
  - Maximize collaboration

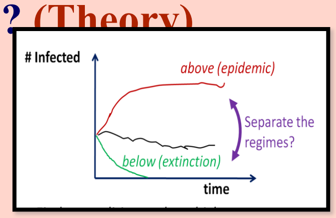
  …..



Carnegie Mellon

Outline

- Motivation
- **Epidemics: what happens? (Theory)**
  - **Background**
  - Result (Static Graphs)
  - Bonus : Competing Viruses
- Action: Who to immunize? (Algorithms)



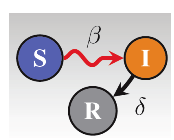
15-826 21

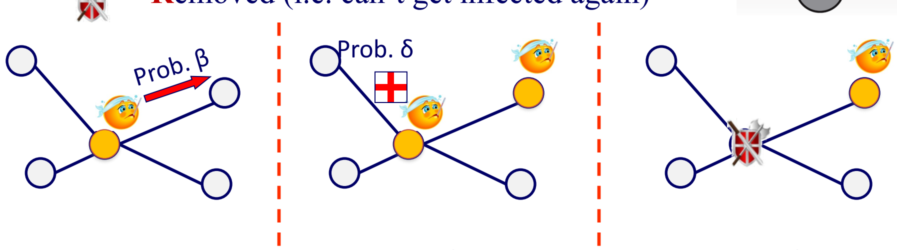
21

Carnegie Mellon **Background**

"SIR" model: life immunity (mumps)

- Each node in the graph is in one of three states
  - **S**usceptible (i.e. healthy)
  - **I**nfected
  - **R**emoved (i.e. can't get infected again)





15-826 22

22

Carnegie Mellon **Background**

Terminology: continued

- Other virus propagation models ("VPM")
  - SIS : susceptible-infected-susceptible, flu-like
  - SIRS : temporary immunity, like pertussis
  - SEIR : mumps-like, with virus incubation (E = Exposed)
  - ....
- Underlying contact-network – 'who-can-infect-whom'

Carnegie Mellon **Background**

Related Work

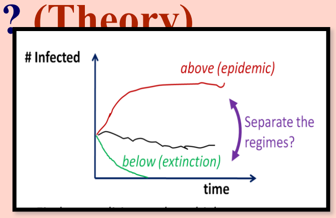
- R. M. Anderson and R. M. May. Infectious Diseases of Humans. Oxford University Press, 1991.
- A. Barrat, M. Barthélemy, and A. Vespignani. Dynamical Processes on Complex Networks. Cambridge University Press, 2010.
- F. M. Bass. A new product growth for model consumer durables. Management Science, 15(5):215–227, 1969.
- D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. ACM TISSEC, 10(4), 2008.
- D. Easley and J. Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010.
- A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology in spread of epidemics. IEEE INFOCOM, 2005.
- Y. Hayashi, M. Minoura, and J. Matsukubo. Recoverable prevalence in growing scale-free networks and the effective immunization. arXiv:cond-at/0305549 v2, Aug. 6 2003.
- H. W. Hethcote. The mathematics of infectious diseases. SIAM Review, 42, 2000.
- H. W. Hethcote and J. A. Yorke. Gonorrhea transmission dynamics and control. Springer Lecture Notes in Biomathematics, 46, 1984.
- J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. IEEE Computer Society Symposium on Research in Security and Privacy, 1991.
- J. O. Kephart and S. R. White. Measuring and modeling computer virus prevalence. IEEE Computer Society Symposium on Research in Security and Privacy, 1993.
- R. Pastor-Santorras and A. Vespignani. Epidemic spreading in scale-free networks. Physical Review Letters 86, 14, 2001.
- ....
- ....
- ....

All are about *either*:

- **Structured topologies** (cliques, block-diagonals, hierarchies, random)
- **Specific virus propagation models**
- **Static graphs**

Outline

- Motivation
- **Epidemics: what happens? (Theory)**
  - Background
  - **Result (Static Graphs)**
  - Bonus: Competing Viruses
- Action: Who to immunize? (Algorithms)

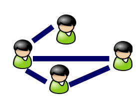


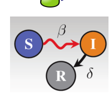
15-826 25

25

How should the answer look like?

- Answer should depend on:
  - Graph
  - Virus Propagation Model (VPM)





- But how??
  - Graph – average degree? max. degree? diameter?
  - VPM – which parameters?
  - How to combine – linear? quadratic? exponential?

$\beta d_{avg} + \delta\sqrt{diameter}$? $(\beta^2 d^2{}_{avg} - \delta d_{avg}) / d_{\max}$? .....

15-826 26

26

Carnegie Mellon

Static Graphs: Our Main Result

w/ Deepay Chakrabarti

For,	
➢ any arbitrary topology (adjacency matrix A)	**λ**
➢ any virus propagation model (VPM) in standard literature	C_{VPM}
the epidemic threshold depends only 1.on the λ, first eigenvalue of A, *and* 2.some constant C_{VPM} determined by the virus propagation model	**No epidemic if $\lambda * C_{VPM} < 1$**

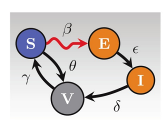
In Prakash+ ICDM 2011 (Selected among **best papers**).

27

Carnegie Mellon

Our thresholds for some models

- s = *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	$s = \lambda \cdot \left(\frac{\beta\gamma}{\delta(\gamma+\theta)}\right)$	
$SI_1I_2V_1V_2$ (H.I.V.)	$s = \lambda \cdot \left(\frac{\beta_1 v_2 + \beta_2 \varepsilon}{v_2(\varepsilon + v_1)}\right)$	

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

28

Carnegie Mellon

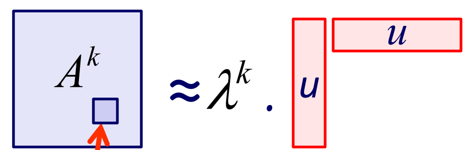
Our result: Intuition for λ

"Official" definition:

- *Let A be the adjacency matrix. Then λ is the root with the largest magnitude of the characteristic polynomial of A [det(A – xI)].*
- Doesn't give much intuition!

"Un-official" Intuition ☺

- *λ ~ # paths in the graph*



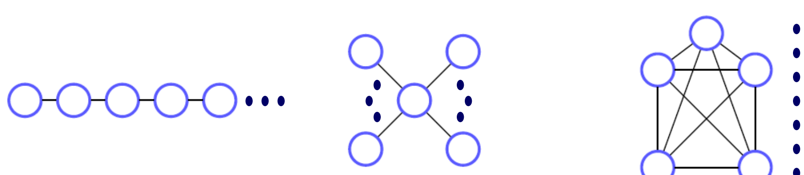
A^k(i, j) = # of paths i → j of length k

29

Carnegie Mellon

Largest Eigenvalue (λ)

better connectivity ⟶ higher λ



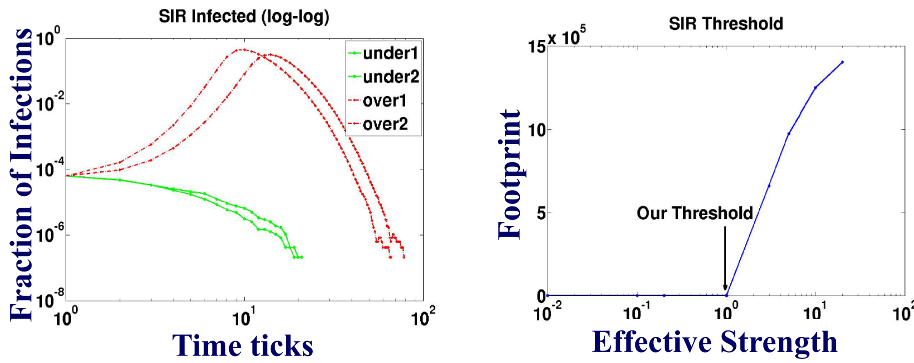
(a)Chain	(b)Star	(c)Clique
$\lambda \approx 2$	$\lambda = \sqrt{N}$	$\lambda = N-1$
$\lambda \approx 2$	λ= 31.67	λ= 999

N = 1000

30

Carnegie Mellon

Examples: Simulations – SIR (mumps)



(a) Infection profile **(b) "Take-off" plot**

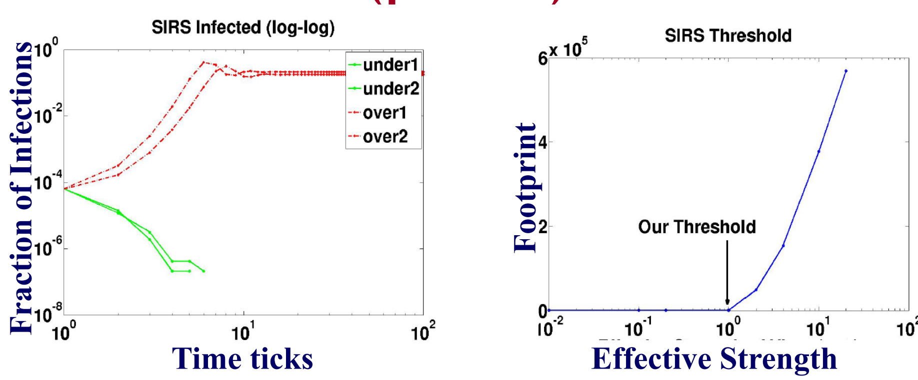
PORTLAND graph: *synthetic population,*
31 million links, 6 million nodes

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

31

Carnegie Mellon

Examples: Simulations – SIRS (pertusis)



(a) Infection profile **(b) "Take-off" plot**

PORTLAND graph: *synthetic population,*
31 million links, 6 million nodes

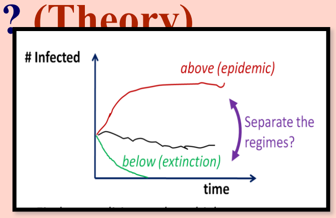
15-826 Copyright (c) 2019 A. Prakash and C. Faloutsos 32

32

Carnegie Mellon

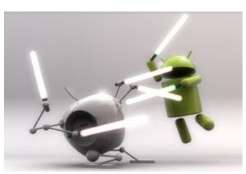
Outline

- Motivation
- **Epidemics: what happens? (Theory)**
  - Background
  - Result (Static Graphs)
  - Bonus: Competing Viruses
- Action: Who to immunize? (Algorithms)



Carnegie Mellon

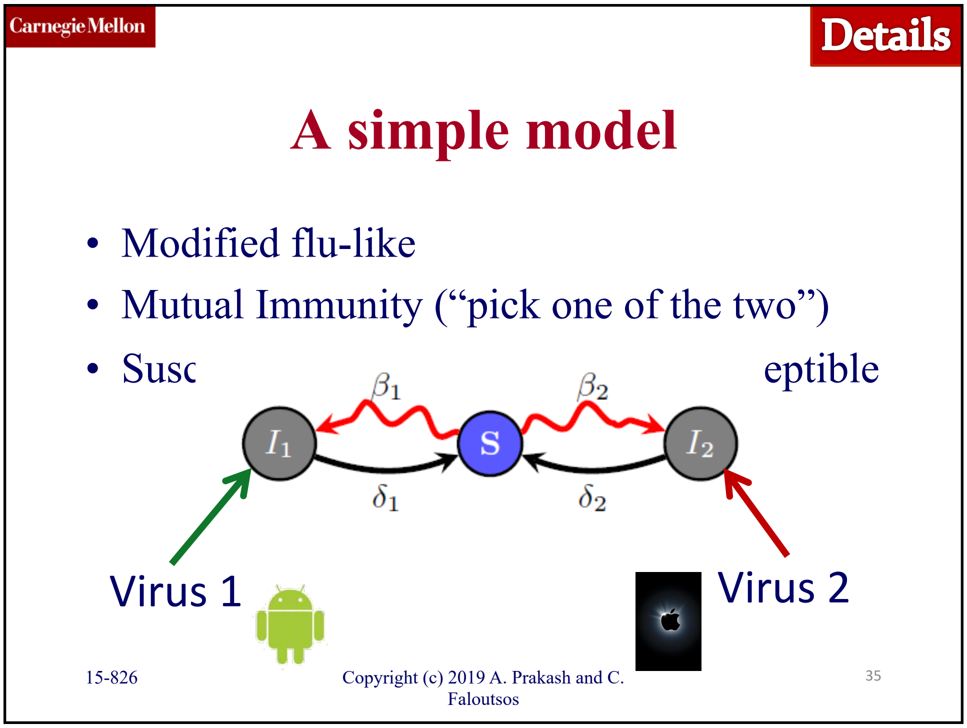
Competing Contagions



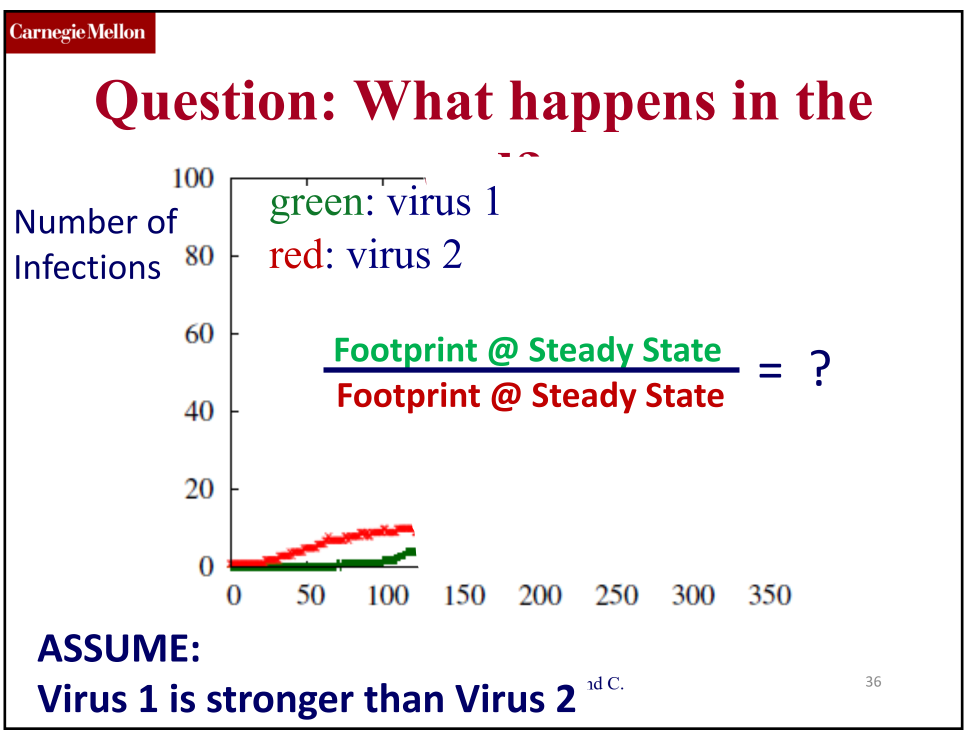
iPhone v **Android**

Blu-ray v **HD-DVD**

Biological common flu/avian flu, pneumococcal inf etc


Carnegie Mellon
Details
A simple model
• Modified flu-like
• Mutual Immunity ("pick one of the two")
• Susc
eptible
β1
β2
I1
S
I2
δ1
δ2
Virus 1
Virus 2
15-826
Copyright (c) 2019 A. Prakash and C. Faloutsos
35

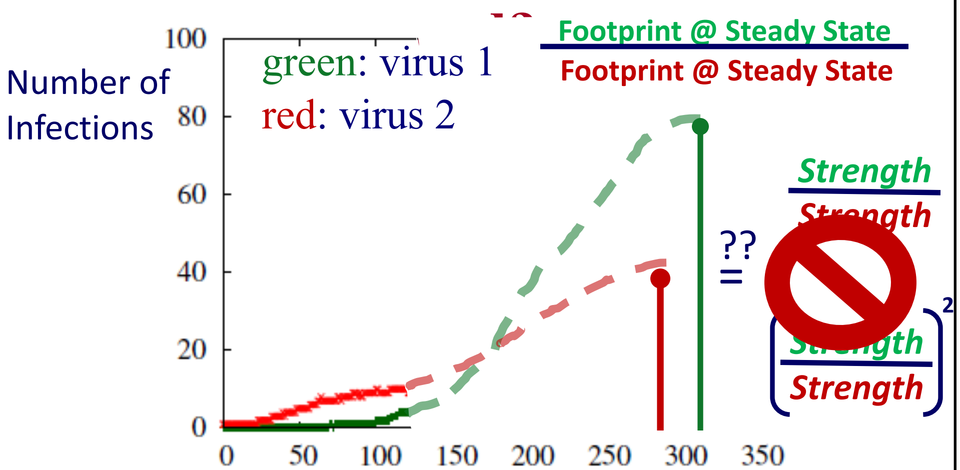
35


Carnegie Mellon
Question: What happens in the
Number of Infections
100
80
60
40
20
0
0 50 100 150 200 250 300 350
green: virus 1
red: virus 2
Footprint @ Steady State
Footprint @ Steady State
= ?
ASSUME:
Virus 1 is stronger than Virus 2
36

36

Carnegie Mellon

Question: What happens in the



ASSUME:

Virus 1 is stronger than Virus 2

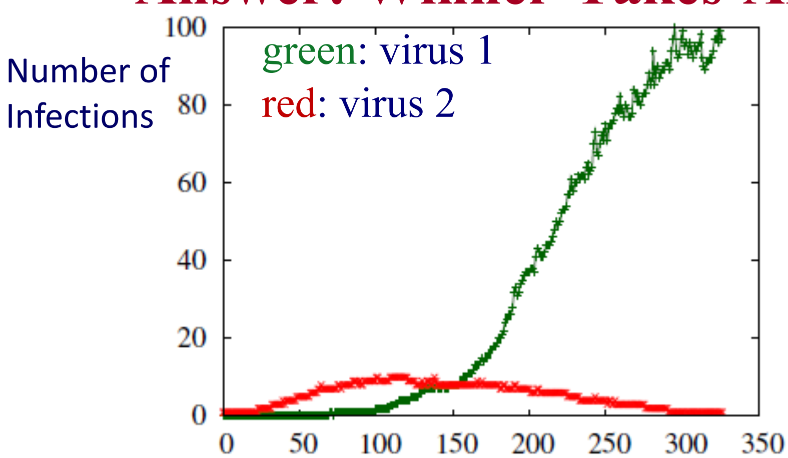
ıd C.

37

37

Carnegie Mellon

Answer: Winner-Takes-All



ASSUME:

Virus 1 is stronger than Virus 2

ıd C.

38

38

Carnegie Mellon

Our Result: Winner-Takes-All

Given our model, and ***any graph***, the weaker virus always **dies-out completely**

Details

1. The stronger survives only if it is above threshold
2. Virus 1 is stronger than Virus 2, if:
   strength(Virus 1) > strength(Virus 2)
3. Strength(Virus) = λ β / δ → same as before!

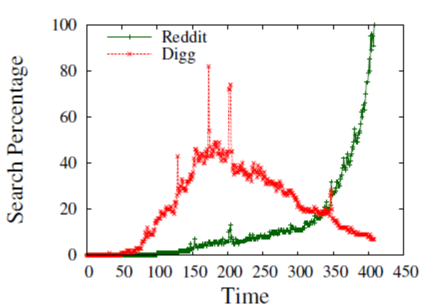
In Prakash+ WWW 2012

39

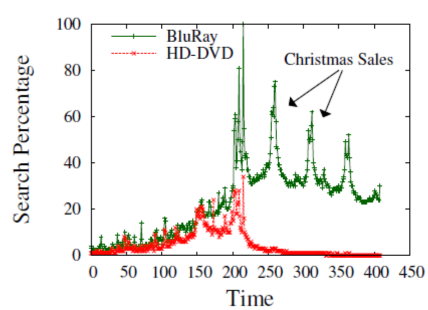
Carnegie Mellon

Real Examples

[Google Search Trends data]



Reddit v Digg



Blu-Ray v HD-DVD

15-826 40

40

Carnegie Mellon

Outline

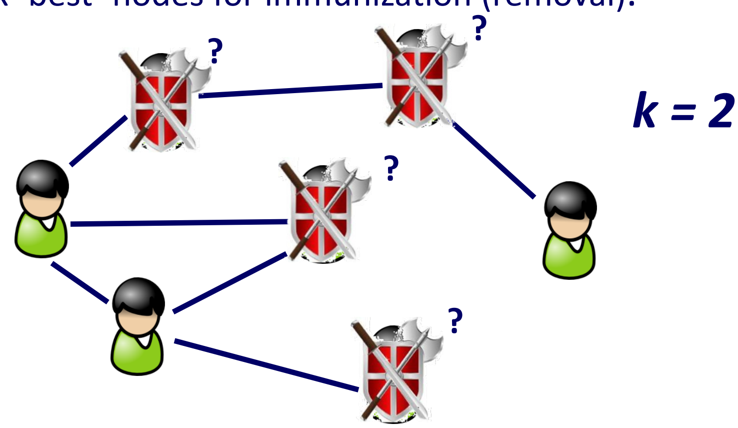
- Motivation
- Epidemics: what happens? (Theory)
- **Action: Who to immunize? (Algorithms)**

41

Carnegie Mellon

Immunization

Given: a graph *A*, virus prop. model and budget *k*;
Find: *k* 'best' nodes for immunization (removal).



42

Challenges

- Given a graph A, budget k,

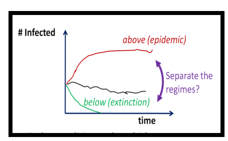
Q1 (Metric) How to measure the 'shield-value' for a set of nodes (S)?

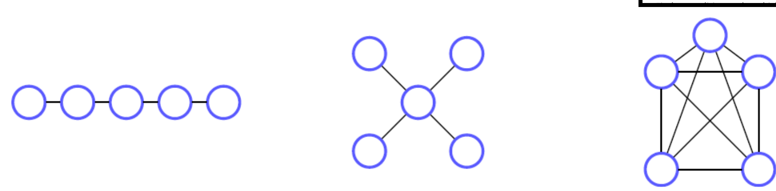
Q2 (Algorithm) How to find a set of k nodes with highest 'shield-value'?

15-826 43

43

Proposed vulnerability measure: λ





"Safe" *"Vulnerable"* *"Deadly"*

(a)Chain(λ = 1.73) **(b)Star**(λ = 2) **(c)Clique**(λ = 4)

higher λ, higher vulnerability

15-826 44

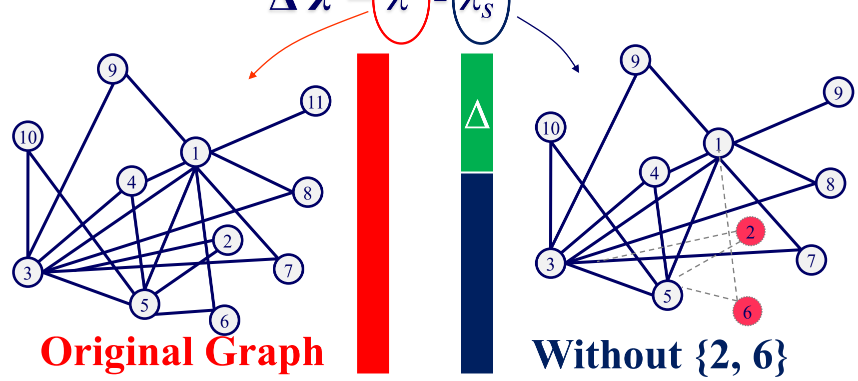
44

Carnegie Mellon

A1: "Eigen-Drop": an ideal shield value

Eigen-Drop(S)

$$\Delta \lambda = \lambda - \lambda_S$$



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

45

Carnegie Mellon

Details

Challenges

- Given a graph A, budget k,

Q1 (Metric) How to measure the 'shield-value' for a set of nodes (S)?

Q2 (Algorithm) How to find a set of k nodes with highest 'shield-value'?

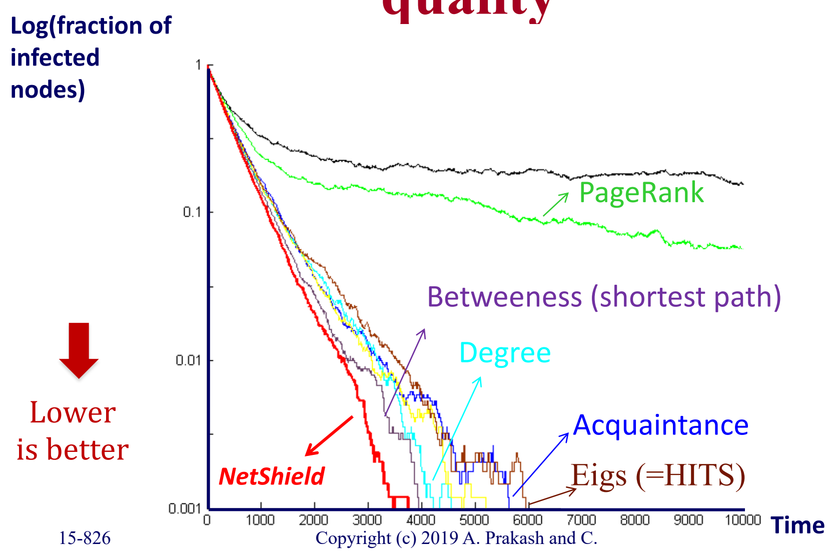
A2: greedy

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

46

Carnegie Mellon

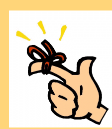
Experiment: Immunization quality



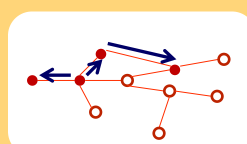
15-826 Copyright (c) 2019 A. Prakash and C. Faloutsos 47

47

Carnegie Mellon



Short answers



- Q1: epidemic?
- A1: tipping point: eigenvalue
- Q2: whom to immunize
- A2: eigen-drop
- (Q3: 2 competing viruses – end result?)
- A3: winner takes all!

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

48