Propagation and Immunization on Networks: Theory and Tools

B. ADITYA PRAKASH
Computer Sc. Dept.
Carnegie Mellon University

Guest Lecture
15-826 – Multimedia Databases and Data Mining

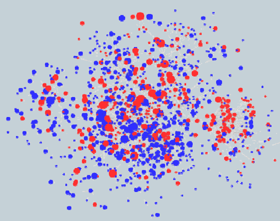
15 September 2011

Motivation

2

One liner: *"Contagions spreading over a network"*

- Disease Propagation
- Blog Cascades
- Viral Marketing
- Influence Propagation



Reading Material

3

- *Primary:* Chapter 21, Networks, Crowds, and Markets: Reasoning about a Highly Connected World.
  by David Easley and Jon Kleinberg. Cambridge University Press, 2010.
  http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch21.pdf
- *Additional:* The Mathematics of Infectious Diseases
  by H. W. Hethcote. SIAM Review Volume 42, Issue 4, 2000.
  http://www.maths.usyd.edu.au/u/marym/populations/hethcote.pdf

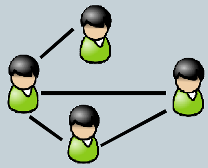
Underlying phenomenon

4

- Contagions over a network

Human contact-networks

Biological Viruses e.g. H1N1



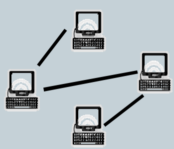
Contagions over networks

5

- examples…

Computer Networks

Computer Viruses e.g. trojans



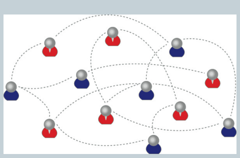
Contagions over a network

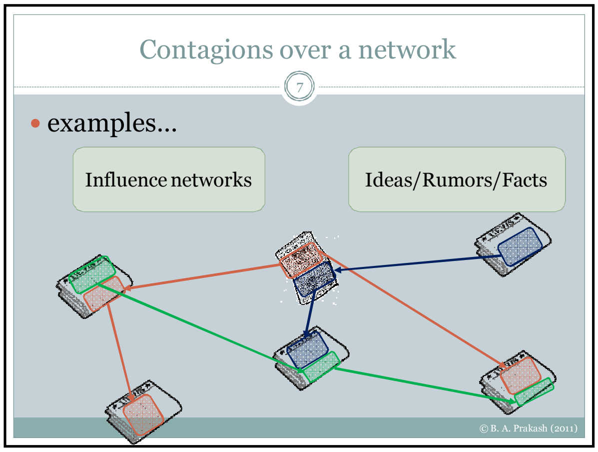
6

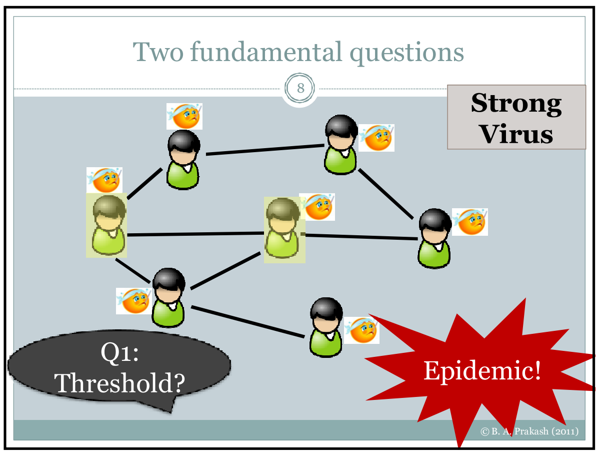
- examples…

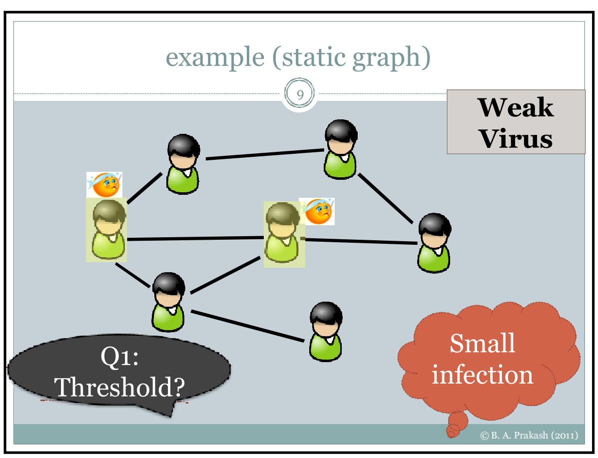
Influence networks

Products



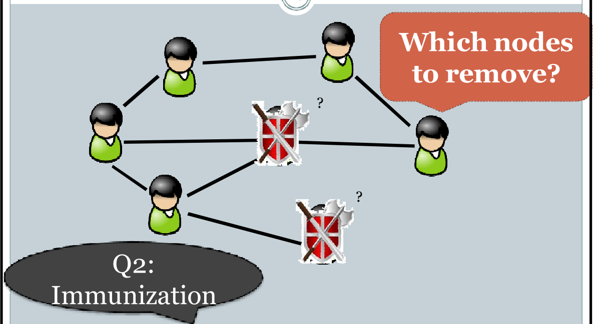

Contagions over a network
7
examples…
Influence networks
Ideas/Rumors/Facts
© B. A. Prakash (2011)


Two fundamental questions
8
Strong Virus
Q1: Threshold?
Epidemic!
© B. A. Prakash (2011)


example (static graph)
9
Weak Virus
Q1: Threshold?
Small infection
© B. A. Prakash (2011)

Questions…

10



SARS costs 700+ lives; $40+ Bn; H1N1 costs Mexico $2.3bn

This lecture…

11

- Focus on propagation processes on arbitrary static and dynamic networks and temporal evolution of data.
  - Theory:
    - Tipping point? Footprint? Etc.
  - Algorithms:
    - Which nodes to immunize? Etc.

Lecture: Overview

12

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs		
Propagation Models on Dynamic Graphs		

Lecture: Overview

13

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs	Tipping point?	Immunization Algorithms
Propagation Models on Dynamic Graphs		

Outline

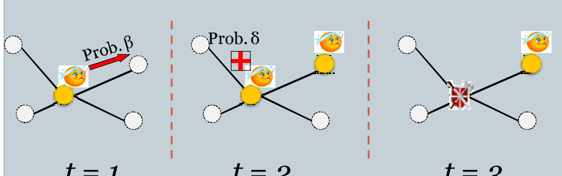
14

- Introduction
- Terminology
- Theory
- Algorithms
- Open Problem
- Conclusion

"SIR" model (mumps-like)

15

- 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)



Terminology: continued

16

- Other virus models
  - 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'
  - Nodes (*people/computers*)
  - Edges (*links between nodes*)

Related Work

17

- 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 models**

➢**Static graphs**

Outline

18

- Introduction
- Terminology
- Theory
- Algorithms
- Open Problem
- Conclusion

Lecture: Overview

19

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs	(TS1) G2-threshold theorem	(AS1) Full Symmetric Static Immunization (NetShield)
Propagation Models on Dynamic Graphs	(TD1) Thresholds	(AD1) Full Symmetric Dynamic Immunization

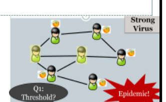
Lecture: Overview

20

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs	**(TS1) G2-threshold theorem**	(AS1) Full Symmetric Static Immunization (NetShield)
Propagation Models on Dynamic Graphs	(TD1) Thresholds	(AD1) Full Symmetric Dynamic Immunization

(TS1) Static Graphs: G2-threshold Theorem

21



Problem Statement

- Given:
  - An undirected unweighted graph G, *and*
  - A virus propagation model (VPM) and its parameters (e.g., β and δ for SIR)
- Find:
  - A condition for virus extinction/invasion

In Prakash+ ICDM 2011. Also posted as *arXiv:1004.0060*

Intuitively

22

- 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?

Static Graphs: G2-threshold Theorem

23

- Main result (informal):

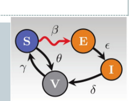
"G2": two orthogonal generalizations

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** $\lambda * C_{VPM} < 1$

Thresholds for some models

24

- *s = effective strength*
- *s* < 1 : *below threshold*



Models	Effective Strength (s)	Threshold (tipping point)
SIS, SIR, SIRS, SEIR	$s = \lambda . \left(\frac{\beta}{\delta}\right)$	$s = 1$
SIV, SEIV	$s = \lambda . \left(\frac{\beta\gamma}{\delta(\gamma+\theta)}\right)$	
$SI_1I_2V_1V_2$(**H.I.V.**)	$s = \lambda . \left(\frac{\beta_1 v_2 + \beta_2 \varepsilon}{v_2(\varepsilon + v_1)}\right)$	

Largest Eigenvalue (λ)

25

- "Official" definition:

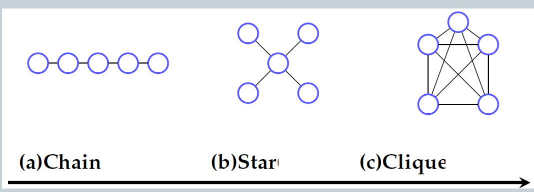
  Let A be the adjacency matrix. Then λ is the root with the largest magnitude of the characteristic polynomial of A [det(A – xI)].

- Yeah, doesn't give too much intuition! ☺

Largest Eigenvalue (λ)

26

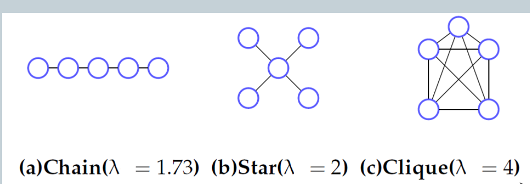
better connectivity ⟶ higher λ



Largest Eigenvalue (λ)

27

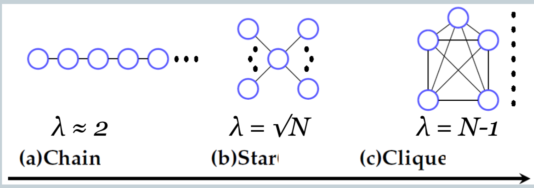
better connectivity ⟶ higher λ



Largest Eigenvalue (λ)

28

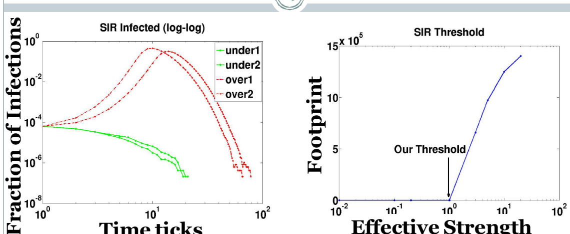
$\lambda \approx$ average degree, only better!



N nodes

Examples: Simulations - SIR

29

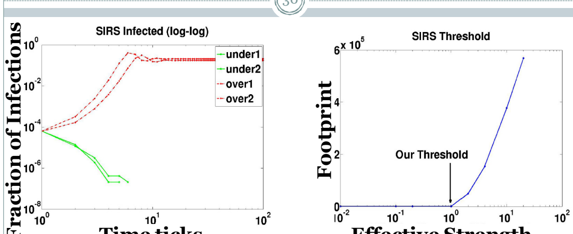


(a) Infection profile **(b) "Take-off" plot**

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

Examples: Simulations - SIRS

30



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

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

So how did we do that?

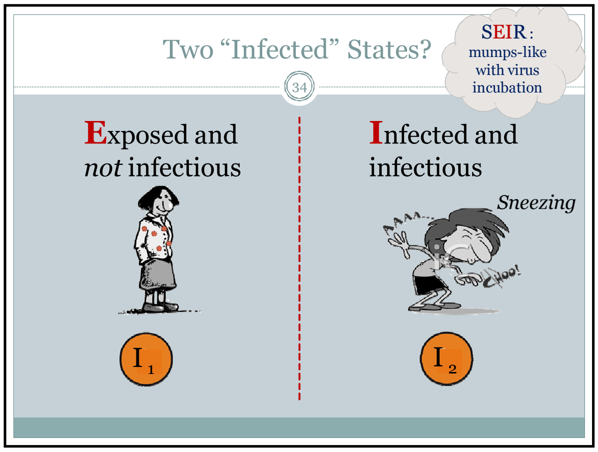
Some trivia

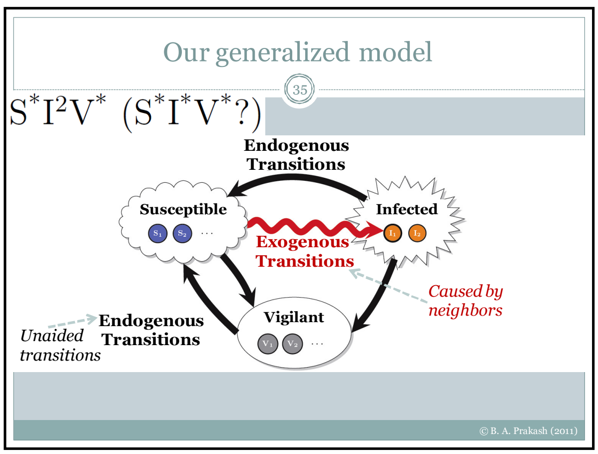
- first person in the US identified as a healthy carrier of the pathogen associated with typhoid fever.
- infected some 53 people, over the course of her career as a cook!
- forcibly quarantined by public health authorities

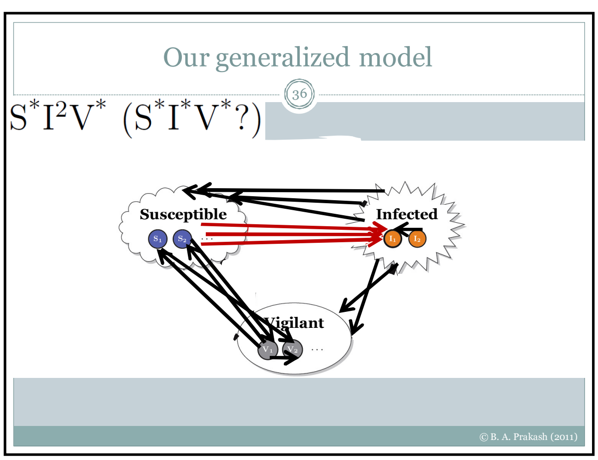
Two "Infected" States?

SICR: with a carrier

Asymtomatic	Symptomatic
	*Sneezing* 
I_1	I_2


Two "Infected" States?
SEIR : mumps-like with virus incubation
34
Exposed and not infectious
Infected and infectious
Sneezing
I_1
I_2

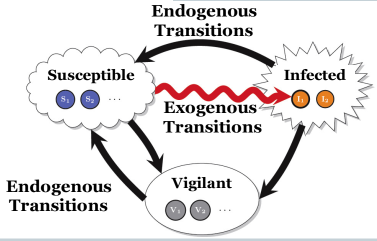

Our generalized model
35
$S^*I^2V^*$ ($S^*I^*V^*$?)
Endogenous Transitions
Susceptible
Infected
Exogenous Transitions
Caused by neighbors
Vigilant
Unaided transitions
Endogenous Transitions
© B. A. Prakash (2011)


Our generalized model
36
$S^*I^2V^*$ ($S^*I^*V^*$?)
Susceptible
Infected
Vigilant
© B. A. Prakash (2011)

Our generalized model

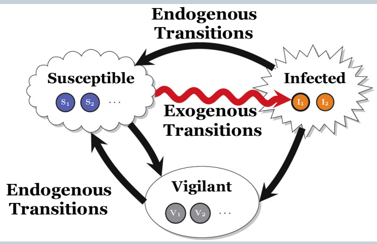
37

$S^*I^2V^*$ ($S^*I^*V^*$?)



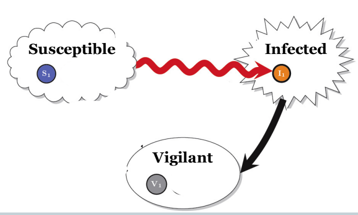
Special case

38

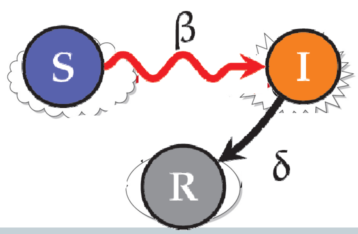


Special case

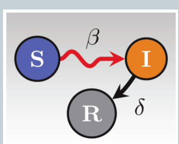
39

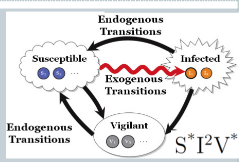


Special case: SIR



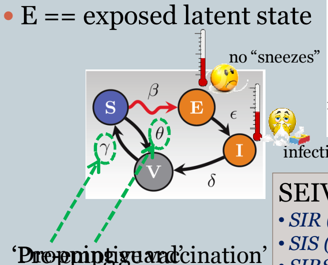
Special case: SIR (permanent immunity)

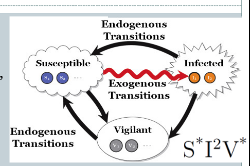




Another special case: SEIV

- E == exposed latent state





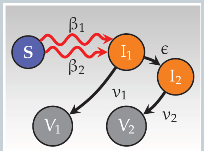
SEIV: *itself generalizes*

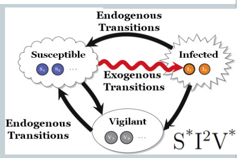
- *SIR (mumps)*
- *SIS (flu-like)*
- *SIRS (pertussis)*
- *SEIR (varicella) etc.*

Special case: H.I.V.

43

$SI_1I_2V_1V_2$





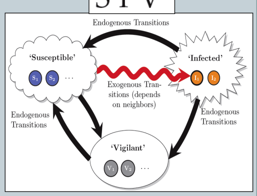
Multiple Infectious, Vigilant states

Models and more models

44

Model	Used for
SIR	**Mumps**
SIS	**Flu**
SIRS	**Pertussis**
SEIR	**Varicella**
……..	
SICR	**Tuberculosis**
MSIR	**Measles**
SIV	**Sensor Stability**
$SI_1I_2V_1V_2$	**H.I.V.**
……….	

$S^*I^2V^*$



Details

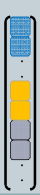
Proof: Main Idea

45

- View as a NLDS $\vec{P}_{t+1} = \mathcal{G}(\vec{P}_t)$
  - discrete time
  - non-linear dynamical system (NLDS)

$\vec{P}_t$

size mN x 1



Probability vector
Specifies the state of the system at time *t*

Details

Proof: Main Idea

46

- View as a NLDS $\vec{P}_{t+1} = \mathcal{G}(\vec{P}_t)$
  - discrete time
  - non-linear dynamical system (NLDS)

$\mathcal{G} : \mathbb{R}^{mN} \rightarrow \mathbb{R}^{mN}$

$\vec{P}_t$

size *mN* x *1*

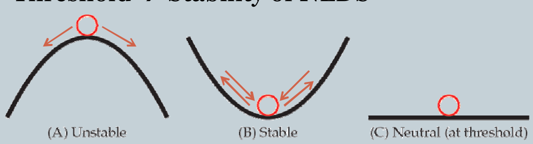
Non-linear function

Explicitly gives the evolution of system

Proof: Main Idea

47

- View as a NLDS $\vec{P}_{t+1} = \mathcal{G}(\vec{P}_t)$
  - discrete time
  - non-linear dynamical system (NLDS)
- Threshold → Stability of NLDS



Details

Special case: SIR

48

$\mathcal{G} : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3N}$

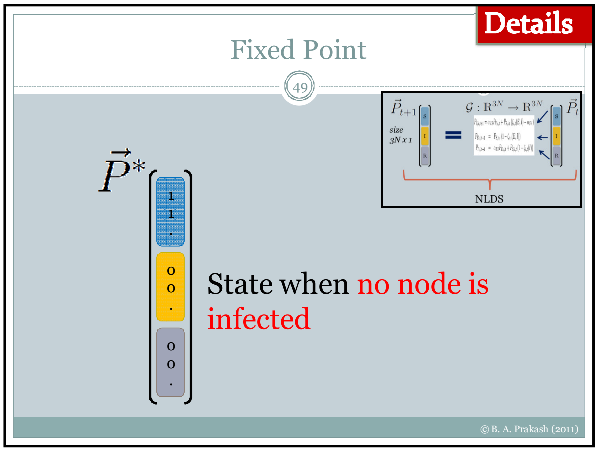
$\vec{P}_{t+1}$ = [S, I, R] (*size* *3N* x *1*) $=$

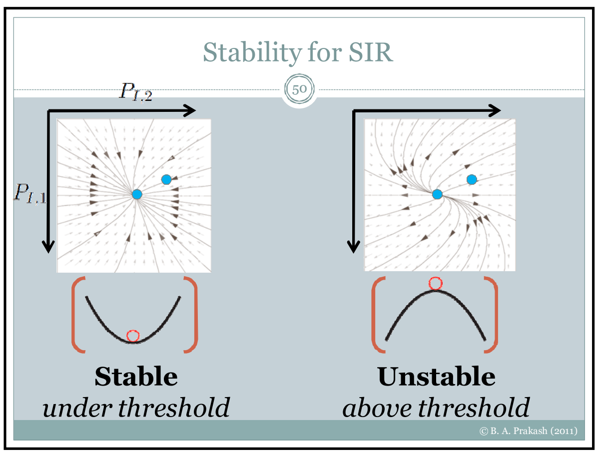
$$\begin{aligned} P_{S,i,t+1} &= P_{S,i,t}\zeta_{i,t}(I) \\ P_{I,i,t+1} &= P_{S,i,t}(1-\zeta_{i,t}(I)) + (1-\delta)P_{I,i,t} \\ P_{R,i,t+1} &= \delta P_{I,i,t} + P_{R,i,t} \end{aligned}$$

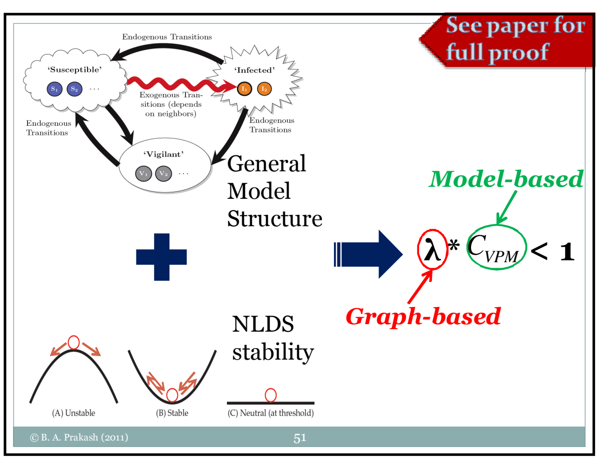
← [S, I, R] $\vec{P}_t$

NLDS

$\zeta_{i,t}(I)$ = probability that node i is not attacked by any of its infectious neighbors


Details
Fixed Point
49
$\vec{P}^*$
1
1
.
0
0
.
0
0
.
State when no node is infected
NLDS
© B. A. Prakash (2011)


Stability for SIR
50
$P_{I,2}$
$P_{I,1}$
Stable
under threshold
Unstable
above threshold
© B. A. Prakash (2011)


See paper for full proof
Endogenous Transitions
'Susceptible'
'Infected'
Exogenous Transitions (depends on neighbors)
Endogenous Transitions
Endogenous Transitions
'Vigilant'
General Model Structure
Model-based
$\lambda * C_{VPM} < 1$
Graph-based
NLDS stability
(A) Unstable
(B) Stable
(C) Neutral (at threshold)
© B. A. Prakash (2011)
51

Lecture: Overview

52

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs	(TS1) G2-threshold Theorem	(AS1) Full Symmetric Static Immunization (NetShield)
Propagation Models on Dynamic Graphs	**(TD1) Thresholds**	(AD1) Full Symmetric Dynamic Immunization

(TD1) Dynamic Graphs: Thresholds

53

- What about time-varying dynamic graphs?

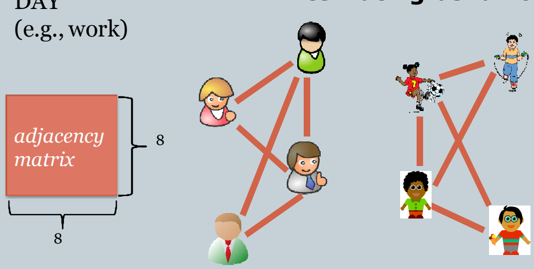
© B. A. Prakash (2011)

(TD1) Dynamic Graphs: Thresholds

54

DAY
(e.g., work)

Alternating behaviors

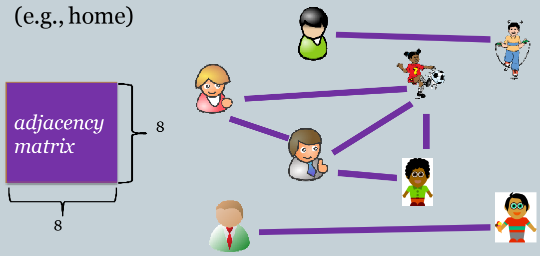


© B. A. Prakash (2011)

Alternating Behaviors

55

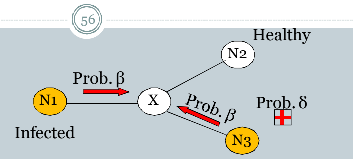
NIGHT
(e.g., home)



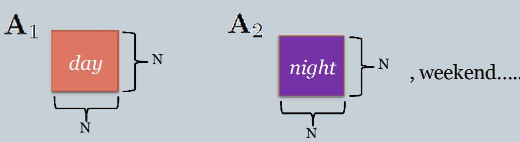
Model Description

56

- SIS model
  - recovery rate δ
  - infection rate β

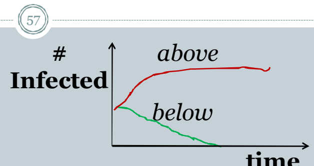


- Set of *T* **arbitrary** graphs $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_T\}$



Problem Statement

57



Find, a condition under which

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

Should we,

58

- Add the adjacency matrices? +
- Multiply? x
- "Superimpose"?
- or some other crazy combination?

Our result: Threshold

59

- **Informally, *NO* epidemic if**

$$eig\ (\mathbf{S}) = \lambda_{\mathbf{S}} < 1$$

Single number!
Largest eigenvalue of
the "*system matrix*" (dfn. next)

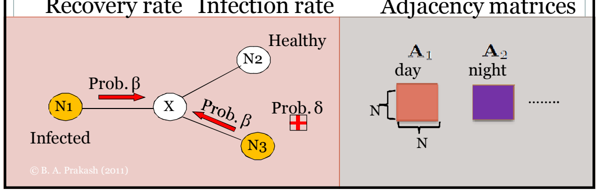
In Prakash+, ECML-PKDD 2010

***NO* epidemic if**
***eig* (S) = $\lambda_{\mathbf{S}}$ < 1**
$\mathbf{S} = \prod_i \mathbf{S}_i$

60

$$\mathbf{S}_i = (1-\delta)\mathbf{I} + \beta \mathbf{A}_i$$

Recovery rate Infection rate Adjacency matrices



Simulation Examples

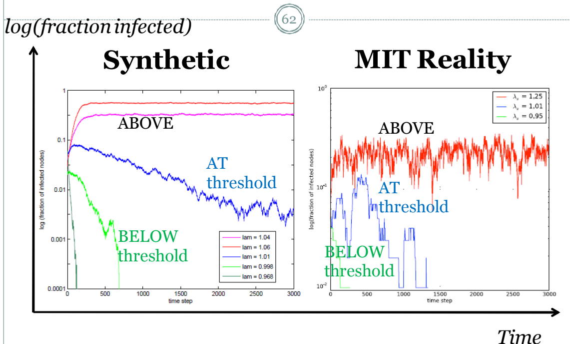
61

- Synthetic
  - 100 nodes
  - $\mathbf{A}_1$ - Clique $\mathbf{A}_2$ - Chain
- MIT Reality Mining
  - 104 mobile devices
  - September 2004 – June 2005
  - 12-hr adjacency matrices

$\mathbf{A}_1$ (day) $\mathbf{A}_2$ (night)

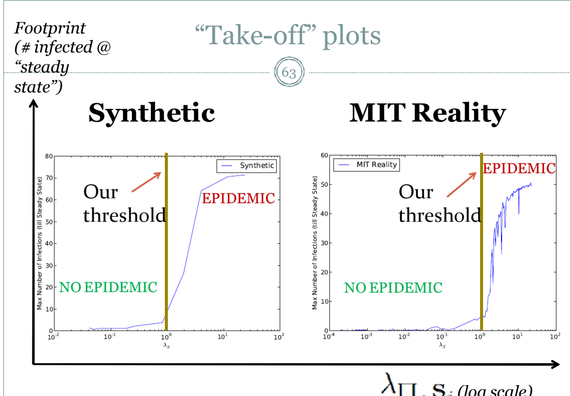
Infection-profile

62



"Take-off" plots

63



Outline

64

- Introduction
- Terminology
- Theory
- Algorithms
- Open Problem
- Conclusion

Lecture: Overview

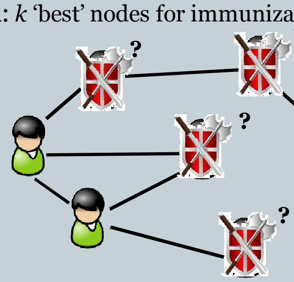
65

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs	(TS1) G2-threshold Theorem	**(AS1) Full Symmetric Static Immunization (NetShield)**
Propagation Models on Dynamic Graphs	(TD1) Thresholds	(AD1) Full Symmetric Dynamic Immunization

(AS1): Full Static Immunization

66

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



$k = 2$

In Tong, Prakash+ ICDM 2010

Challenges

67

- Given a graph A, budget k,

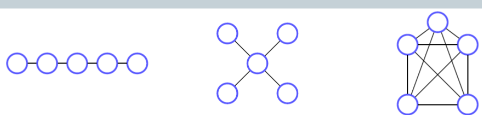
Q1 (Metric) How to measure the 'shield-value' for a set of nodes (S)?

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

Proposed vulnerability λ

68

λ is the epidemic threshold!



"Safe" (a)Chain($\lambda = 1.73$)

"Vulnerable" (b)Star($\lambda = 2$)

"Deadly" (c)Clique($\lambda = 4$)

Increasing λ

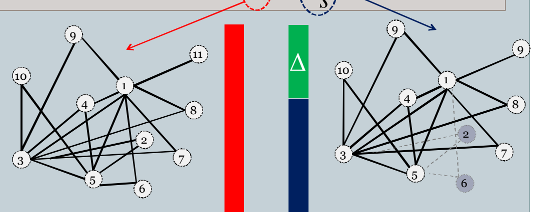
Increasing vulnerability

A1: Eigen-Drop: an ideal shield value

69

Eigen-Drop(S)

$$\Delta \lambda = \lambda - \lambda_s$$



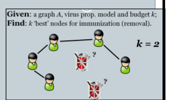
Original Graph **Without {2, 6}**

(Q2) - Direct Algorithm too expensive!

70

- Immunize k nodes which maximize $\Delta\lambda$

$$S = \text{argmax } \Delta\lambda$$



- Combinatorial!
- Complexity: $O\left(\binom{n}{k} \cdot m\right)$
  - Example:
    - 1,000 nodes, with 10,000 edges
    - It takes 0.01 seconds to compute λ
    - It takes **2,615 years** to find 5-best nodes!

A2: Our Solution

71

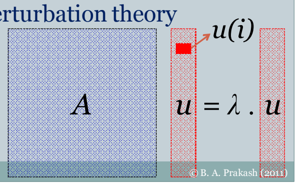
- Part 1
  - Approximate Eigen-drop ($\Delta\lambda$)
- Not enough!
- Part 2
  - Greedily pick best node at each step
- Algorithm: *NetShield*
  - $O(nk^2+m)$

Our Solution: Part 1

Details

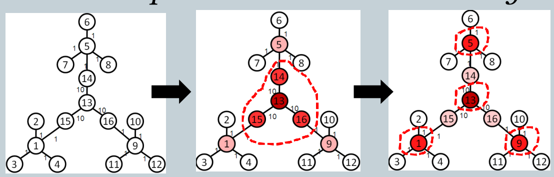
72

- Approximate Eigen-drop ($\Delta\lambda$)
- $\Delta\lambda \approx \widehat{SV}(S) = \sum_{i \in S} 2\lambda \mathbf{u}(i)^2 - \sum_{i,j \in S} \mathbf{A}(i,j)\mathbf{u}(i)\mathbf{u}(j)$
  - Result using Matrix perturbation theory
  - $u(i)$ == '*eigenscore*' ~~ *pagerank(i)*



Details

$$\underbrace{\sum_{i\in\mathcal{S}} 2\lambda\mathbf{u}(i)^2}_{\text{P1: } node\ importance} - \underbrace{\sum_{i,j\in\mathcal{S}} \mathbf{A}(i,j)\mathbf{u}(i)\mathbf{u}(j)}_{\text{P2: } set\ diversity}$$



Original Graph | Select by P1 | Select by P1+P2

Details

Done?

74

- S = argmax $\widehat{\text{SV}}(S)$
- Still combinatorial
- Complexity: $O(\binom{n}{k}\cdot k^2)$
  - Example:
    - 1,000 nodes, with 10,000 edges
    - It takes 0.00001 seconds to compute $\widehat{\text{SV}}(S)$
    - It takes **3 months** to find 5-best nodes

Our Solution: Part 2: NetShield

75

- We prove that:
  $\widehat{\text{SV}}(S)$ is sub-modular (& monotone non-decreasing)

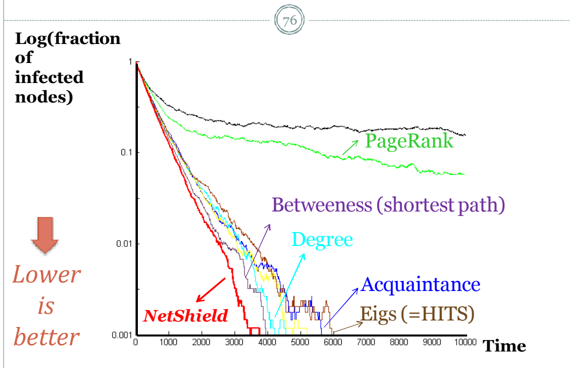
Corollary: Greedy algorithm works
1. NetShield is near-optimal (w.r.t. max $\widehat{\text{SV}}(S)$)
2. NetShield is $O(nk^2+m)$

- NetShield: Greedily add best node at each step

Footnote: near-optimal means $\widehat{\text{SV}}(S^{NetShield}) >= (1-1/e)\widehat{\text{SV}}(S^{Opt})$

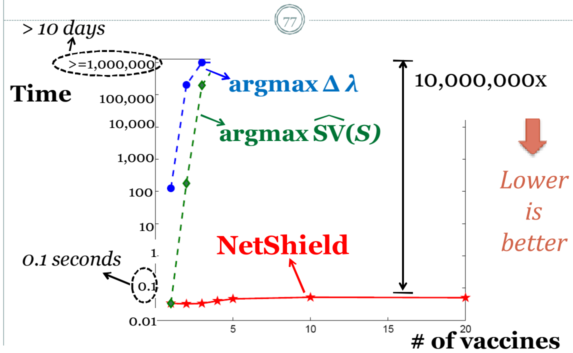
Experiment: Immunization quality

76



NetShield Speed

77



Lecture: Overview

78

	Theory: Observations and Results	Algorithms: Tools
Propagation Models on Static Graphs	(TS1) G2-threshold Theorem	(AS1) Full Symmetric Static Immunization (NetShield)
Propagation Models on Dynamic Graphs	(TD1) Thresholds	**(AD1) Full Symmetric Dynamic Immunization**

(AD1) Full Dynamic Immunization

79

- *Given:*
  Set of T **arbitrary** graphs $\{\mathbf{A}_1, \mathbf{A}_2 \ldots, \mathbf{A}_T\}$

$\mathbf{A}_1$ *day* N × N $\mathbf{A}_2$ *night* N × N , weekend.....

- *Find:*
  k 'best' nodes to immunize (remove)

In Prakash+ ECML-PKDD 2010

(AD1) Full Dynamic Immunization

80

$\mathbf{A}_1$ day $\mathbf{A}_2$ night

- Our solution
  - Recall theorem
  - Simple: reduce $\lambda(\prod_i S)$ $(=\lambda)$ *Matrix Product*
- Goal: max eigendrop $\Delta\lambda$

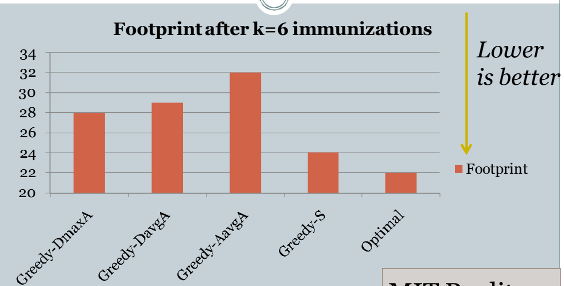
$$\Delta\lambda = \lambda_{before} - \lambda_{after}$$

- No competing policy for comparison
- We propose and evaluate many policies

© B. A. Prakash (2011)

Performance of Policies

81



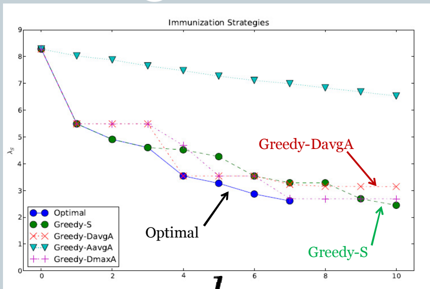
MIT Reality Mining

λ vs k

82

Lower is better

λ



k (# vaccines)

Outline

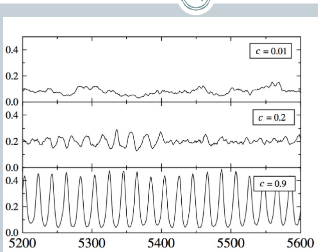
83

- Introduction
- Terminology
- Theory
- Algorithms
- Open problem
- Conclusion

Periodicities in SIRS (temp. immunity)

84

Fraction of Infected



Time

c = # "shortcuts"

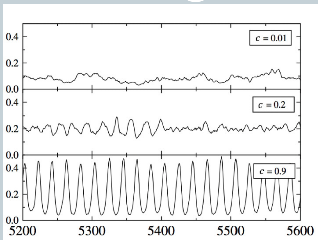
Small-World Graphs (WS Model)

Easley + Kleinberg 2010

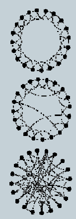
Periodicities in SIRS (temp. immunity)

85

Fraction of Infected



Time



Small-World Graphs (WS Model)

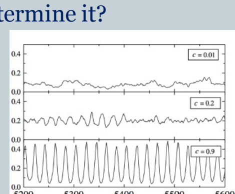
c = # "shortcuts"

Easley + Kleinberg 2010

Periodicities in SIRS (temp. immunity)

86

- Questions (for a class project? ☺)
  - Can we observe it on arbitrary graphs?
  - What properties of graphs cause it?
  - What SIRS parameters determine it?



Outline

87

- Introduction
- Terminology
- Theory
- Algorithms
- Open Problem
- Conclusion

Conclusion

88

- Threshold depends (for any VPM)
  - on λ (largest eigenvalue) of suitable matrix
  - a constant C_{VPM}
- Use Δλ to guide immunization policies
  - fast, scalable, provably near-optimal algorithms

Conclusion

89

- Real, compelling applications in diverse domains
- Lots of interesting research problems!
  - we saw only some of them
    - multiple competing viruses? (Iphone vs andriod)
    - information diffusion (Twitter) etc. etc.

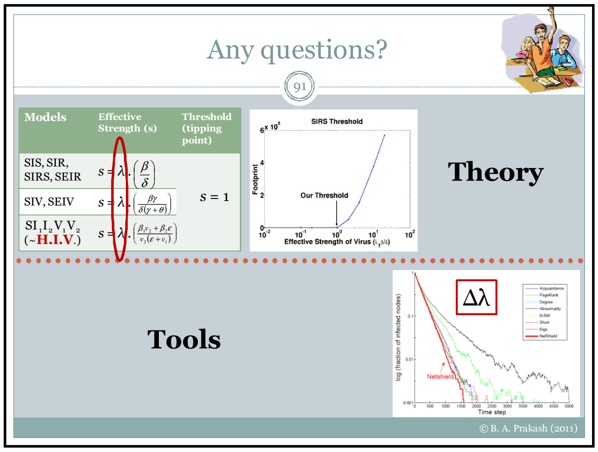
References

90

- *Got the Flu (or mumps)? Check the Eigenvalue!* by B. Aditya Prakash, Deepayan Chakrabarti, Michalis Faloutsos, Nicholas Valler, Christos Faloutsos. *ICDM 2011, posted on arXiv.*
- *Worst-case footprints in the SIS model* by B. Aditya Prakash, Varun Gupta, Christos Faloutsos. *In preparation.*
- *Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms* by B. Aditya Prakash, Hanghang Tong, Nicholas Valler, Michalis Faloutsos, Christos Faloutsos. *In ECML-PKDD 2010, Barcelona*
- *Epidemic Spread in Mobile Ad Hoc Networks: Determining the Tipping Point* by Nicholas Valler, B. Aditya Prakash, Hanghang Tong, Michalis Faloutsos, Christos Faloutsos. *In IFIP NETWORKING 2011, Valencia*
- *On the Vulnerability of Large Graphs* by Hanghang Tong, B. Aditya Prakash, Charalampos Tsourakakis, Tina Eliassi-Rad, Christos Faloutsos, Duen Horng Chau. *In IEEE ICDM 2010, Sydney*

ACKNOWLEDGEMENTS

Christos Faloutsos, Roni Rosenfeld, Michalis Faloutsos, Lada Adamic, Jack Iwashnya, Deepayan Chakrabarti, Hanghang Tong, Varun Gupta, Polo Chau, Nicholas Valler



Any questions?
91
Models
Effective Strength (s)
Threshold (tipping point)
SIS, SIR, SIRS, SEIR
SIV, SEIV
(~H.I.V.)
s = 1
SIRS Threshold
Footprint
Our Threshold
Effective Strength of Virus
Theory
Tools
Δλ
Netshield
Time step
© B. A. Prakash (2011)