Large Graph Mining: Power Tools and a Practitioner's guide

Task 6: Virus/Influence Propagation

Faloutsos, Miller, Tsourakakis

CMU

Outline

- Introduction – Motivation
- Task 1: Node importance
- Task 2: Community detection
- Task 3: Recommendations
- Task 4: Connection sub-graphs
- Task 5: Mining graphs over time
- **Task 6: Virus/influence propagation**
- Task 7: Spectral graph theory
- Task 8: Tera/peta graph mining: hadoop
- Observations – patterns of real graphs
- Conclusions

Detailed outline

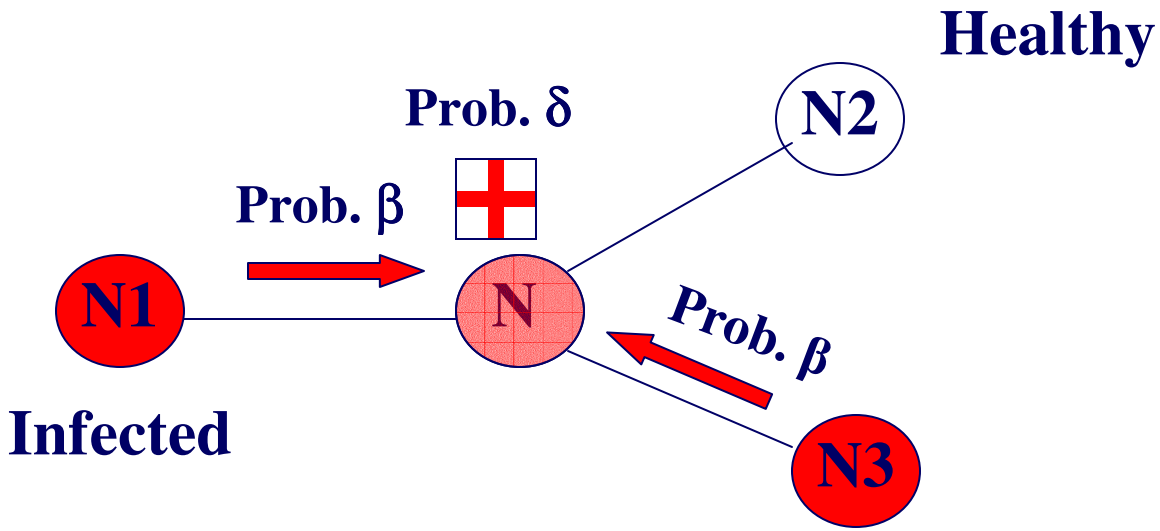
- Epidemic threshold
  - Problem definition
  - Analysis
  - Experiments
- Fraud detection in e-bay

Virus propagation

- How do viruses/rumors propagate?
- Blog influence?
- Will a flu-like virus linger, or will it become extinct soon?

The model: SIS

- 'Flu' like: Susceptible-Infected-Susceptible
- Virus 'strength' s= β/δ



Epidemic threshold τ

of a graph: the value of τ, such that

$$\text{if} \quad \text{strength } s = \beta / \delta < \tau$$

an epidemic can not happen

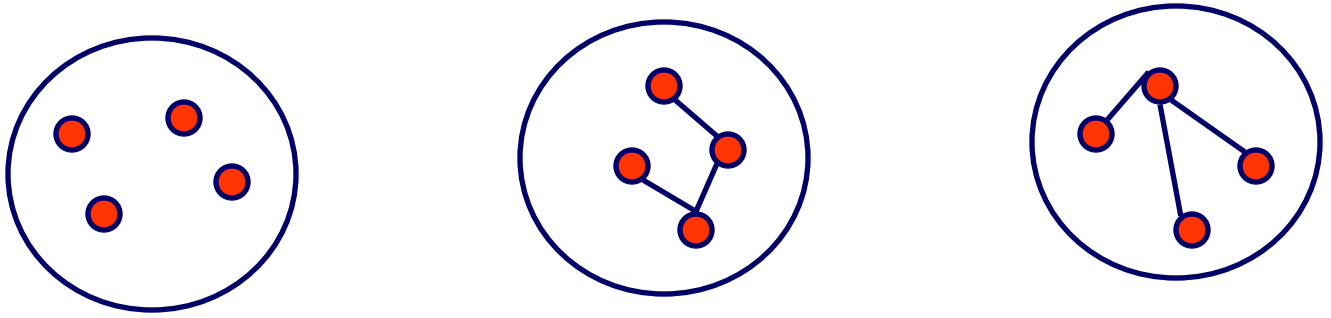
Thus,

- given a graph
- compute its epidemic threshold

Epidemic threshold τ

What should τ depend on?

- avg. degree? and/or highest degree?
- and/or variance of degree?
- and/or third moment of degree?
- and/or diameter?



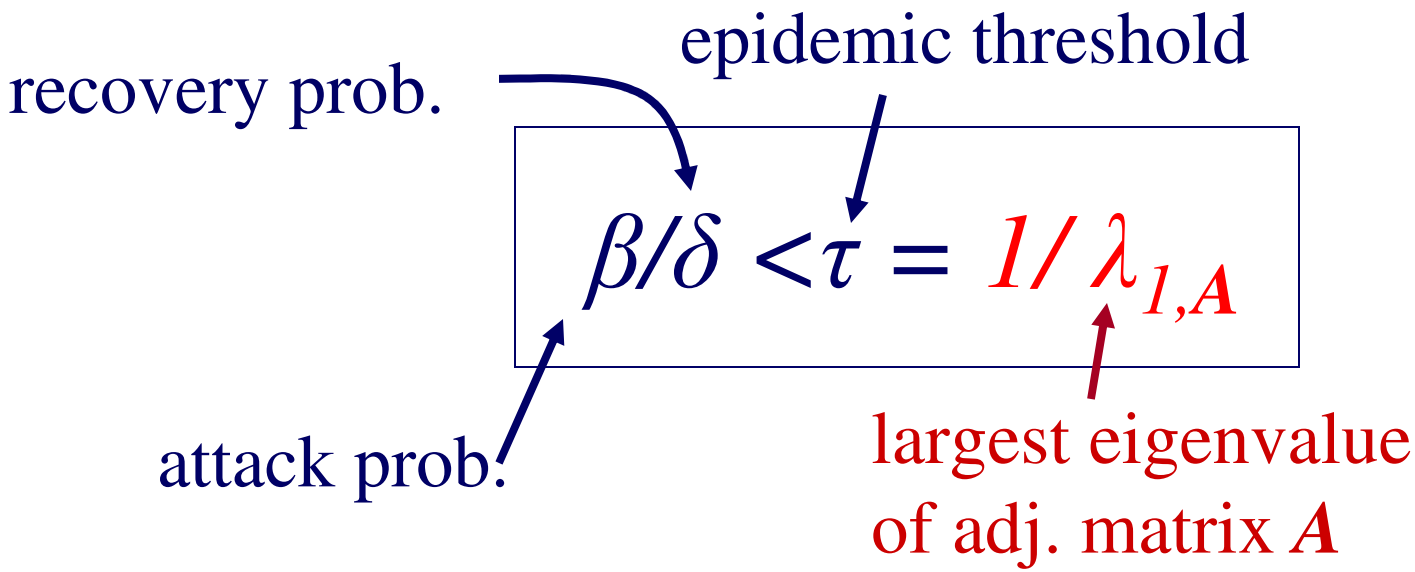
Epidemic threshold

- [Theorem 1] We have no epidemic, if

$$\beta/\delta < \tau = 1/\lambda_{1,A}$$

Epidemic threshold

- [Theorem 1] We have no epidemic (*), if



Proof: [Wang+03]

(*) under mild, conditional-independence assumptions

Beginning of proof

Healthy @ t+1:

- (healthy or healed)
- and not attacked @ t

Let: p(i , t) = Prob node i is sick @ t+1

$$1 - p(i, t+1) = (1 - p(i, t) + p(i, t) * \delta) * \prod_j (1 - \beta\, aji * p(j, t))$$

Below threshold, if the above *non-linear dynamical system* above is 'stable' (eigenvalue of Hessian < 1)

Epidemic threshold for various networks

Formula includes older results as special cases:

- Homogeneous networks [Kephart+White]
  - $\lambda_{1,A} = <k>$; $\tau = 1/<k>$ (<k> : avg degree)
- Star networks (d = degree of center)
  - $\lambda_{1,A} = sqrt(d)$; $\tau = 1/ sqrt(d)$
- Infinite power-law networks
  - $\lambda_{1,A} = \infty$; $\tau = 0$; [Barabasi]

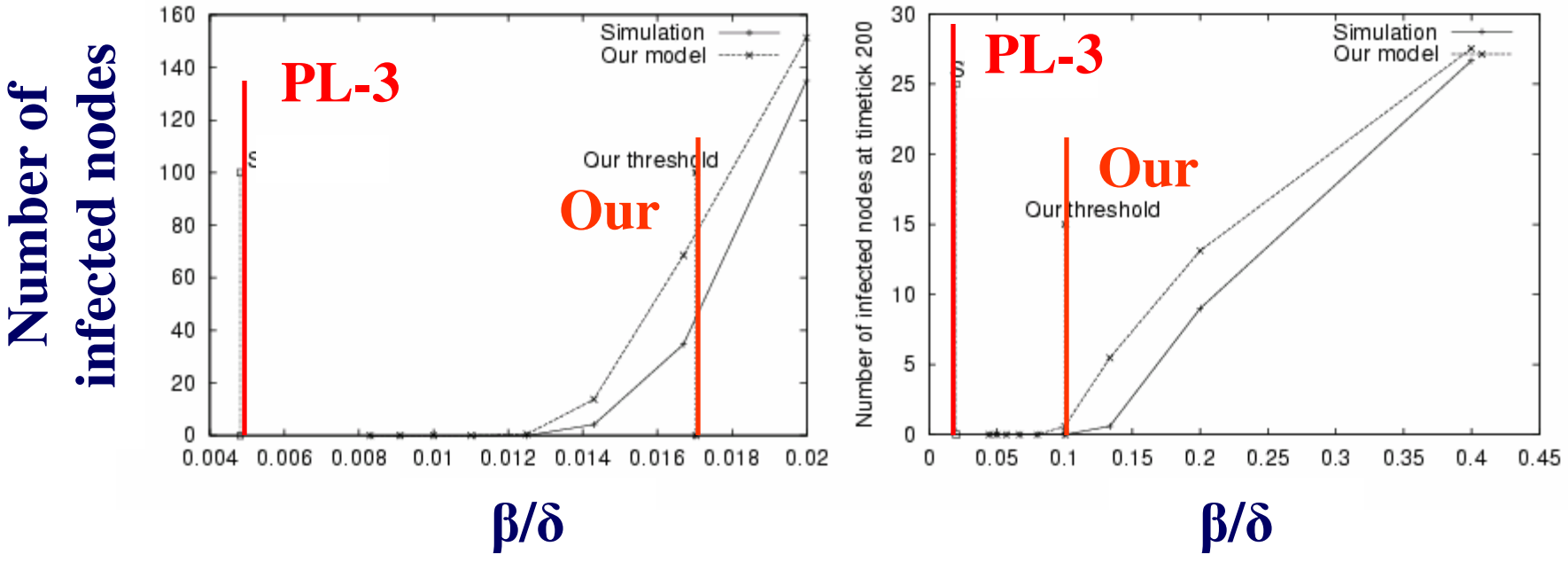
Epidemic threshold

- [Theorem 2] Below the epidemic threshold, the epidemic dies out exponentially

Detailed outline

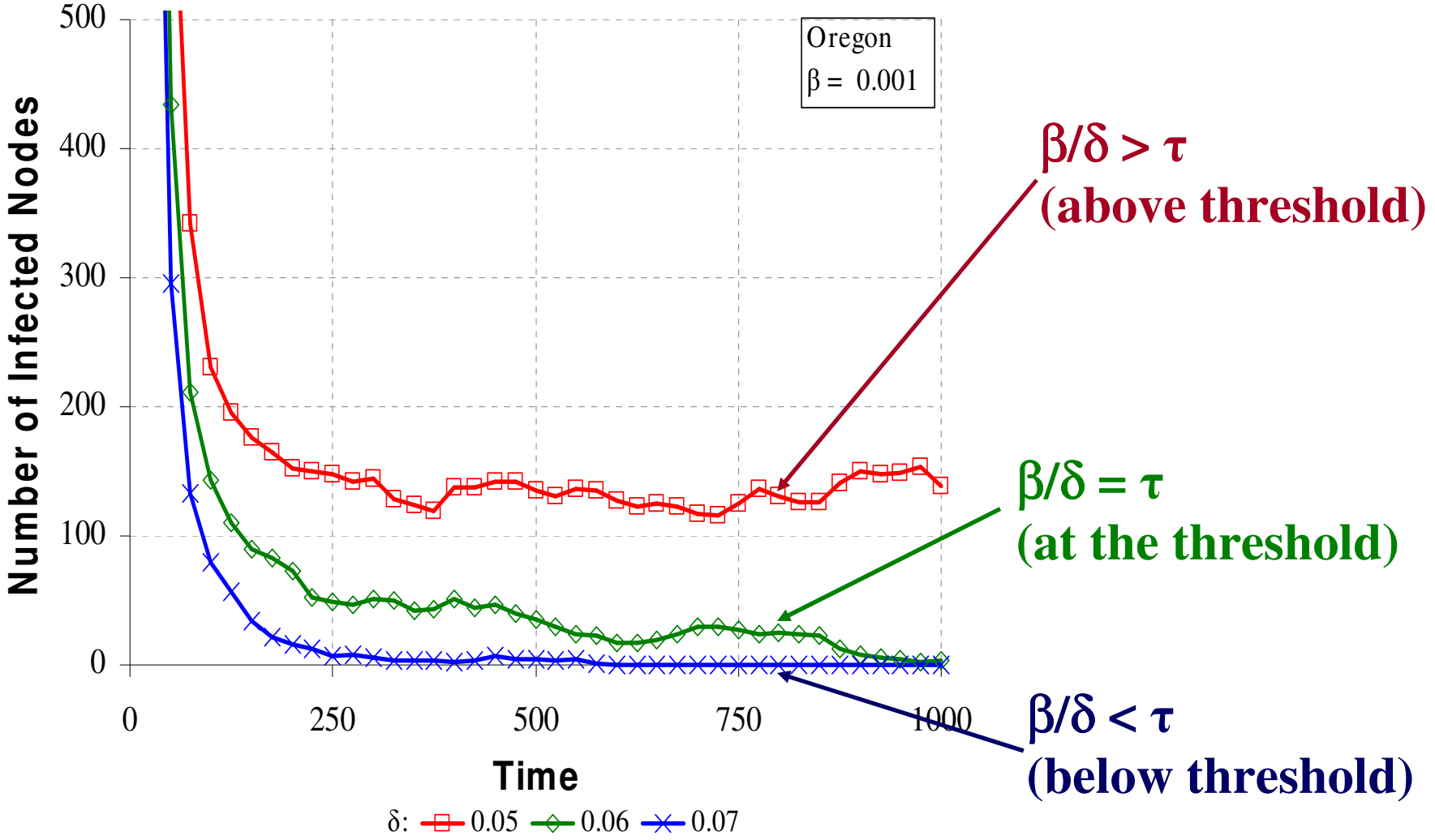
- Epidemic threshold
  - Problem definition
  - Analysis
  - Experiments
- Fraud detection in e-bay

Current prediction vs. previous



The formula's predictions are more accurate

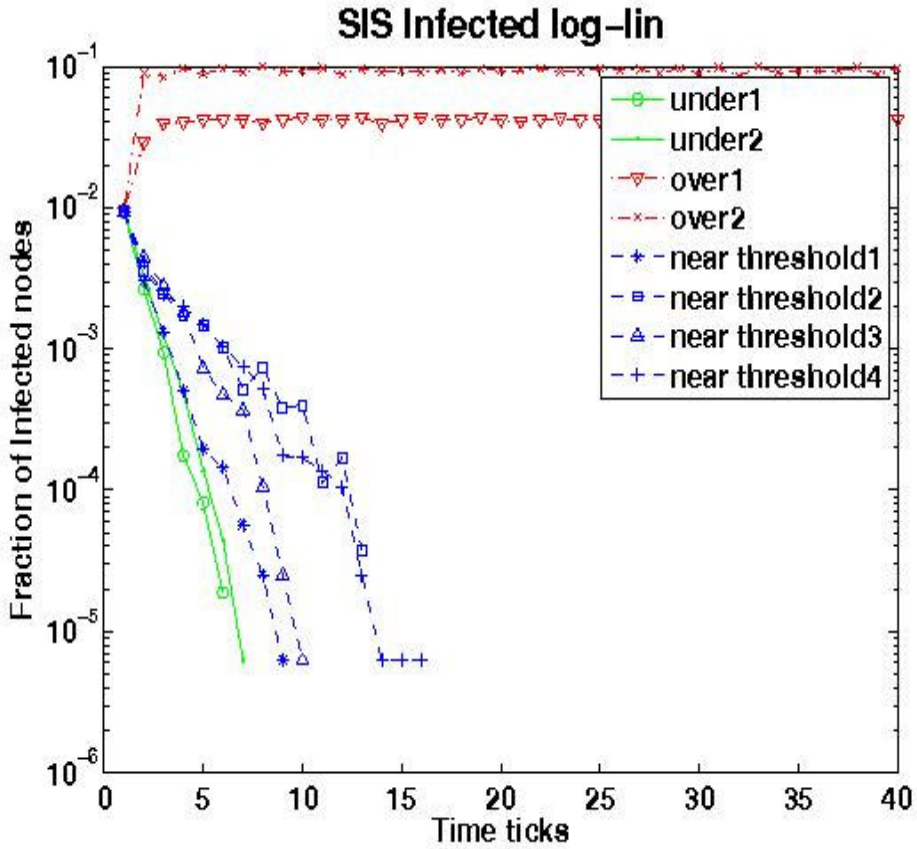
Experiments (Oregon)



SIS simulation - # infected nodes vs time

Log - Lin

#inf.
(log scale)



Time (linear scale)

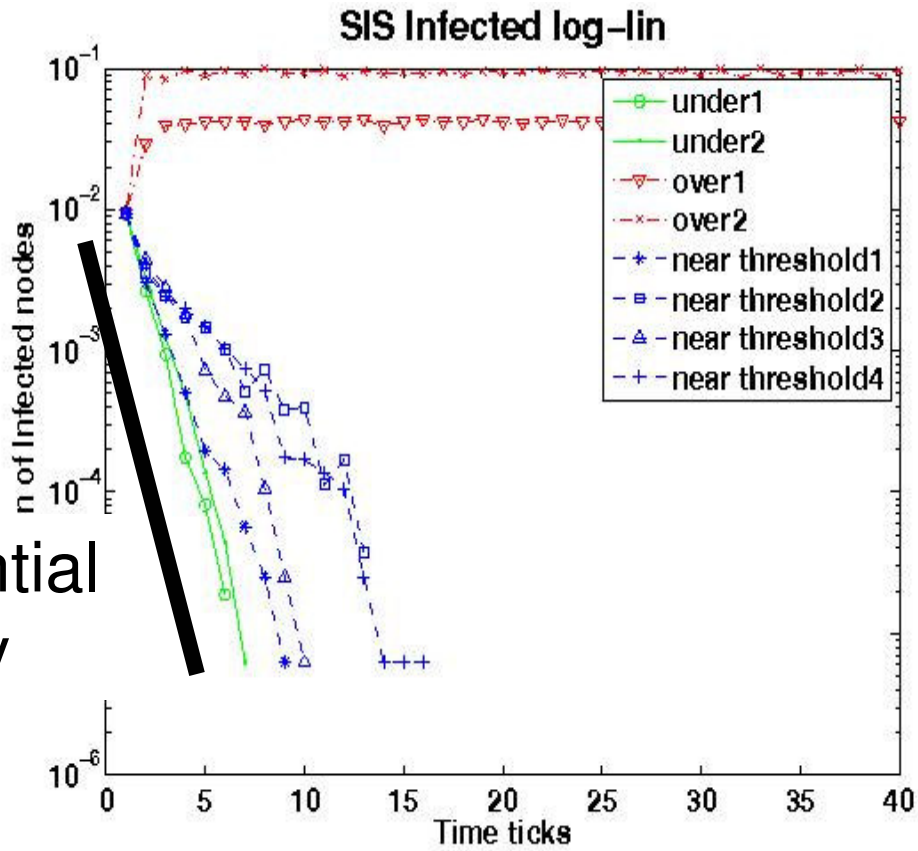
above
at
below

SIS simulation - # infected nodes vs time

Log - Lin

#inf. (log scale)

Exponential decay



Time (linear scale)

above

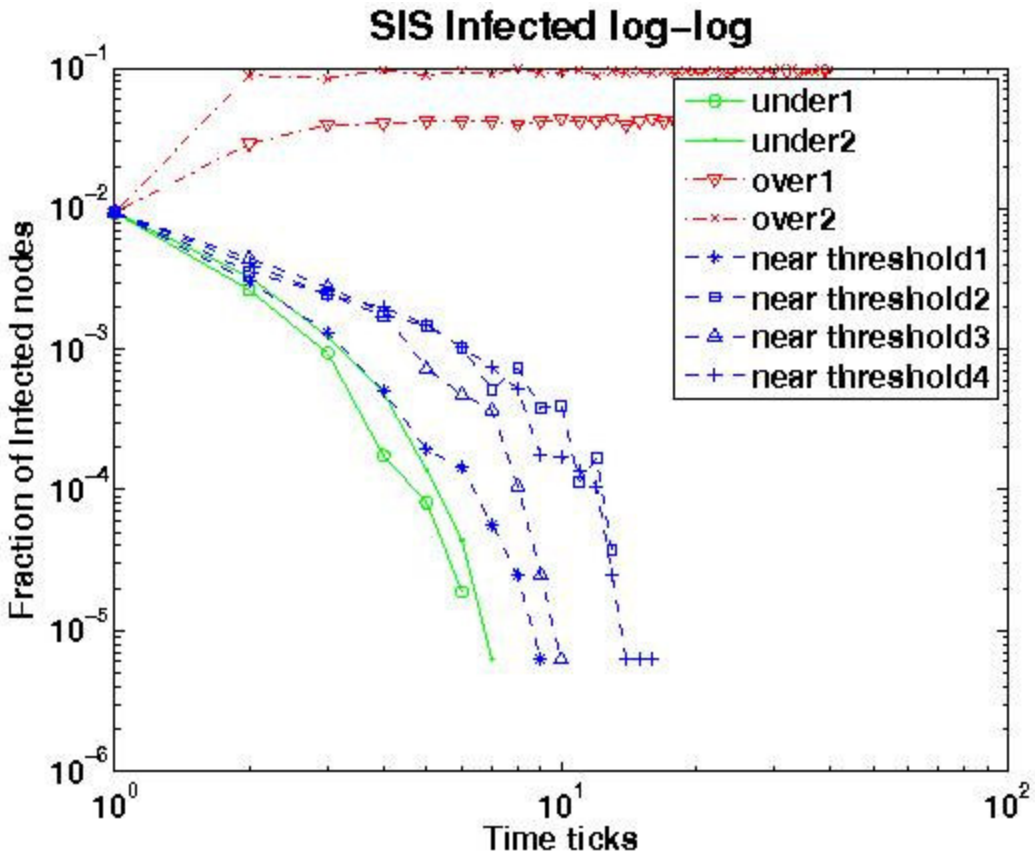
at

below

SIS simulation - # infected nodes vs time

Log - Log

#inf. (log scale)



Time (log scale)

above

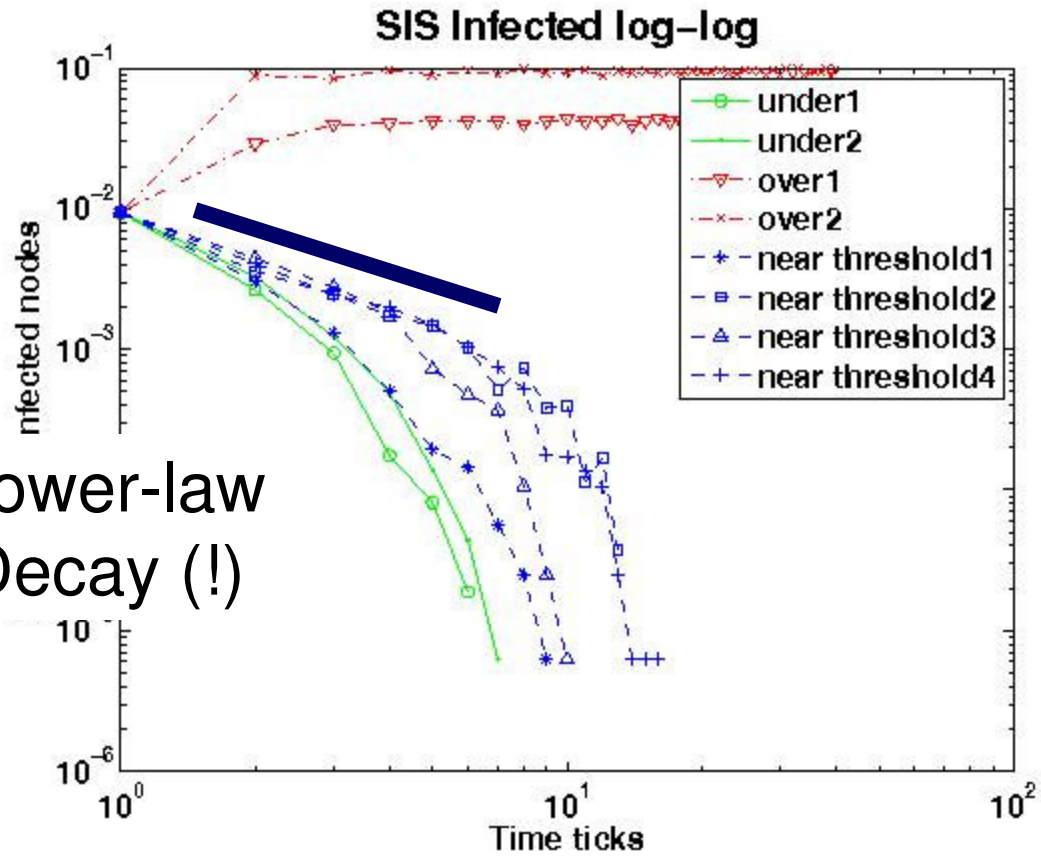
at

below

SIS simulation - # infected nodes vs time

Log - Log

#inf. (log scale)



Power-law Decay (!)

Time (log scale)

above

at

below

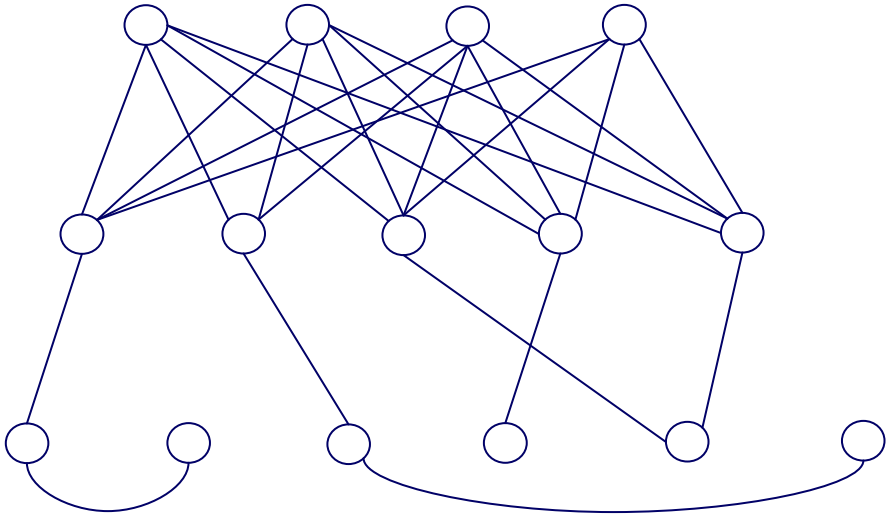
Detailed outline

- Epidemic threshold
- Fraud detection in e-bay

E-bay Fraud detection

extra

w/ Polo Chau &
Shashank Pandit, CMU

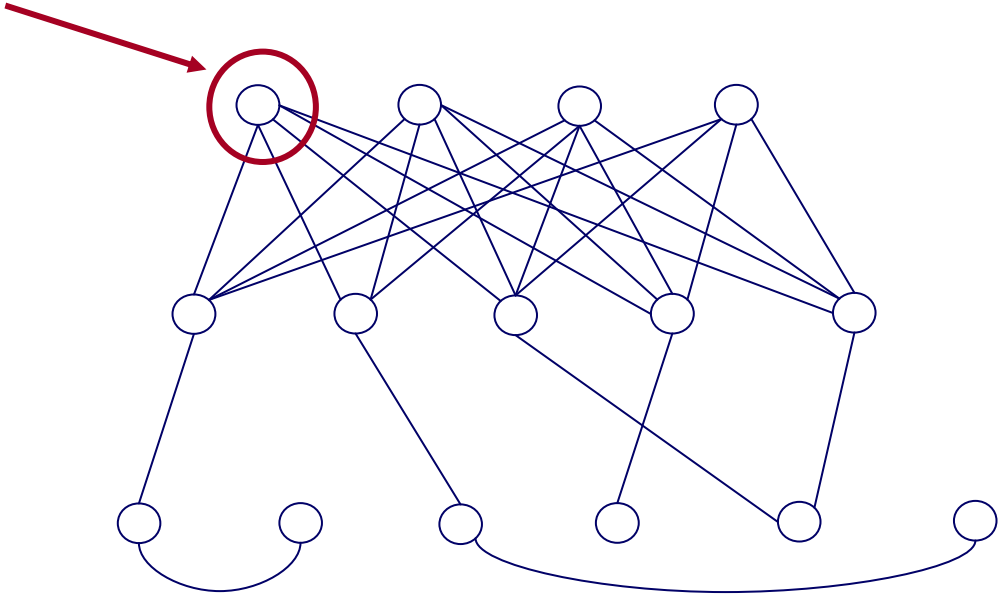


NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks, S. Pandit, D. H. Chau, S. Wang, and C. Faloutsos *(WWW'07)*, pp. 201-210

E-bay Fraud detection

extra

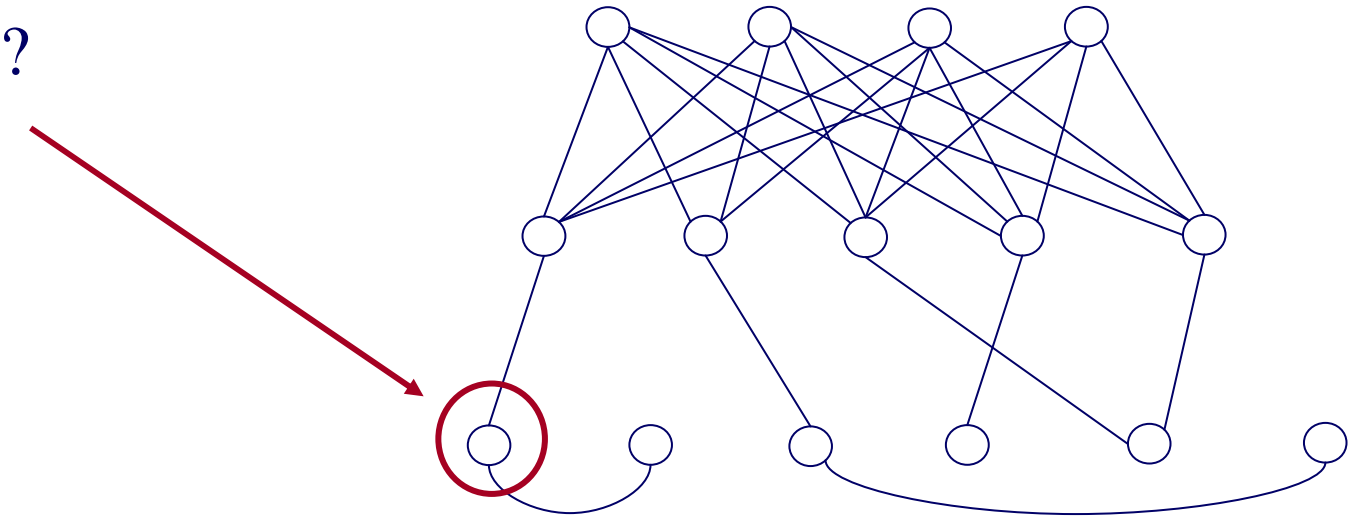
- lines: positive feedbacks
- would you buy from him/her?



E-bay Fraud detection

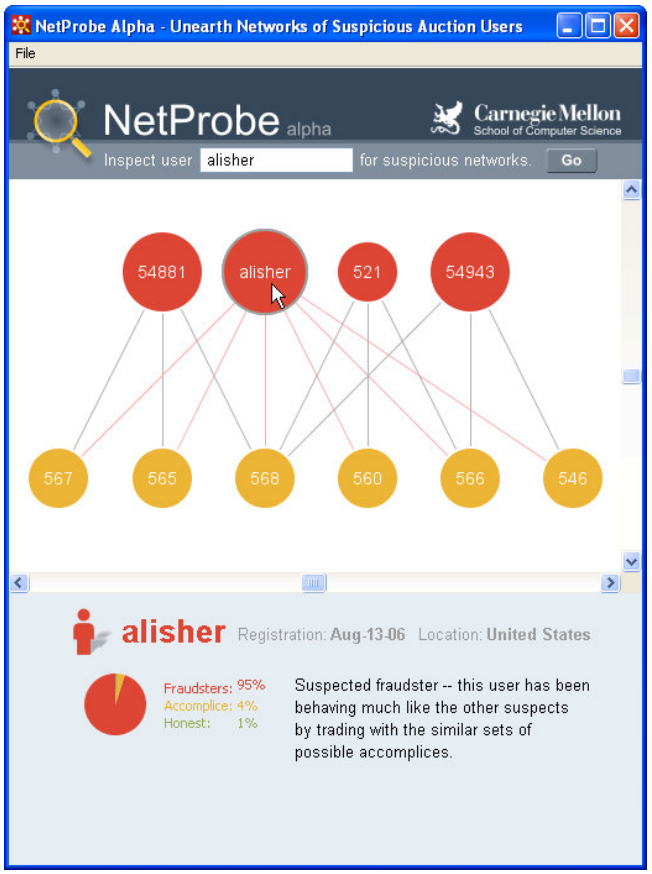
- lines: positive feedbacks
- would you buy from him/her?

- or him/her?

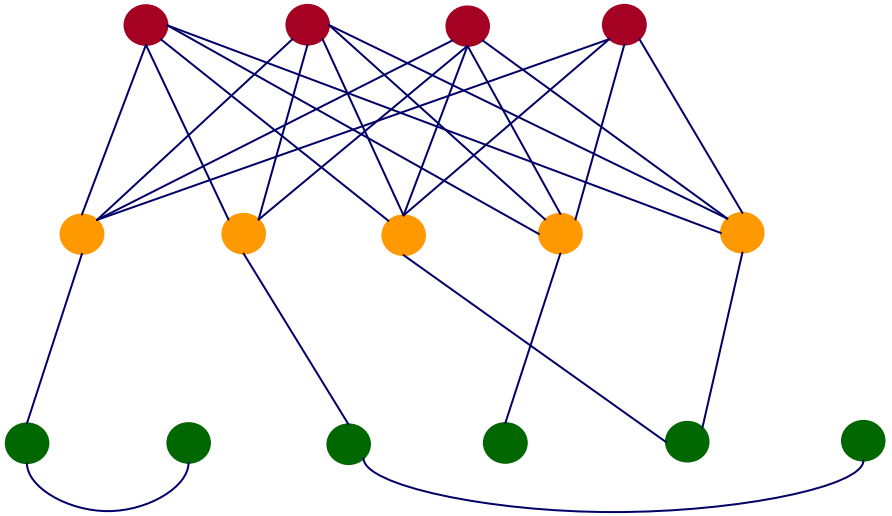


E-bay Fraud detection - NetProbe

extra



Belief Propagation gives:



Conclusions

- $\lambda_{1,A}$: Eigenvalue of adjacency matrix determines the survival of a flu-like virus
  - It gives a measure of how well connected is the graph (~ # paths – see Task 7, later)
  - May guide immunization policies
- [Belief Propagation: a powerful algo]

References

- D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, *Epidemic Thresholds in Real Networks*, in ACM TISSEC, 10(4), 2008
- Ganesh, A., Massoulie, L., and Towsley, D., 2005. The effect of network topology on the spread of epidemics. In *INFOCOM*.

References (cont'd)

- Hethcote, H. W. 2000. The mathematics of infectious diseases. *SIAM Review 42*, 599–653.
- Hethcote, H. W. AND Yorke, J. A. 1984. *Gonorrhea Transmission Dynamics and Control*. Vol. 56. Springer. Lecture Notes in Biomathematics.

References (cont'd)

- Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos, *Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint*, in SRDS 2003 (pages 25-34), Florence, Italy