Influence propagation in large graphs - theorems and algorithms

B. Aditya Prakash
http://www.cs.cmu.edu/~badityap

Christos Faloutsos
http://www.cs.cmu.edu/~christos

Carnegie Mellon University

GraphEx’12
Thank you!

• Nadya Bliss
• Lori Tsoulas
Networks are everywhere!

Facebook Network [2010]

Gene Regulatory Network [Decourty 2008]

Human Disease Network [Barabasi 2007]

The Internet [2005]

GraphEx '12

Prakash and Faloutsos 2012
Dynamical Processes *over* networks are also everywhere!
Why do we care?

- Information Diffusion
- Viral Marketing
- Epidemiology and Public Health
- Cyber Security
- Human mobility
- Games and Virtual Worlds
- Ecology
- Social Collaboration
Why do we care? (1: Epidemiology)

- Dynamical Processes over networks

"Diseases over contact networks"

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

[AJPH 2007]
Why do we care? (1: Epidemiology)

• Dynamical Processes over networks

Diseases over contact networks

CDC data: Visualization of the first 35 tuberculosis (TB) patients and their 1039 contacts
Why do we care? (1: Epidemiology)

- Dynamical Processes over networks

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

Diseases over contact networks
Why do we care? (1: Epidemiology)

• Dynamical Processes over networks

• Each circle is a hospital
• ~3000 hospitals
• More than 30,000 patients transferred

Problem: Given $k$ units of disinfectant, whom to immunize?
Why do we care? (1: Epidemiology)

~6x fewer!

CURRENT PRACTICE

OUR METHOD

Hospital-acquired inf. took 99K+ lives, cost $5B+ (all per year)
Why do we care? (2: Online Diffusion)

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

~100m active users

> 50m users
Why do we care? (2: Online Diffusion)

• Dynamical Processes over networks

Followers ➔ Celebrity ➔ Buy Versace™!

Social Media Marketing
High Impact – Multiple Settings

Q. How to squash **rumors** faster?

Q. How do **opinions** spread?

Q. How to **market** better?

epidemic out-breaks

products/viruses

transmit s/w patches
Research Theme

DATA
Large real-world networks & processes

ANALYSIS
Understanding

POLICY/
ACTION
Managing

Prakash and Faloutsos 2012
In this talk

Given propagation models:

Q1: Will an epidemic happen?

ANALYSIS
Understanding
In this talk

Q2: How to immunize and control out-breaks better?

POLICY/ ACTION
Managing
Outline

• Motivation
• Epidemics: what happens? (Theory)
• Action: Who to immunize? (Algorithms)
A fundamental question

Strong Virus

Epidemic?
A fundamental question

Epidemic?

Strong Virus

Prakash and Faloutsos 2012
A fundamental question

Epidemic?

Strong Virus

Prakash and Faloutsos 2012
A fundamental question

Epidemic?

Strong Virus

Prakash and Faloutsos 2012
A fundamental question

Strong Virus

Epidemic?
A fundamental question

Strong Virus

Epidemic?
example (static graph)

Weak Virus

Epidemic?
example (static graph)

Epidemic?

Weak Virus

Prakash and Faloutsos 2012
example (static graph)

Weak Virus

Epidemic?
example (static graph)

Weak Virus

Epidemic?
example (static graph)

Epidemic?

Weak Virus

Prakash and Faloutsos 2012
Problem Statement

Find, a condition under which

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

…..
Outline

- Motivation
- **Epidemics: what happens? (Theory)**
  - Background
  - Result (Static Graphs)
  - Proof Ideas (Static Graphs)
  - Bonus 1: Dynamic Graphs
  - Bonus 2: Competing Viruses
- Action: Who to immunize? (Algorithms)
“SIR” model: life immunity (mumps)

- Each node in the graph is in one of three states
  - Susceptible (i.e. healthy)
  - Infected
  - Removed (i.e. can’t get infected again)
Terminology: continued

• Other virus propagation models (“VPM”)
  – SIS: susceptible-infected-susceptible, flu-like
  – SIRS: temporary immunity, like pertussis
  – SEIR: mumps-like, with virus incubation
    \((E = \text{Exposed})\)

• Underlying contact-network – ‘who-can-infect-whom’
Related Work


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)
  - Proof Ideas (Static Graphs)
  - Bonus 1: Dynamic Graphs
  - Bonus 2: Competing Viruses

• Action: Who to immunize? (Algorithms)
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?

\[ \beta d_{avg} + \delta \sqrt{\text{diameter}} \ ? \ (\beta^2 d_{avg}^2 - \delta d_{avg}) / d_{max} \ ? \ ..... \]
### Static Graphs: Our Main Result

- **Informally,**

<table>
<thead>
<tr>
<th>For,</th>
<th>(\lambda)</th>
<th>(C_{VPM})</th>
</tr>
</thead>
<tbody>
<tr>
<td>any arbitrary topology (adjacency matrix (A))</td>
<td>[\lambda]</td>
<td>(C_{VPM})</td>
</tr>
<tr>
<td>any virus propagation model (VPM) in standard literature</td>
<td>[\lambda] * (C_{VPM})</td>
<td>(&lt; 1)</td>
</tr>
</tbody>
</table>

- The epidemic threshold depends only on:
  1. the \(\lambda\), **first eigenvalue of \(A\)**, and
  2. some **constant** \(C_{VPM}\), determined by the virus propagation model

---

In Prakash+ ICDM 2011 (Selected among **best papers**).
### Our thresholds for some models

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

<table>
<thead>
<tr>
<th>Models</th>
<th>Effective Strength ( (s) )</th>
<th>Threshold (tipping point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIS, SIR, SIRS, SEIR</td>
<td>( s = \lambda \cdot \left( \frac{\beta}{\delta} \right) )</td>
<td></td>
</tr>
<tr>
<td>SIV, SEIV</td>
<td>( s = \lambda \cdot \left( \frac{\beta \gamma}{\delta (\gamma + \theta)} \right) )</td>
<td>( s = 1 )</td>
</tr>
<tr>
<td>SI_1 I_2 V_1 V_2 (H.I.V.)</td>
<td>( s = \lambda \cdot \left( \frac{\beta_1 v_2 + \beta_2 \epsilon}{v_2 (\epsilon + v_1)} \right) )</td>
<td></td>
</tr>
</tbody>
</table>

Prakash and Faloutsos 2012

GraphEx '12
Our result: Intuition for $\lambda$

“Official” definition:
- Let $A$ be the adjacency matrix. Then $\lambda$ is the root with the largest magnitude of the characteristic polynomial of $A$ \([\det(A - xI)]\).
- Doesn’t give much intuition!

“Un-official” Intuition 😊
- $\lambda \sim \# $ paths in the graph
- $A^k(i, j) = \# $ of paths $i \rightarrow j$ of length $k$

Prakash and Faloutsos 2012

GraphEx '12
Largest Eigenvalue ($\lambda$)

Better connectivity $\rightarrow$ higher $\lambda$

\( \lambda \approx 2 \)  
(a) Chain

\( \lambda = \sqrt{N} \)  
(b) Star

\( \lambda = N - 1 \)  
(c) Clique

$N = 1000$

Prakash and Faloutsos 2012
Examples: Simulations – SIR (mumps)

(a) Infection profile
(b) “Take-off” plot

PORTLAND graph: synthetic population, 31 million links, 6 million nodes
Examples: Simulations – SIRS (pertussis)

(a) Infection profile
PORTLAND graph: synthetic population, 31 million links, 6 million nodes

(b) “Take-off” plot
Outline

• Motivation

• Epidemics: what happens? (Theory)
  - Background
  - Result (Static Graphs)
  - Proof Ideas (Static Graphs)
  - Bonus 1: Dynamic Graphs
  - Bonus 2: Competing Viruses

• Action: Who to immunize? (Algorithms)
General VPM structure

Model-based

Graph-based

\[ \lambda^* \hat{C}_{VPM} < 1 \]
Outline

• Motivation

• Epidemics: what happens? (Theory)
  – Background
  – Result (Static Graphs)
  – Proof Ideas (Static Graphs)
  – Bonus 1: Dynamic Graphs
  – Bonus 2: Competing Viruses

• Action: Who to immunize? (Algorithms)
Dynamic Graphs: Epidemic?

DAY
(e.g., work)

Alternating behaviors

adjacency matrix

Prakash and Faloutsos 2012
Dynamic Graphs: Epidemic?

NIGHT
(e.g., home)

Alternating behaviors

adjacency
matrix
Model Description

• SIS model
  - recovery rate $\delta$
  - infection rate $\beta$

• Set of $T$ arbitrary graphs

\[ \{ A_1, A_2 \ldots, A_T \} \]

\[ A_1 \quad \text{day} \]
\[ A_2 \quad \text{night} \]

\[ \text{, weekend} \ldots \]
Informally, no epidemic if $eig(S) = \lambda_s < 1$

Our result: Dynamic Graphs Threshold

- Single number!
- Largest eigenvalue of The system matrix $S$

$$S = \prod_i S_i$$

$$S_i = (1 - \delta)I + \beta A_i$$

In Prakash+, ECML-PKDD 2010
Infection-profile

$log(fraction \text{ infected})$

Synthetic

MIT Reality

GraphEx '12
Prakash and Faloutsos 2012
“Take-off” plots

Synthetic

MIT Reality

Footprint (# infected @ “steady state”)

\[ \lambda \prod_i S_i \] (log scale)
Outline

• Motivation

• Epidemics: what happens? (**Theory**)
  - Background
  - Result (Static Graphs)
  - Proof Ideas (Static Graphs)
  - Bonus 1: Dynamic Graphs
  - Bonus 2: Competing Viruses

• Action: Who to immunize? (**Algorithms**)
Competing Contagions

iPhone v Android

Blu-ray v HD-DVD

Biological: common flu/avian flu, pneumococcal inf etc
A simple model

- Modified flu-like
- Mutual Immunity ("pick one of the two")
- Susceptible-Infected1-Infected2-Susceptible
Question: What happens in the end?

ASSUME:
Virus 1 is stronger than Virus 2
Question: What happens in the end?

ASSUME:
Virus 1 is stronger than Virus 2
Answer: Winner-Takes-All

ASSUME:
Virus 1 is stronger than Virus 2
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:
   \[ \text{strength(Virus 1)} > \text{strength(Virus 2)} \]
3. Strength(Virus) = \( \lambda \beta / \delta \) → same as before!

In Prakash+ WWW 2012
Real Examples

[Google Search Trends data]

Reddit vs Digg

Blu-Ray vs HD-DVD

GraphEx '12
Prakash and Faloutsos 2012
Outline

• Motivation
• Epidemics: what happens? (Theory)
• Action: Who to immunize? (Algorithms)
Full Static Immunization

**Given**: a graph $A$, virus prop. model and budget $k$;

**Find**: $k$ ‘best’ nodes for immunization (removal).

$k = 2$
**Full Static Immunization**

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

$k = 2$

GraphEx '12  Prakash and Faloutsos 2012
Given: a graph $A$, virus prop. model and budget $k$;
Find: $k$ ‘best’ nodes for immunization (removal).
Outline

• Motivation
• Epidemics: what happens? (Theory)
• Action: Who to immunize? (Algorithms)
  – Full Immunization (Static Graphs)
  – Fractional Immunization

Prakash and Faloutsos 2012
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’?
Proposed vulnerability measure $\lambda$

$\lambda$ is the epidemic threshold

- **“Safe”**
  - Chain($\lambda = 1.73$)
- **“Vulnerable”**
  - Star($\lambda = 2$)
- **“Deadly”**
  - Clique($\lambda = 4$)

Increasing $\lambda$

Increasing vulnerability

Prakash and Faloutsos 2012
A1: “Eigen-Drop”: an ideal shield value

Eigen-Drop(S)

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

**Original Graph**

**Without \{2, 6\}**

Prakash and Faloutsos 2012
(Q2) - Direct Algorithm too expensive!

• 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 $\lambda$
    • It takes 2,615 years to find 5-best nodes!
A2: Our Solution

• Part 1: Shield Value
  – Carefully approximate Eigen-drop ($\Delta \lambda$)
  – Matrix perturbation theory

• Part 2: Algorithm
  – Greedily pick best node at each step
  – Near-optimal due to submodularity

• NetShield (linear complexity)
  – $O(nk^2 + m)$  \( n = \# \) nodes; \( m = \# \) edges

In Tong, Prakash+ ICDM 2010
Experiment: Immunization quality

Log(fraction of infected nodes)

NetShield
Degree
PageRank
Eigs (=HITS)
Acquaintance
Betweenness (shortest path)

Lower is better

Prakash and Faloutsos 2012

GraphEx '12
Outline

• Motivation
• Epidemics: what happens? (Theory)
• Action: Who to immunize? (Algorithms)
  – Full Immunization (Static Graphs)
  – Fractional Immunization
Fractional Immunization of Networks
B. Aditya Prakash, Lada Adamic, Theodore Iwashyna (M.D.), Hanghang Tong, Christos Faloutsos

Under review
Fractional Asymmetric Immunization

Drug-resistant Bacteria (like XDR-TB)

Hospital → Another Hospital

Prakash and Faloutsos 2012
Fractional Asymmetric Immunization

Drug-resistant Bacteria (like XDR-TB)

Hospital

Another Hospital
Fractional Asymmetric Immunization

**Problem:** Given $k$ units of disinfectant, how to distribute them to maximize hospitals saved?
Fractional Asymmetric Immunization

**Problem:** Given \( k \) units of disinfectant, how to distribute them to maximize hospitals saved?
Fractional Asymmetric Immunization

**Problem**: Given $k$ units of disinfectant, how to distribute them to maximize hospitals saved?

Hospital

Another Hospital
Our Algorithm “SMART-ALLOC”

~6x fewer!

[US-MEDICARE NETWORK 2005]
• Each circle is a hospital, ~3000 hospitals
• More than 30,000 patients transferred

CURRENT PRACTICE

SMART-ALLOC

GraphEx ’12
Prakash and Faloutsos 2012

79
Running Time

Wall-Clock Time

Simulations

SMART-ALLOC

≈

> 1 week

> 30,000x speed-up!

14 secs

Lower is better

Prakash and Faloutsos 2012
Experiments

Lower is better

PENN-NETWORK ~5 x
SECOND-LIFE ~2.5 x

K = 200
K = 2000

Prakash and Faloutsos 2012
Acknowledgements

Funding
References

1. Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks (B. Aditya Prakash, Deepayan Chakrabarti, Michalis Faloutsos, Nicholas Valler, Christos Faloutsos) - In IEEE ICDM 2011, Vancouver (Invited to KAIS Journal Best Papers of ICDM.)

2. Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms (B. Aditya Prakash, Hanghang Tong, Nicholas Valler, Michalis Faloutsos and Christos Faloutsos) – In ECML-PKDD 2010, Barcelona, Spain

3. Epidemic Spreading on Mobile Ad Hoc Networks: Determining the Tipping Point (Nicholas Valler, B. Aditya Prakash, Hanghang Tong, Michalis Faloutsos and Christos Faloutsos) – In IEEE NETWORKING 2011, Valencia, Spain

4. Winner-takes-all: Competing Viruses or Ideas on fair-play networks (B. Aditya Prakash, Alex Beutel, Roni Rosenfeld, Christos Faloutsos) – In WWW 2012, Lyon

5. On the Vulnerability of Large Graphs (Hanghang Tong, B. Aditya Prakash, Tina Eliassi-Rad and Christos Faloutsos) – In IEEE ICDM 2010, Sydney, Australia

6. Fractional Immunization of Networks (B. Aditya Prakash, Lada Adamic, Theodore Iwashyna, Hanghang Tong, Christos Faloutsos) - Under Submission

7. Rise and Fall Patterns of Information Diffusion: Model and Implications (Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, Christos Faloutsos) - Under Submission

http://www.cs.cmu.edu/~badityap/
Analysis

**Our thresholds for some models**

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

<table>
<thead>
<tr>
<th>Models</th>
<th>Effective Strength ($s$)</th>
<th>Threshold (tipping point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIS, SIR, SIRS, SEIR</td>
<td>$s = \lambda \left( \frac{\beta r}{\alpha} \right)$</td>
<td>$s = 1$</td>
</tr>
<tr>
<td>SIV, SEIV</td>
<td>$s = \lambda \left( \frac{\beta r}{\alpha r + \theta} \right)$</td>
<td></td>
</tr>
<tr>
<td>SI, I, V, I, V</td>
<td>$s = \lambda \left( \frac{\beta I V}{V I + \alpha I} \right)$</td>
<td></td>
</tr>
</tbody>
</table>

**Our Algorithm “SMART-ALLOC”**

- ~6x fewer!

**Real Examples**

- [US-MEDICARE NETWORK 2005]
  - Each circle is a hospital, ~3000 hospitals
  - More than 30,000 patients transferred

- [Google Search Trends data]