Lecture #29: Graph mining - virus propagation & immunization

Christos Faloutsos

Must-read material

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

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

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?)
Problem

- Q1: epidemic?
- Q2: whom to immunize
- (Q3: 2 competing viruses – end result?)

Short answers

- Q1: epidemic?
- A1: tipping point: eigenvalue
- Q2: whom to immunize
- A2: eigen-drop
- (Q3: 2 competing viruses – end result?)
Influence propagation in large graphs - theorems and algorithms

Prof. B. Aditya Prakash
http://people.cs.vt.edu/~badityap/

Networks are everywhere!

Facebook Network [2010]
Gene Regulatory Network [Decourty 2008]
Human Disease Network [Barabasi 2007]
The Internet [2005]
Dynamical Processes *over* networks are also everywhere!

**Why do we care?**
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 [AJPH 2007]

Diseases over contact networks

CDC data: Visualization of the first 35 tuberculosis (TB) patients and their 1039 contacts
Why do we care? (2: Online Diffusion)

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

~100m active users

> 50m users

• Dynamical Processes over networks

Social Media Marketing
Outline

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

A fundamental question

Strong Virus

Epidemic?
Problem Statement

Find, a condition under which

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

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

“SIR” model: life immunity (mumps)

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

\[ t = 1 \quad t = 2 \quad t = 3 \]
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’

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

For,

- any arbitrary topology (adjacency matrix $A$)
- any virus propagation model ($VPM$) in standard literature

the epidemic threshold depends only

1. on the $\lambda$, first eigenvalue of $A$, and
2. some constant $C_{VPM}$ determined by the virus propagation model

No epidemic if $\lambda \times C_{VPM} < 1$

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>$s = \lambda \left( \frac{\beta}{\delta} \right)$</td>
<td>$s = 1$</td>
</tr>
<tr>
<td>SIV, SEIV</td>
<td>$s = \lambda \cdot \left( \frac{\beta \gamma}{\delta (\gamma + \theta)} \right)$</td>
<td></td>
</tr>
<tr>
<td>$SI_1I_2V_1V_2$ (H.I.V.)</td>
<td>$s = \lambda \left( \frac{\beta_1v_2 + \beta_2\varepsilon}{\delta (v_2 + v_1)} \right)$</td>
<td></td>
</tr>
</tbody>
</table>

Copyright (c) 2019 A. Prakash and C. Faloutsos
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 $A$ (i.e., $\text{det}(A - \lambda I)$).
- Doesn’t give much intuition!

**“Un-official” Intuition**
- $\lambda \approx \# \text{ paths in the graph}$
- $A^k(i, j) = \# \text{ of paths } i \to j \text{ of length } k$

$Largest\ Eigenvalue\ (\lambda)$

better connectivity $\rightarrow$ higher $\lambda$

\[
\begin{align*}
\lambda &\approx 2 \\
\lambda &\approx 2 \\
\lambda &= \lambda = \sqrt{N} \\
\lambda &= \lambda = N-1 \\
N &= 1000
\end{align*}
\]
Examples: Simulations – SIR (mumps)

(a) Infection profile
(b) “Take-off” plot
PORTLAND graph: synthetic population, 31 million links, 6 million nodes

Examples: Simulations – SIRS (pertusis)

(a) Infection profile
(b) “Take-off” plot
PORTLAND graph: synthetic population, 31 million links, 6 million nodes
Outline

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

Competing Contagions

- iPhone vs Android
- Blu-ray vs 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

\[ \begin{align*}
I_1 &\quad \leftrightarrow \quad S \quad \leftrightarrow \quad I_2 \\
\beta_1 &\quad \delta_1 &\quad \beta_2 &\quad \delta_2 \\
\text{Virus 1} &\quad \text{Virus 2} 
\end{align*} \]

Question: What happens in the end?

Footprint @ Steady State

ASSUME:
Virus 1 is stronger than Virus 2

\[ \text{Footprint @ Steady State} = ? \]
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]
Outline

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

Immunization

**Given:** a graph \( A \), virus prop. model and budget \( k \);

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

\[ k = 2 \]
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: $\lambda$

“Safe”  “Vulnerable”  “Deadly”

(a)Chain($\lambda = 1.73$)  (b)Star($\lambda = 2$)  (c)Clique($\lambda = 4$)

$\textit{higher } \lambda, \textit{ higher vulnerability}$
A1: “Eigen-Drop”: an ideal shield value

\[ \Delta \lambda = \lambda - \hat{\lambda}_S \]

Original Graph

Without \{2, 6\}

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
Experiment: Immunization

Log(fraction of infected nodes)

Lower is better

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!