

Optimization vs Epidemics

Paul Beaujean

joint work with Cristina Bazgan & Éric Gourdin

LAMSADE - Université Paris-Dauphine & Orange Labs

COCOA 2018

December 16, 2018



WannaCry (May 2017)



113,068

ONLINE



113,732

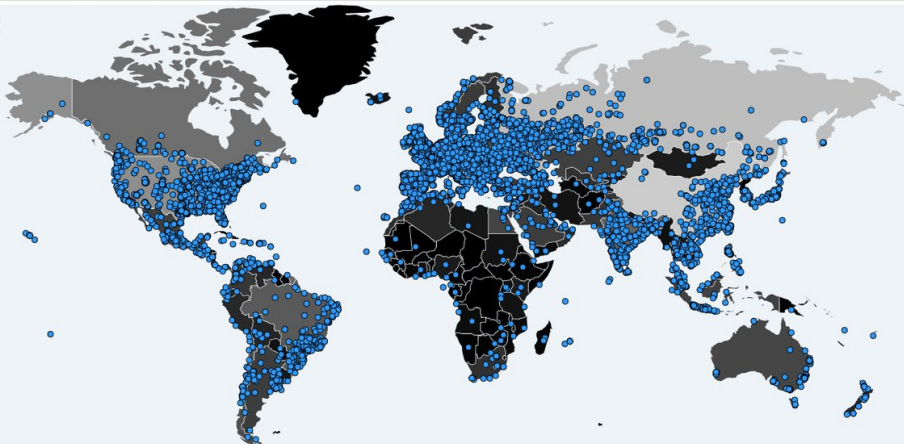
OFFLINE



226,800

TOTAL

📍 Infection Map (age: 0h 26m 58s)



Petya (from early 2016 to June 2017)




CHRONIQUES
DES (R)ÉVOLUTIONS NUMÉRIQUES

VIE EN LIGNE

Le virus Petya a coûté plus d'un milliard d'euros aux entreprises

Des ports de marchandise à l'arrêt, des usines immobilisées et des entreprises ralenties... la facture de ce faux « rançongiciel » est salée, selon un décompte du « Monde ».

SDN: software-defined networking

 OPENDAYLIGHT

Devices

Flows

Troubleshoot

admin

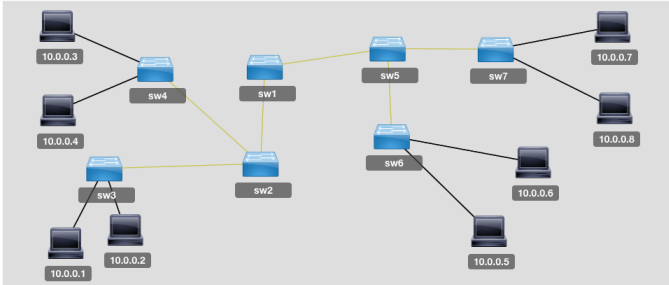
Nodes Learned

Nodes Learned

Node Name	Node ID	Ports
sw2	OF 00:00:00:00:00:00:02	3
sw3	OF 00:00:00:00:00:00:03	3
sw1	OF 00:00:00:00:00:00:01	2
sw4	OF 00:00:00:00:00:00:04	3
sw7	OF 00:00:00:00:00:00:07	3

1-5 of 7 items

Page 1 of 2



```
graph LR; sw1 --- sw2; sw1 --- sw3; sw1 --- sw4; sw1 --- sw5; sw1 --- sw7; sw2 --- sw3; sw2 --- sw4; sw2 --- sw5; sw2 --- sw7; sw3 --- sw4; sw3 --- sw5; sw3 --- sw7; sw4 --- sw5; sw4 --- sw7; sw5 --- sw7; h1[10.0.0.1] --- sw3; h2[10.0.0.2] --- sw3; h3[10.0.0.3] --- sw4; h4[10.0.0.4] --- sw4; h5[10.0.0.5] --- sw6; h6[10.0.0.6] --- sw6; h7[10.0.0.7] --- sw7; h8[10.0.0.8] --- sw7
```

Static Route Configuration

Connection Manager

Static Route Configuration

[Add Static Route](#) [Remove Static Route](#)

<input type="checkbox"/>	Name	Static Route	Next Hop Address
0 items			

Subnet Gateway Configuration

SPAN Port Configuration

Subnet Gateway Configuration

[Add Gateway IP Address](#) [Remove Gateway IP Address](#) [Add Ports](#)

<input type="checkbox"/>	Name	Gateway IP Address/Mask	Ports
<input type="checkbox"/>	default (cannot be modified)	0.0.0.0/0	

1-1 of 1 item

Page 1 of 1

Poseidon: machine learning for node health



This repository

Search

Pull requests

Issues

Marketplace

Explore



CyberReboot / poseidon

Watch

20

Star

76

Fork

40

Code

Issues 22

Pull requests 0

Projects 1

Wiki

Insights

Poseidon is a python-based application that leverages software defined networks (SDN) to acquire and then feed network traffic to a number of machine learning techniques. The machine learning algorithms classify and predict both the type of device and if the device is acting normally or abnormally.

machine-learning

sdn

network-analysis

networking

software-defined-network

network-forensics

pcap

pcap-files

pcap-analyzer

1,580 commits

3 branches

6 releases

23 contributors

Branch: master

New pull request

Create new file

Upload files

Find file

Clone or download



cglewis Merge pull request #497 from cglewis/master

Latest commit 46f6e3c 12 hours ago

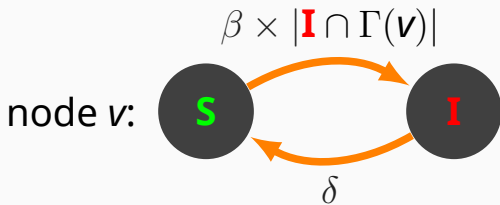
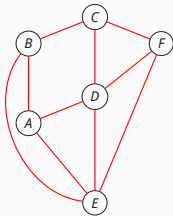
.circlec	more coverage	4 months ago
config	fix tests	7 days ago
docs	bump version back to dev	13 days ago

SDN against malware

- transparent and unified data analysis
- network management by algorithms
- ▷ opportunity to design autonomous security systems for networks
- our focus: stopping malware epidemics

Networked epidemic models

- undirected graph $G = (V, E)$ represents neighborhood of $n := |V|$ nodes
- a node is either **S**usceptible or **I**nfected



- Markov chain with 2^n states
- unique absorbing state: every node is **S**

sufficient condition for
 $O(\log n)$ time to extinction:

$$\lambda_1(\mathbf{G}) < \frac{\delta}{\beta}$$

[Prakash et al. '12]

From a theorem to a control system

How to achieve $\lambda_1(\mathbf{G}) < \frac{\delta}{\beta}$?

note: β and δ can be inferred in real time with e.g. maximum likelihood estimators [Ruhi et al. '17]

A) deploy software patches to increase δ/β

→ hand-crafted response against specific malware

→ hard to predict the effect of a patch on β and/or δ

B) modify topology to decrease $\lambda_1(\mathbf{G})$

→ generic response against any epidemic

From a theorem to a control system

How to achieve $\lambda_1(\mathbf{G}) < \frac{\delta}{\beta}$?

note: β and δ can be inferred in real time with e.g. maximum likelihood estimators [Ruhi et al. '17]

A) deploy software patches to increase δ/β

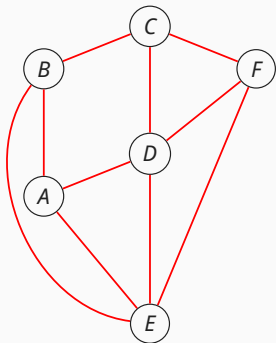
→ hand-crafted response against specific malware

→ hard to predict the effect of a patch on β and/or δ

B) modify topology to decrease $\lambda_1(\mathbf{G})$

→ generic response against any epidemic

What is $\lambda_1(G)$?



$$\begin{matrix} & A & B & C & D & E & F \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \end{matrix}$$

spectrum of the above adjacency matrix:

$$\{ \mathbf{3.39}, 0.19, 0.81, -0.55, -1.57, -2.27 \}$$

The spectral radius of $G = (V, E)$

- structural param. related to connectivity

$$\max \left(\sqrt{\Delta(G)}, \frac{2m}{n} \right) \leq \lambda_1(G) \leq \Delta(G)$$

- a graph with no edges has $\lambda_1(\mathbf{0}) = 0$

Theorem: λ_1 is monotone

if H is a subgraph of G then $\lambda_1(H) \leq \lambda_1(G)$

Theorem: λ_1 is positive homogeneous

if $\alpha > 0$ then $\lambda_1(\alpha A) = \alpha \lambda_1(A)$

A natural optimization problem

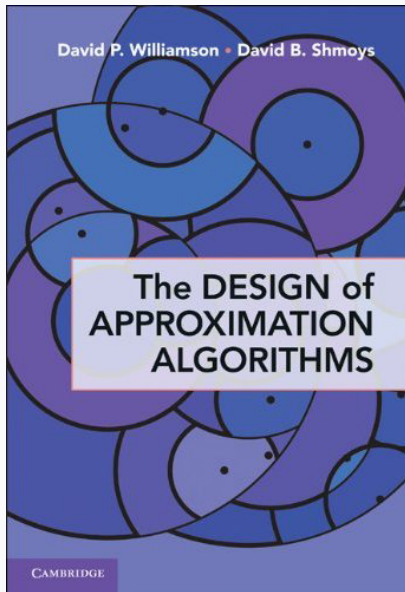
Problem: MAXIMUM λ -SPECTRAL SUBGRAPH

Instance: undirected graph $G = (V, E)$ and spectral threshold $1 \leq \lambda < \lambda_1(G)$

Task: find a subgraph $H = (V, E')$ of G with maximum number of edges and $\lambda_1(H) \leq \lambda$

- Set $\lambda = \frac{\delta}{\beta}$ to end a given epidemic
- NP-hard even in graphs with $\Delta \leq 3$

I always have trouble with



MλSSP as a mathematical program

$$\begin{aligned} \max \quad & \sum_{ij \in E} y_{ij} \\ \text{s.t.} \quad & \sum_{ij \in E} y_{ij} A_{ij} \preceq \lambda I \quad (\mathcal{P}) \\ & y_{ij} \in \{0, 1\}, \forall ij \in E \end{aligned}$$

- recall that $M \succeq N \iff \forall i, \lambda_i(M - N) \geq 0$
- hence $A \preceq \lambda I \iff \lambda_1(A) \leq \lambda$
- binary SDP amenable to MISOCP solvers

A polytime solvable relaxation

$$\begin{aligned} \max \quad & \sum_{ij \in E} y_{ij} \\ \text{s.t.} \quad & \sum_{ij \in E} y_{ij} A_{ij} \preceq \lambda I \quad (\mathcal{S}) \\ & \sum_{j \in \Gamma(i)} y_{ij} \leq \lambda^2, \forall i \in V \\ & y_{ij} \in [0, 1], \forall ij \in E \end{aligned}$$

- valid inequality because: $\lambda_1 \leq \lambda \Rightarrow \Delta \leq \lambda^2$
- solvable by fast SDP solvers e.g. SuperSCS

Relaxation & randomized rounding

- solve relaxation and let $y^* = \arg \mathcal{S}$
- define r.v. $\mathbf{x}_{ij} \sim \text{Bernoulli}(y_{ij}^*/r)$
- let $H = \{ij \in E \mid x_{ij} = 1\}$
- determine $r > 1$ s.t. H is feasible w.h.p.
 $\Pr(\lambda_1(\sum \mathbf{x}_{ij} \mathbf{A}_{ij}) \leq \lambda) = 1 - 1/n$
- ▷ randomized r -approximation algorithm
- oldie but goodie [Raghavan & Thompson '87]

Approximate maximization

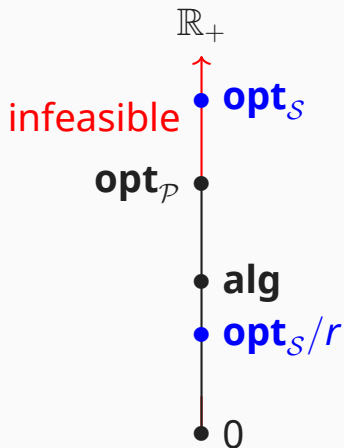


Figure: **alg** is the expected number of edges in H

Determining the approx. ratio r

- instead of designing algorithms...
 - ▷ prove concentration for a random object!
- here: random adjacency matrices
- use concentration inequalities for sum of random matrices $\mathbf{X} = \sum_e \mathbf{X}_e$ [Tropp '15]

$$\Pr(\lambda_1(\mathbf{X}) \geq a) \leq \min_{t>0} e^{-ta} \operatorname{tr} \exp \left(\sum_e \log (\mathbb{E} e^{t\mathbf{X}_e}) \right)$$

Matrix Bernstein gives

- for $\mathbf{X} = \sum_e \mathbf{X}_e$ with $\mathbb{E}\mathbf{X} = \mathbf{0}$ and $\lambda_1(\mathbf{X}_e) \leq L$
- matrix variance: $v(\mathbf{X}) \stackrel{\text{def}}{=} \lambda_1(\sum_e \mathbb{E}\mathbf{X}_e^2)$

$$\Pr(\lambda_1(\mathbf{X}) \geq a) \leq n \exp\left(-\frac{a^2}{2v(\mathbf{X}) + 2La/3}\right)$$

- apply to $\bar{\mathbf{A}} = \sum \mathbf{x}_{ij}\mathbf{A}_{ij} - \mathbb{E}\mathbf{x}_{ij}\mathbf{A}_{ij}$, $a = (1 - 1/r)\lambda$

Goal: $\Pr(H \text{ is infeasible}) \leq 1/3$

Matrix Bernstein gives

- for $\mathbf{X} = \sum_e \mathbf{X}_e$ with $\mathbb{E}\mathbf{X} = \mathbf{0}$ and $\lambda_1(\mathbf{X}_e) \leq L$
- matrix variance: $v(\mathbf{X}) \stackrel{\text{def}}{=} \lambda_1(\sum_e \mathbb{E}\mathbf{X}_e^2)$

$$\Pr(\lambda_1(\mathbf{X}) \geq a) \leq n \exp\left(-\frac{a^2}{2v(\mathbf{X}) + 2La/3}\right)$$

- apply to $\bar{\mathbf{A}} = \sum \mathbf{x}_{ij}A_{ij} - \frac{y_{ij}^*}{\mathbf{r}}A_{ij}$, $a = (1 - 1/\mathbf{r})\lambda$

Goal: $\Pr(H \text{ is infeasible}) \leq 1/3$ for $\mathbf{r} = ?$

$r = O(\log n)$ when $\lambda \geq \log n$

- key property: $v(\bar{\mathbf{A}}) \leq \lambda^2/r$ (cf. valid ineq.)
- crank out the math...

Result: $\Pr(H \text{ is infeasible}) \leq 1/3$ for $r = 1 + 3 \log n$ if $\lambda \geq \log n$

- from $1/3$ to $1/n$ with amplification

What about $\lambda \leq \log n$?

- spectral ver. of classical result on Δ vs ν

$$\nu(G) \geq \frac{m}{\lambda_1^2(G) - 1}$$

- implies that a maximum matching is a $(\lambda^2 - 1)$ -approximation of M λ SSP

Result: $O(\log^2 n)$ -approx. if $\lambda \leq \log n$

Conclusion

- randomized $O(\log^2 n)$ -approximation algorithm
- parallel approximate SDP solver + parallel independent rounding + power method
- super fast in practice, can exploit GPUs

Perspectives

- hardness of approx. (Max k -Cut?)
- Erdős-Rényi graphs limit our approach:

$$\lambda_1(G(n, p)) = \Theta \left(\sqrt{\frac{\log n}{\log \log n}} \right)$$

even when $np = O(1)$ [Krivelevich et al. '01]

- maximum degree-constrained subgraph for $\lambda \leq \log n$ (from λ^2 to λ -approx?)
- better concentration bounds? [Le et al., '15]



Any questions?

Problem variants

- spectral subgraph setting with two parameters: number of edges vs spectral radius

	constraint	objective
v. Miegheem et al. '12	$ E' = k$	$\min \lambda_1$
Saha et al. '15	$\lambda_1 \leq \lambda$	$\min E - E' $
Zhang et al. '15	$ E' \leq k$	$\min \lambda_1$
this work	$\lambda_1 \leq \lambda$	$\max E' $

The Laplacian spectrum

- $\mu \leq \mu_2(H)$ instead of $\lambda_1(H) \leq \lambda$
- related to expander construction:

$$\begin{aligned} \min \quad & \sum_{ij \in \bar{E}} y_{ij} \\ \text{s.t.} \quad & L(G) + \sum_{ij \in \bar{E}} y_{ij} L_{ij} \succeq \mu P_{\mathbf{1}^\perp} \quad (\mathcal{Q}) \\ & y_{ij} \in \{0, 1\}, \forall ij \in \bar{E} \end{aligned}$$

- some results [Ghosh & Boyd '06, Kolla et al. '09]

