Parcours Intelligence Artificielle

Machine Learning for Network Modeling

Argyris Kalogeratos

name.surname@ens.paris-saclay.fr

LIPS research group

Center Giovanni Borelli

ENS Paris-Saclay

Course plan

Short course on **Machine Learning** for **Network** Modeling

Planning: 4 dense sessions, 2.5 hours each

1. Introduction to Graph Theory and Network Science
2. Network models - Static and dynamic graphs*
3. Structure and topology inference
4. Processes and signals over graphs

* by Fabian Tarissan, CNRS, ENS Paris-Saclay

fabien.tarissan@ens-paris-saclay.fr

How is this going to work?

Attend the courses

Do a short project

- Using the tools of the course for a problem of your main discipline or a thematic you'd like to pursue in the future
- The subject and perimeter of each project should be discussed
- Deliverables: report + codes (Matlab, R, Python, …)

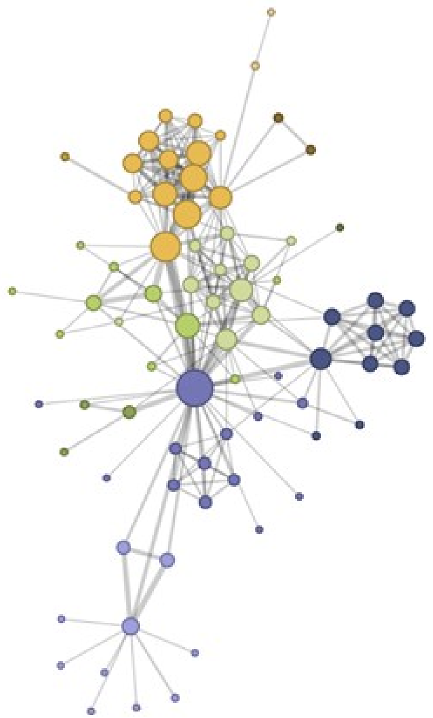
.:: In this lecture

1. Diffusion processes on networks
2. Control of diffusion processes
3. Control of competitive diffusion processes
4. Adding restrictions to the control problem

Diffusion Processes on Networks

Diffusion Processes on Networks

Basics



DPs arise in systems with interconnected agents (real or electronic networks)

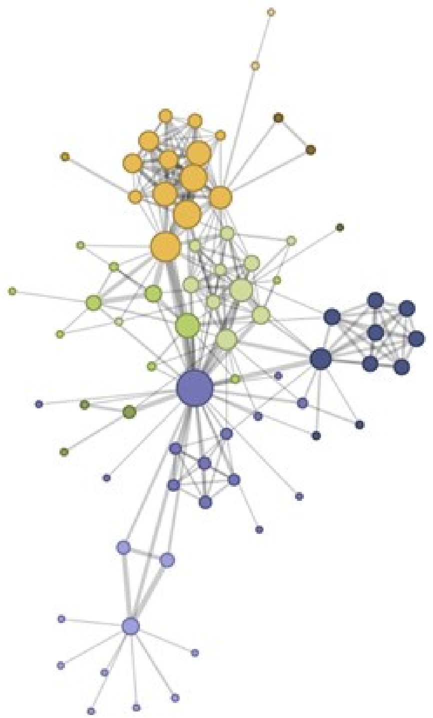
- each agent has a *variable state*
- agent behavior depends on, and propagates to, its close environment
- the propagation causes changes in agents' state according to some "*rules*"

Propagating entities: ***from disease epidemics to… digital and social epidemics***

- *Epidemiology*: diseases/viruses
- *Computer systems*: computer viruses, fault cascade, computational errors (e.g. sensor networks)
- *Social and information networks*: information, ideas, rumors, social behaviors…

Diffusion Processes on Networks

Basics



This is what we will talk about

Propagating entities: ***from disease epidemics to… digital and social epidemics***

- *Epidemiology*: diseases/viruses
- *Computer systems*: computer viruses, fault cascade, computational errors (e.g. sensor networks)
- *Social and information networks*: information, ideas, rumors, social behaviors…

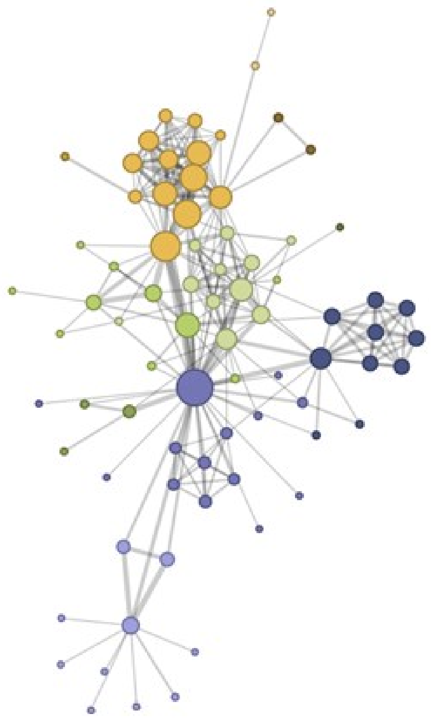
Diffusion Processes on Networks

Directions of research

Depending on the situation, a DP can be desired or undesired

Roughly three directions of research

- **Network assessment:** worst case analysis, risk/vulnerability assessment
- **DP engineering:** influence maximization, (viral) marketing
- **DP suppression and control:** containment of viruses, rumors, social behaviors, etc., using ***control actions***



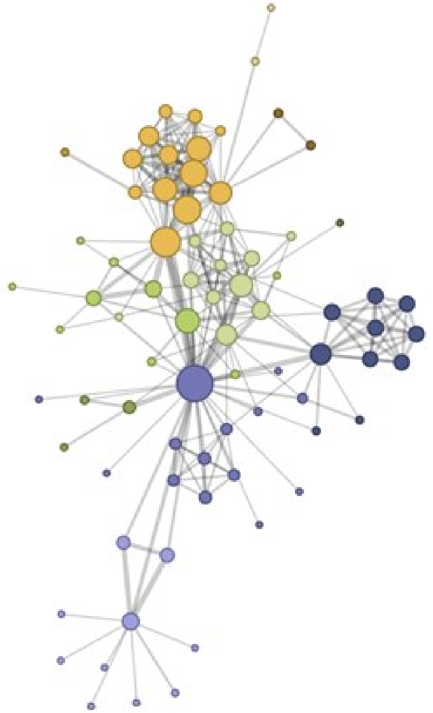
Diffusion Processes on Networks

Directions of research

Depending on the situation, a DP can be desired or undesired

Roughly three directions of research

- **Network assessment:** structure inference, worst case analysis, risk/vulnerability assessment
- **DP engineering:** influence maximization, (viral) marketing
- **DP suppression and control:** containment of viruses, rumors, social behaviors, etc., using ***control actions***

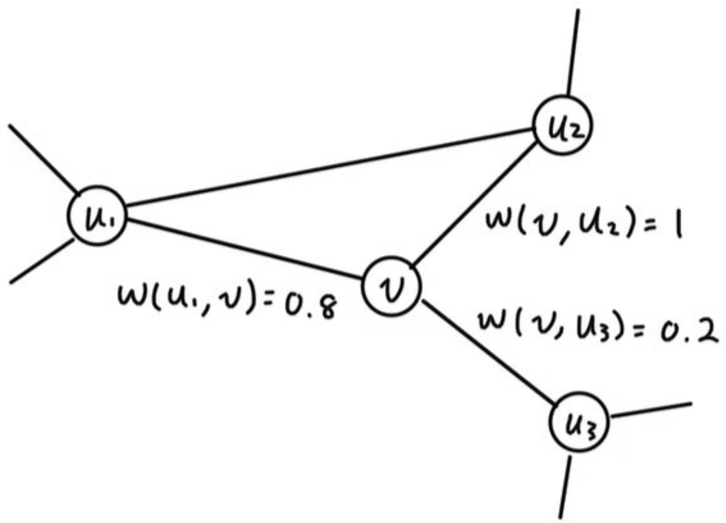


Random Walks

Definition

Random Walk (RW) is a stochastic process on a graph that each time $t = 1,2\ldots$, jumps from the current node u to one of its neighbors based on the associated edge weights

- The process is typically Markovian, so it does not have memory about the previous steps and they do not affect the jumps
- The probability for the process to be at node v after *maaaaany steps* (at the limit) is proportional to the degree of v



Random Walks

As a diffusion mechanism

- RWs simulate a diffusion mechanism
- By sampling several RWs of given length, starting from the same seed node v (or even more seeds), we can estimate the expected diffusion effect encoded in a *visiting probability vector* p_v
- p_v is `characteristic' for the starting conditions of the RWs and can be used to assess the graph structure

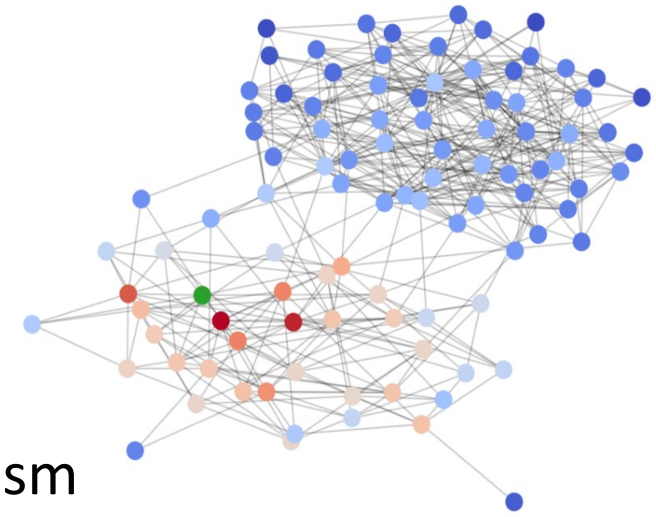


Figure: $t = 10$

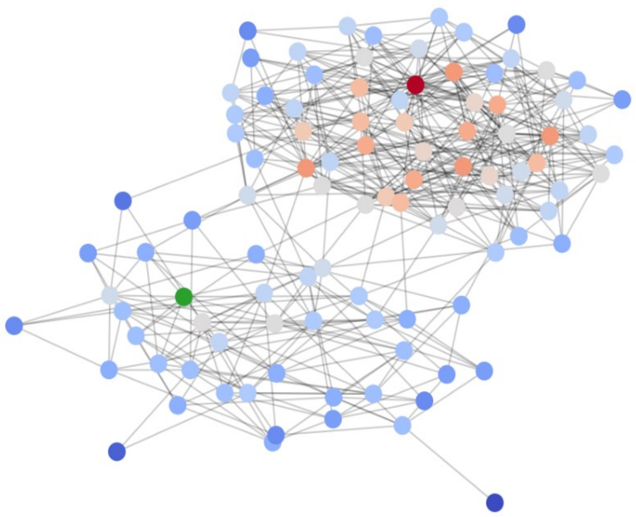
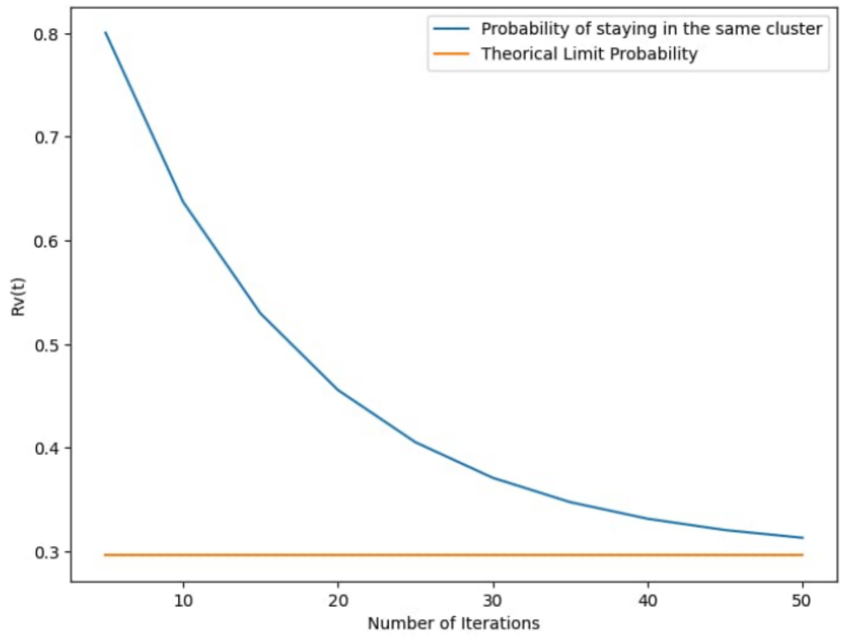


Figure: $t \geq 50$



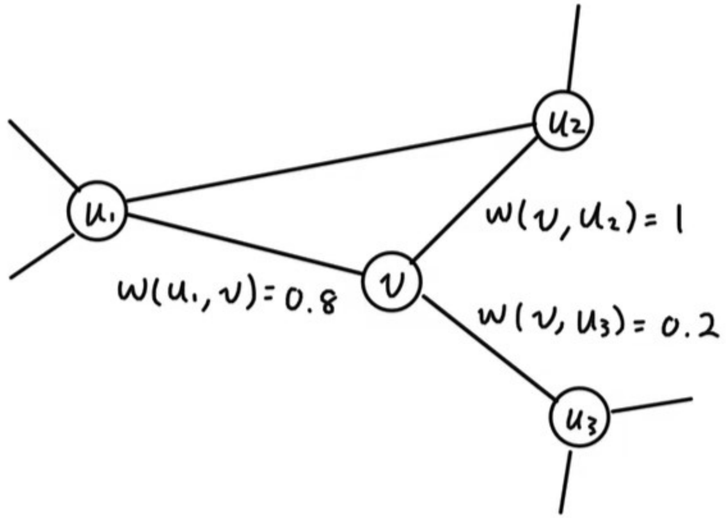
Random Walks

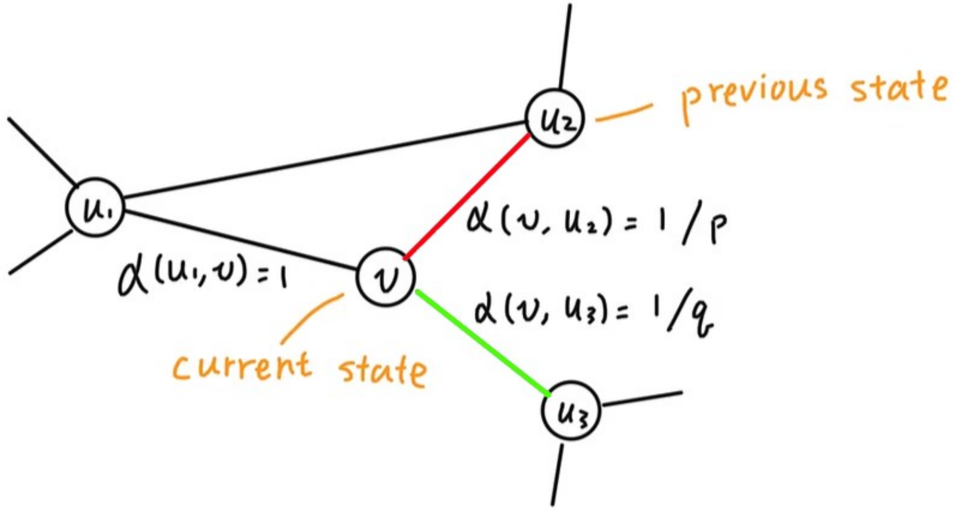
Node2vec and biased walks

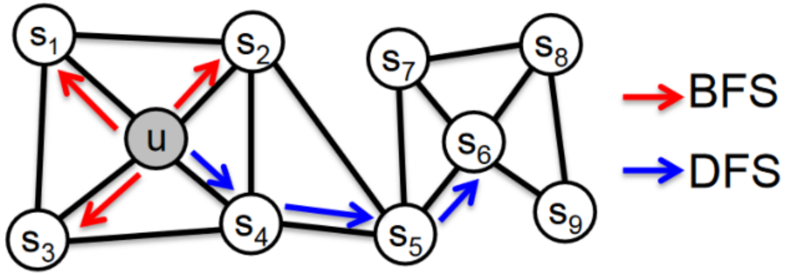
Node2vec is a second order RW that introduces a bias related to the previous node of the walk

$$a_{pq}(t,v) = \begin{cases} \frac{1}{p} & \text{if } t = v \\ 1 & \text{if } \text{dist}(t,v) = 1 \\ 1/q & \text{if } \text{dist}(t,v) = 2 \end{cases}$$

- New edge weights: $w'_{uv} = a_{pq}(u,v)\, w_{uv}$
- BFS is obtained with small p, large q (vice-versa for DFS)
- We cannot use the matrix multiplication trick to compute the visiting expectation… Individual walks should be simulated







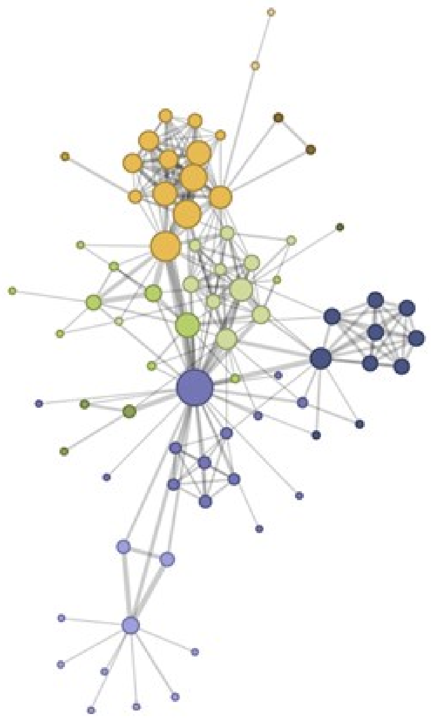
Diffusion Processes on Networks

Directions of research

Depending on the situation, a DP can be desired or undesired

Roughly three directions of research

- **Network assessment:** worst case analysis, risk/vulnerability assessment
- **DP engineering:** influence maximization, (viral) marketing
- **DP suppression and control:** containment of viruses, rumors, social behaviors, etc., using ***control actions***



Influence Maximization

Definition

Definition: Let a given graph $G = (V, E)$ on which a diffusion process can take place. Find the subset of $k < |V|$ influential users $S \subseteq V$ (called *seeds*) that can maximize the spread of the influence measured by function $\sigma(S)$:

$$S^* = \arg \max_{S \subseteq V, |S|=k} \sigma(S)$$

Estimating $\sigma(S)$ is hard (usually through Monte-Carlo simulations).

The problem is overall NP-hard.

Properties: $\sigma(\cdot)$ *is monotonic and submodular function*

- *Monotonic increase* of influence w.r.t. seed set size: $\sigma(S) \leq \sigma(S'), \ S \subset S'$
- *Diminishing returns*: $\sigma(S \cup u) - \sigma(S) \geq \sigma(S' \cup u) - \sigma(S'), \ \forall u \in S', S \subset S'$

Influence Maximization

Greedy approximation

We can approximate the solution in a greedy fashion: by choosing each time the node that increases maximally the influence of the augmented seed set

Competitive ratio ~ $\left(1 - \frac{1}{e}\right)$-competitive

That means ~ 63.2% of the optimal

Several approaches

- Time-aware
- Location-aware
- Context-aware
- Online vs Offline

Algorithm : Greedy Algorithm

Input: Social graph: G; Seed set: size: k;
Influence Function: $\sigma(S)$
Output: Seed set: S

$S \leftarrow \phi$;
for $i = 1, 2, \ldots, k$ **do**
  $u^* \leftarrow \text{argmax}_{u \in V \setminus S} \left(\sigma\left(S \cup \{u\}\right) - \sigma(S)\right)$
  $S \leftarrow S \cup \{u^*\}$
return Seed set S

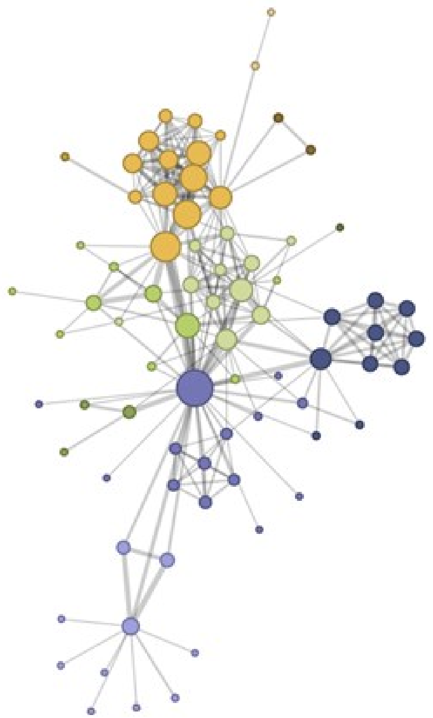
Diffusion Processes on Networks

Directions of research

Depending on the situation, a DP can be desired or undesired

Roughly three directions of research

- **Network assessment:** worst case analysis, risk/vulnerability assessment
- **DP engineering:** influence maximization, (viral) marketing
- **DP suppression and control:** containment of viruses, rumors, social behaviors, etc., using ***control actions***



Diffusion Processes on Networks

Diffusion Models

Multitude of diffusion models

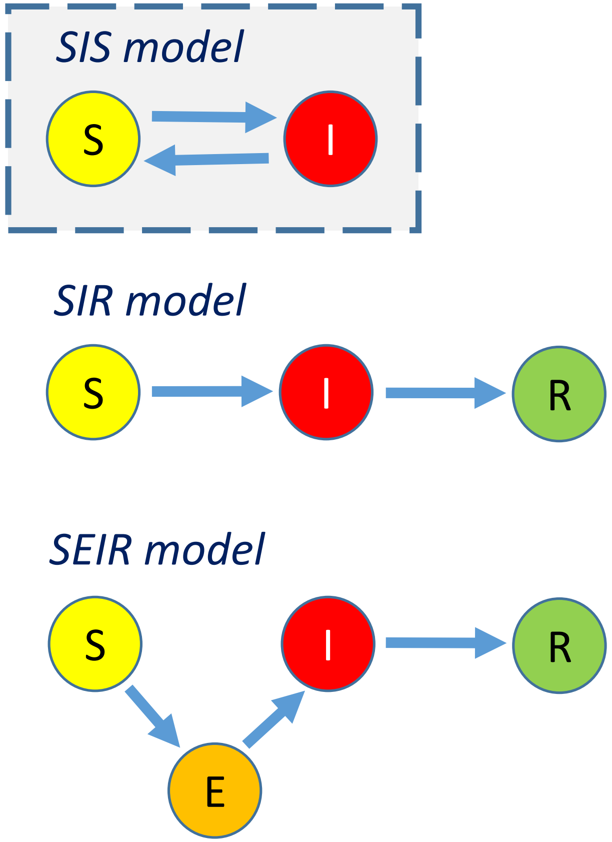
- no single model describes all possible complex diffusion phenomena

Well-studied models

- compartmental models from epidemiology (SIS, SIR, SEIR, …)
- other models from statistical physics (e.g. Percolation)
- common characteristic: constant propagation rates

Modern information-oriented models

- Information Cascades, Hawks Processes, …
- Common direction: propagation rates variable in time to model user interest



S: susceptible | E: exposed
I: infected | R: recovered

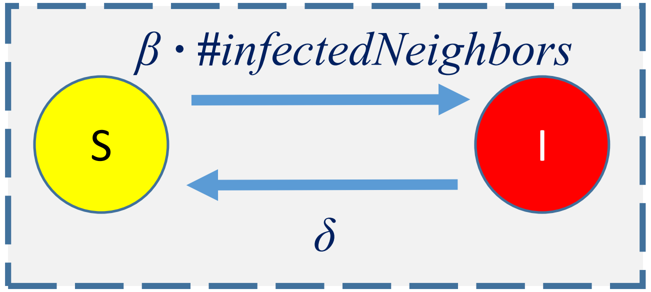
Diffusion Processes on Networks

Diffusion Models – SIS demo

Example

- uncontrolled SIS process on contact network

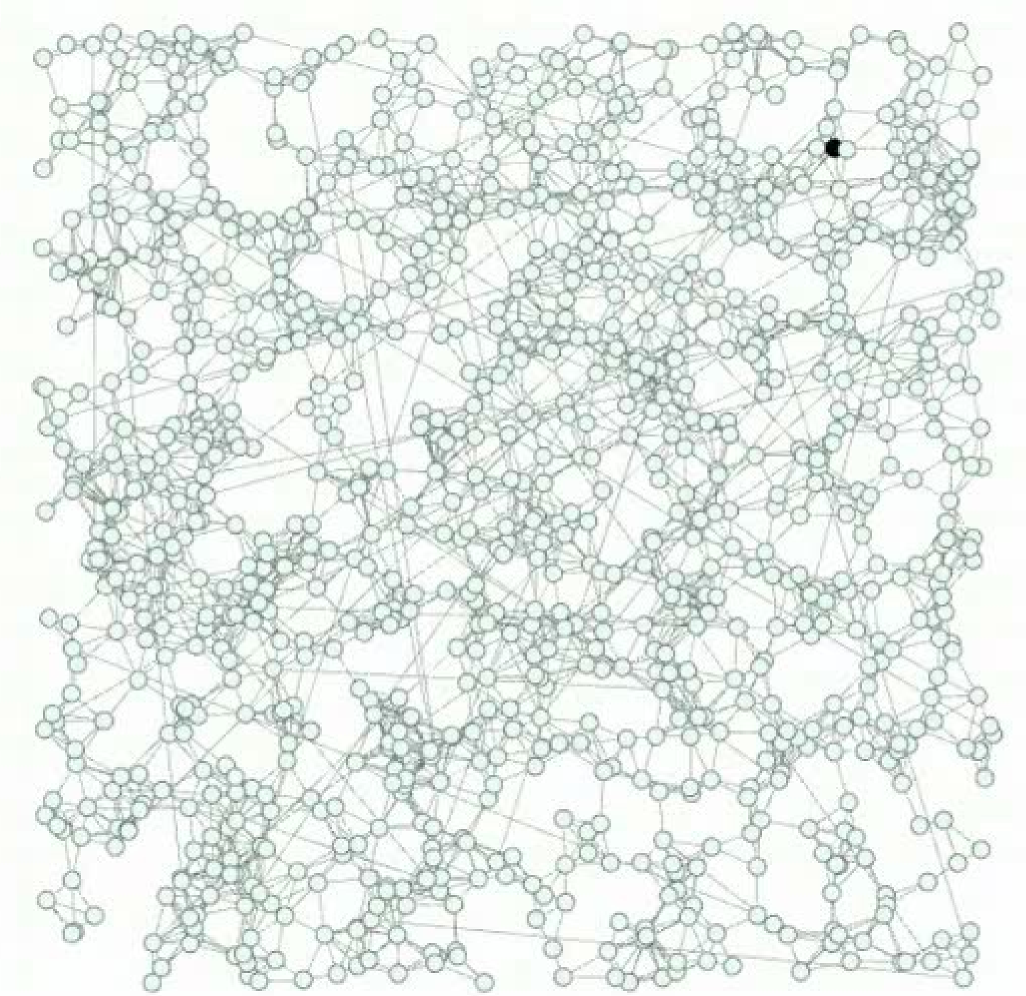
Homogeneous continuous-time SIS model for one node



$$X_i(t): 0 \to 1 \text{ at rate } \beta \sum_j A_{ji} X_j(t)$$
$$X_i(t): 1 \to 0 \text{ at rate } \delta$$

- spreading rate β
- node self-recovery rate δ
- adjacency matrix A
- network state X
- **two possible events each time:** infection or recovery

SIS diffusion process in a contact nework

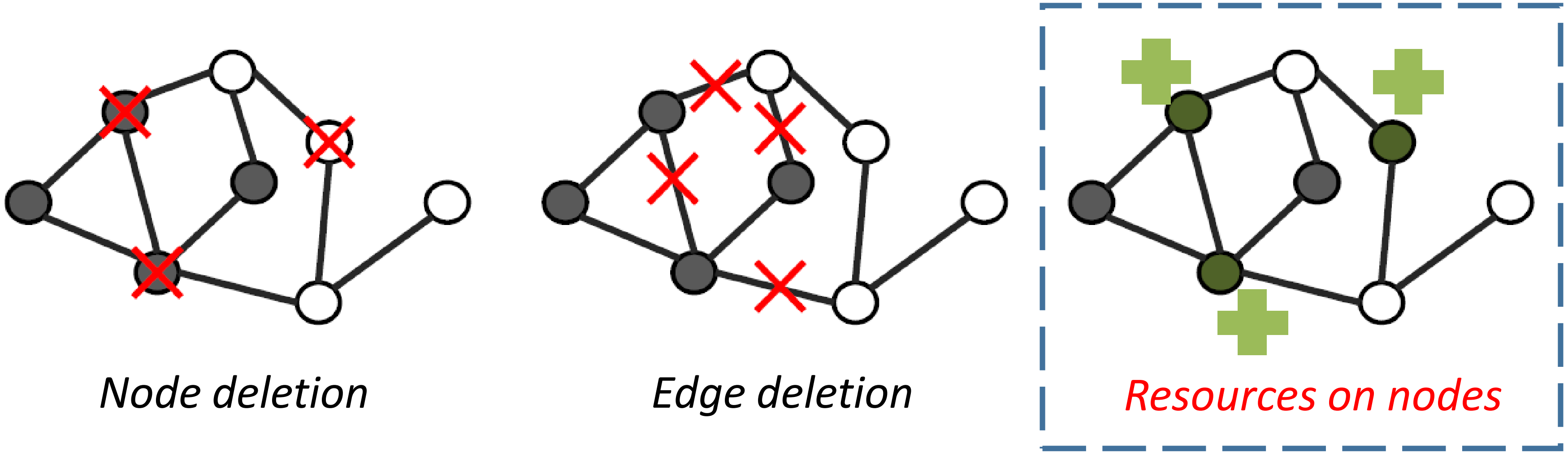


Watch online: http://www.youtube.com/watch?v=fGSKHxSD-40

Diffusion Suppression and control

Possible control actions

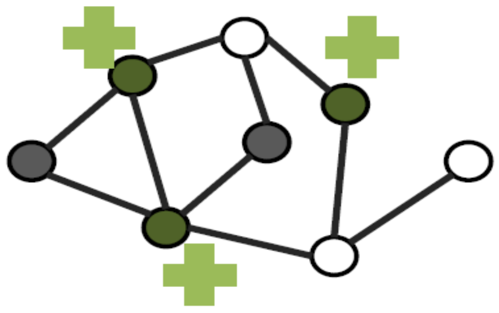
DP suppression and control using ***control actions*** on <u>nodes</u> or <u>edges</u>



Diffusion Suppression and control

Healing resources on nodes

DP suppression and control using ***control actions*** on <u>nodes</u>



Resources on nodes

:

Dynamic Resource Allocation

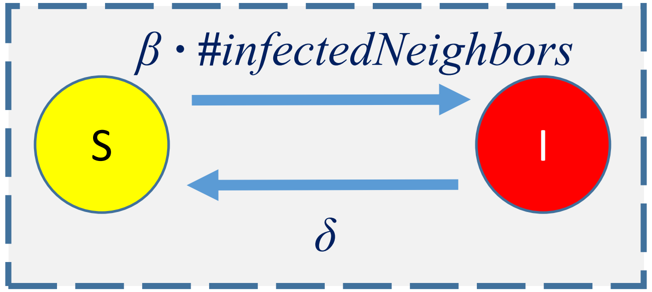
preventive	preparatory	corrective
vaccines	*antidotes*	*treatments*

Diffusion Suppression and control

Introducing resources

Homogeneous continuous-time SIS model for one node

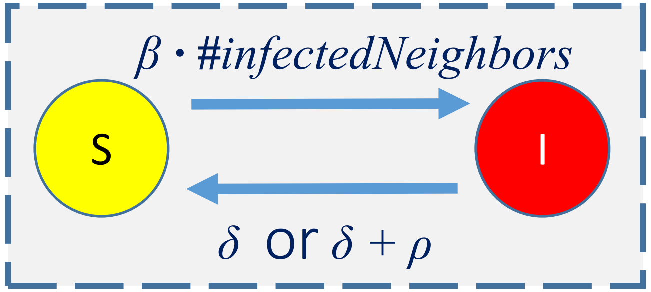
without control



$X_i(t): 0 \to 1$ at rate $\beta \sum_j A_{ji} X_j(t)$
$X_i(t): 1 \to 0$ at rate δ

- **two possible events each time:** infection or recovery
- spreading rate β
- node self-recovery rate δ
- adjacency matrix A
- network state X

with control



$X_i(t): 0 \to 1$ at rate $\beta \sum_j A_{ji} X_j(t)$
$X_i(t): 1 \to 0$ at rate $\underline{\delta + \rho R_i(t)}$

- treatment efficiency ρ
- resource allocation R

Dynamic Resource Allocation (DRA)

A modelling and control framework

DRA objective

$$\min_R C_\gamma(R) = \int_{t=0}^{+\infty} e^{-\gamma t}\mathbb{E}[N_I(t)]dt$$

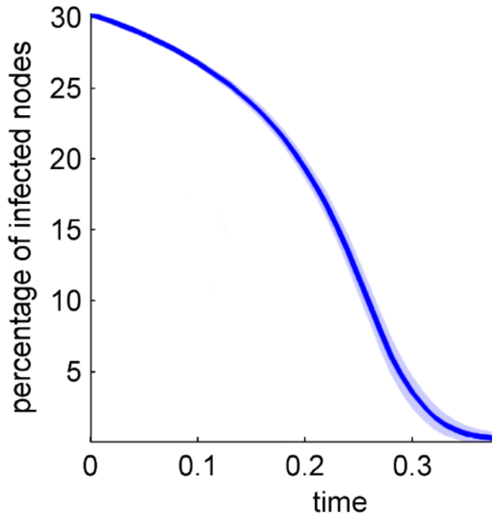
$$\min_R C_\gamma(R) = \int_{t=0}^{+\infty} e^{-\gamma t}\mathbb{E}[N_I(t+u)|X(t)=X]du$$

Formally a DRA strategy

$$R : \mathbb{R}_+ \to \{0,1\}^N$$
$$\text{s.t. } \forall t \in \mathbb{R}_+, \sum_i R_i(t) \le b(t)$$

Constraints

- unlimited resources, disposed at limited constant rate
- limited accumulation of resources on single node
- inability to store resources



Dynamic Resource Allocation (DRA)

Score-based strategies

Score-based DRA strategies

$$R_i(t) = \begin{cases} 1 & \text{if } S_i(X(t)) \geq \theta_t \\ 0 & \text{otherwise} \end{cases}$$

where $\sum_i R_i(t) = b_{tot}$

Heal the nodes with the top-b_{tot} ranked scores…

Baseline heuristics and LRIE

Strategy	Scoring function $S^i(X)$ for node i
RAND	$\sigma(X_i)+R_i$, where R_i is i.i.d. uniform in [0, 1]
MN	$\sigma(X_i)+\sum_j A_{ij}$
PRC	$\sigma(X_i)+P_i$, where P_i is the PageRank score for node i
LRSR	$\sigma(X_i)+(\lambda_1-\lambda_1^{G\backslash i})$, where λ_1 is the largest eigenvalue of A, and $\lambda_1^{G\backslash i}$ the largest eigenvalue of the matrix $A^{G\backslash i}$ for the network without node i
MSN	$\sigma(X_i)+\sum_j A_{ij}\overline{X}_j$
LIN	$\sigma(X_i)-\sum_j A_{ji}X_j$
LRIE	$\sigma(X_i)+\sum_j [A_{ij}\overline{X}_j - A_{ji}X_j]$, sums MSN and LIN

- $\sigma(1) = 0$ and $\sigma(0) = -\infty$

A Greedy Approach for Dynamic Control of Diffusion Processes in Networks,
K. Scaman, A. Kalogeratos, and N. Vayatis, ICTAI 2015.

OPTIMAL Greedy DRA

LRIE - Largest Reduction of Infectious Edges

Derivation

- rewrite the DRA objective according to the Markovian property

$$\min_R C_\gamma(R) = \int_{t=0}^{+\infty} e^{-\gamma t}\mathbb{E}[N_I(t)]dt$$

$$\min_R C_\gamma(R,t,X) = \int_{u=0}^{+\infty} e^{-\gamma u}\underbrace{\mathbb{E}[N_I(t+u)|X(t)=X]}_{\Phi_{t,X}(u)}du$$

- then, a second order approximation

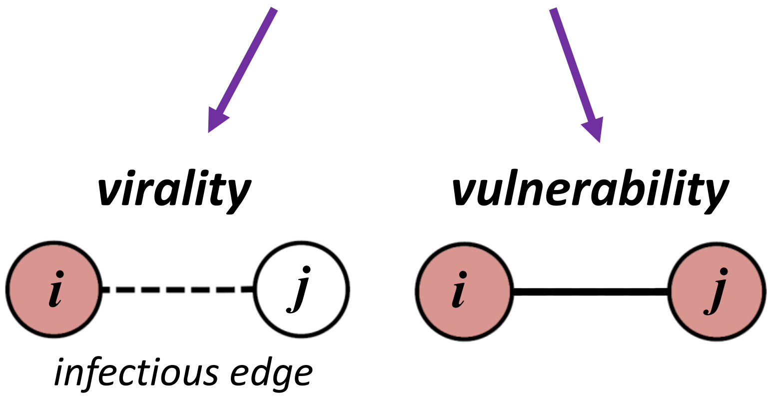
$$C_\gamma(R,t,X) = \frac{1}{\gamma}\sum_i X_i + \frac{1}{\gamma^2}\Phi'_{t,X}(0) + \frac{1}{\gamma^3}\Phi''_{t,X}(0) + O(\tfrac{1}{\gamma^4})$$

⋮

$$S_{\mathrm{LRIE}}(X(t)) = A\overline{X}(t) - A^\top X(t) = \left[\sum_j [A_{ij}\overline{X}_j(t) - A_{ji}X_j(t)]\right]_{i=1}^N$$

For an infected node i

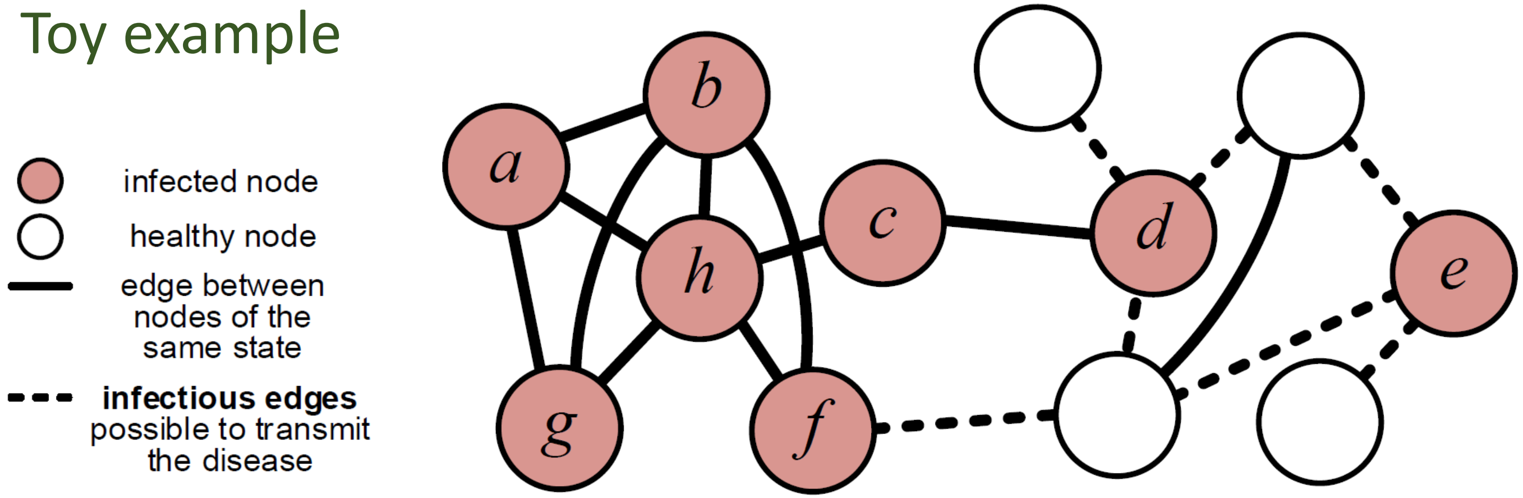
$$\sum_j [A_{ij}\overline{X}_j(t) - A_{ji}X_j(t)]$$



OPTIMAL Greedy DRA

LRIE - Largest Reduction of Infectious Edges

Toy example



- Node ***h*** is the most central
- Node ***e*** and ***d*** are the most viral
- Node ***e*** is the least vulnerable (safest)

LRIE node ranking

Priority 1: e | S_e=3-0

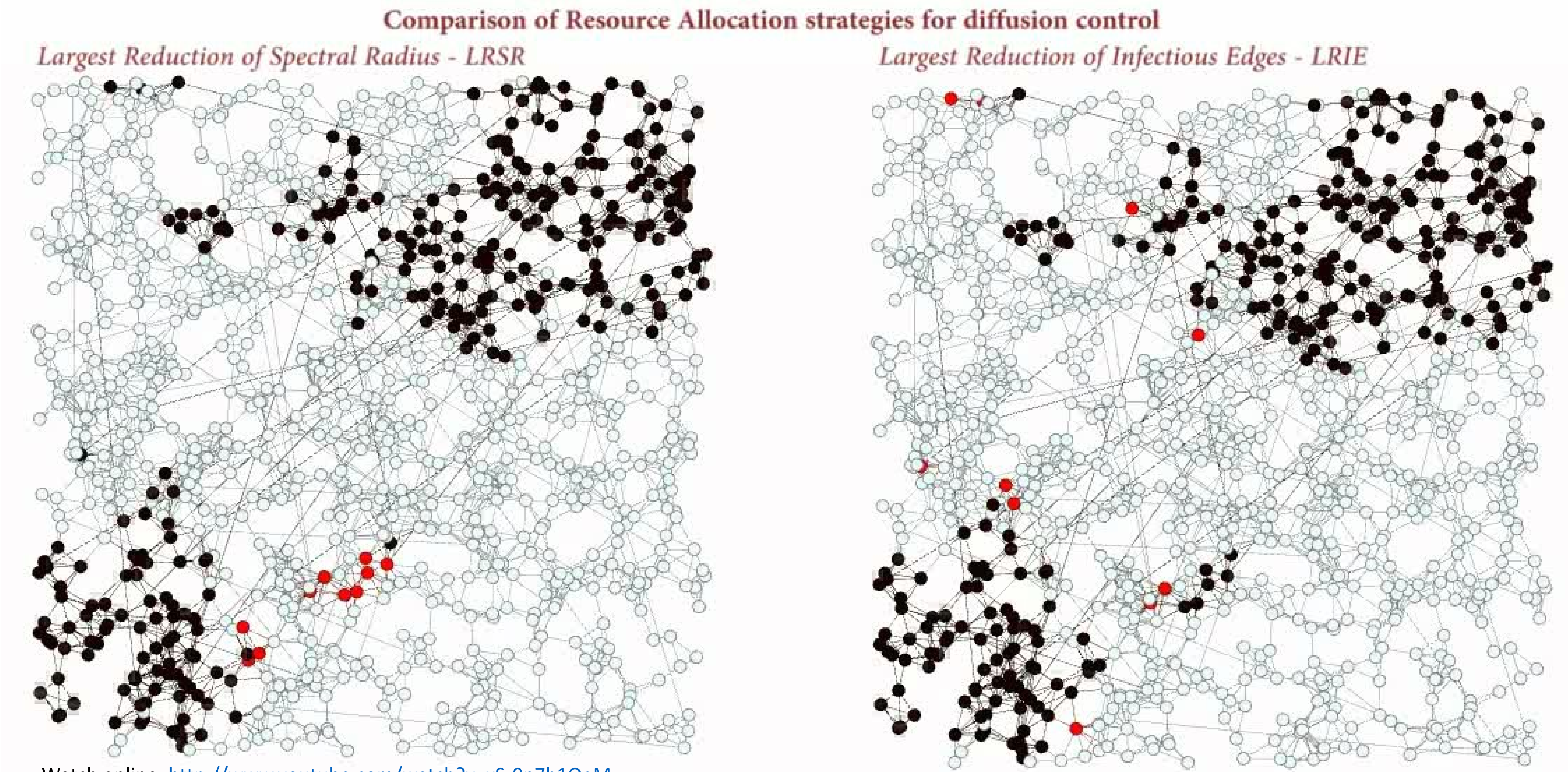
Priority 2: d | S_d=3-1

Priority 3: f | S_f=1-2

⋮

OPTIMAL Greedy DRA

Demonstration on an artificial contact network



Watch online: http://www.youtube.com/watch?v=xS-0p7h1OeM

LRIE: pros & Cons

Advantages

- brings the intuitive idea of reduction of infectious edges (front)
- optimal greedy, fast and quite efficient
- can adapt to network and/or budget changes
- not difficult to imagine a distributed version

Disadvantages

- ignores macroscopic network properties (e.g. clusters)
- cannot apply coordinated actions

Question to answer

LRIE is particularly elegant but greedy?

Can we do better?

(Global) Priority Planning

Definitions

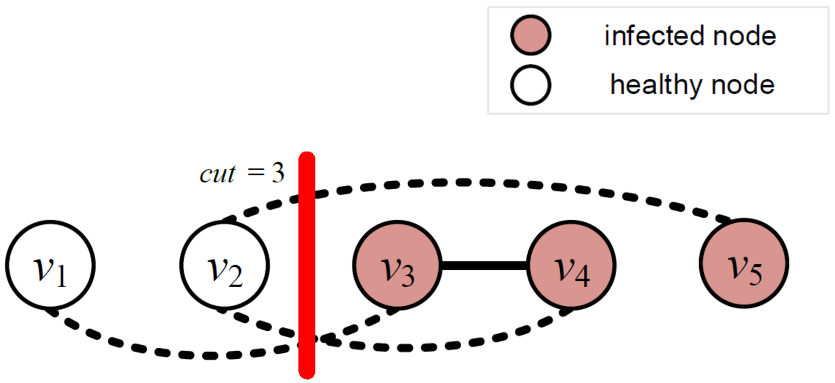
Priority-order: a bijection $\ell : \mathcal{V} \rightarrow \{1, ..., N\}$ s.t. $\ell(v)$ the position of node v in the order

Priority planning: DRA strategies that are based on a priority-order

- limited budget r, max resource per node ρ, healing top-$q(t)$ nodes (i.e. left-most)

$$q(t) = \min\left\{\lceil \tfrac{r}{\rho} \rceil, \textstyle\sum_i X_i(t)\right\}$$

$$\rho_i(t) = \begin{cases} \frac{r}{\lceil \frac{r}{\rho} \rceil} & \text{if } X_i(t) = 1 \text{ and } \ell(v_i) \leq \theta(t); \\ 0 & \text{otherwise} \end{cases}$$



Priority-order $\ell: \mathcal{V} \rightarrow \{1, 2, 3, 4, 5\}$

Global Priority Planning

Graph theoretic properties of a priority-order

Cut at position c: $C_c(\ell) = \sum_{i,j} A_{ij} \mathbb{1}_{\{\ell(v_i) < c \leq \ell(v_j)\}}$

MaxCut of ℓ $\mathcal{C}^*(\ell) = \max_{c=1,\ldots,N} C_c(\ell)$

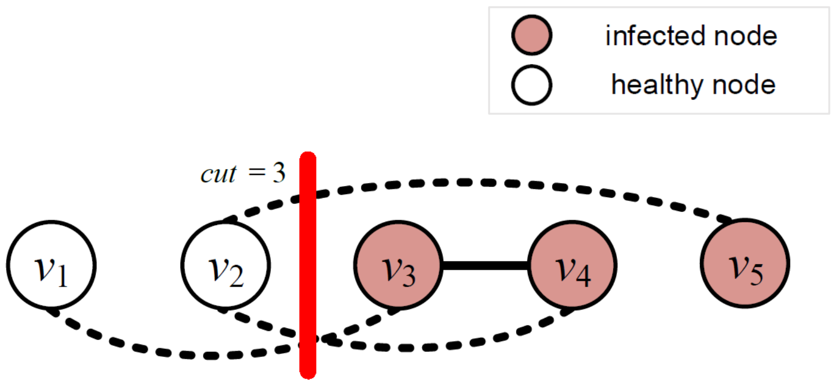
Cutwidth of G: $\mathcal{W} = \min_{\ell} \mathcal{C}^*(\ell)$

Extinction time: $\tau_x = \min\{t \in \mathbb{R}_+ | X(0) = x,\ X(t) = \mathbf{0}\}$

- non-inf random quantity depending on the DRA strategy
- *sub-critical* behavior: $\mathbb{E}[\tau_x]$ ≤ polynomial function
- *super-critical* behavior: $\mathbb{E}[\tau_x]$ > exponential function

Requirement for designing a strategy:

- connect the properties of the order ℓ to $\mathbb{E}[\tau_x]$



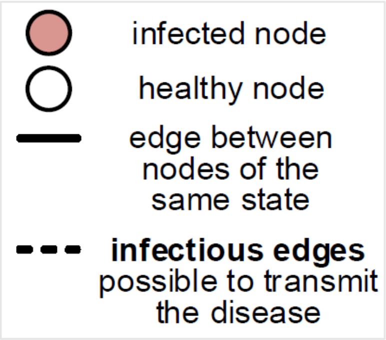
Priority-order $\ell : \mathcal{V} \rightarrow \{1, 2, 3, 4, 5\}$

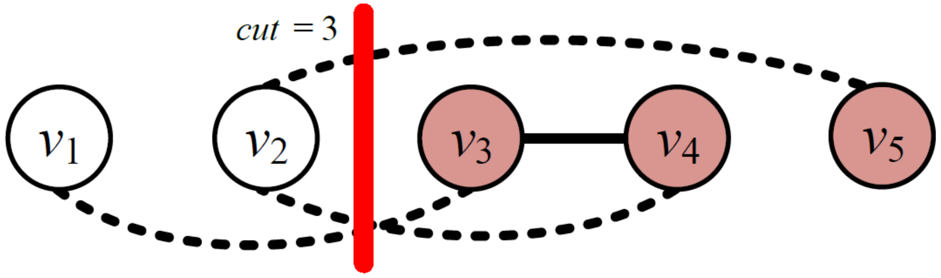
Suppressing epidemics on Networks with Treatments of Limited Efficiency, K. Scaman, A. Kalogeratos, and N. Vayatis, IEEE Trans. on Network Science and Engineering, 2016.

Priority Planning

MaxCut Minimization strategy (MCM)

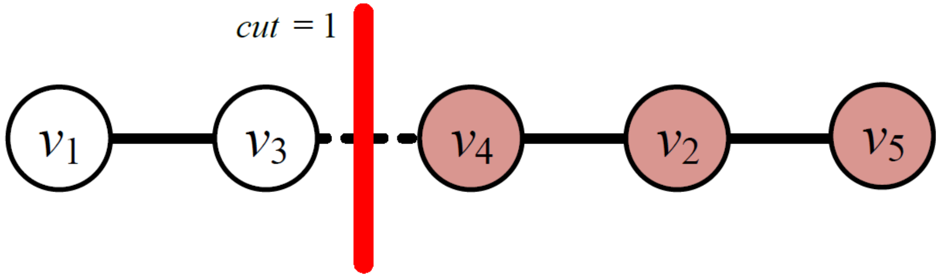
Toy example





(a) Priority-order $\ell : \mathcal{V} \rightarrow \{1, 2, 3, 4, 5\}$

Priority-order with MaxCut = 3



(b) Priority-order $\ell' : \mathcal{V} \rightarrow \{1, 3, 4, 2, 5\}$

Priority-order with minimal MaxCut = 1

- Red vertical line: the ***front*** separating the healthy (left) from the infected part (right) of the network
- The MaxCut indicates highest vulnerability for the healthy part and is the most difficult step of the priority plan

Theoretical results

How good priority-orders are?

UPPER BOUND

Let d the maximum number of neighbors, $q = \lceil \frac{r}{\rho} \rceil$ the number of treated nodes, and $\epsilon = \frac{d(3+2\ln N+4q)}{\mathcal{C}^*(\ell)}$. Assume that:

$$r + \delta q > \beta \mathcal{C}^*(\ell) \left(1 + 2\sqrt{\epsilon} + \epsilon\right)$$

Then the following upper bound holds for the expected extinction time $\mathbb{E}[\tau_1]$:

$$\mathbb{E}[\tau_1] \leq \frac{6N}{\beta}.$$

Theoretical results

How good priority-orders are?

LOWER BOUND

Let $\delta = 0$ (no self- healing), $\eta \in [0, \frac{1}{2}]$, and d, q and ϵ defined as before. Assume that $q < \frac{\mathcal{C}^*(\ell)}{d}$ and

$$r < (1-\eta)\beta\mathcal{C}^*(\ell)(1 - \frac{dq}{\mathcal{C}^*(\ell)})$$

Then the following lower bound holds for the expected extinction time $\mathbb{E}[\tau_1]$:

$$\mathbb{E}[\tau_1] \geq \frac{1}{r}\exp\left(\frac{\eta^2}{12}\Big(\frac{\mathcal{C}^*(\ell)}{d} - q\Big)\right).$$

Maxcut Minimization (MCM)

MCM Strategy

MCM strategy

- seeks for the priority-order ℓ with the ***minimum MaxCut*** $C^*(\ell)$ of edges
- heals the $q(t)$ leftmost infected nodes in ℓ
- uses a relaxation of $\ell_{MCM}(\mathcal{G}) = \underset{\ell}{\operatorname{argmin}}\, \mathcal{C}^*(\ell)$ by

$$\text{MpLA:}\ \ \phi(\mathcal{G}, \ell) = \Big(\sum_{i,j} A_{ij} \, |\ell(v_i) - \ell(v_j)|^p \Big)^{1/p}$$

Algorithm 1 MCM strategy

▷ *Prior to he diffusion process:*

Compute the priority-order $\ell = \ell_{MCM}(\mathcal{G})$ by minimizing the maxcut $\mathcal{C}^*(\ell)$

Order the nodes of $\mathcal{G}$ according to ℓ, i.e. compute the node list $(v_1, \dots, v_N)$ s.t. $\forall i \in \{1, \dots, N\}$, $\ell(v_i) = i$

▷ *During the diffusion process:*

Input: network $\mathcal{G}$, state vector $X(t)$, resource budget r, resource threshold ρ

Output: the resource allocation vector $\rho(t)$

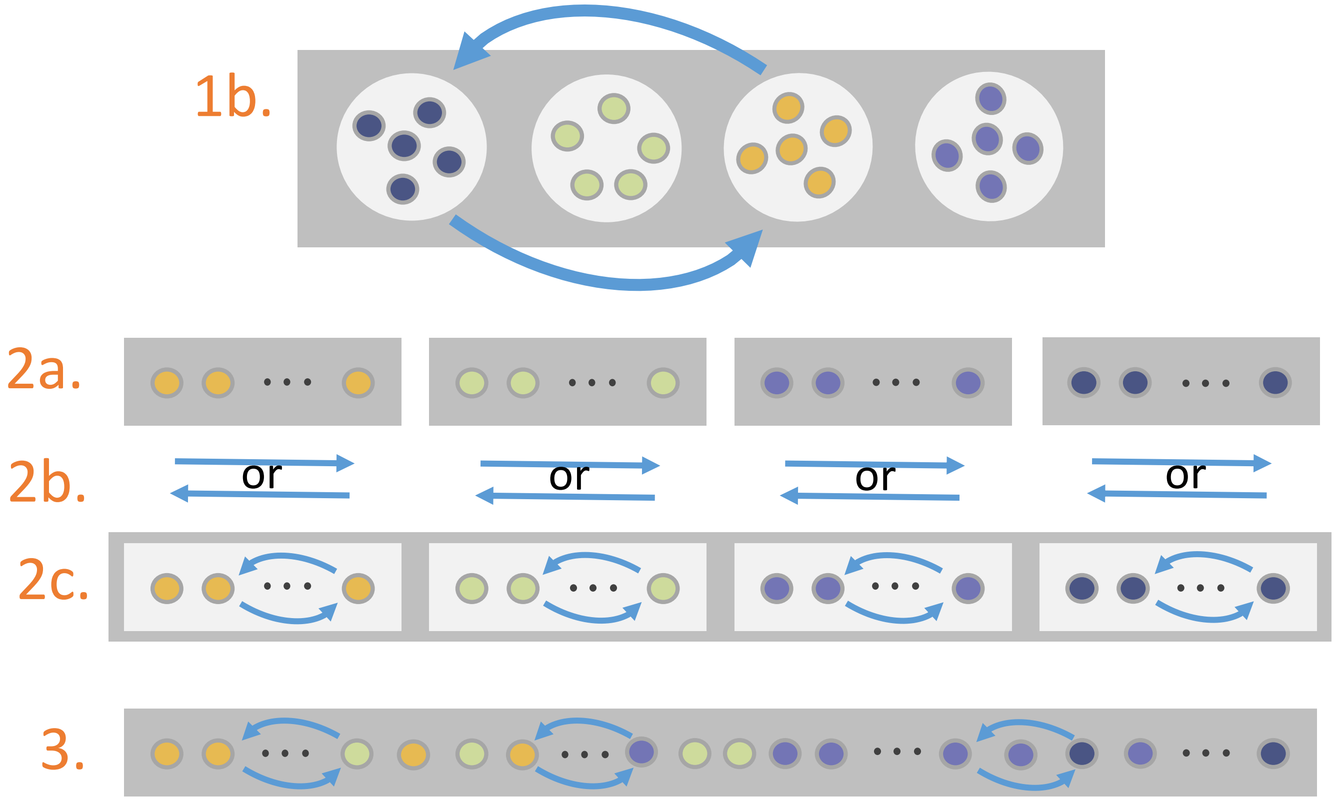
```
q ← ⌈r/ρ⌉
if Σ_i X_i(t) < q then
    return (r/q) X(t)
end if
ρ(t) ← 0                                  // a zero vector in R^N
budget ← q
i ← 1
while budget > 0 do
    if X_{v_i}(t) = 1 then
        ρ_{v_i}(t) ← r/q
        budget ← budget − 1
    end if
    i ← i + 1
end while
return ρ(t)
```

Maxcut MinimiZation (MCM)

Solving the MLA problem

Learning an ordering for a network

1. find communities in G and order them (high-level nodes) with *spectral sequencing*
2. order nodes inside each cluster with *spectral sequencing*, *orient* to each other, and then optimize with *node swaps* internally to clusters
3. apply the swap-based approach again to the overall node ordering

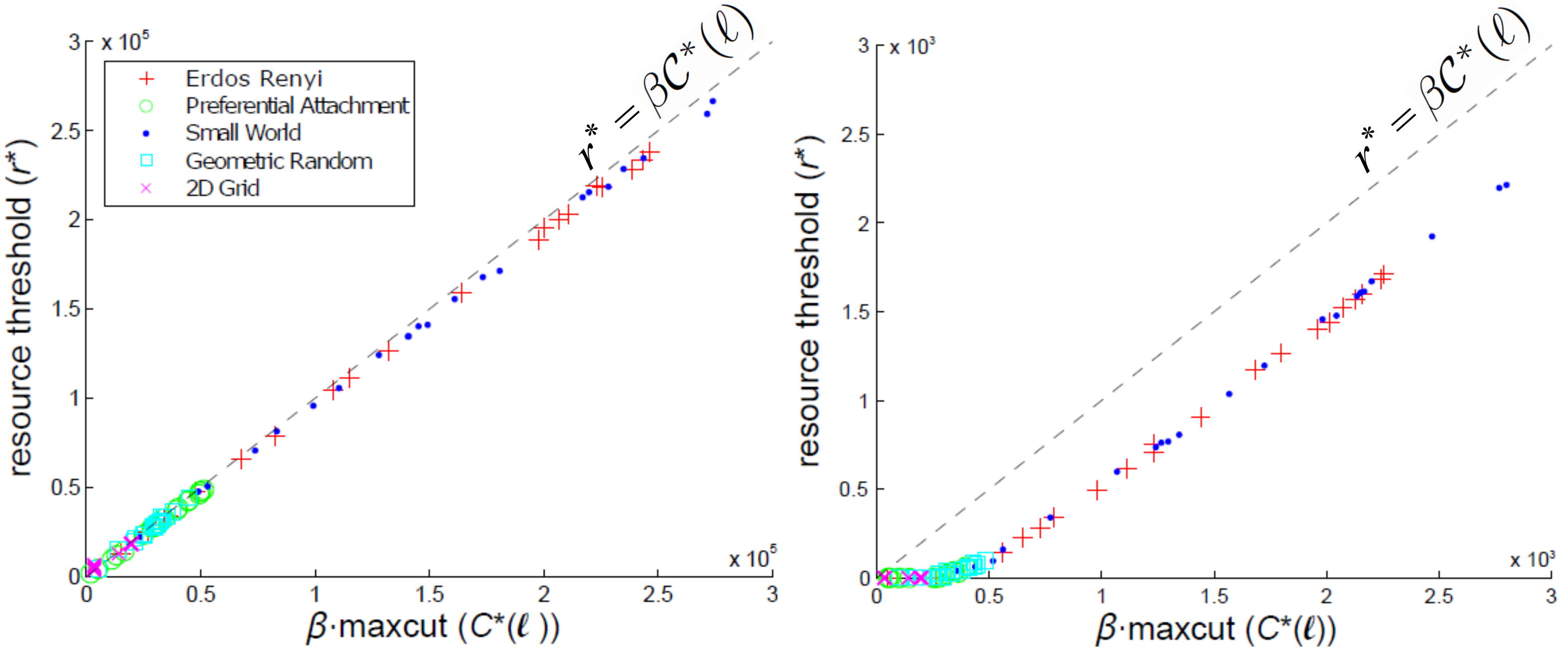


Results

Quality of the theoretical bound

Verifying

$r^* \approx \beta \mathcal{C}^*(\ell)$

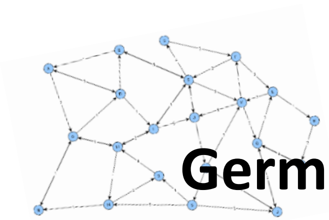


(a) high infectivity: $\beta = 10$ (b) low infectivity: $\beta = 0.1$

- picks orderings at random out of MCM, RAND, MN, LN, LRSR
- various random network models, $N = 1{,}000$, $q = \{1, \ldots 100\}$
- r^* was estimated empirically with simulations

Results

Experiments on real-networks



GermanSpeedway

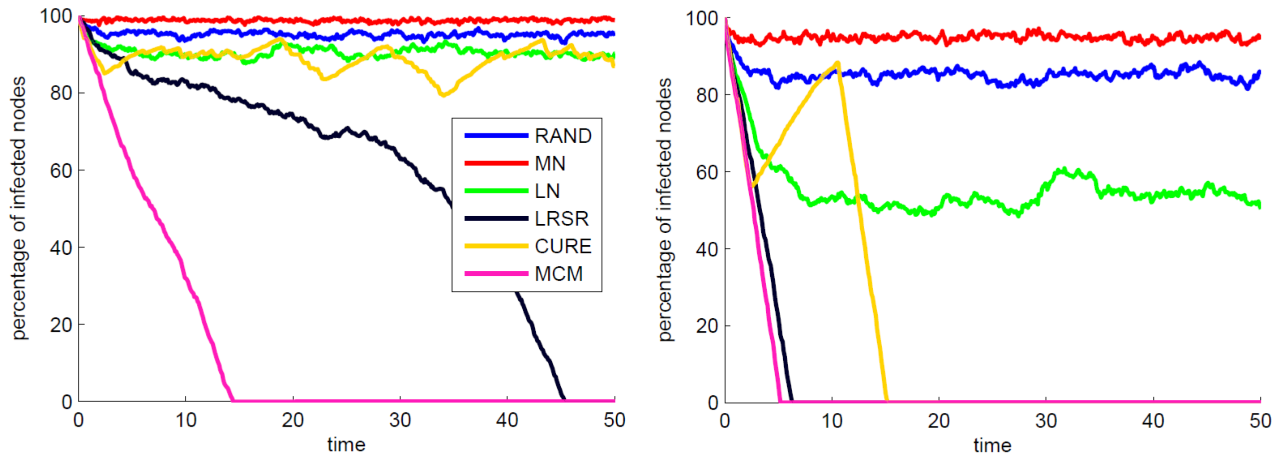
$N = 1{,}168$ nodes, $E = 1{,}243$ edges, $max(d) = 12$, β=1, $\delta = 0$, $q = 1$

MaxCut: 650+/-50 RAND, 379 MN and LN, 104 LRSR, 29 CURE and MCM

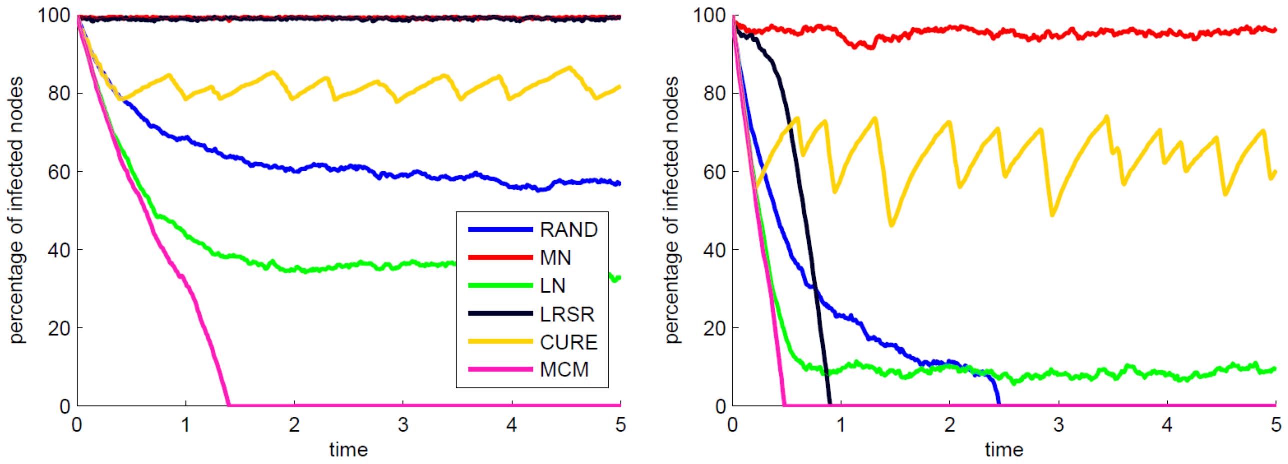
OpenFlights

$N = 2{,}939$ nodes, $E = 30{,}501$ edges, $max(d) = 242$, β=1, $\delta = 0$, $q = 1$

MaxCut: 7,800+/-100 RAND, 7,504 MN and LN, 6,223 LRSR, 2,231 CURE and MCM



(a) low resource budget: $r = 100$ (b) high resource budget: $r = 250$



(a) low resource budget: $r = 3000$ (b) high resource budget: $r = 7000$

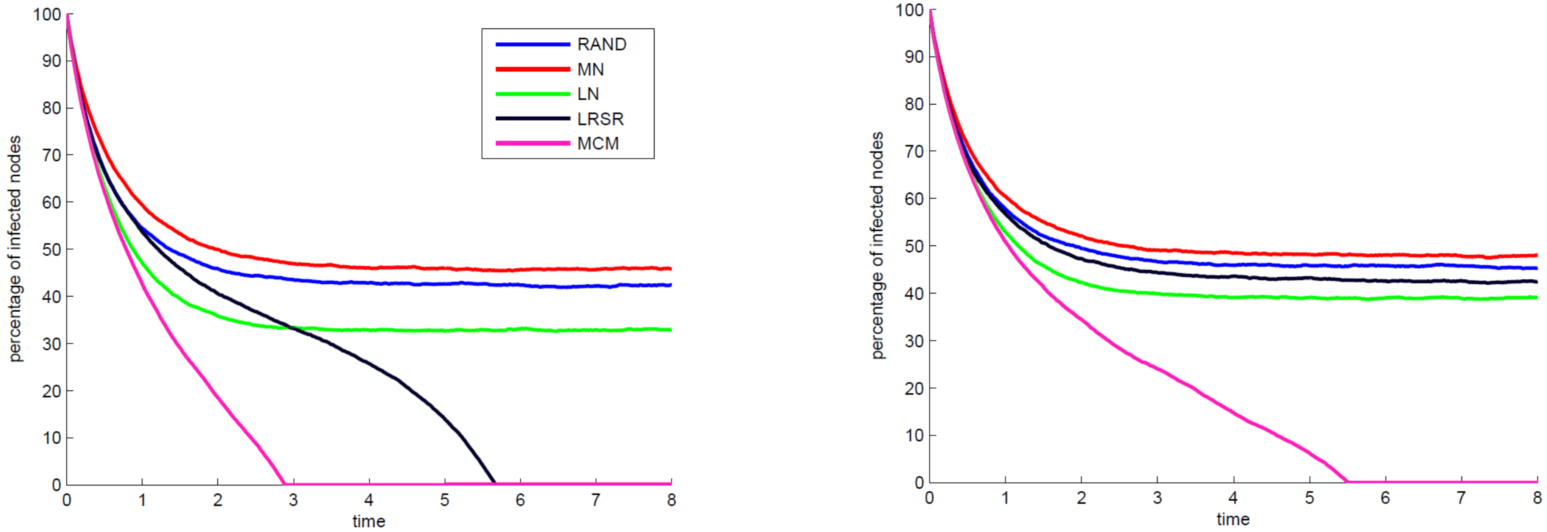
Global Priority Planning

Experiments on real-networks

Subset of Twitter network with 81.306 nodes

Strategy	Maxcut	Maxcut % w.r.t. RAND	Expected resource threshold ($\delta=1$, $\beta=0.1$, $q=100$)
RAND	$670{,}000 \pm 1000$	100.0 %	67,000
MN	628,571	93.8 %	62,957
LN	628,571	93.8 %	62,957
LRSR	349,440	52.2 %	34,944
MCM	**71,956**	**10.7 %**	**7,196**

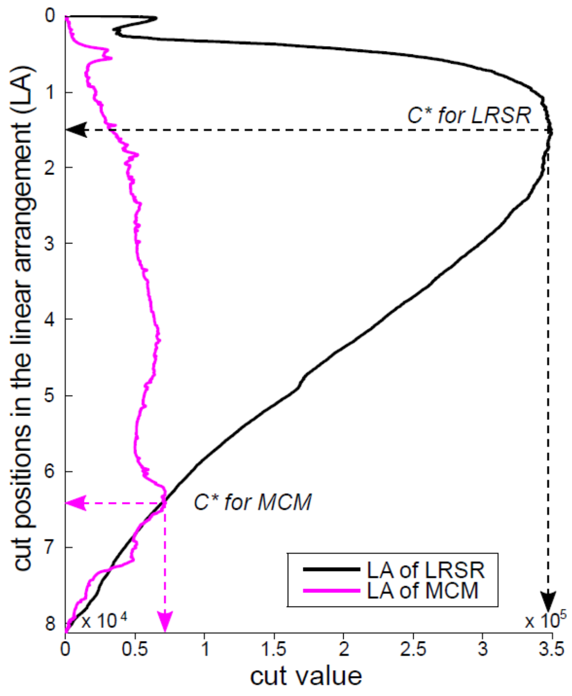
MCM can remove the contagion with ~5 times less resources than its best competitor !!



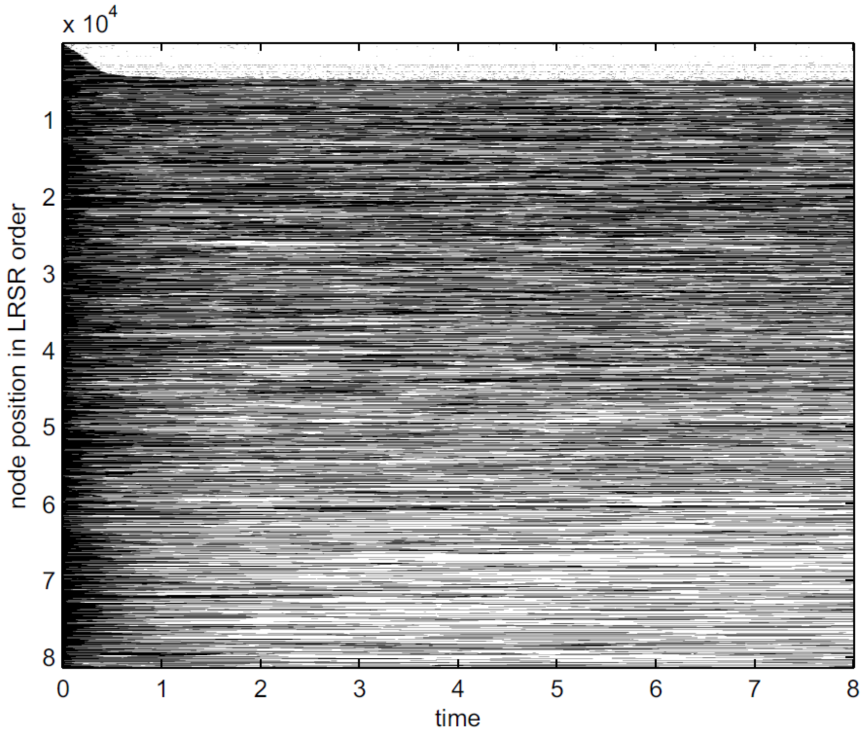
(a) high resource budget: $r = 20,000$ (b) low resource budget: $r = 12,000$

Global Priority Planning

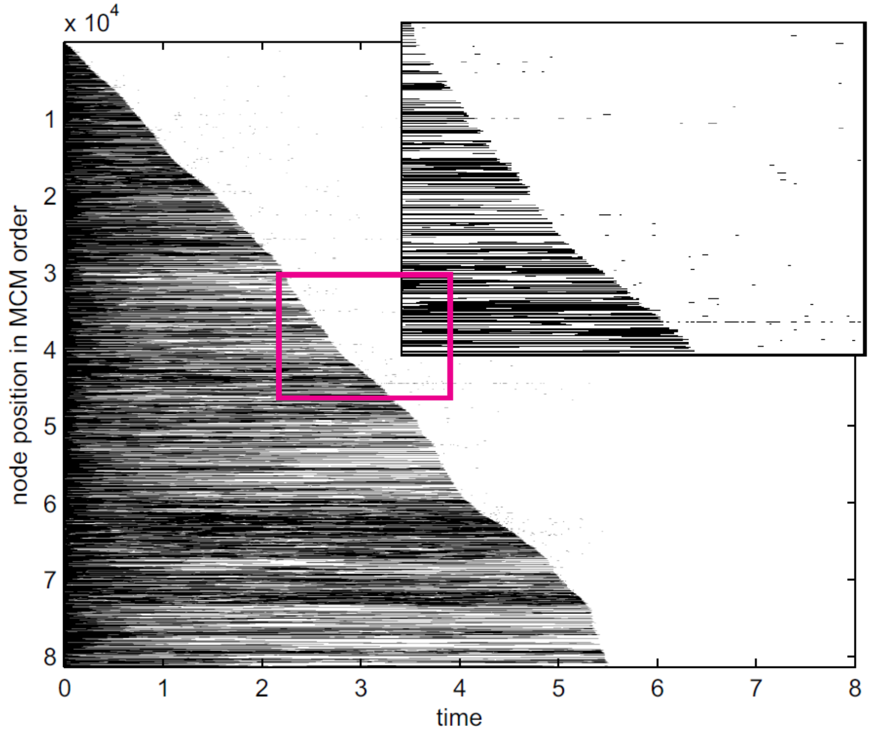
Experiments on real network (TwitterNet)



(c) cuts and maxcuts



(d) network state under LRSR



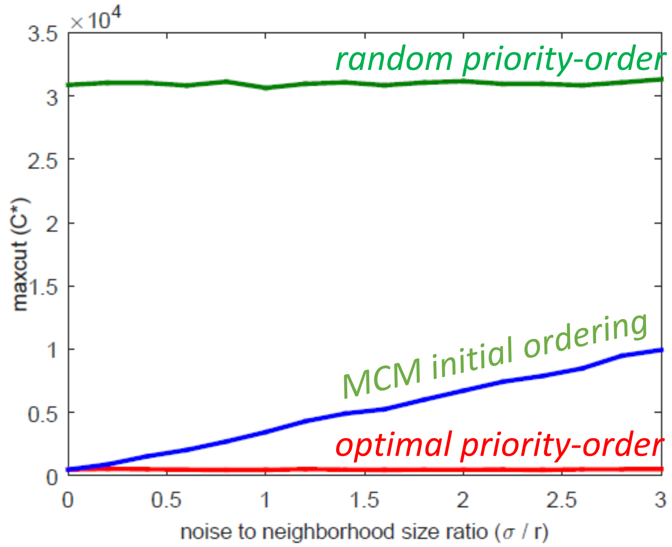
(e) network state under MCM

Robustness analysis

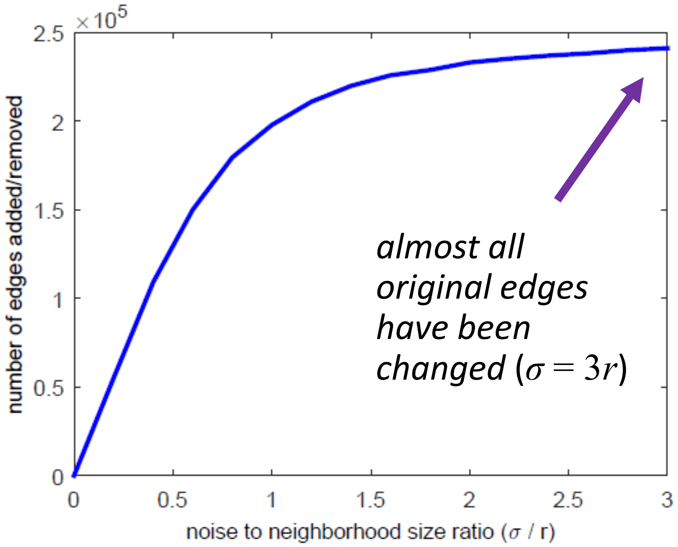
Experiments on an increasingly perturbed contact network

Contact network in $[0,1]^2$ where each node is connected with all nodes in radius r

The priority ordering remains valid after local modifications of the network connectivity



(a) $\mathcal{C}^*(\ell)$ value as a function of noise



(b) number of edges added or removed as a function of noise

Question to answer

From disease epidemics to… digital and social epidemics

Are the strategies we have at hand efficient

in the presence of ***competition****?*

Can we do better?

Competitive social Diffusion Processes

Motivation

Various studies have identified diffusion and competition in lifestyle and social behavior

- Obesity [Wing et al. 2009, Christakis et al. 2007, Hill et al. 2010]
- Smoking [Poulsen et. al. 2002, Christakis et al. 2009]
- Alcohol consumption [Rosenquist et al. 2010]
- Emotions in social networks [Fowler et al. 2009, Hill et al. 2010]

⋮

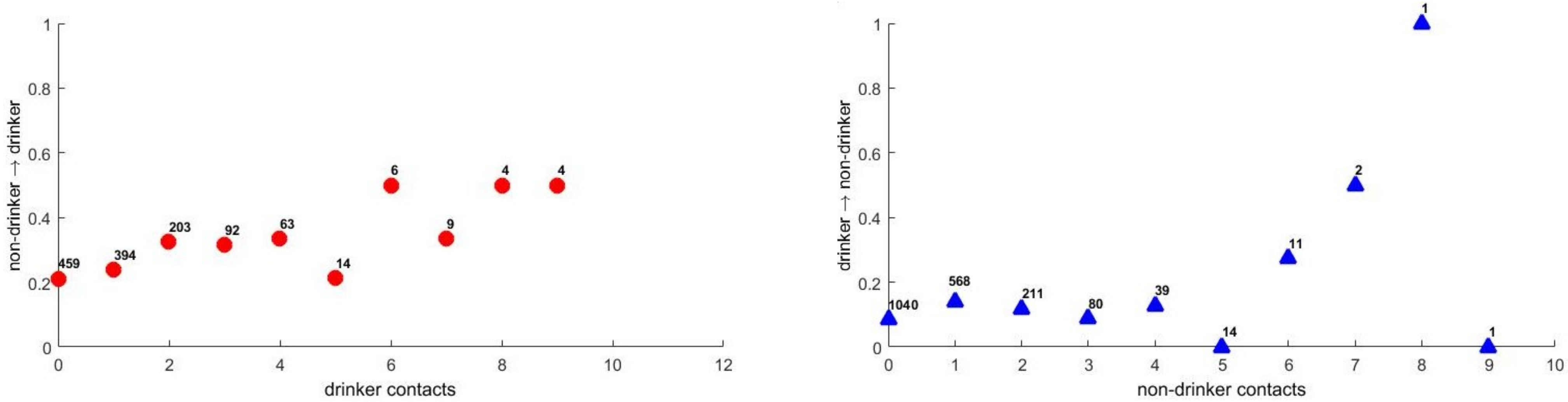
Features / challenging properties

- SIS-like processes with evidence of competition
- “evidently” complex propagation functions
  - non-linear
  - with saturation points

Competitive social Diffusion Processes

Motivation

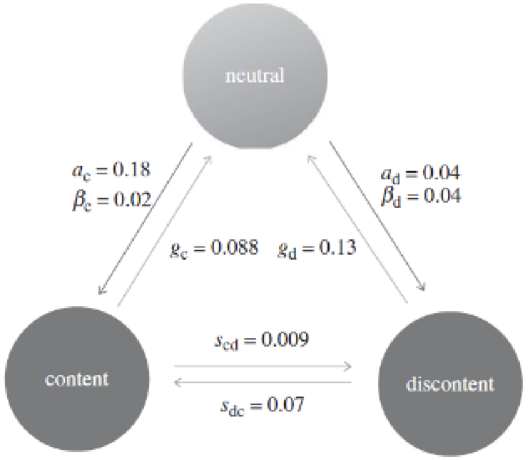
Example: human behavior for **alcohol consumption**



Christakis et al. 2010.

Modeling competition with SIS

Competitive models from literature

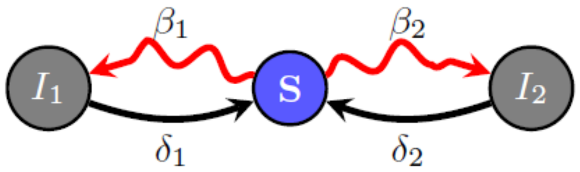


Related epidemic models

- SISa: includes spontaneous infection

  [Hill et. al. 2010]

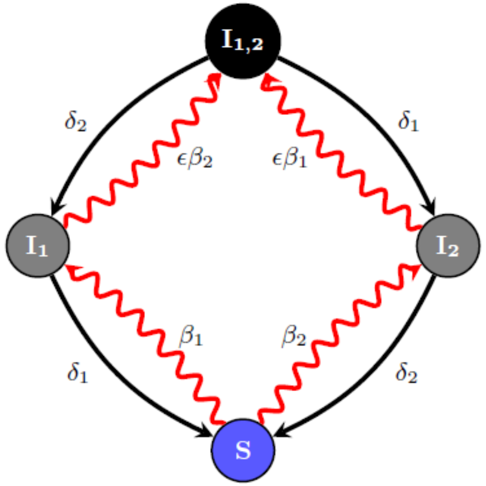
- SI_1I_2S:



  [Prakash et. al. 2012]

- $SI_{1|2}S$

  [Beutel et. al. 2012]

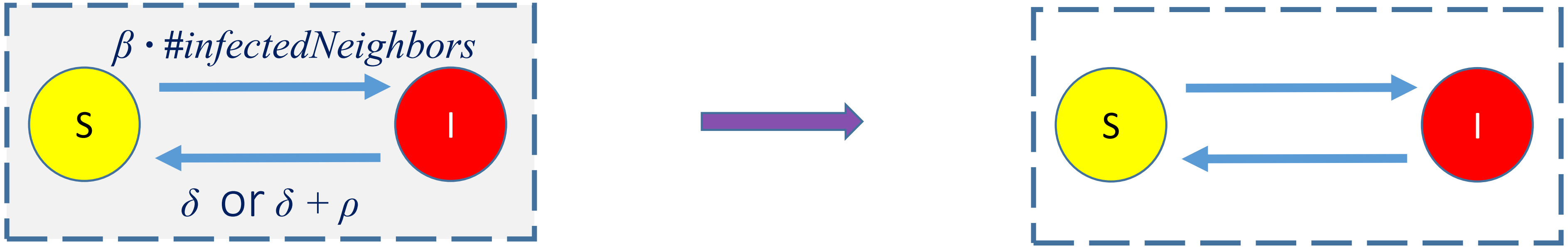


A NOVEL Competitive SIS model

Introducing arbitrary propagation functions and competition

SIS with control

Generalized SIS with competition and control



$X_i(t): 0 \rightarrow 1$ at rate $\beta \sum_j A_{ji} X_j(t)$

$X_i(t): 1 \rightarrow 0$ at rate $\delta + \rho R_i(t)$

$X_i(t): 0 \rightarrow 1$ at rate $\mathcal{I}_i(X(t))$

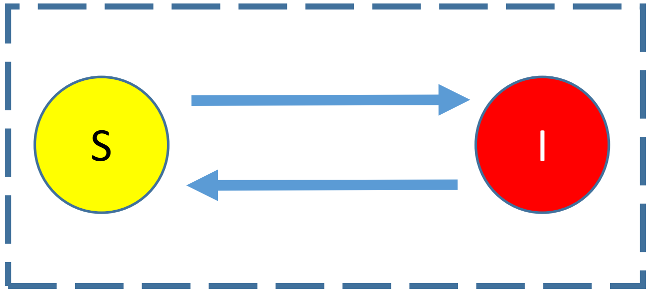
$X_i(t): 1 \rightarrow 0$ at rate $\mathcal{H}_i(X(t)) + \rho R_i(t)$

- with $\mathcal{I}_i$ and $\mathcal{H}_i$ two node-specific memoryless propagation functions
- … that represent the competing positive and negative diffusions

Our Competitive SIS model

Relaxation

Generalized SIS



$$X_i(t): 0 \to 1 \quad \text{at rate} \quad \mathcal{I}_i(X(t))$$
$$X_i(t): 1 \to 0 \quad \text{at rate} \quad \mathcal{H}_i(X(t)) + \rho R_i(t)$$

Relaxation

Assumptions for the propagation functions:

- *locality*
- *exchangeability*
- *Invariance*

➔ They depend only on the node degree d_i and on the number infected neighbors n_i :

$$\mathcal{H}_i = \mathcal{H}(n_i, d_i)$$
$$\mathcal{I}_i = \mathcal{I}(n_i, d_i)$$

Optimal greedy strategy for competition

Generalized LRIE (gLRIE)

Derivation: By minimizing $C_\gamma(R, t, X)$ we obtain the **gLRIE scoring function**

For an infected node i :

concerns i — *the healing of i affects its neighborhood*

$$S_i = -\left[(\mathcal{H}_i + \mathcal{I}_i) + \sum_j A_{ij}\left[X_j(\Delta\mathcal{H}_j^-) - \bar{X}_j(\Delta\mathcal{I}_j^-)\right]\right]$$

Interpretation

- $\mathcal{H}_i$: *self-recovery + receiving local "social" healing*
- $\mathcal{I}_i$: *vulnerability*
- $X_j(\Delta\mathcal{H}_j^-)$: *contribution to local "social" healing*
- $\bar{X}_j(\Delta\mathcal{I}_j^-)$: *virality*

Recovering LRIE: $\mathcal{H}_i = \delta$ $\quad \mathcal{I}_i = \beta \sum_j A_{ji} X_j$ $\quad \Delta\mathcal{H}_i^- = 0$ $\quad \Delta\mathcal{I}_i^- = \beta$

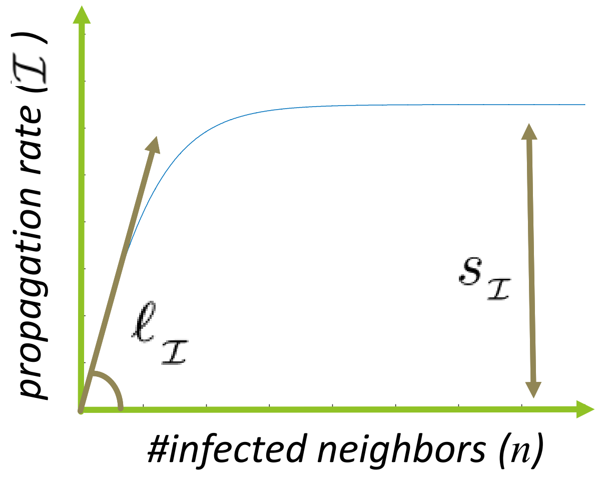
Numerical simulations

Examples of arbitrary propagation functions

Sigmoid propagation functions for experimentation :

- adjustable non-linearity
- adjustable saturation level

Case A



Two variations:

- **Case A:** Dependency only on #infected neighbors (n)
- **Case B:** Dependency only on the infection ratio (n/d)

$$\begin{cases} \mathcal{I}(n,d) = s_{\mathcal{I}}\left[1 - \frac{2}{1+\exp\left(4\ell_{\mathcal{I}} *\right)}\right] \\ \mathcal{H}(n,d) = s_{\mathcal{H}}\left[1 - \frac{2}{1+\exp\left(4\ell_{\mathcal{H}}(d- *)\right)}\right] + \delta \end{cases}$$

Numerical simulations

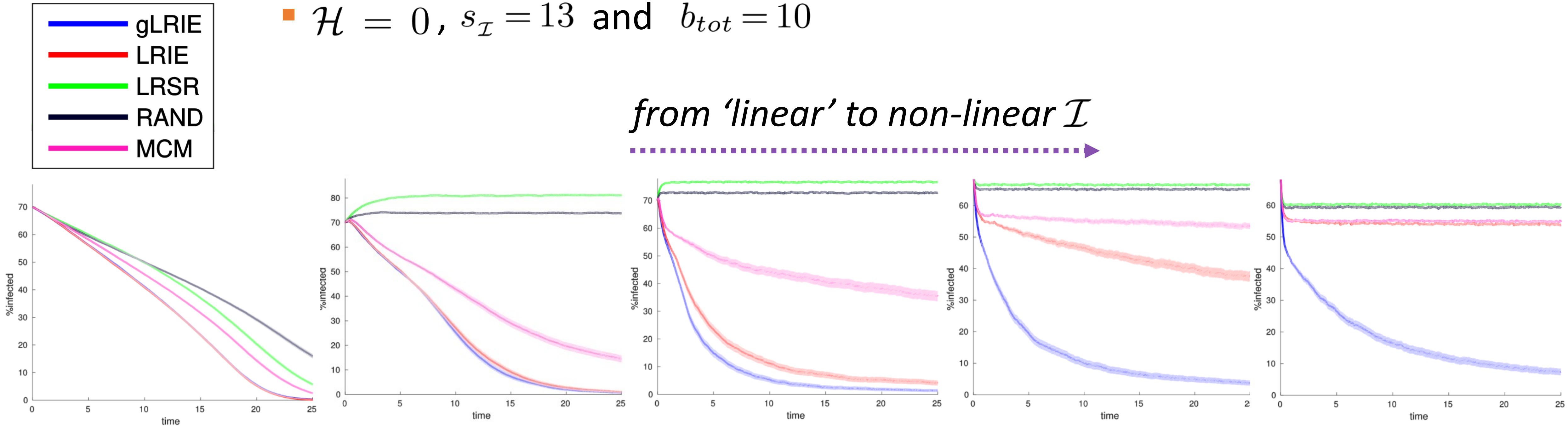
Evolution plots for random graphs

Case A: ~ *to #infected neighbors* + no competition

Erdös-Rényi networks: $N = 1000$ nodes, $p = 0.001$, mean degree 8, 10^4 simulations

- $\mathcal{H} = 0$, $s_{\mathcal{I}} = 13$ and $b_{tot} = 10$

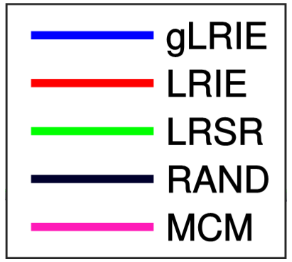
from 'linear' to non-linear $\mathcal{I}$



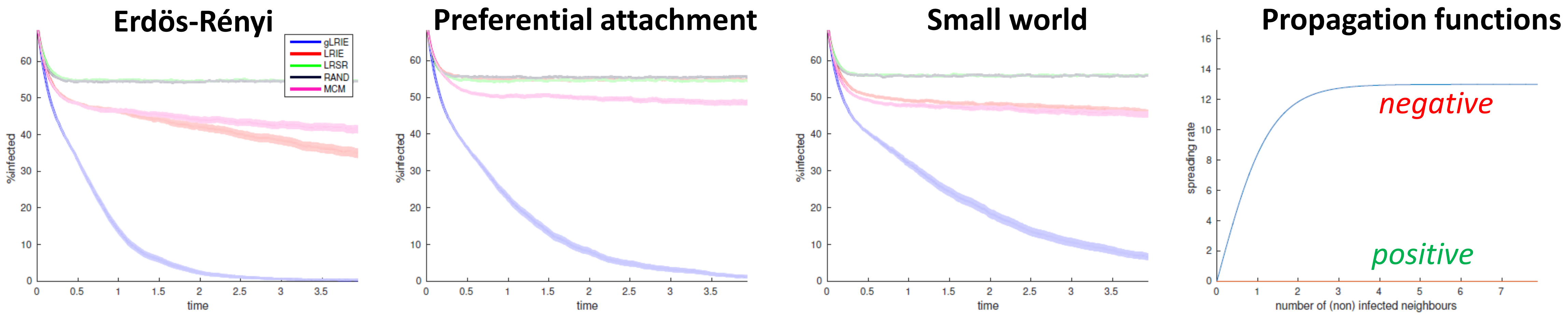
(a) $\ell_{\mathcal{I}} = 0.01, \rho = 1.6$ (b) $\ell_{\mathcal{I}} = 0.1, \rho = 8$ (c) $\ell_{\mathcal{I}} = 1, \rho = 63$ (d) $\ell_{\mathcal{I}} = 3, \rho = 120$ (e) $\ell_{\mathcal{I}} = 5, \rho = 140$

Numerical simulations

Evolution plots for random graphs

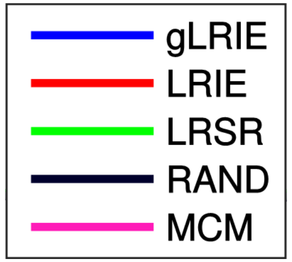


Case A: *~ to #infected neighbors*



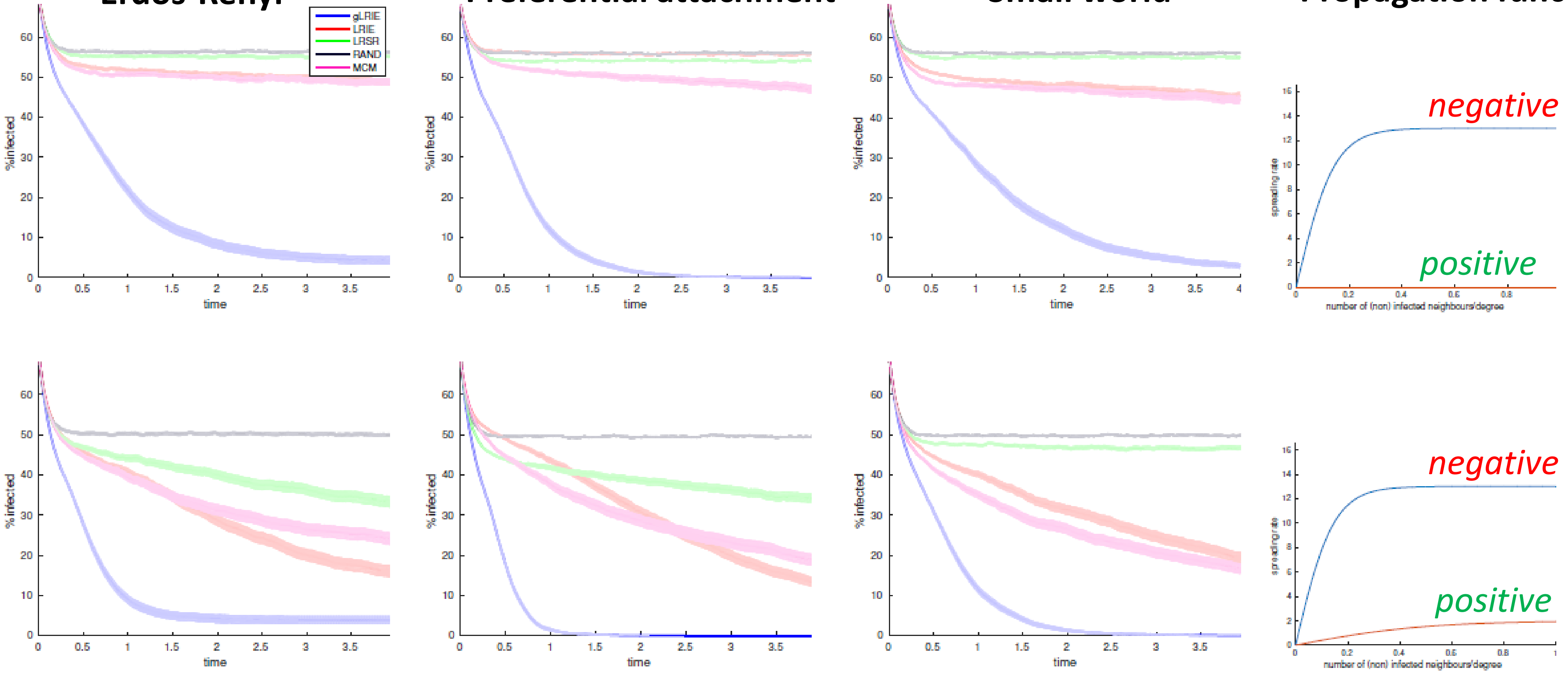
Numerical simulations

Evolution plots for random graphs



Case B: *~ to %infected neighbors*

Erdös-Rényi | **Preferential attachment** | **Small world** | **Propagation functions**



Results

Real-world networks

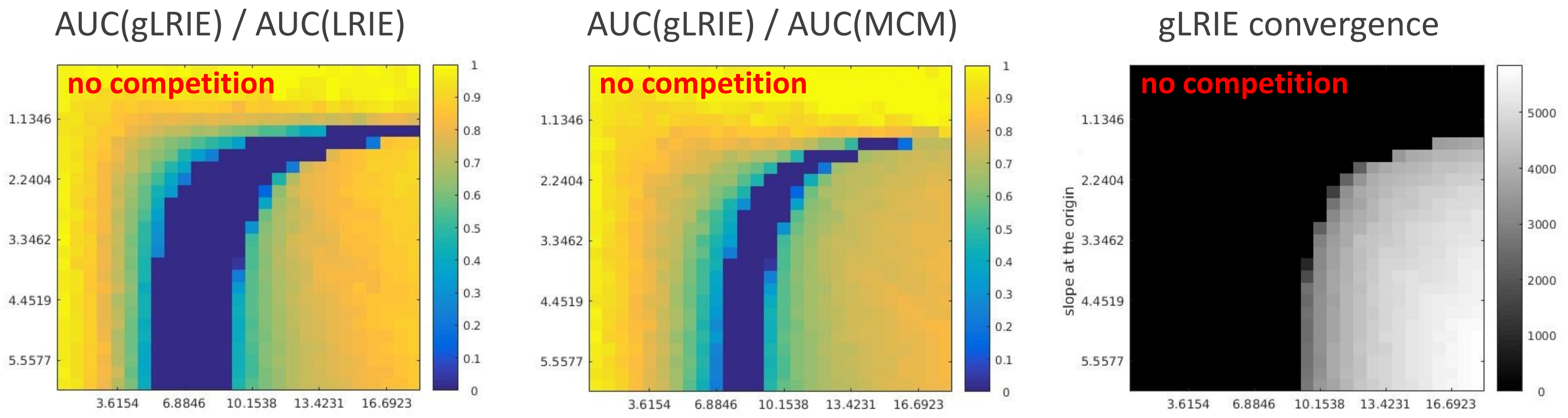
Heatmaps in the (s_I, ℓ_I)-space

saturation →, *slope* ↓

Gnutella

8846 nodes

31839 edges



Results

Real-world networks

Heatmaps in the (s_I, ℓ_I)-space

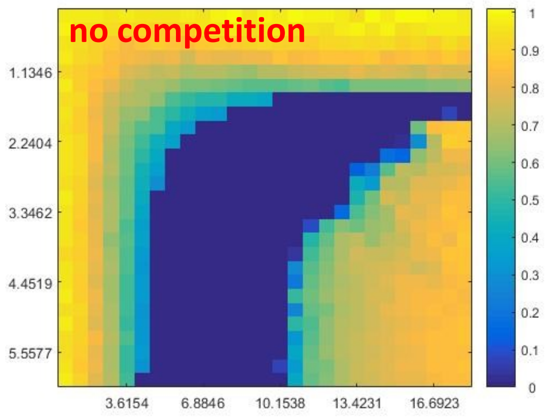
saturation → / *slope* ↓

arXiv H.E.Physics

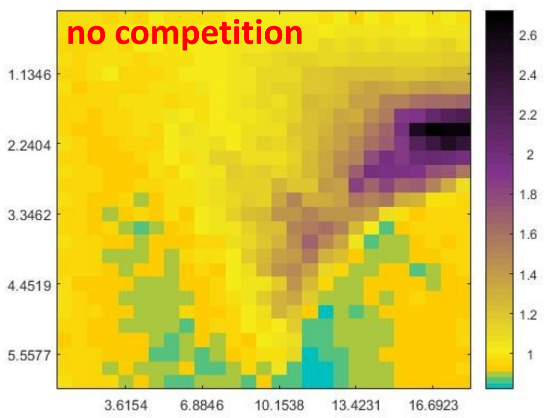
8637 nodes

24803 edges

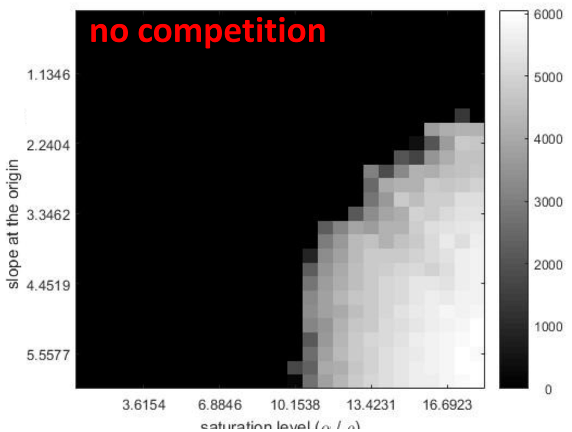
AUC(gLRIE) / AUC(LRIE)



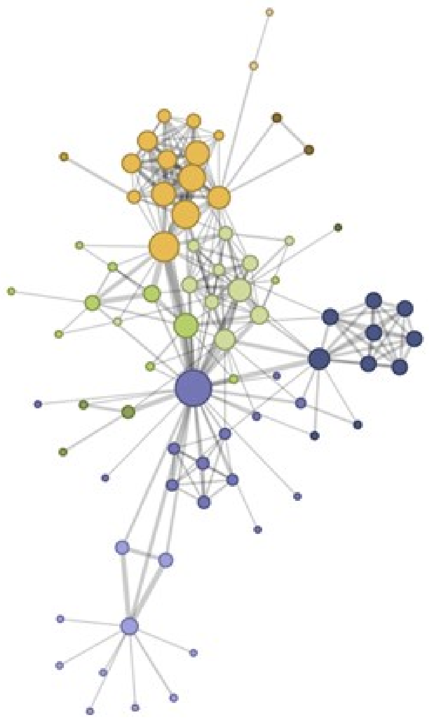
AUC(gLRIE) / AUC(MCM)



gLRIE convergence



gLRIE pros and cons



Pros

- motivated by social contagions scenarios
- takes into account competition
- arbitrary propagation functions
- Inherits the adaptivity and elegance of LRIE

Cons

- inherits the greediness and lack of co-ordination of LRIE

Question to answer

What about relaxing the requirements of the DRA class of resource allocation strategies

Standard DRA

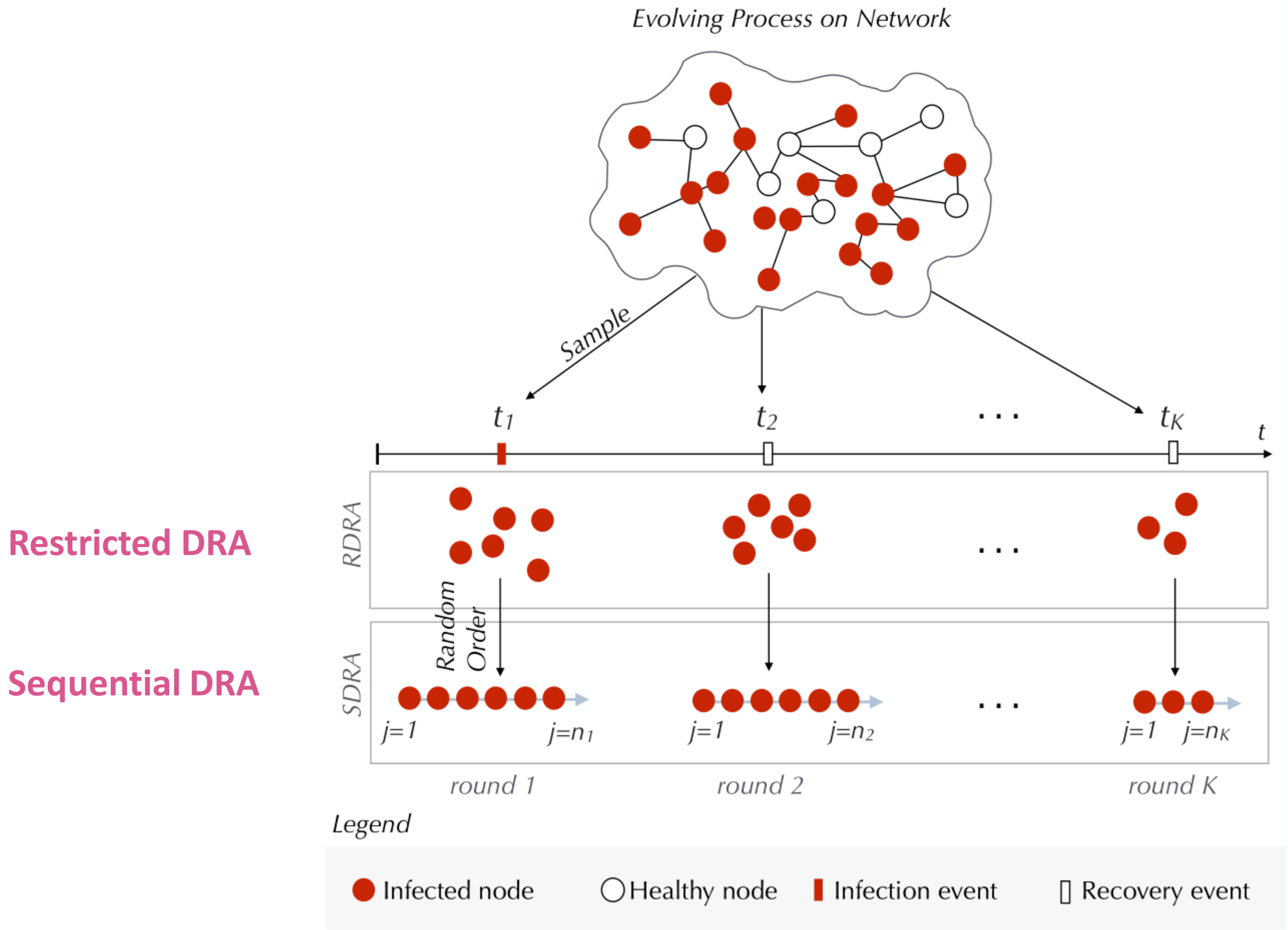
Motivations

- Unrealistic ‘power’ of the administrator
- Play with access and information

Assumption

- access and information are inextricable

Restricted and Sequential DRA



The sequential selection problem (SSP)

BASIC VERSION

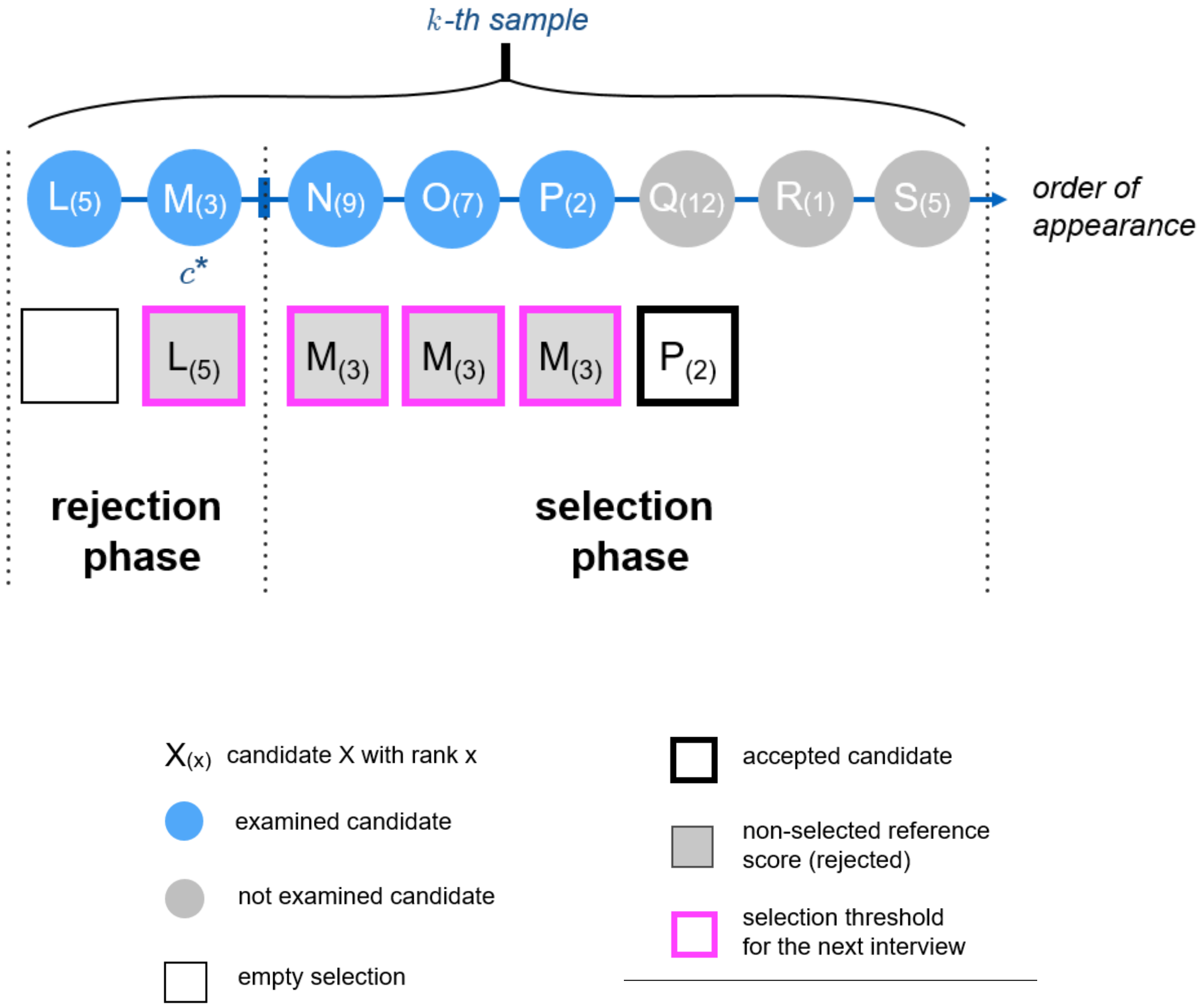
Constraints

Immediate and **irrevocable** decision after each interview

No info about, or control to, the input

<ins>**Limitations**</ins> of the classical SSP setting

- *Cold start*: zero prior knowledge
- Single-shot problem…



The Warm-starting Sequential Selection Problem

- **Warm-starting Sequential Selection Problem (WSSP)**

  – *Background*

  $\mathcal{B} = (b, n, s, \mathbf{C}^R)$, where
    - $n \in \mathbb{N}$: nb. of nodes to evaluate,
    - $b \in \mathbb{N}$: nb. of resources,
    - $s : \mathcal{V} \to \mathbb{R}_+$: scoring function,
    - $\mathbf{C}^R \subset \mathcal{V}$: *preselection*, i.e. set of currently treated nodes

  – *Process & Decisions*

  $\mathbf{C} = (C_1, \ldots, C_n) \in \mathcal{P}_n(\mathcal{V} \backslash \mathbf{C}^R)$ and $(R_{C_1}, \ldots, R_{C_n}) \in \{0,1\}^n$

  – *Evaluation*

  The regret function is defined as:

$$\phi_{\mathcal{B}} = \max_{R_i^*, i \in \mathcal{C}} (\mathbf{S} \cdot \mathbf{R}^*) - (\mathbf{S} \cdot \mathbf{R}) \in \mathbb{R}_+,$$

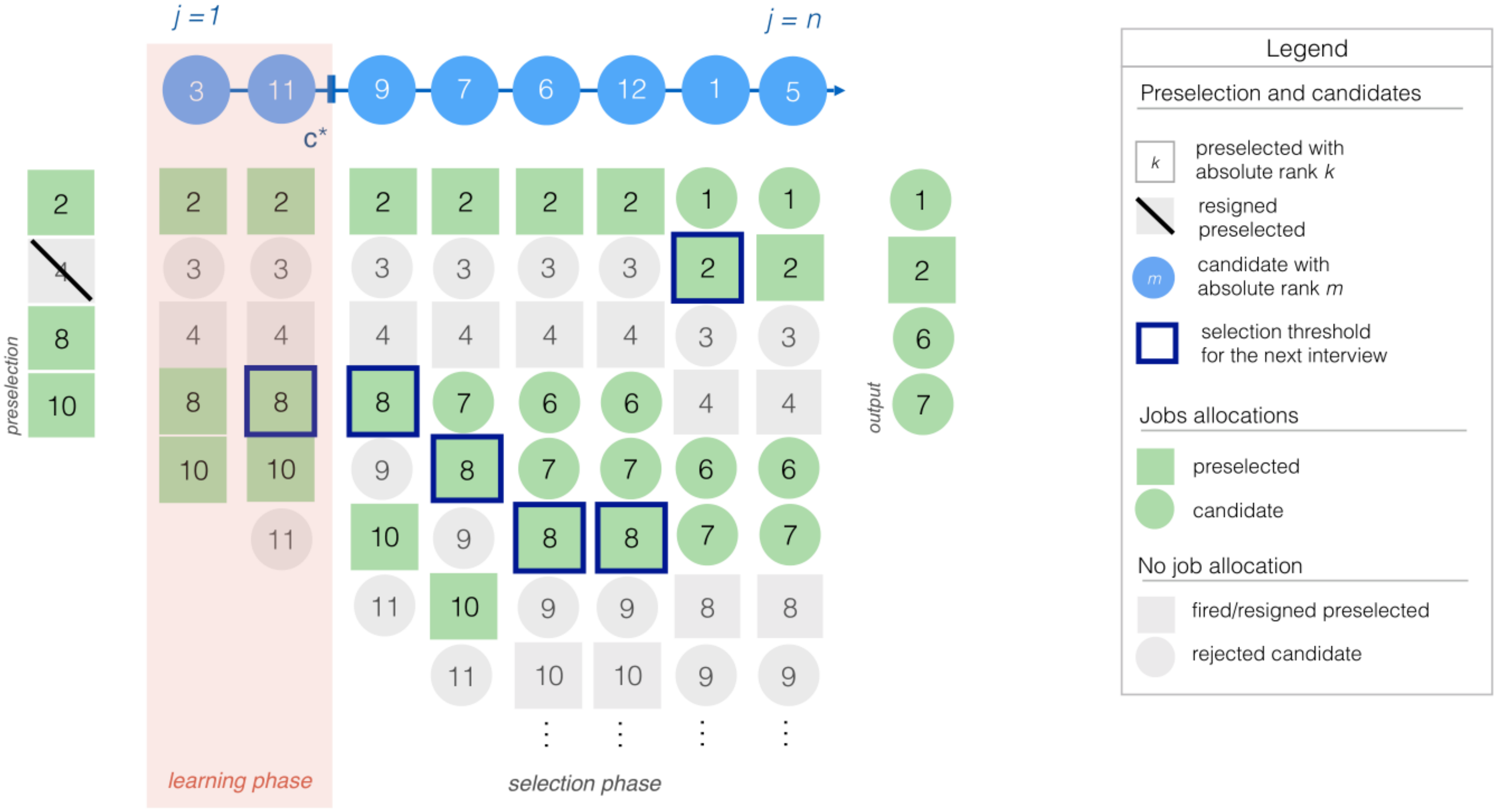
  where $\mathcal{C} = (\mathbf{C}^R, C_1, \ldots, C_n) \subset \mathcal{V}$.

Goal: minimize $\mathbb{E}[\phi_{\mathcal{B}}]$.

The Multi-Round Sequential Selection Problem, Mathilde Fekom, Argyris Kalogeratos, Nicolas Vayatis, arxiv preprint, 2018

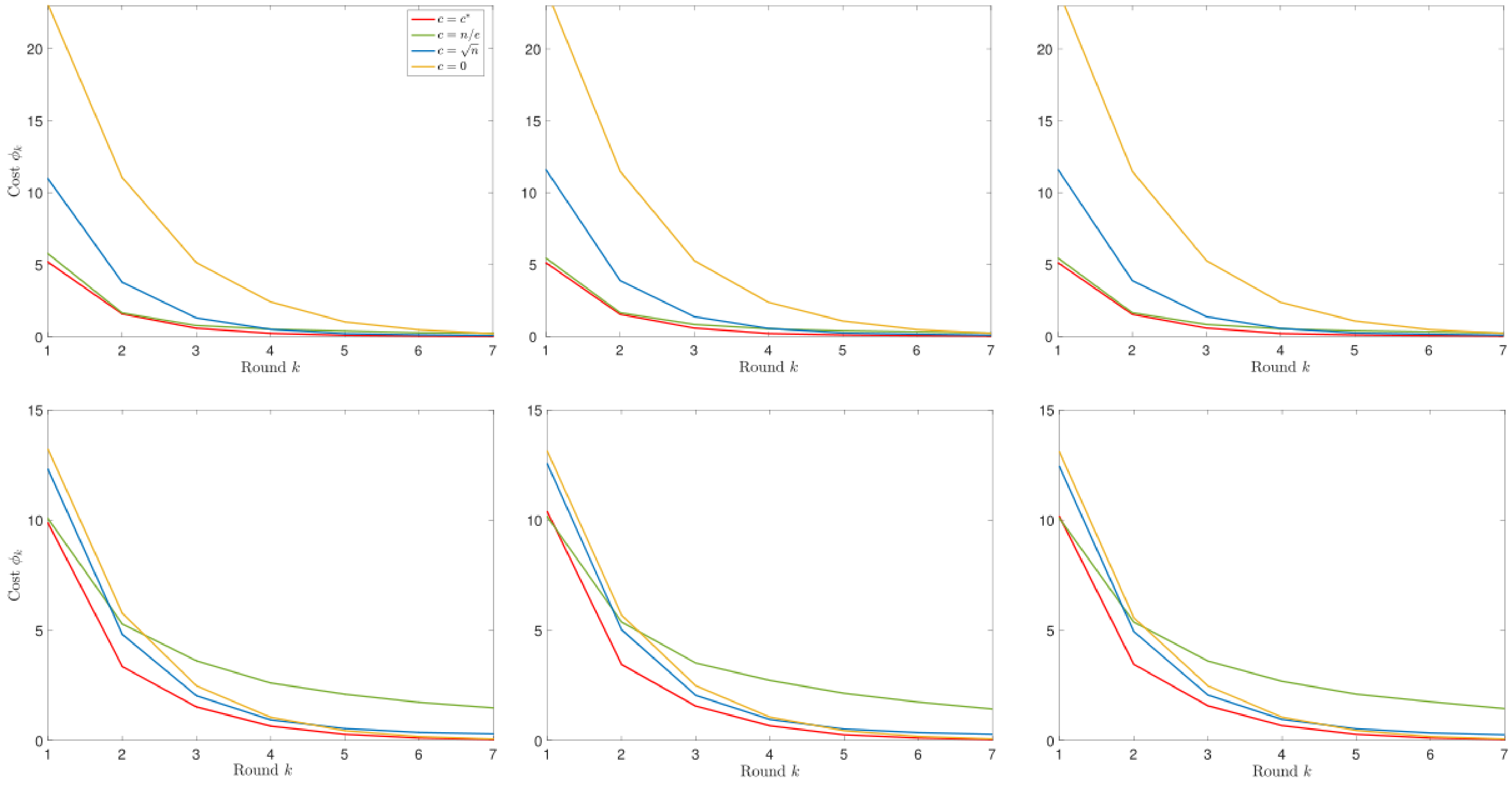
The Multi-Round Sequential Selection Problem, Mathilde Fekom, Argyris Kalogeratos, Nicolas Vayatis, arxiv preprint, 2018

Warm-starting & multi-round sequential selection processes



The Multi-Round Sequential Selection Problem, Mathilde Fekom, Argyris Kalogeratos, Nicolas Vayatis, arxiv preprint, 2018

Warm-starting & multi-round sequential selection processes



(a) $S_{i,k} \sim \mathrm{U}(0,1)$ (b) $S_{i,k} \sim \mathrm{N}(0,5)$ (c) $S_{i,k} \sim \mathrm{Exp}(1)$

Pluging warm-starting into Sequential DRA

Example on a scale-free network

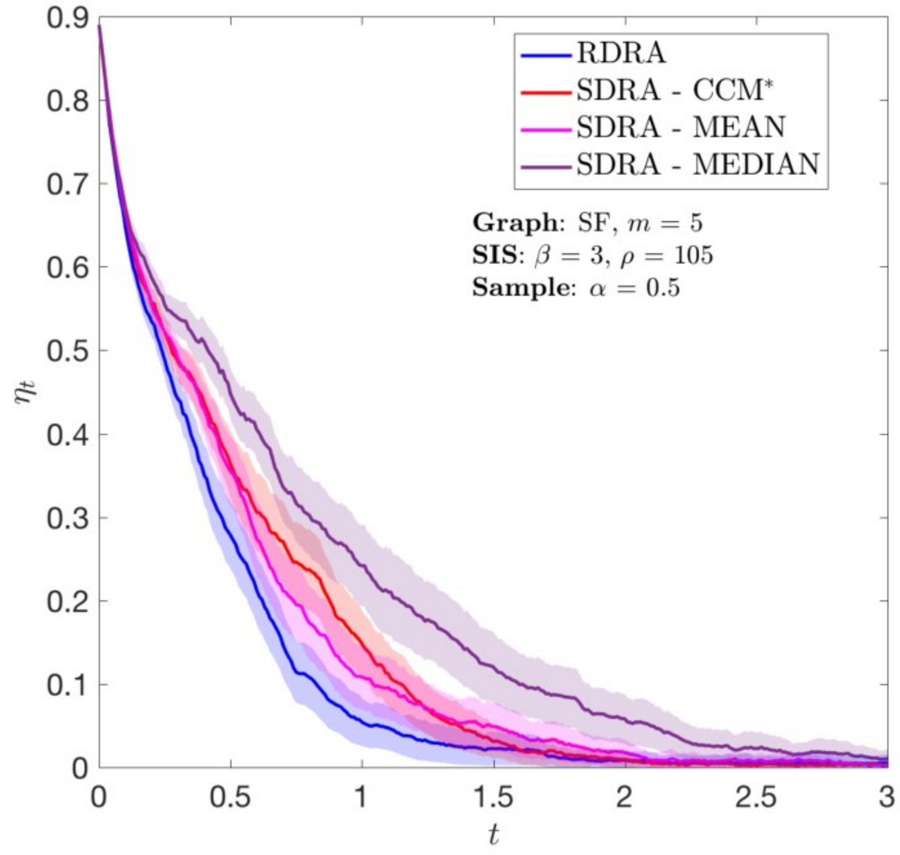


Figure: Percentage of infected nodes w.r.t. time.

<u>**Various strategies**</u>

- **Hiring-above-the-mean** (MEAN) [Broder et al., 2009]:
  Acceptance threshold is the mean of employees
  Goal: grow the company as much as possible while keeping maximal the average score of the employee
- **Cutoff-based Cost Minimization** (CCM) [Fekom, Vayatis, Kalogeratos 2019]
  Generic algorithm i.e. works with any scoring function
  Goal: minimize the expectation of the ranks of the selected
- **Warm-starting Dynamic Thresholding** (WDT)) [Fekom, Vayatis, Kalogeratos 2019]
  Assumes score distribution is known
  Optimal acceptance threshold
  Goal: maximize the expectation of the scores of the selected

Discussion

?!