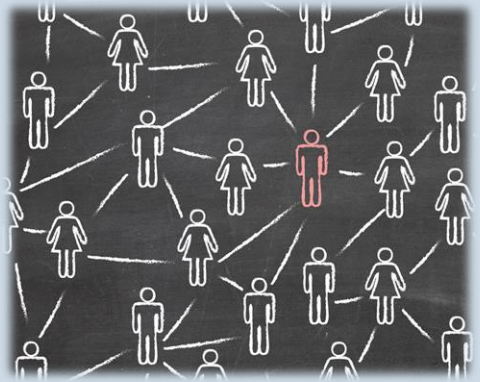
SUPPRESSING EPIDEMICS ON ARBITRARY NETWORKS USING TREATMENT RESOURCES OF LIMITED EFFICIENCY



APPLICATIONS IN SOCIAL / HEALTH SCIENCES AND MARKETING

Argyris Kalogeratos

Joint work with

Kevin Scaman and Nicolas Vayatis

ENS CACHAN — CMLA Centre de Mathématiques et de Leurs Applications — MLMDA MACHINE LEARNING AND MASSIVE DATA ANALYSIS

Seminar talk

Complex Networks Group - LIP6

13 June 2016

ML+NETWORKS @ CMLA

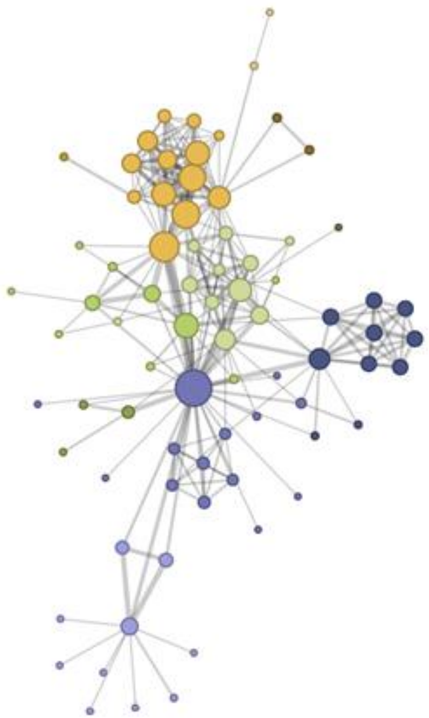
- **CMLA:** Centre de Mathématique et de Leurs Applications (ENS Cachan)
- **MLMDA**: Machine Learning and Massive Data Analysis
  - Director: Nicolas Vayatis, 2 researchers. 3 post-docs,
  - 8 PhD students, 3 MSc interns, 2 BSc interns
- **Machine Learning on Networks**
  - 1 researcher, 2 PhD students, 3 MSc interns

TALK SUMMARY

- Diffusion processes and epidemic models
- Overview of diffusion suppression control approaches
- A greedy method with dynamic resource allocation
- A dynamic method based on priority planning
- Conclusions

DIFFUSION PROCESSES ON NETWORKS

Basics



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

- 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…

MOTIVATION:

FROM DISEASE EPIDEMICS TO… DIGITAL AND SOCIAL EPIDEMICS

[5] Brockmann et al. *The Hidden Geometry of Complex, Network-Driven Contagion Phenomena, Science,* 2013.

DIFFUSION PROCESSES ON NETWORKS

Diffusion Models

Diffusion model

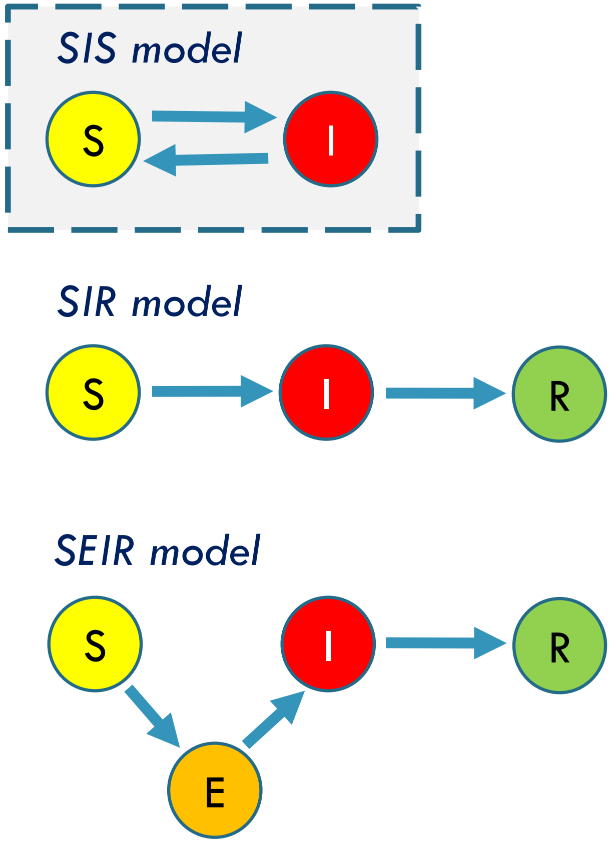
- a mathematical model that encodes the "*propagation rules*"
- no single model able to describe all possible complex diffusion phenomena

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

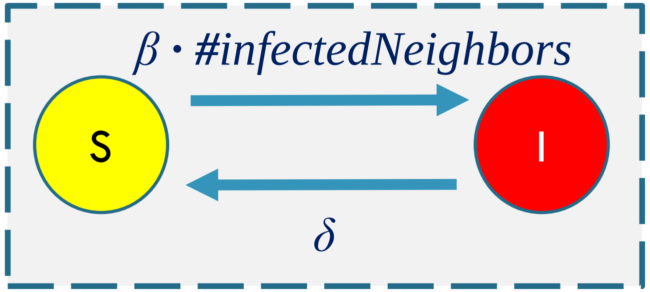
DIFFUSION PROCESSES ON NETWORKS

Diffusion Models – SIS demo

Example

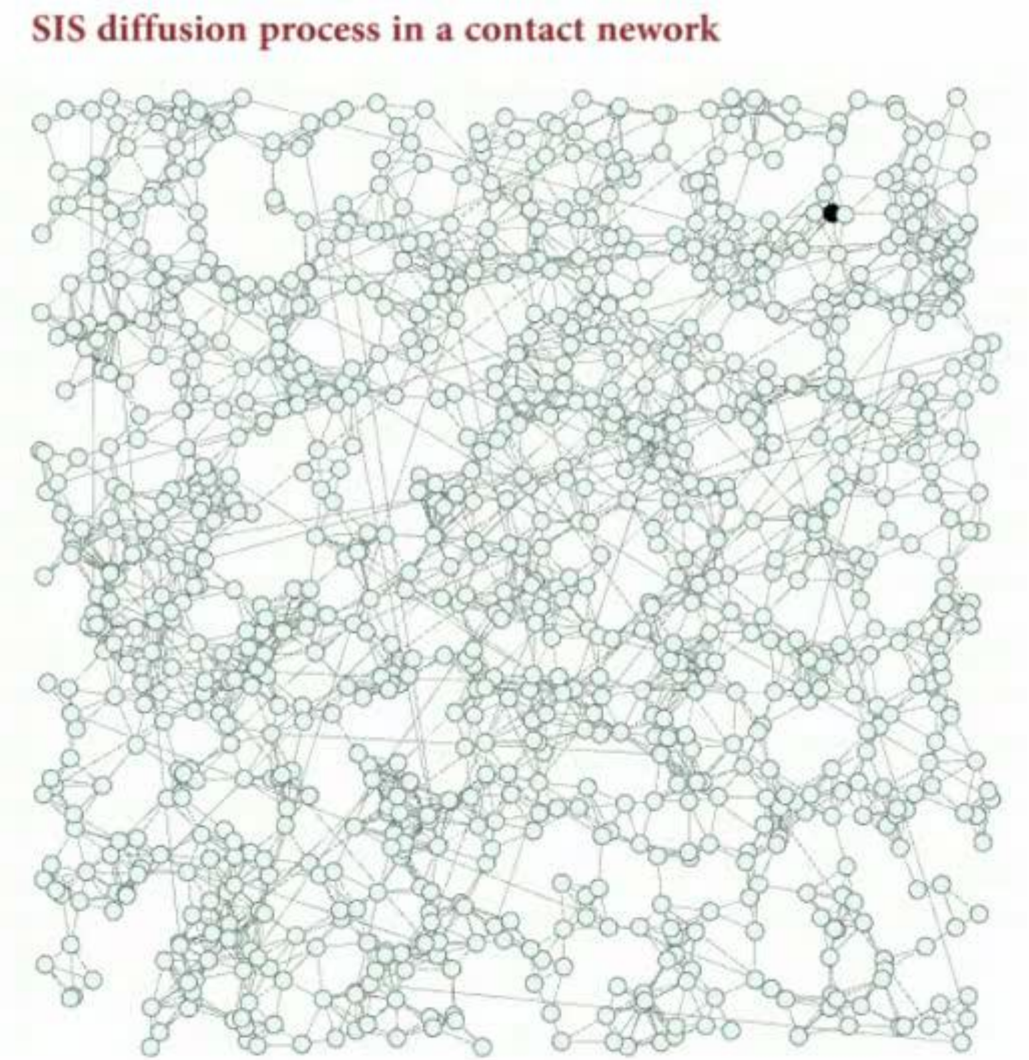
- uncontrolled SIS process on contact network

Homogeneous continuous-time SIS model for one node



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

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



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

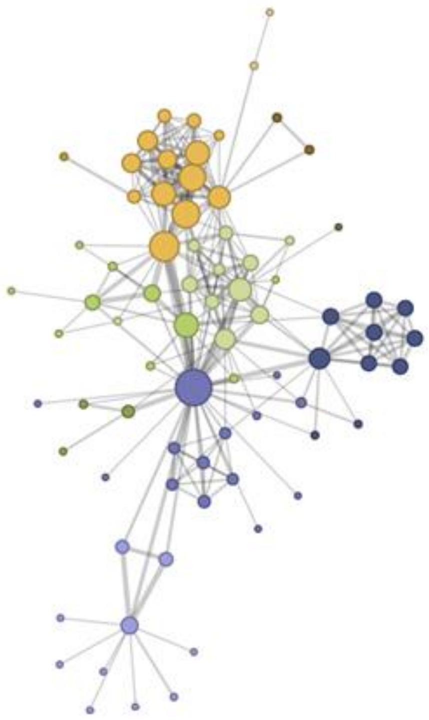
DIFFUSION PROCESSES ON NETWORKS

Directions of research

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

Roughly three directions of research

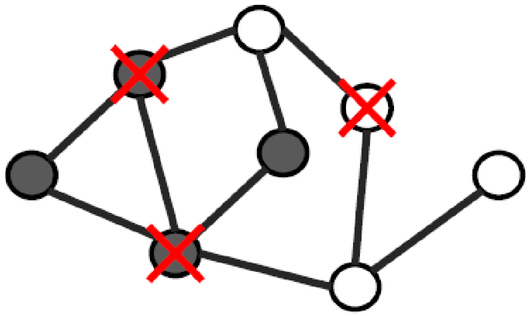
- **Network assessment:** worst case analysis, risk/vulnerability assessment
- **DP engineering:** influence maximization, (viral) marketing
- **DP suppression and control:** containment of viruses, rumors, social behaviors, etc., using ***control actions***



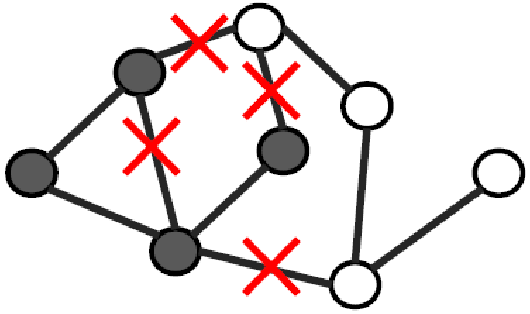
DIFFUSION SUPPRESSION AND CONTROL

Possible control actions

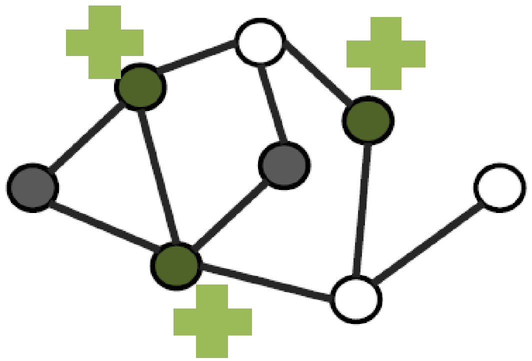
DP suppression and control using ***control actions*** on <u>nodes</u> or <u>edges</u>



Node deletion



Edge deletion

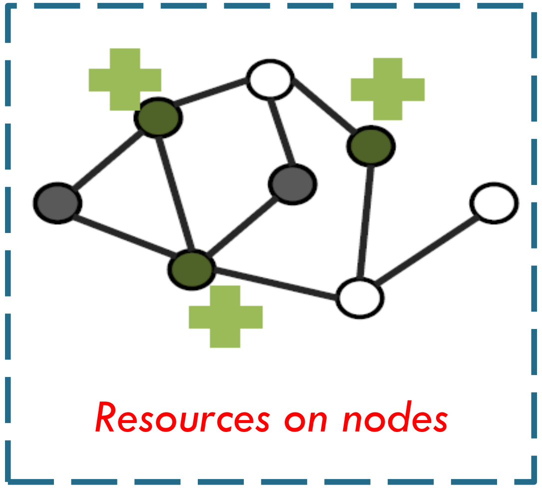


Resources on nodes

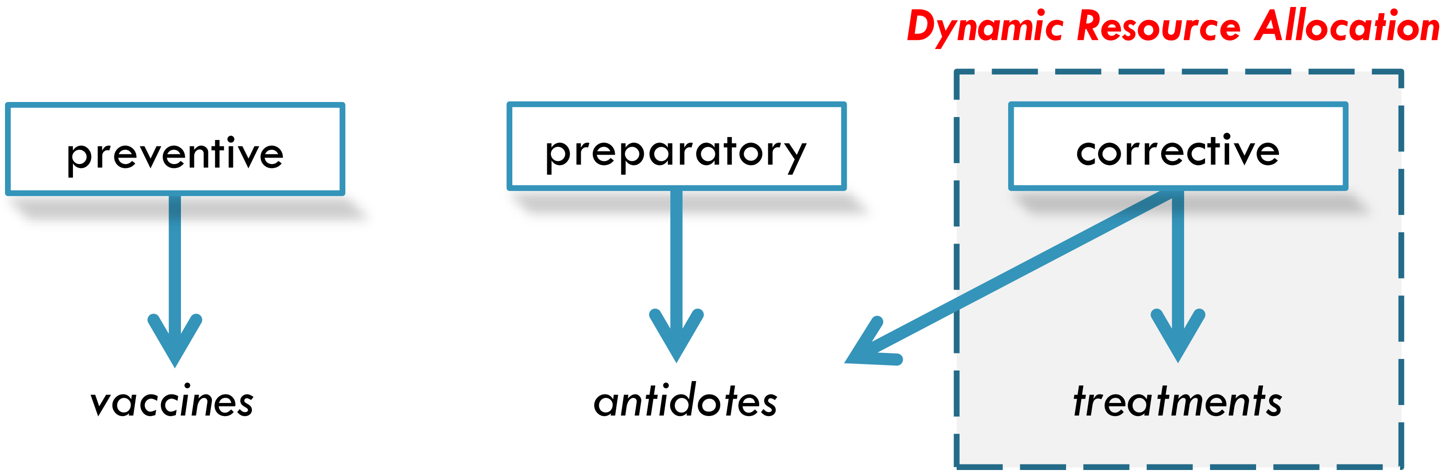
DIFFUSION SUPPRESSION AND CONTROL

Healing resources on nodes – more variations

DP suppression and control using ***control actions*** on nodes



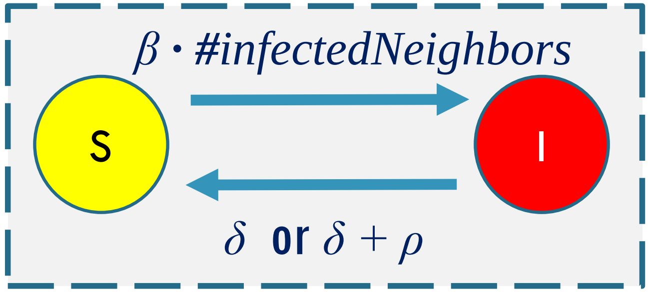
:



DYNAMIC RESOURCE ALLOCATION (DRA)

Modelling and control framework

SIS model for one node



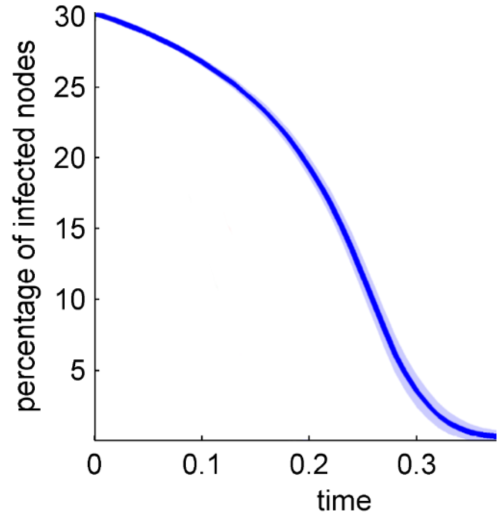
$X_i(t): 0 \rightarrow 1$ at rate $\beta \sum_j A_{ji} X_j(t)$
$X_i(t): 1 \rightarrow 0$ at rate $\delta + \rho R_i(t)$

DRA objective

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

Formally a DRA strategy

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



Continuous-time SIS model

- treatment efficiency ρ
- resource allocation R

Constraints for tractability

- unlimited resources disposed at constant rate
- inability to store resources

DYNAMIC RESOURCE ALLOCATION (DRA)

Modelling and control framework

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

Complexity

- update $O(E+N\log N)$
- <u>but</u> much lower for scores that are based on local graph properties

---

Algorithm Applying a score-based DRA strategy

Input : infection state vector $X(t)$, budget size b_{tot}, scoring function S.
Output: the resource allocation vector $R(t)$.

if $\sum_i X_i(t) < b_{tot}$ **then**
  return $X(t)$
end if
Let $R(t)$ a zero N-dimensional vector
Let $V \leftarrow \{S_i(X(t))\}_{i=1}^{N}$ a vector containing the node scores
Sort the elements of V in *descending* order
  and let I the node indexes of the ranking
for $i = 1$ **to** b_{tot} **do**
  $R_{I(i)}(t) \leftarrow 1$
end for
return $R(t)$

DYNAMIC RESOURCE ALLOCATION (DRA)

Modelling and control framework

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

Complexity

- update $O(E+N\log N)$
- but much lower for scores that are based on local graph properties

Examples

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	$\sigma(X_i)+\sum_j A_{ij}\bar{X}_j$
LIN	$\sigma(X_i)-\sum_j A_{ji}X_j$
LRIE	$\sigma(X_i)+\sum_j [A_{ij}\bar{X}_j - A_{ji}X_j]$, sums MSN and LIN

- $\sigma(1) = 0$ and $\sigma(0) = -\infty$
- $X(t)$ the infection state, $\bar{X}(t) = \mathbb{1} - X(t)$

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

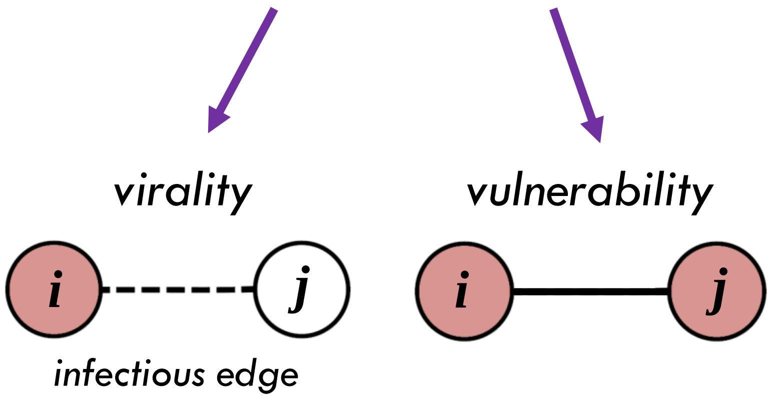
$$C_\gamma(R,t,X) = \frac{1}{\gamma}\sum_i X_i + \frac{1}{\gamma^2}\Phi'_{t,X}(0) + \frac{1}{\gamma^3}\Phi''_{t,X}(0) + O(\frac{1}{\gamma^4})$$

⋮

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

For an infected node i

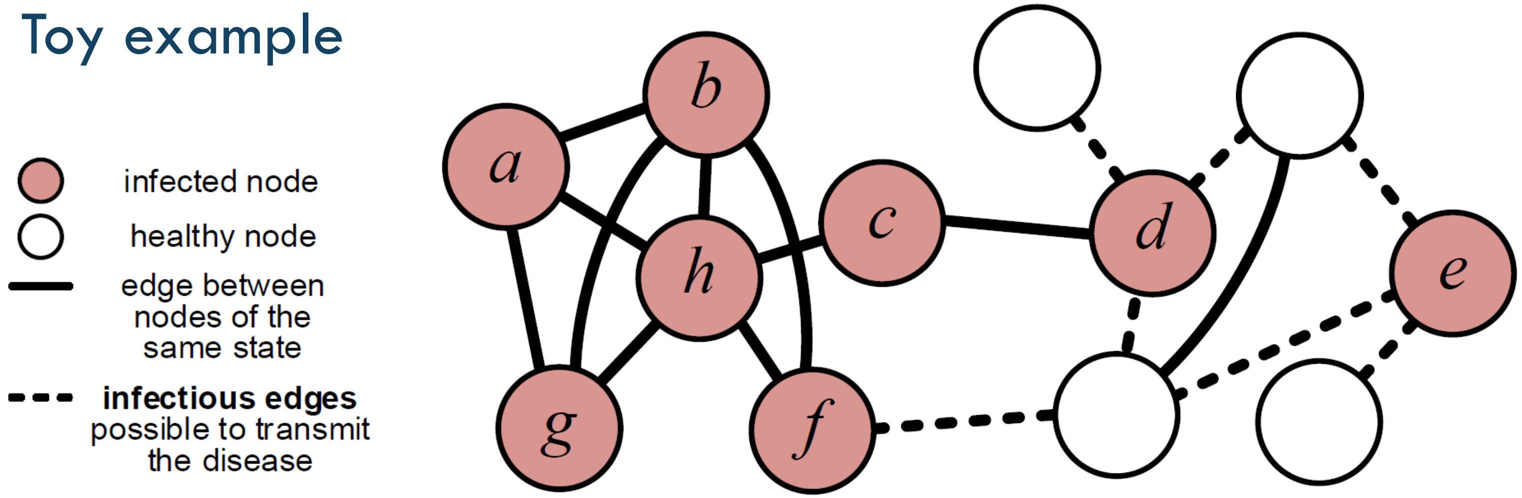
$$\sum_j [A_{ij}\overline{X}_j(t) - A_{ji}X_j(t)]$$



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

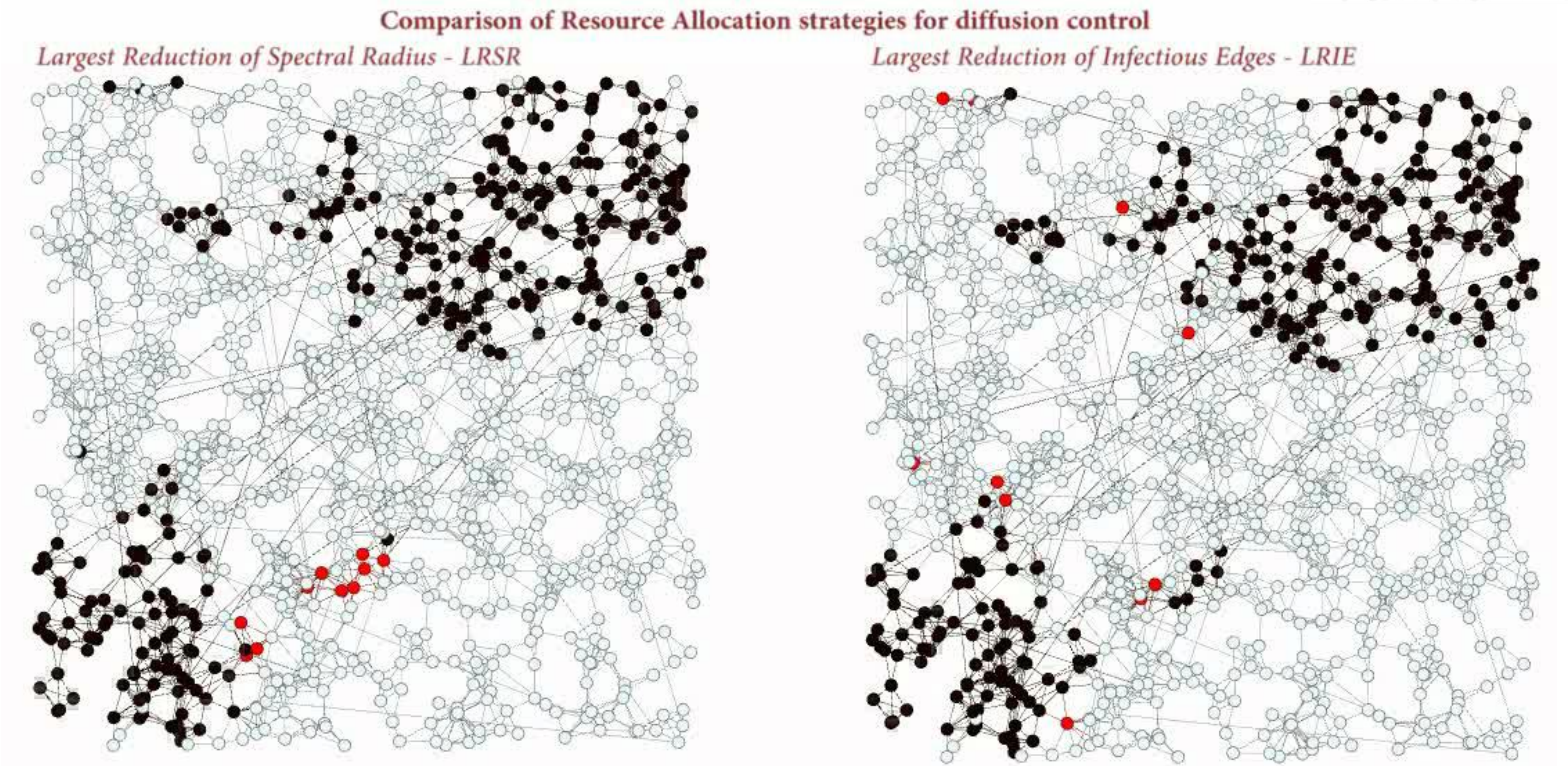
Priority 3: f | S_f=1-2

⋮

OPTIMAL GREEDY DRA

Demonstration on an artificial contact network

LRIE
Largest Reduction in Infectious Edges



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

RESULTS

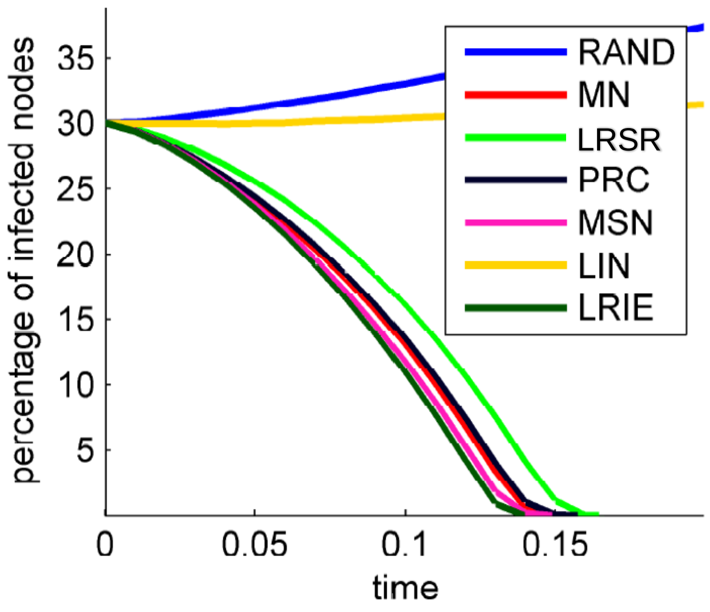
Random graph model: scale-free

Scale-free network:

$N = 10^4$ nodes

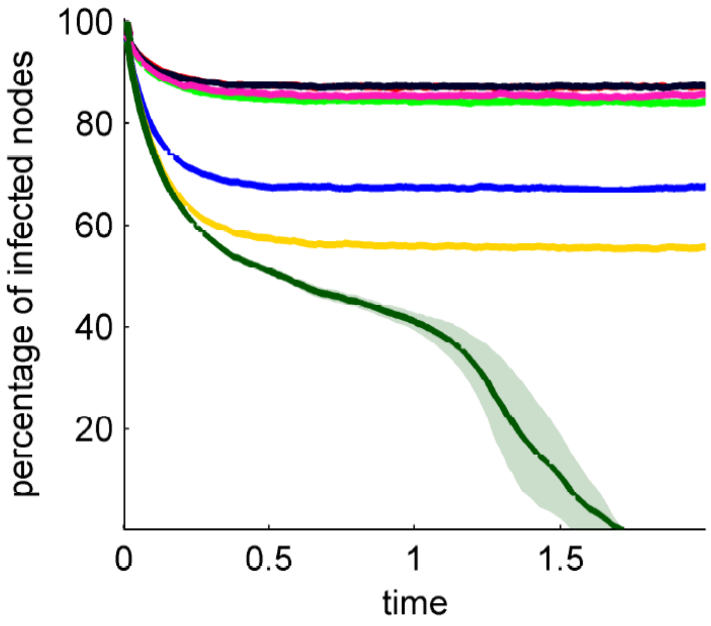
$p = 0.001$

$m = 5$



(a) $e = 4000$

$r = 2, b_{tot} = 10$



(b) $e = 3000$

$r = 2, b_{tot} = 10$

RESULTS

Random graph model: Erdös-Rényi

Heatmaps of avg. AUC ratio

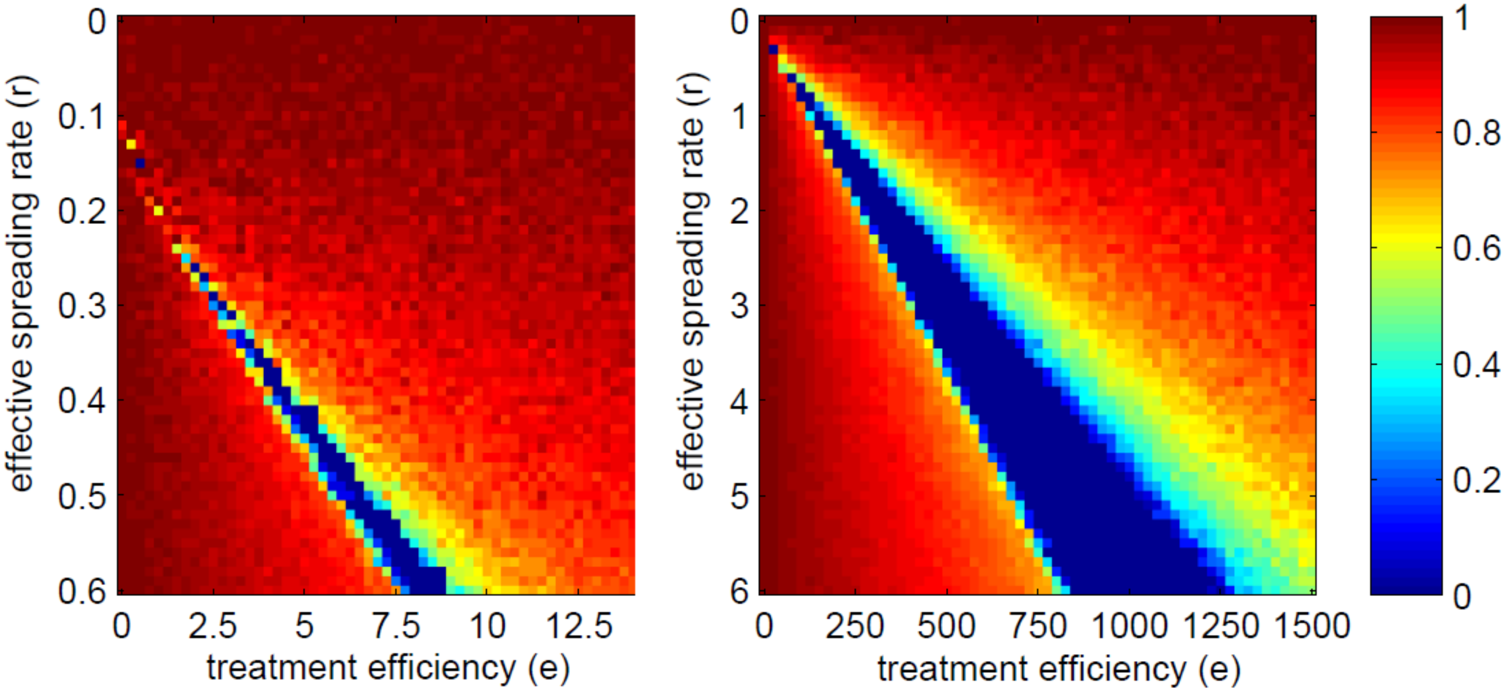
AUC(**LRIE**) / AUC(**LRSR**)

Erdös-Rényi networks:

$N = 1{,}000$ nodes, $p = 0.01$

Small and *large values for*

$r = \beta / \delta$ and $e = \rho / \delta$



(a) $b_{tot} = 100$; small r, e (b) $b_{tot} = 10$; large r, e values

RESULTS

Random graph model: scale-free

Heatmaps of avg. AUC ratio

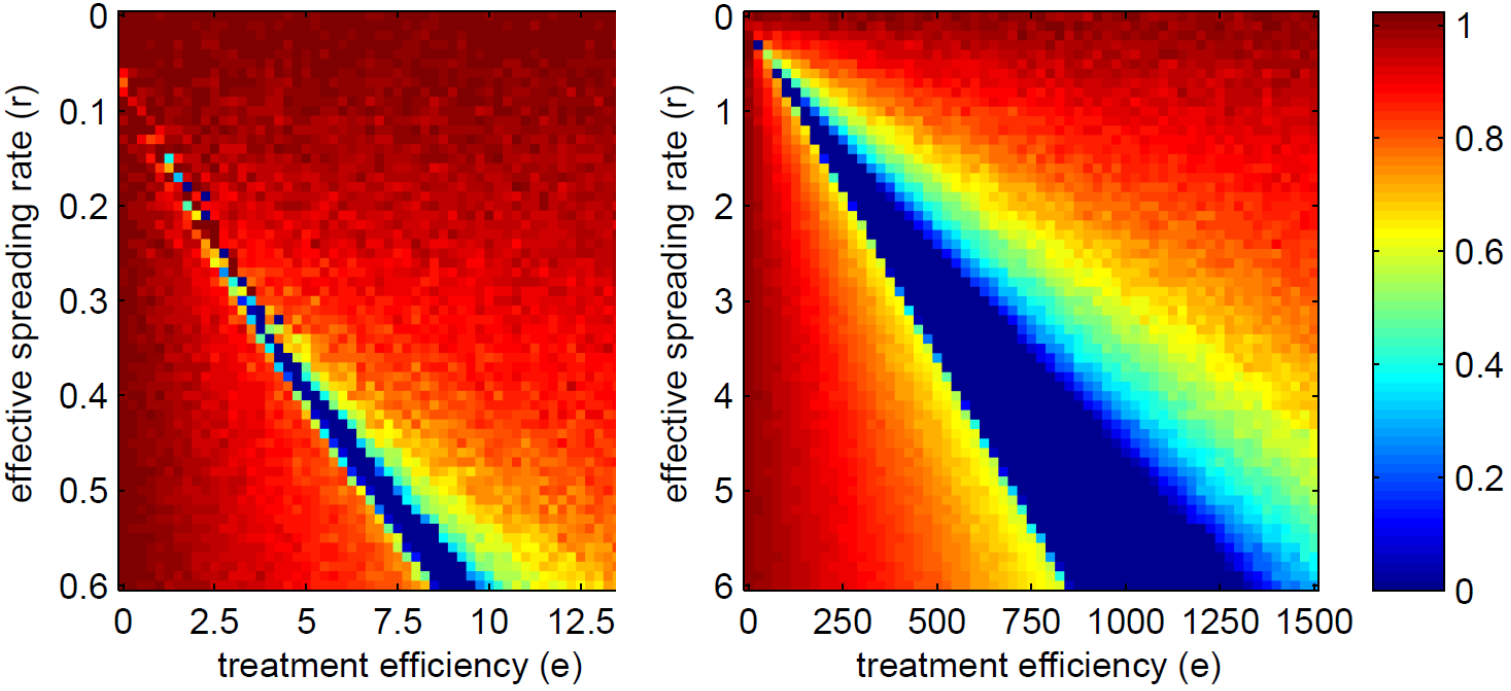
AUC(**LRIE**) / AUC(**LRSR**)

Scale-free networks:

N = 1,000 nodes, p = 0.01

Small and *large values for*

$r = \beta / \delta$ and $e = \rho / \delta$



(a) $b_{tot} = 100$; small r, e (b) $b_{tot} = 10$; large r, e values

RESULTS

Real-world networks

Twitter subgraph

1,000 ego-networks

N = 81,306 nodes, E = 1,342,303 edges

US air traffic

N = 2,939 nodes, E = 30,501 edges

Network	DP scenario δ	r	e	b_{tot}	Strategy	AUC$_\downarrow$	$T_{ext\downarrow}$	$N_I(T)_\downarrow$
Twitter subgraph	1	0.2	300	100	RAND	∞	∞	30.6%
					MN	∞	∞	33.4%
					LRSR	246,476	7.70	0%
					MSN	89,671	2.52	0%
					LRIE	**64,425**	**2.07**	**0%**
	1	0.2	200	100	RAND	∞	∞	37.3%
					MN	∞	∞	42.3%
					LRSR	161,195	5.11	0%
					LRIE	**87,600**	**3.03**	**0%**
	1	0.2	50	100	RAND	∞	∞	46.4%
					MN	∞	∞	48.5%
					LRSR	∞	∞	48.9%
					MSN	∞	∞	44.4%
					LRIE	∞	∞	**29.2%**
US air traffic	1	2	210	50	RAND	∞	∞	26.1%
					MN	∞	∞	73.8%
					LRSR	3,723	1.81	0%
					MSN	3,235	1.65	0%
					LRIE	**493**	**0.43**	**0%**
	1	2	150	50	RAND	∞	∞	38.9%
					MN	∞	∞	76.6%
					LRSR	∞	∞	76.5%
					MSN	∞	∞	76.4%
					LRIE	**863**	**1.08**	**0%**
	1	2	100	50	RAND	∞	∞	49.7%
					MN	∞	∞	79.0%
					LRSR	∞	∞	79.2%
					MSN	∞	∞	77.4%
					LRIE	∞	∞	**23.1%**

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 co-ordinated actions

PROBLEM SOLVED?

Question

Is there a way to make an efficient plan that respects the network properties, and follow it persistently throughout the whole process?

What kind of guarantees could be provided?

GLOBAL PRIORITY PLANNING

Definitions

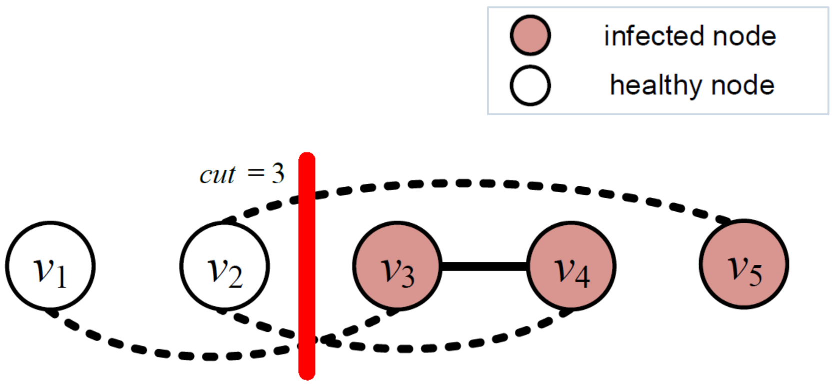
Priority-order: a bijection $\ell : \mathcal{V} \to \{1, ..., N\}$ s.t. $\ell(v)$ the position of node v in the order

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 $\ell: \mathcal{V} \to \{1, 2, 3, 4, 5\}$

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

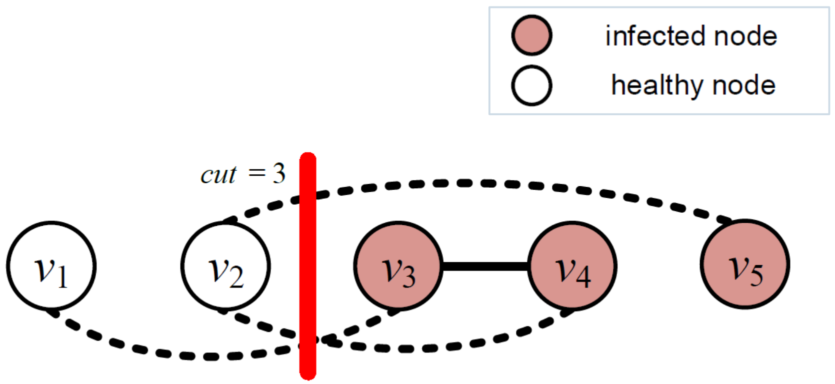
Cutwidth of G: $\mathcal{W} = \min_{\ell} \mathcal{C}^*(\ell)$

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: $\mathbb{E}[\tau_x] \leq$ polynomial function
- *super-critical* behavior: $\mathbb{E}[\tau_x] >$ exponential function

Requirement for designing a strategy:

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

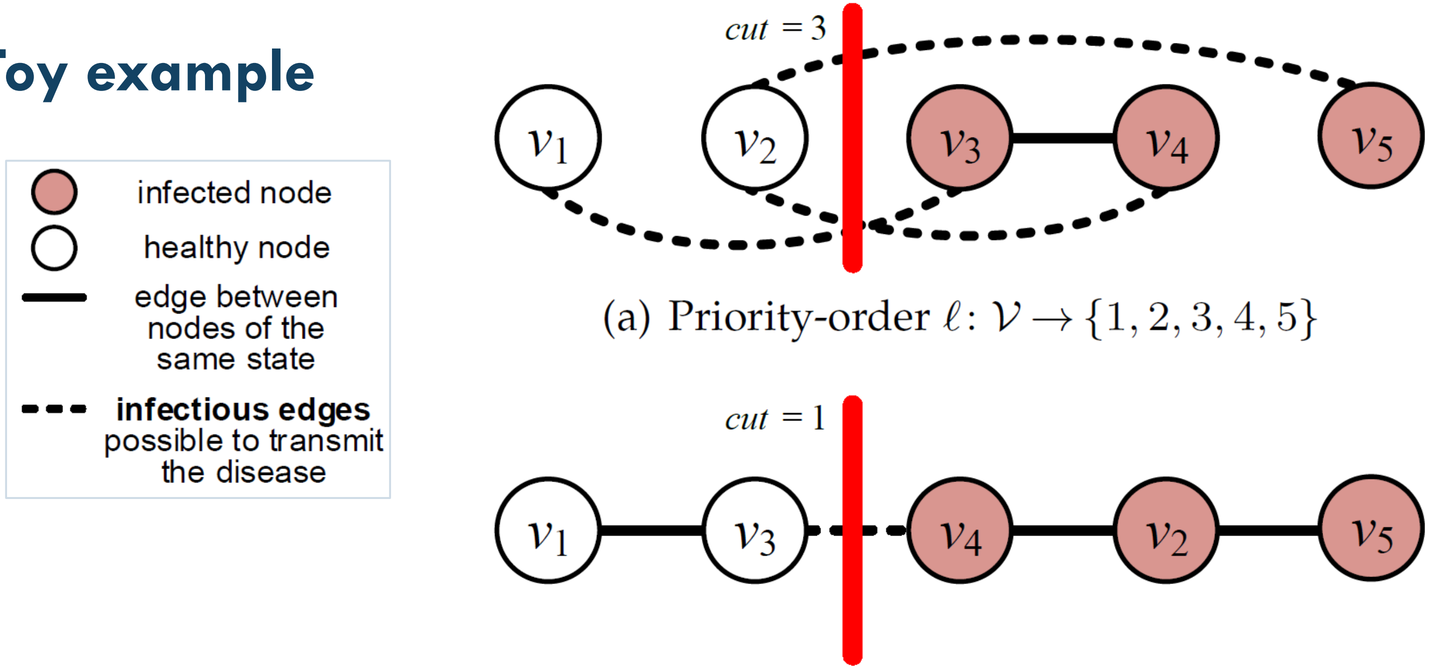


Priority-order $\ell : \mathcal{V} \to \{1, 2, 3, 4, 5\}$

GLOBAL PRIORITY PLANNING

Explaining the role of MaxCut

Toy example



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

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

Priority-order with MaxCut = 3

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 $\mathbb{E}[\tau_1]$:

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

MAXCUT MINIMIZATION (MCM)

MCM Strategy [2, 3, 4]

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 $\ell_{MCM}(\mathcal{G}) = \operatorname{argmin}_{\ell} \mathcal{C}^*(\ell)$

  by

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

Algorithm 1 MCM strategy

▷ *Prior to the 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 $(v_1, \ldots, v_N)$ s.t. $\forall i \in \{1, \ldots, N\}, \ell(v_i) = i$

▷ *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)

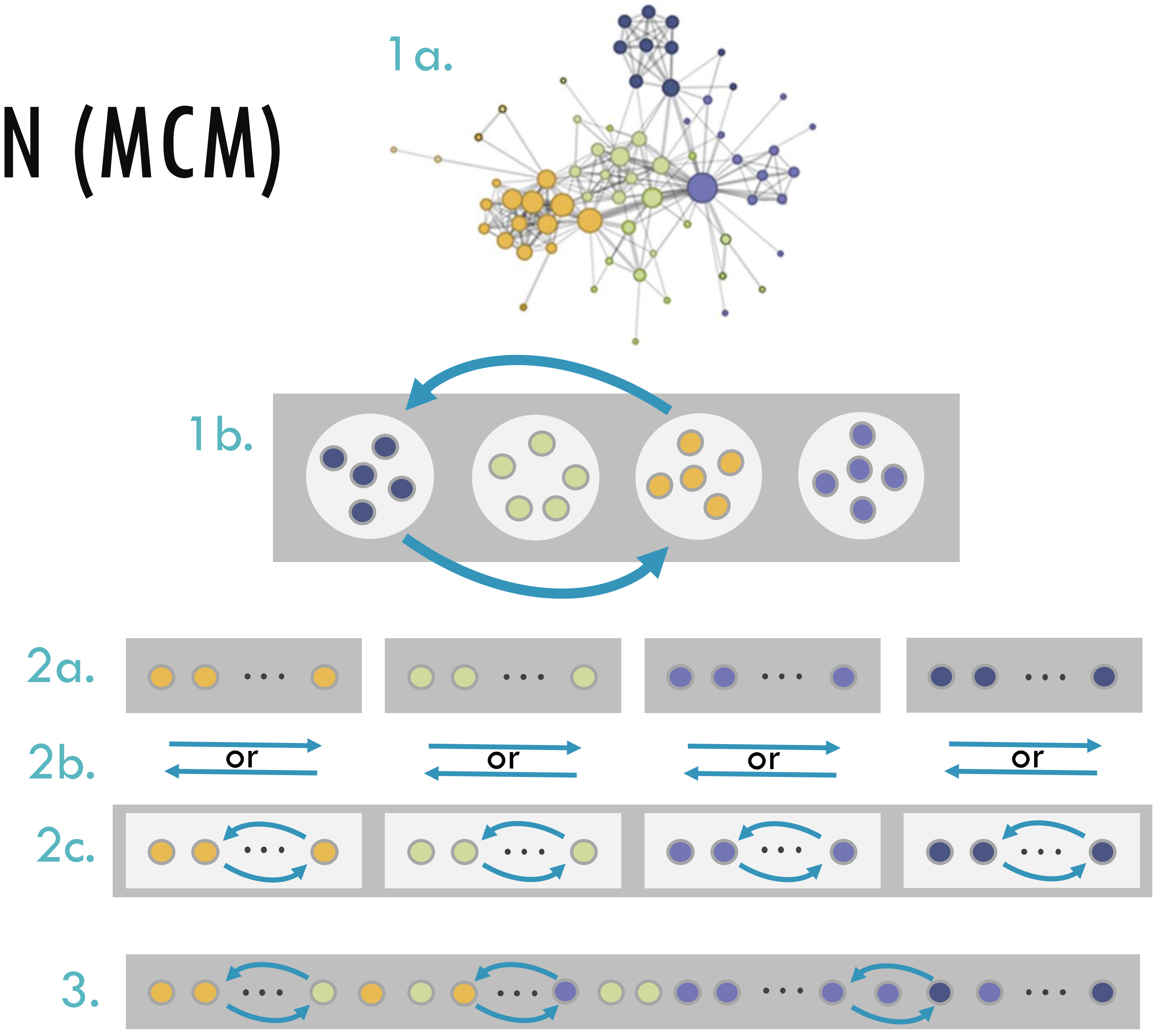
Solving the MLA problem

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

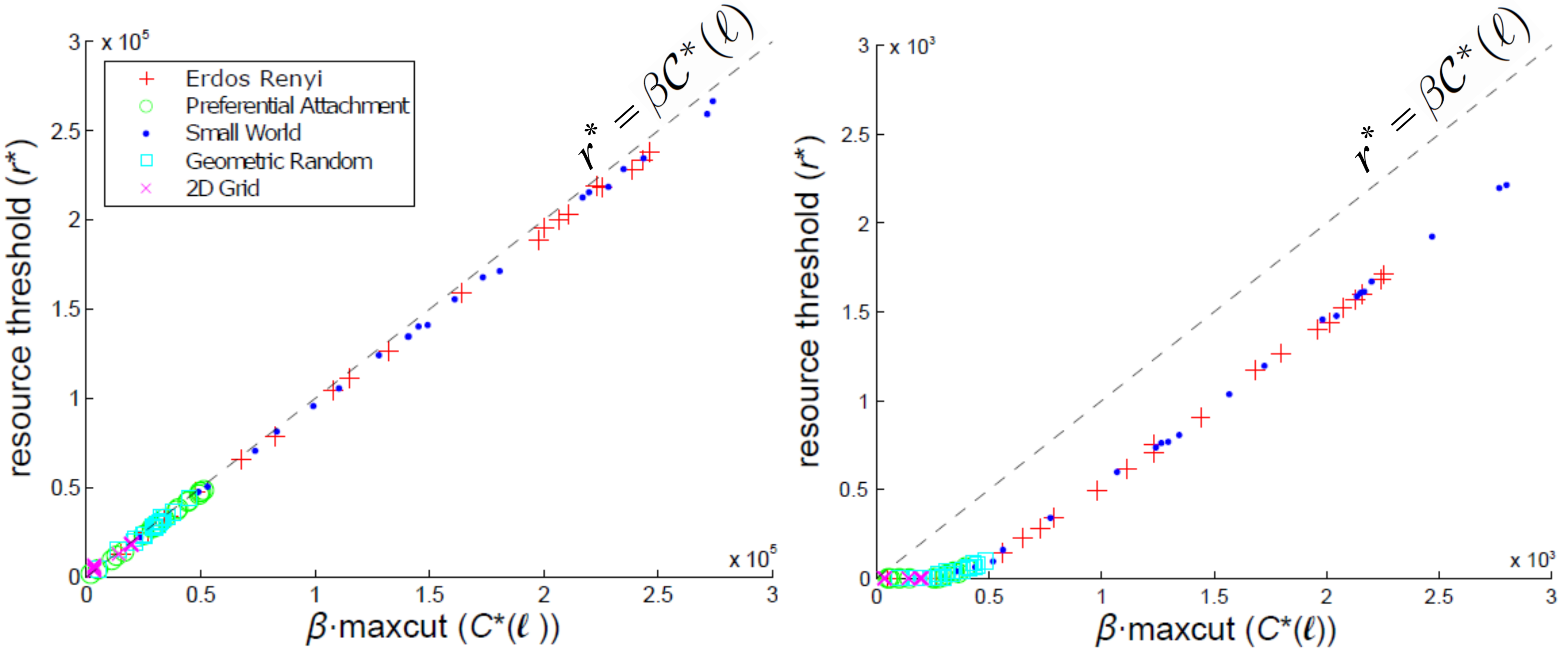


RESULTS

Quality of the theoretical bound

Verifying

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

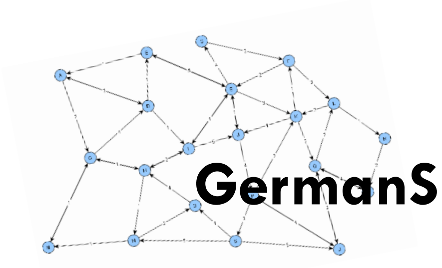


(a) high infectivity: $\beta = 10$ (b) low infectivity: $\beta = 0.1$

- 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

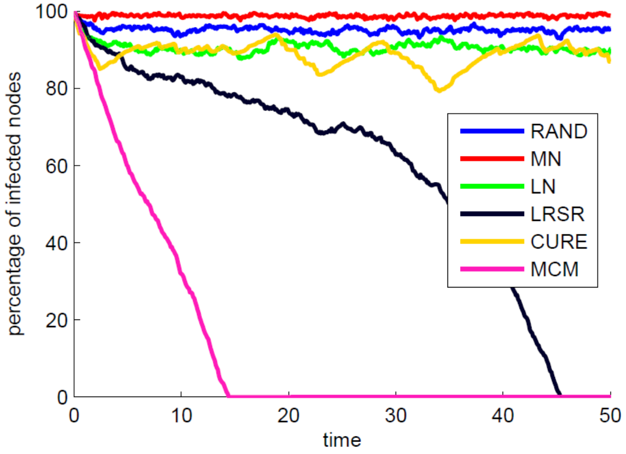


GermanSpeedway

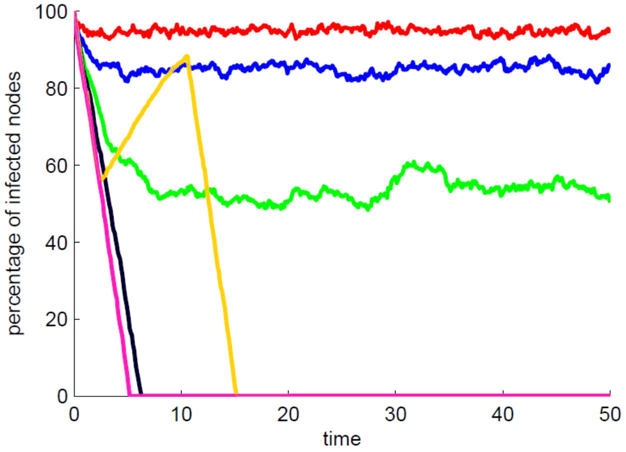
N = 1,168 nodes, E = 1,243 edges,
$max(d) = 12, \beta=1, \delta = 0, q = 1$
MaxCut: 650+/-50 RAND, 379 MN and LN, 104 LRSR, 29 CURE and MCM

OpenFlights

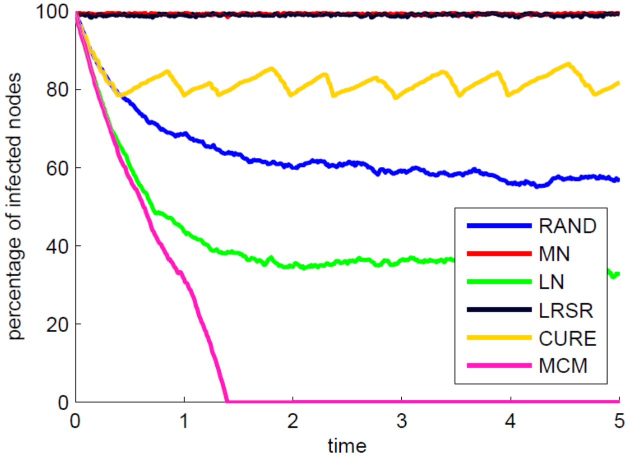
N = 2,939 nodes, E = 30,501 edges,
$max(d) = 242, \beta=1, \delta = 0, q = 1$
MaxCut: 7,800+/-100 RAND, 7,504 MN and LN, 6,223 LRSR, 2,231 CURE and MCM



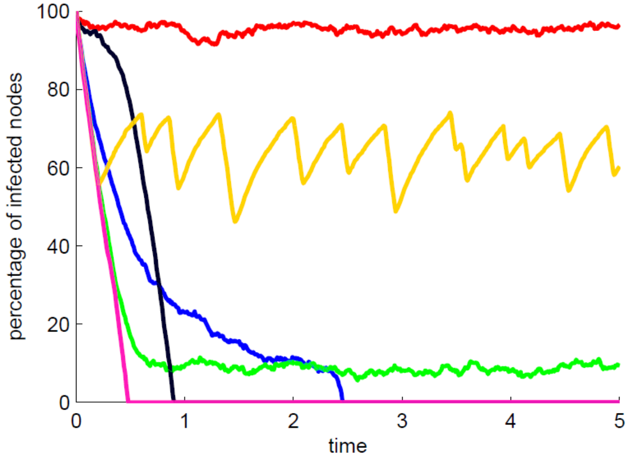
(a) low resource budget: $r = 100$



(b) high resource budget: $r = 250$



(a) low resource budget: $r = 3000$



(b) high resource budget: $r = 7000$

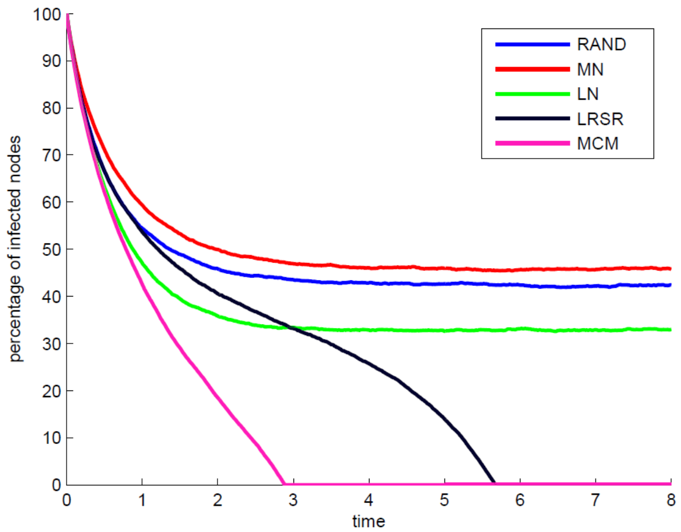
GLOBAL PRIORITY PLANNING

Experiments on real-networks

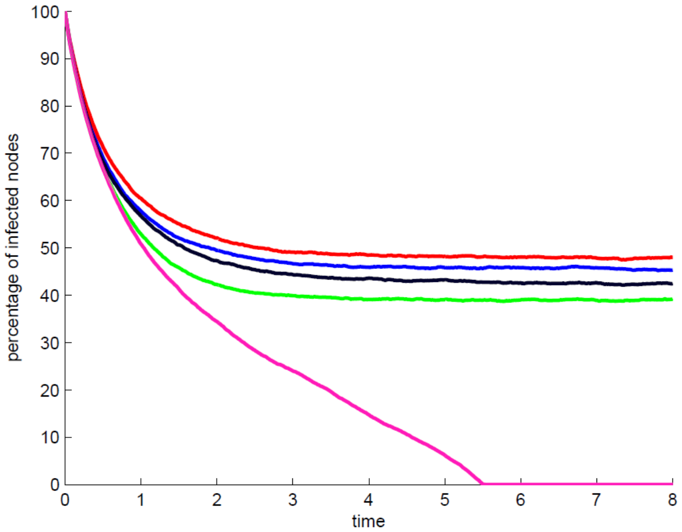
Subset of Twitter network with 81.306 nodes

Strategy	Maxcut	Maxcut % w.r.t. RAND	Expected resource threshold ($\delta = 1$, $\beta = 0.1$, $q = 100$)
RAND	670,000 ± 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 !!



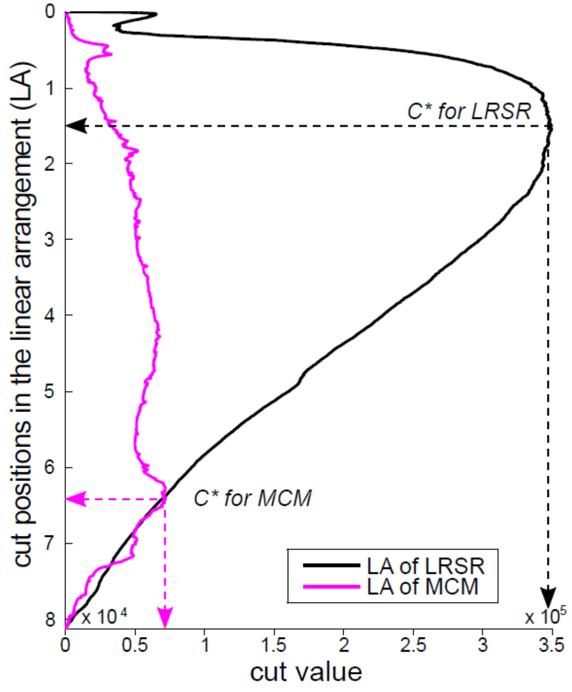
(a) high resource budget: $r = 20,000$



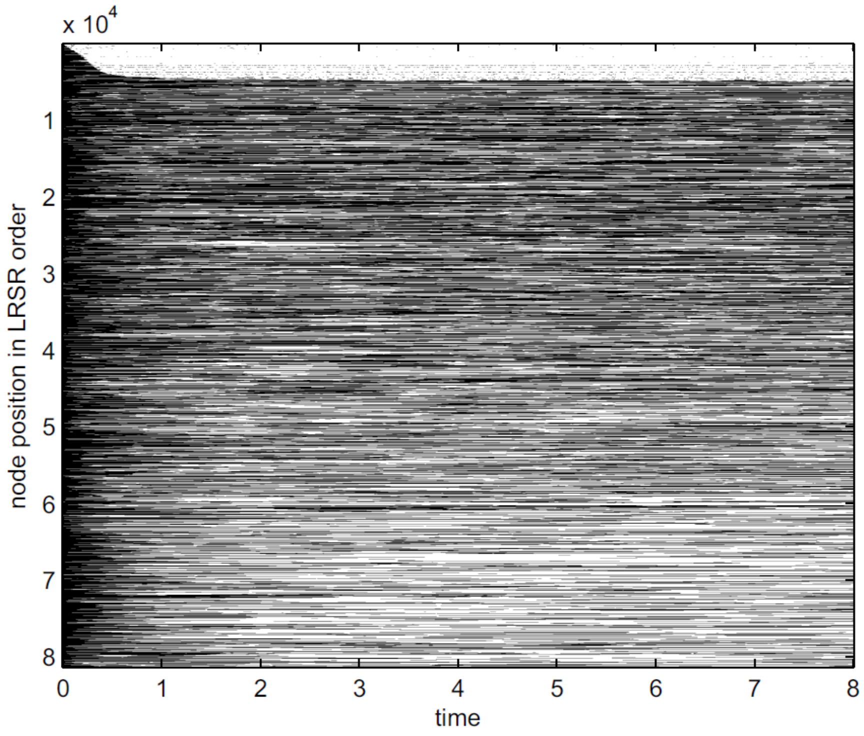
(b) low resource budget: $r = 12,000$

GLOBAL PRIORITY PLANNING

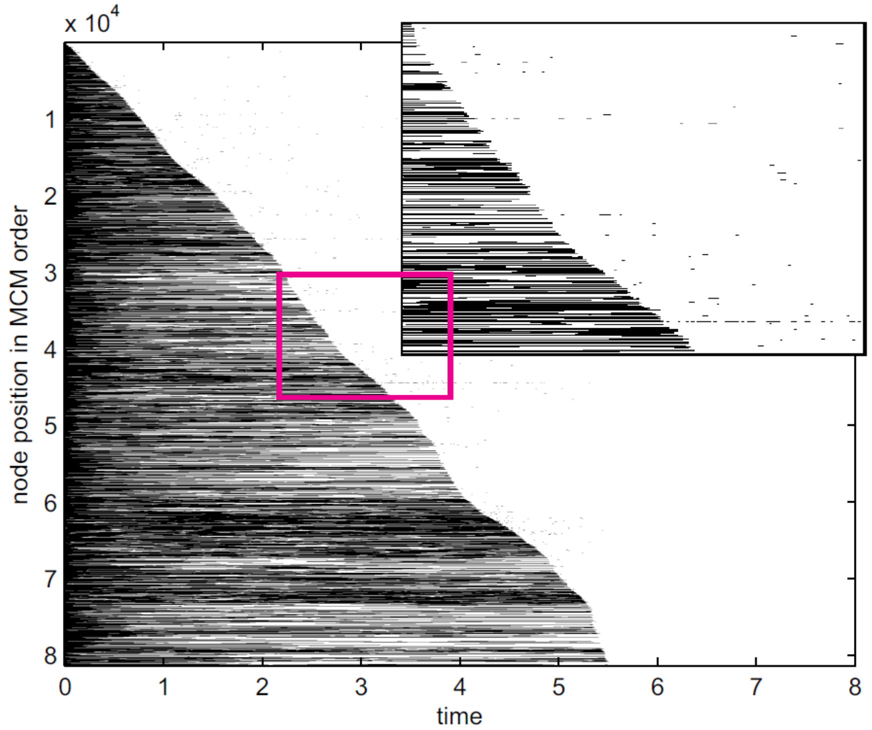
Experiments on real network (TwitterNet)



(c) cuts and maxcuts



(d) network state under LRSR



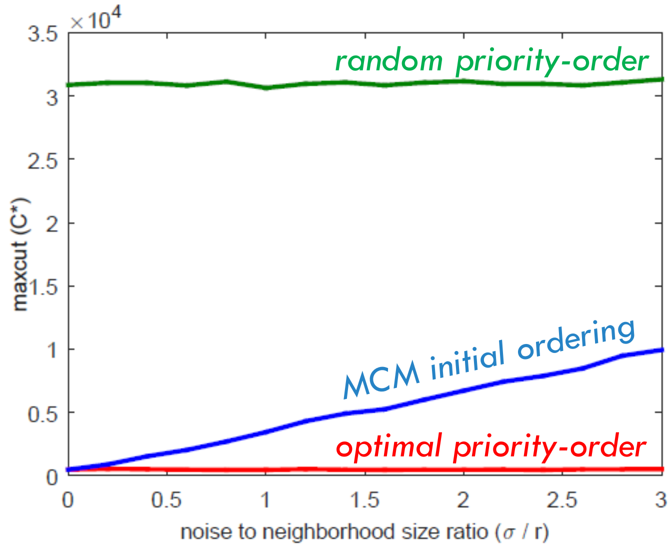
(e) network state under MCM

ROBUSTNESS ANALYSIS

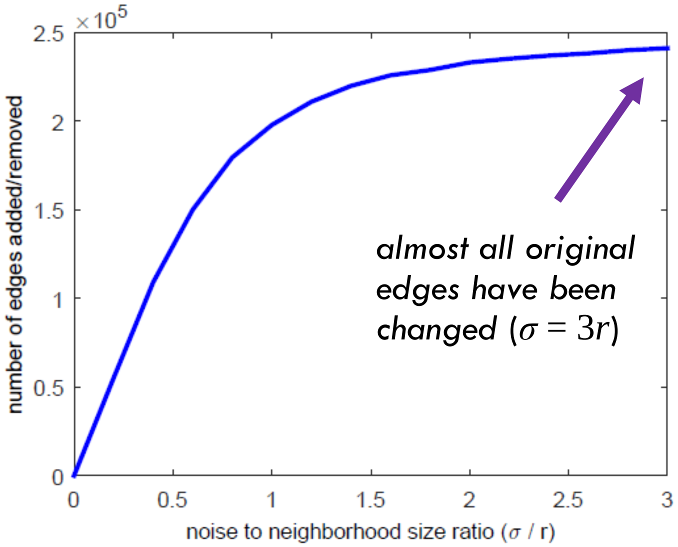
Experiments on an increasingly perturbed contact network

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

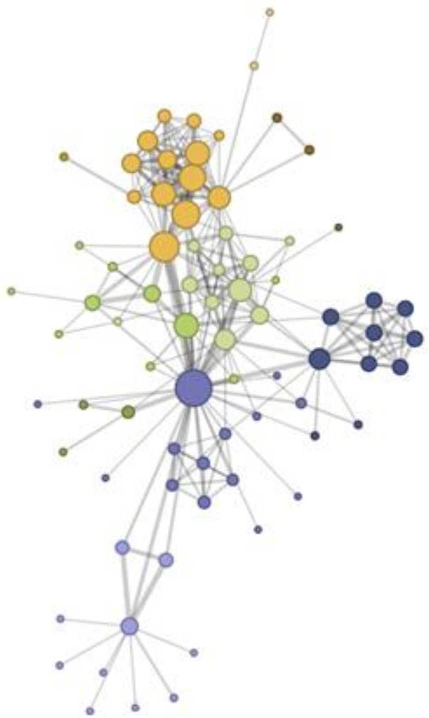
CONCLUSION

- Diffusion processes and control… introduced
  - DPs are super-significant in the new socio-economic context

- Two efficient methods were presented for dynamic resource allocation
  - Computational approaches which can be applied in multiple network resolutions
  - They can be used for epidemic control, MCM also as an assessment tool

a. The **MaxCut** assesses the quality of a plan

b. The **Minimum MaxCut** assess the resource needs of a network



REFERENCES

[1] ***A Greedy Approach for Dynamic Control of Diffusion Processes in Networks**,*
K. Scaman, A. Kalogeratos, and N. Vayatis, to appear in the IEEE International Conference on Tools with Artificial Intelligence (ICTAI), November 9-12, 2015 -- [pdf][suppl. material] .::||::. [software].

[2] ***What Makes a Good Plan? An Efficient Planning Approach to Control Diffusion Processes in Networks**,*
K. Scaman, A. Kalogeratos, and N. Vayatis, CORR arXiv:1407.4760, 18 pages, July 17, 2014 -- [arXiv link][pdf].

[3] ***Suppressing epidemics on Networks with Treatments of Limited Efficiency**,*
K. Scaman, A. Kalogeratos, and N. Vayatis, to appear in IEEE Trans. on Network Science and Engineering, 2015.

[4] ***Learning to Suppress SIS Epidemics in Networks***,
A. Kalogeratos, K. Scaman and N. Vayatis, "Networks in the Social and Information Sciences" workshop in conjunction with Neural Information Processing Systems (NIPS 2015), 2015. [pdf][poster][bib].

[5] ***The Hidden Geometry of Complex, Network-Driven Contagion Phenomena***,
D. Brockmann and D. Helbing, Science, (2013)., 342 (6164), 1337-1342.

[6] ***An efficient curing policy for epidemics on graphs*,**
K. Drakopoulos, A. Ozdaglar, and J.N. Tsitsiklis, IEEE Trans. on Network Science and Engineering, vol. 1, no. 2, pp. 67–75, 2014.

QUESTIONS

Thank you!