Laplacian Spectra of Graphs and Cyber-Insurance Protection

Nabil Kazi-Tani
IECL, Université de Lorraine
En collaboration avec Thierry Cohignac (CCR, Paris)

Séminaire, LIMOS, Saint-Etienne, 26 janvier 2023

Introduction

- A given graph has to be protected against communication interruption, via prevention measures, or via insurance.
- Graph = Company's computer network, susceptible to exterior attacks.

Introduction

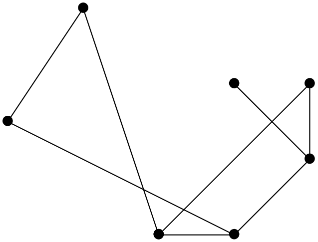
- A given graph has to be protected against communication interruption, via prevention measures, or via insurance.
- Graph = Company's computer network, susceptible to exterior attacks.
- Or... Interconnected data, in the form of a graph $\rightarrow$ susceptible to data breaches, which can be insured.

Introduction

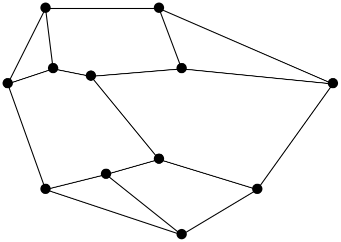
- A given graph has to be protected against communication interruption, via prevention measures, or via insurance.
- Graph = Company's computer network, susceptible to exterior attacks.
- Or... Interconnected data, in the form of a graph $\rightarrow$ susceptible to data breaches, which can be insured.
- Question: **impact of the topology of the network on the chosen protection?**

Introduction

- **Nodes** can represent: computers, servers, softwares, routers, data sets, etc.
- **Edges** can be: internet connections, distant access to servers, links between softwares etc.



(a) A graph on $n = 7$ vertices, with 8 edges.



(b) A 3-regular graph on $n = 12$ vertices.

Outline of the talk

- The SIS model.
- An optimization criteria: the **algebraic connectivity**.
- Solution and numerical illustrations.

Literature

- *Economics of information security* (Gordon and Loeb 2002, Böhme 2013).
- Insurability of cyber risk (Eling 2020's)
- Cyber claims pricing (Fahrenwaldt & al 2018, Xu and Hua 2019)
- Impact of the network's topology (Hillairet & al 2021, Hillairet and Lopez 2021)
- Statistical inference (Bessy-Rolland & al 2020 , Farkas & al 2021)
- And in practice ? (Romanosky &al 2019, Malavasi & al 2021)
- Eigenvalues optimization, epidemic models (refs. in the paper).

Outline of the talk

- **The SIS model**.
- An optimization criteria: the algebraic connectivity.
- Solution and numerical illustrations.

Inhomogeneous Suseptible-Infected-Susceptible (SIS) model on Edges.

- State at time t of edge i is Bernoulli: $X_i(t) = 0$ when edge i is **healthy** and $X_i(t) = 1$ when edge i is infected.

Inhomogeneous Suseptible-Infected-Susceptible (SIS) model on Edges.

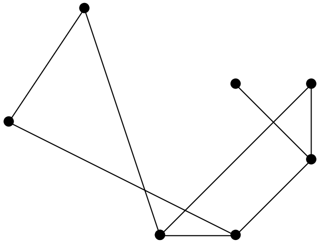
- State at time t of edge i is Bernoulli: $X_i(t) = 0$ when edge i is **healthy** and $X_i(t) = 1$ when edge i is infected.
- When infected, edge i can infect each edge j sharing a common vertex with rate $\beta_{ij} > 0$.

Inhomogeneous Suseptible-Infected-Susceptible (SIS) model on Edges.

- State at time t of edge i is Bernoulli: $X_i(t) = 0$ when edge i is **healthy** and $X_i(t) = 1$ when edge i is infected.
- When infected, edge i can infect each edge j sharing a common vertex with rate $\beta_{ij} > 0$.
- Each edge i can be cured, i.e. its state X_i jumps from 1 to 0 at rate $\delta_i > 0$.
- The Poisson infection and curing processes are assumed independent.

Epidemic model

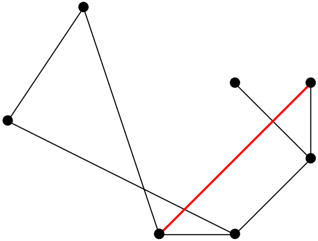
Inhomogeneous Suseptible-Infected-Susceptible (SIS) model.



No edge is infected.

Epidemic model

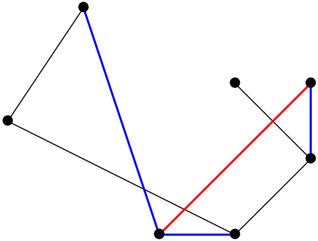
Inhomogeneous Suseptible-Infected-Susceptible (SIS) model.



Red edge i is infected, and can be cured at rate $\delta_i > 0$.

Epidemic model

Inhomogeneous Suseptible-Infected-Susceptible (SIS) model.



Edge i infects each blue edge j at rate $\beta_{ij} > 0$.

Epidemic model

- We use 2 simplifying approximations:

Epidemic model

- We use 2 simplifying approximations:
- N-Intertwined **Mean Field Approximation** (Kooij, Omic, Van Mieghem 2008): upper bounds the infection probabilities.

Epidemic model

- We use 2 simplifying approximations:
- *N*-Intertwined **Mean Field Approximation** (Kooij, Omic, Van Mieghem 2008): upper bounds the infection probabilities.
- **Long time steady state**: which is accurate when infection probabilities are high, corresponding to **cascading infections**.

Epidemic model - Approximations

Let $\widetilde{a}$ = edge adjacency matrix and m = number of edges.

$$\frac{X_i(t+\Delta t) - X_i(t)}{\Delta t} = (1 - X_i(t)) \sum_{j=1}^{m} \tilde{a}_{ij} \beta_{ij} X_j(t) - \delta_i X_i(t).$$

Epidemic model - Approximations

Let $\widetilde{a}$ = edge adjacency matrix and m = number of edges.

$$\frac{X_i(t+\Delta t) - X_i(t)}{\Delta t} = (1 - X_i(t)) \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}X_j(t) - \delta_i X_i(t).$$

leads to

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t) - \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}\mathbb{E}[X_i(t)X_j(t)] - \delta_i v_i(t),$$

where $v_i(t) = \mathbb{P}(X_i(t) = 1)$.

Epidemic model - Approximations

$v_i(t) = \mathbb{P}(X_i(t) = 1)$ and

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t) - \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}\mathbb{E}[X_i(t)X_j(t)] - \delta_i v_i(t).$$

- System of m coupled ODEs.

Epidemic model - Approximations

$v_i(t) = \mathbb{P}(X_i(t) = 1)$ and

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t) - \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}\mathbb{E}[X_i(t)X_j(t)] - \delta_i v_i(t).$$

- System of m coupled ODEs.
- System of ~~m~~ $2^m - 1$ coupled ODEs.
- Curse of dimensionality!

Epidemic model - Approximations

N-Intertwined Mean Field Approximation:

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t) - \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}\mathbb{E}[X_i(t)]\mathbb{E}[X_j(t)] - \delta_i v_i(t),$$

Epidemic model - Approximations

N-Intertwined Mean Field Approximation:

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t) - \sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}\mathbb{E}[X_i(t)]\mathbb{E}[X_j(t)] - \delta_i v_i(t),$$

i.e

$$\frac{dv_i(t)}{dt} = \left(\sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t)\right)(1 - v_i(t)) - \delta_i v_i(t), \quad i = 1, \ldots, m.$$

Epidemic model - Approximations

N-Intertwined Mean Field Approximation and **Long time steady state**:

$$\frac{dv_i(t)}{dt} = 0 = \left(\sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t)\right)(1 - v_i(t)) - \delta_i v_i(t), \quad i = 1, \ldots, m.$$

Epidemic model - Approximations

N-Intertwined Mean Field Approximation and **Long time steady state**:

$$\frac{dv_i(t)}{dt} = 0 = \left(\sum_{j=1}^{m} \tilde{a}_{ij}\beta_{ij}v_j(t)\right)(1 - v_i(t)) - \delta_i v_i(t), \quad i = 1, \ldots, m.$$

i.e.

$$v_i = \frac{\beta \sum_{j=1}^{m} \tilde{a}_{ij} v_j}{\delta_i + \beta \sum_{j=1}^{m} \tilde{a}_{ij} v_j}, \quad i = 1, \ldots m,$$

where $\beta_{ij} = \beta$.

Optimization Problem

- Control the vector of curing rates $(\delta_1, \ldots, \delta_m)$ to make the graph as connected as possible, in a situation of cascading infections.

Optimization Problem

- Control the vector of curing rates $(\delta_1, \ldots, \delta_m)$ to make the graph as connected as possible, in a situation of cascading infections.
- *Stand Alone Cyber Contracts*: clauses of repairing actions, through the intervention of a specified cyber security company.
- Prevention or *cyber hygiene* measures: back-ups, users training, network mapping and inventory etc.

Optimization Problem

- Control the vector of curing rates $(\delta_1, \dots, \delta_m)$ to make the graph as connected as possible, in a situation of cascading infections.
- *Stand Alone Cyber Contracts*: clauses of repairing actions, through the intervention of a specified cyber security company.
- Prevention or *cyber hygiene* measures: back-ups, users training, network mapping and inventory etc.
- How should we measure the "connectedness" of a given graph G?

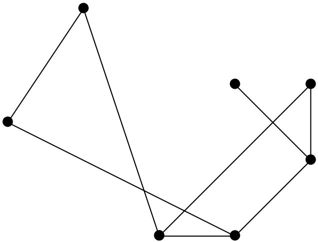
Outline of the talk

- The SIS model.
- **An optimization criteria: the algebraic connectivity.**
- Solution and numerical illustrations.

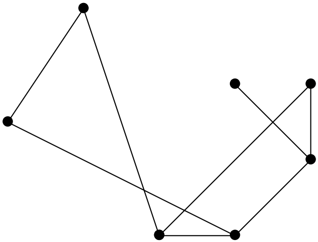
Optimization Criteria

- Given a graph G with adjacency matrix A and degree matrix D, the Laplacian matrix of G is given by $\mathrm{L} := \mathrm{D} - \mathrm{A}$.
- For $x \in \mathbb{R}^n$ (n= number of nodes), $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$.

Laplacian matrix: example

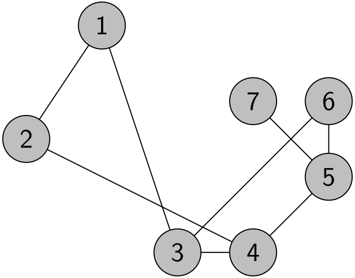


Laplacian matrix: example

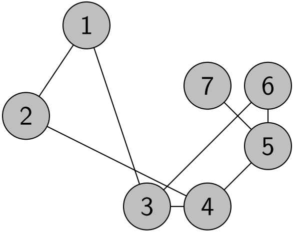


Choose a numbering of the nodes.

Laplacian matrix: example



Laplacian matrix: example



$$L = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

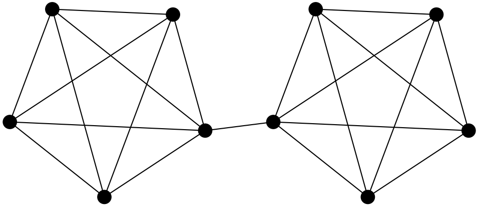
Optimization Criteria

- Given a graph G with adjacency matrix A and degree matrix D, the Laplacian matrix of G is given by $\mathrm{L} := \mathrm{D} - \mathrm{A}$.
- For $x \in \mathbb{R}^n$ (n= number of nodes), $x^T Lx = \sum_{(i,j)\in E}(x_i - x_j)^2$.

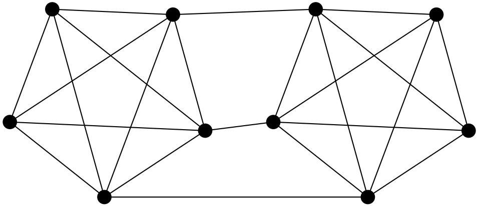
Optimization Criteria

- Given a graph G with adjacency matrix A and degree matrix D, the Laplacian matrix of G is given by $\mathrm{L} := \mathrm{D} - \mathrm{A}$.
- For $x \in \mathbb{R}^n$ (n= number of nodes), $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$.
- **Algebraic connectivity** $\lambda_2(G)$= lowest strictly positive eigenvalue of the Laplacian matrix.
- Known to describe the "connectedness" of a graph (Fiedler 1973).

Optimization Criteria



(a) A graph G, such that $\lambda_2(G) = 0.298$.



(b) A graph $\widetilde{G}$, such that $\lambda_2(\widetilde{G}) = 1$.

Figure: The lower graph $\widetilde{G}$ is "more connected" than the upper graph G.

Optimization Criteria

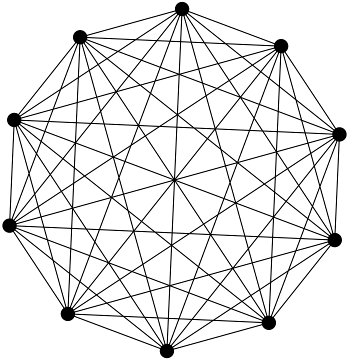


Figure: Complete graph K_{10} with $\lambda_2(K_{10}) = 10$.

Optimization Problem

1. Consider a SIS epidemic model on the edges of a graph G, for a long time.

Optimization Problem

1. Consider a SIS epidemic model on the edges of a graph G, for a long time.
2. We obtain a random graph $\widetilde{G}_\infty$, with less edges.

Optimization Problem

1. Consider a SIS epidemic model on the edges of a graph G, for a long time.
2. We obtain a random graph $\widetilde{G}_\infty$, with less edges.
3. Consider the associated random Laplacian matrix L_∞ and its average $L := \mathbb{E}[L_\infty]$.

Optimization Problem

1. Consider a SIS epidemic model on the edges of a graph G, for a long time.
2. We obtain a random graph $\widetilde{G}_\infty$, with less edges.
3. Consider the associated random Laplacian matrix L_∞ and its average $L := \mathbb{E}[L_\infty]$.
4. Maximize the algebraic connectivity of L:

$$\begin{aligned} \underset{\delta \in \mathcal{A}}{\text{maximize}} \quad & \lambda_2(L) \\ \text{subject to} \quad & c(\delta) \leq B, \end{aligned}$$

where c is a cost function, B a given budget constraint, and $\mathcal{A}$ an admissible set of curing rates.

Budget constraint

1. Edge i interrupted $\rightarrow$ loss Z_i with $\mathbb{E}[Z_i] = c_i$, with $(Z_1, \ldots, Z_m)$ independent and independent of the underlying SIS process.
2. $W(\delta) =$ total loss on the graph, when the curing rates are given by $\delta \in \mathbb{R}^m$.
3. $\delta_0 =$ curing rates before any insurance or prevention.

Budget constraint

1. Edge i interrupted $\rightarrow$ loss Z_i with $\mathbb{E}[Z_i] = c_i$, with $(Z_1, \ldots, Z_m)$ independent and independent of the underlying SIS process.
2. $W(\delta) =$ total loss on the graph, when the curing rates are given by $\delta \in \mathbb{R}^m$.
3. $\delta_0 =$ curing rates before any insurance or prevention.
4. The insurance cost function $c : \mathbb{R}_+^m \rightarrow \mathbb{R}_+$ is given by

$$c(\delta) := \varphi \left\{ \mathbb{E}[W(\delta_0)] - \mathbb{E}[W(\delta)] \right\}, \quad (1)$$

where $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is non decreasing.

Outline of the talk

- The SIS model.
- An optimization criteria: the algebraic connectivity.
- **Solution and numerical illustrations.**

A Solution

Theorem

An optimal solution is given by

$$\delta_i^\star = \left(\beta \sum_{j=1}^{m} \tilde{a}_{ij} v_j^\star\right) \frac{1 - v_i^\star}{v_i^\star},$$

where $(1 - v_1^\star, \ldots, 1 - v_m^\star)$ is the w component of a solution to the following convex problem in $\gamma \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $w \in \mathbb{R}^m$:

A Solution

Theorem

An optimal solution is given by

$$\delta_i^\star = \left(\beta \sum_{j=1}^{m} \tilde{a}_{ij} v_j^\star \right) \frac{1 - v_i^\star}{v_i^\star},$$

where $(1 - v_1^\star, \ldots, 1 - v_m^\star)$ is the w component of a solution to the following convex problem in $\gamma \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $w \in \mathbb{R}^m$:

$$\begin{aligned} \sup \quad & \gamma \\ \text{subject to} \quad & L(w) - \gamma I + \mu 1 1^t \succcurlyeq 0 \\ & 0 \leq w \leq 1 \\ & \sum_{j=1}^{m} c_j w_j \leq B, \end{aligned}$$

where $1^t := (1, \ldots, 1) \in \mathbb{R}^n$, $L(w)$ denotes the Laplacian matrix of the weighted graph (V, E_0, w), $M \succcurlyeq 0$ means that M is positive semidefinite and $(\tilde{a}_{ij})_{a \leq i,j \leq m}$ are the entries of the edge-adjacency matrix.

Idea

Reformulate the problem on δ as a problem on v!

Lemma 1

$L = \mathbb{E}[L_\infty]$ is the Laplacian matrix associated to the weighted graph $G = (V, E_0, \overline{v})$, where each edge ℓ is given the weight $\overline{v}_\ell = 1 - v_\ell$, with $(v_1, \ldots, v_m)$ satisfying the **steady state equation**.

Idea

Reformulate the problem on δ as a problem on v!

Lemma 1

$L = \mathbb{E}[L_\infty]$ is the Laplacian matrix associated to the weighted graph $G = (V, E_0, \overline{v})$, where each edge ℓ is given the weight $\overline{v}_\ell = 1 - v_\ell$, with $(v_1, \ldots, v_m)$ satisfying the **steady state equation**.

Lemma 2

$$\mathbb{E}[W(\delta_0)] - \mathbb{E}[W(\delta)] = \sum_{j=1}^{m} c_j \overline{v}_j.$$

In particular, the cost function c given in (1) is a convex function of v.

Examples

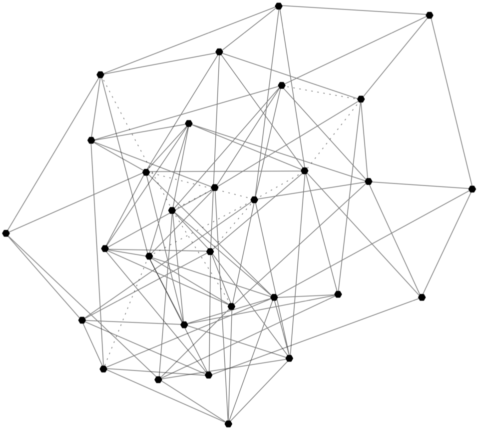


Figure: Erdös-Rényi Random Graph with 102 edges. The dotted edges have optimal δ equal to 0.

Solved with the **CVX Matlab Package (Boyd, Grant)**.

Examples

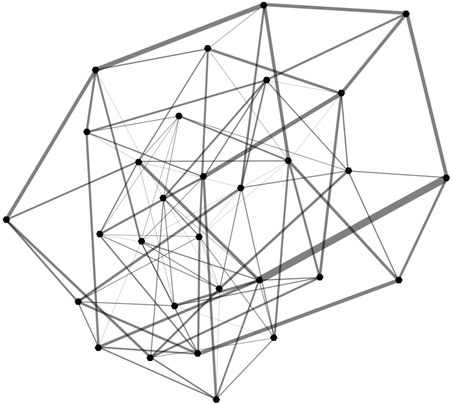


Figure: Each edge width is proportional to the optimal δ.

Examples

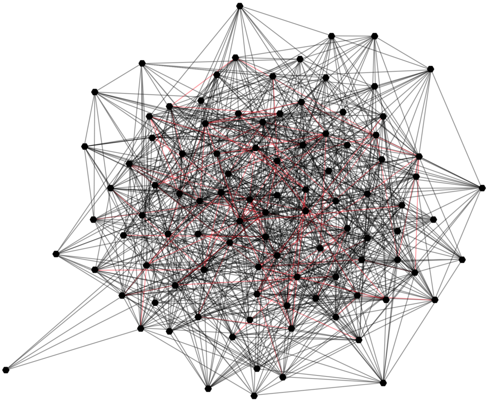


Figure: Erdös-Rényi Random Graph with 1004 edges. Red edges have null $\delta^\star$ values.

Examples

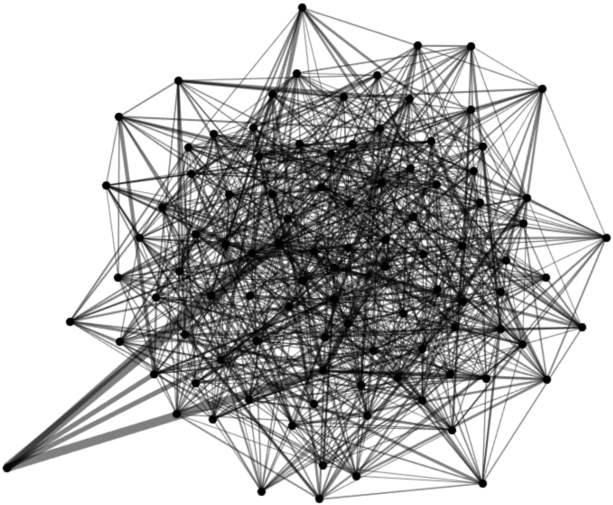


Figure: Each edge width is proportional to the optimal δ.

Examples

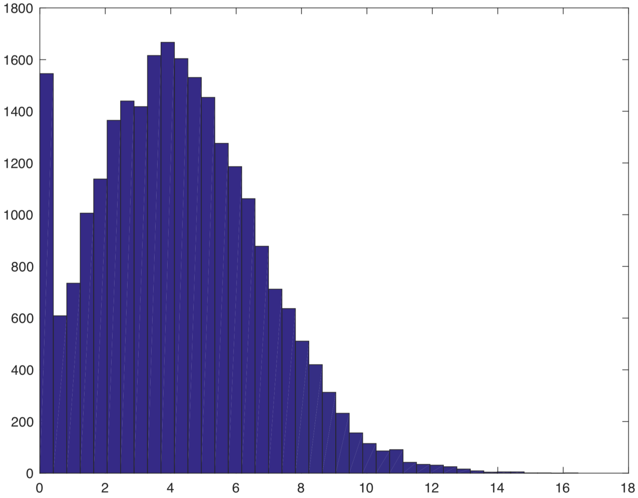


Figure: $\delta^\star$ values, for a graph with 24978 edges.

Perspectives

- Finite horizon problem and dynamic optimization: stochastic control problem! (Avoids both NIMFA and Steady state approximations)

Perspectives

- Finite horizon problem and dynamic optimization: stochastic control problem! (Avoids both NIMFA and Steady state approximations)
- Inhomogeneous fast mixing Markov chains (Boyd, Diaconis, Sun, Xiao).

Perspectives

- Finite horizon problem and dynamic optimization: stochastic control problem! (Avoids both NIMFA and Steady state approximations)
- Inhomogeneous fast mixing Markov chains (Boyd, Diaconis, Sun, Xiao).
- Simulation and (optimal) importance sampling for Markov chains.
- Optimal stochastic control and games on networks.

Merci de votre attention