Sparsity-Promoting Optimal Control of Distributed Systems

Mihailo Jovanović
www.umn.edu/∼mihailo

joint work with:
Makan Fardad
Fu Lin

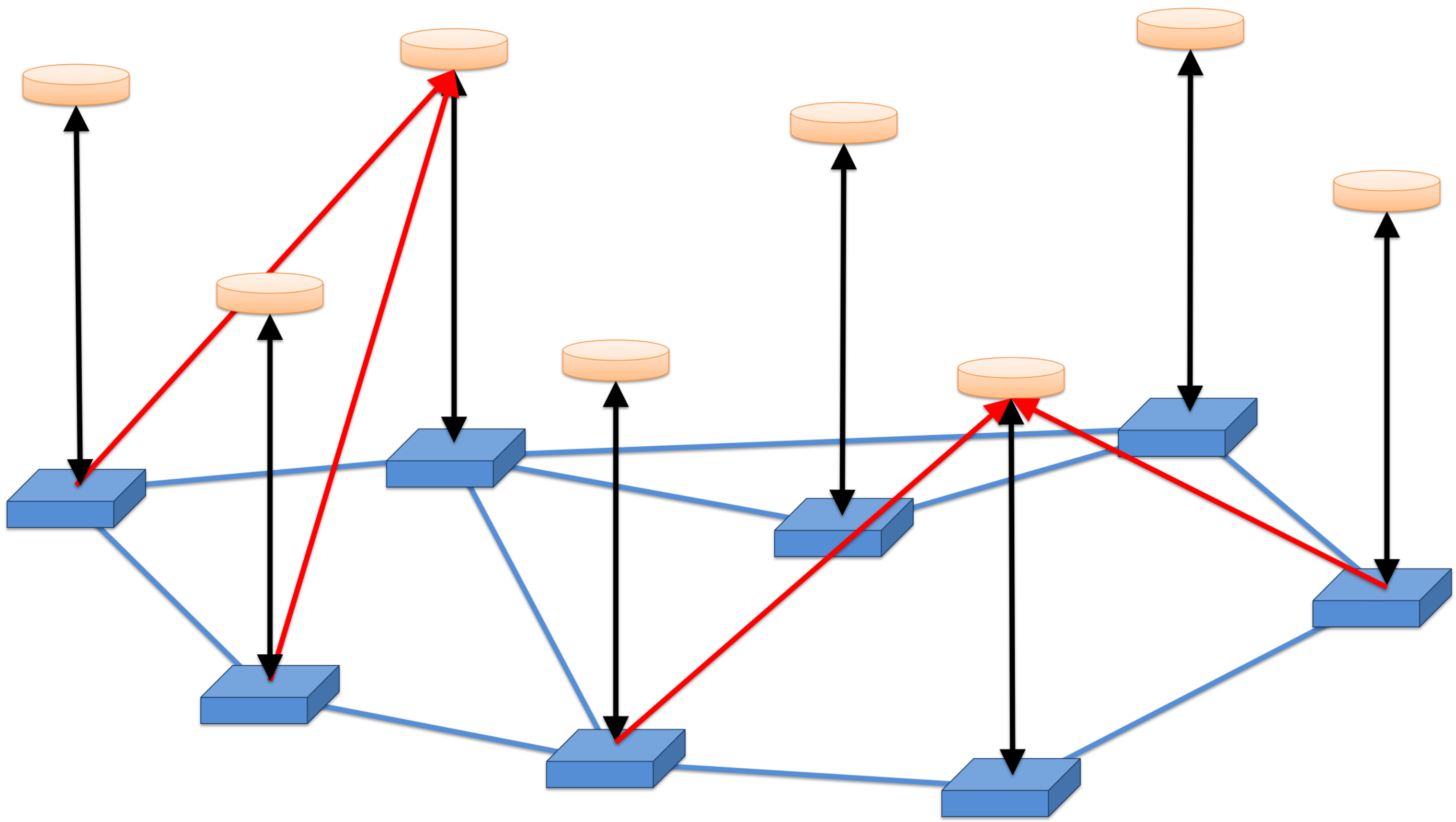
UNIVERSITY OF MINNESOTA

Banff Workshop on Decentralized Systems; Oct. 14, 2014

Structured distributed control

- **Blue layer:** **distributed plant and its interaction links**

structured memoryless controller



KEY CHALLENGE:

identification of a signal exchange network

performance vs **sparsity**

Minimum variance state-feedback problem

dynamics: $\dot{x} \;=\; A\,x \;+\; B_1\,d \;+\; B_2\,u$

objective function: $J \;=\; \lim_{t \,\to\, \infty} \mathcal{E}\left(x^T(t)\,Q\,x(t) \;+\; u^T(t)\,R\,u(t)\right)$

memoryless controller: $u \;=\; -F\,x$

- CLOSED-LOOP VARIANCE AMPLIFICATION

$$J(F) \;=\; \mathrm{trace}\left(\int_0^\infty \mathrm{e}^{(A \,-\, B_2 F)^T t}\left(Q \;+\; F^T R\, F\right) \mathrm{e}^{(A \,-\, B_2 F) t}\,\mathrm{d}t\; B_1 B_1^T\right)$$

⋆ **no structural constraints**

globally optimal controller:

$$A^T P \;+\; P\,A \;-\; P\,B_2\,R^{-1}B_2^T P \;+\; Q \;=\; 0$$
$$F_c \;=\; R^{-1}B_2^T P$$

SDP formulation

$$\begin{array}{ll} \underset{X,\,F}{\text{minimize}} & \text{trace}\big((Q + F^T R\, F)\, X\big) \\ \text{subject to} & (A - B_2 F)\, X + X\, (A - B_2 F)^T = -B_1 B_1^T \\ & X \succ 0 \end{array}$$

- CHANGE OF VARIABLES: $FX = Y$

$$\begin{array}{ll} \underset{X,\,Y \succ 0}{\text{minimize}} & \text{trace}\,(Q\, X) + \text{trace}\big(R\, Y\, X^{-1}\, Y^T\big) \\ \text{subject to} & (A\, X - B_2\, Y) + (A\, X - B_2\, Y)^T + B_1 B_1^T = 0 \end{array}$$

SDP characterization:

$$\begin{array}{ll} \underset{X,\,Y,\,Z}{\text{minimize}} & \text{trace}\,(Q\, X) + \text{trace}\,(R\, Z) \\ \text{subject to} & (A\, X - B_2\, Y) + (A\, X - B_2\, Y)^T + B_1 B_1^T \preceq 0 \\ & \begin{bmatrix} Z & Y \\ Y^T & X \end{bmatrix} \succeq 0 \end{array}$$

- STRUCTURAL CONSTRAINTS $F \in \mathcal{S}$

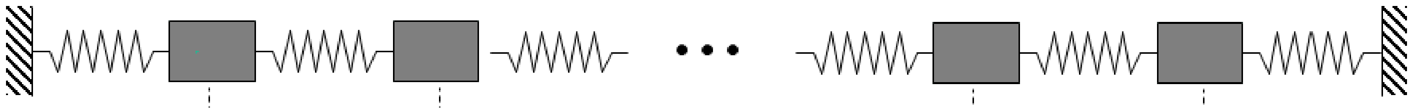
centralized **fully decentralized** **localized**

$$\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \qquad \begin{bmatrix} * & & & \\ & * & & \\ & & * & \\ & & & * \end{bmatrix} \qquad \begin{bmatrix} * & * & & \\ * & * & * & \\ & * & * & * \\ & & * & * \end{bmatrix}$$

CHALLENGE:

convex characterization in the face of structural constraints

An example



$$u(t) \;=\; -\begin{bmatrix} F_p & F_v \end{bmatrix} \begin{bmatrix} p(t) \\ v(t) \end{bmatrix}$$

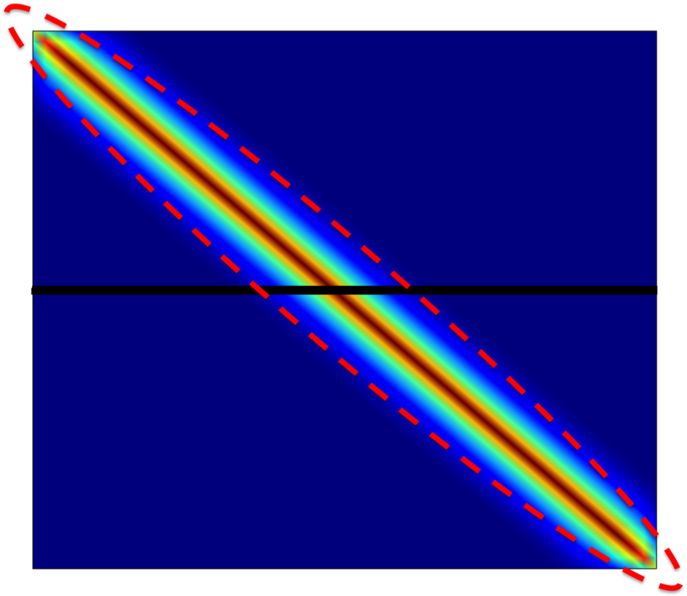
- **Objective:** **design** $\begin{bmatrix} F_p & F_v \end{bmatrix}$ **to minimize steady-state variance of** p, v, u

OPTIMAL CONTROLLER – LINEAR QUADRATIC REGULATOR

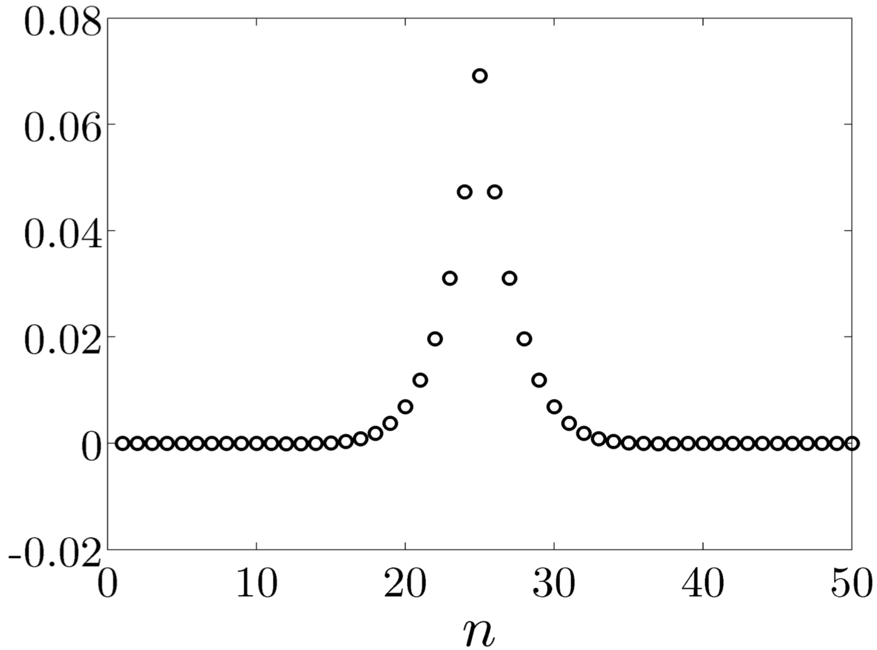
$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \end{bmatrix} = -\underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{F_p} \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \end{bmatrix} - \underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{F_v} \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \\ v_4(t) \end{bmatrix}$$

Structure of optimal controller

position feedback matrix:



position gains for middle mass:



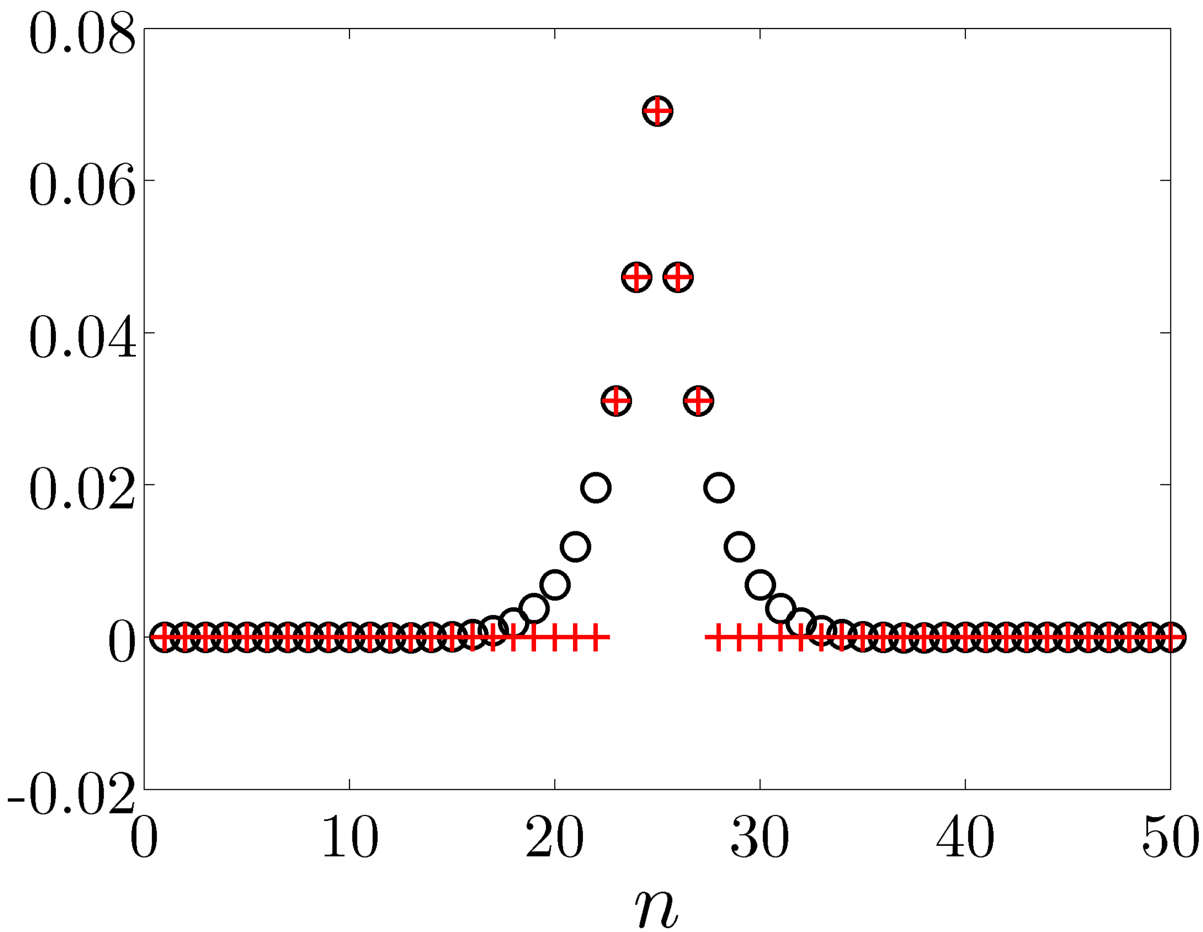
- OBSERVATIONS
  - ⋆ Diagonals almost constant (modulo edges)
  - ⋆ Off-diagonal decay of a centralized gain

Bamieh, Paganini, Dahleh, IEEE TAC '02

Motee & Jadbabaie, IEEE TAC '08

Enforcing localization?

- **One approach: truncating centralized controller**



- POSSIBLE DANGERS
  - ⋆ **Performance degradation**
  - ⋆ **Instability**

Outline

- SPARSITY-PROMOTING OPTIMAL CONTROL
  - ⋆ **identification and design of sparse feedback gains**
  - ⋆ **tools from control theory, optimization, and compressive sensing**
- ALGORITHM
  - ⋆ **Alternating Direction Method of Multipliers**
- EXAMPLES
- SUMMARY AND OUTLOOK

SPARSITY-PROMOTING OPTIMAL CONTROL

Sparsity-promoting optimal control

$$\text{minimize} \quad J(F) \quad + \quad \gamma\, \mathbf{card}\,(F)$$

$J(F)$ ↓ **variance amplification**

$\gamma\, \mathbf{card}\,(F)$ ↓ **sparsity-promoting penalty function**

⋆ $\mathbf{card}\,(F)$ – number of non-zero elements of F

$$F = \begin{bmatrix} 5.1 & -2.3 & 0 & 1.5 \\ 0 & 3.2 & 1.6 & 0 \\ 0 & -4.3 & 1.8 & 5.2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{card}\,(F) = 8$$

⋆ $\gamma > 0$ – performance vs sparsity tradeoff

Fardad, Lin, Jovanović, ACC '11

Lin, Fardad, Jovanović, IEEE TAC '13

Convex relaxations of $\textbf{card}\,(F)$

$$\ell_1 \text{ norm:} \quad \sum_{i,\,j} |F_{ij}|$$

$$\text{weighted } \ell_1 \text{ norm:} \quad \sum_{i,\,j} W_{ij}\,|F_{ij}|, \qquad W_{ij} \;\geq\; 0$$

- CARDINALITY VS WEIGHTED ℓ_1 NORM

$$\{W_{ij} \;=\; 1/|F_{ij}|,\;\; F_{ij} \;\neq\; 0\} \quad\Rightarrow\quad \textbf{card}\,(F) \;=\; \sum_{i,\,j} W_{ij}\,|F_{ij}|$$

RE-WEIGHTED SCHEME

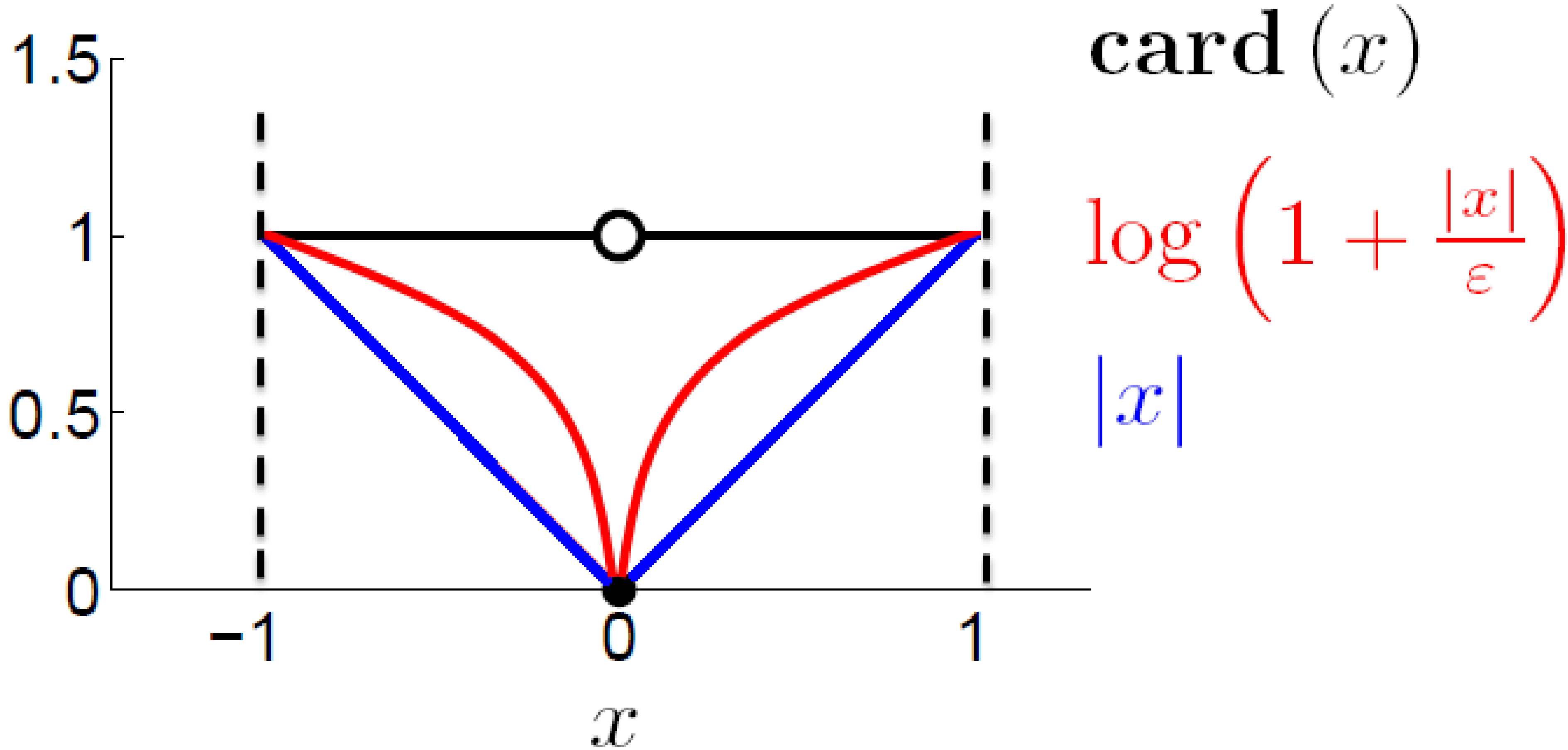
⋆ **Use feedback gains from previous iteration to form weights**

$$W_{ij}^{+} \;=\; \frac{1}{|F_{ij}| \,+\, \varepsilon}$$

Candès, Wakin, Boyd, J. Fourier Anal. Appl. '08

A non-convex relaxation of $\mathbf{card}\,(F)$

sum-of-logs: $$\sum_{i,j} \log\left(1 + \frac{|F_{ij}|}{\varepsilon}\right), \qquad 0 < \varepsilon \ll 1$$

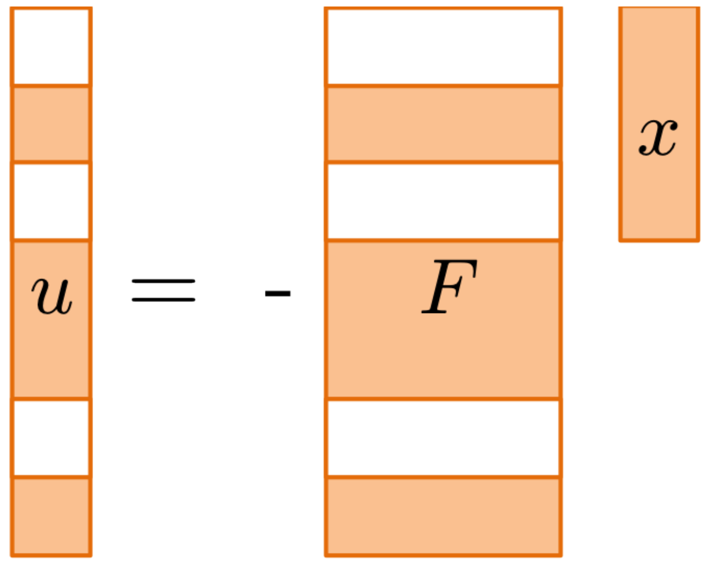


Candès, Wakin, Boyd, J. Fourier Anal. Appl. '08

Classes of convex problems

Optimal actuator/sensor selection

- OBJECTIVE: **identify row-sparse feedback gain**

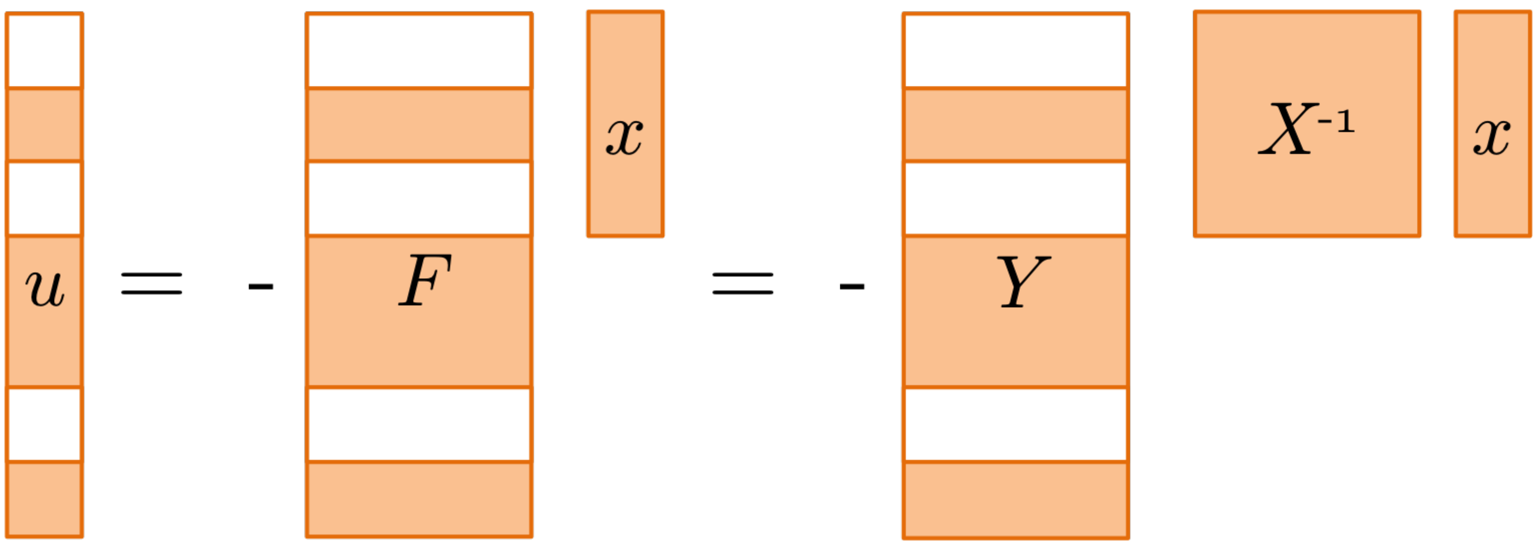


$$\text{minimize} \quad J(F) \quad + \quad \gamma \sum_i \|\mathrm{e}_i^T F\|_2$$

$J(F)$ ↓ **variance amplification**

$\gamma \sum_i \|\mathrm{e}_i^T F\|_2$ ↓ **row-sparsity-promoting penalty function**

- Change of variables: $Y := F X$
  - $\star$ **convex dependence of J on X and Y**
  - $\star$ **row-sparse structure preserved**

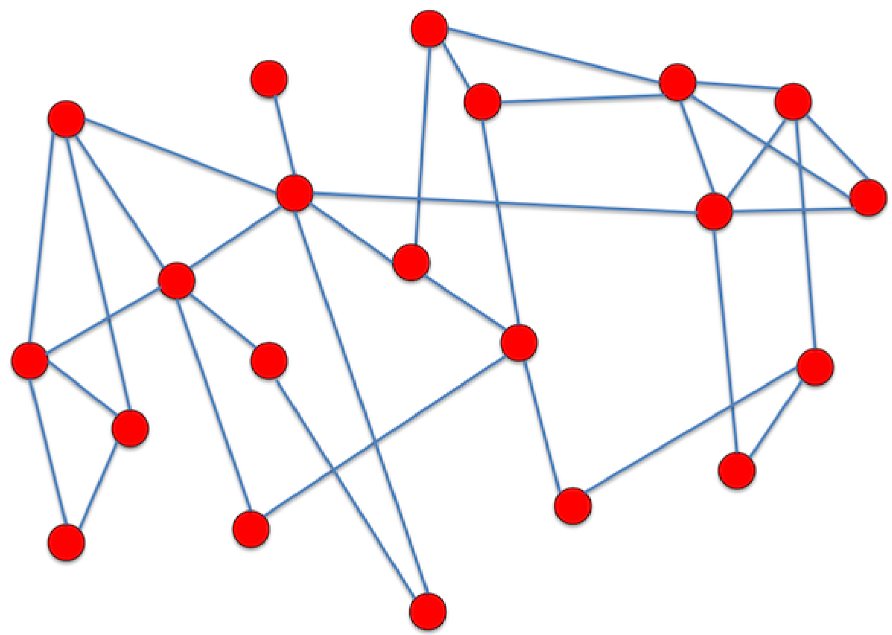


- Optimal control problem
  - $\star$ **admits SDP characterization**

Polyak, Khlebnikov, Shcherbakov, ECC '13

Dhingra, Jovanović, Luo, CDC '14

Consensus by distributed computation



- RELATIVE INFORMATION EXCHANGE WITH NEIGHBORS
  - ⋆ **simplest distributed averaging algorithm**

$$\dot{x}_i(t) \;=\; -\sum_{j\,\in\mathcal{N}_i}\Big(x_i(t)\;-\;x_j(t)\Big)$$

connected network $\Rightarrow$ **convergence to the average value**

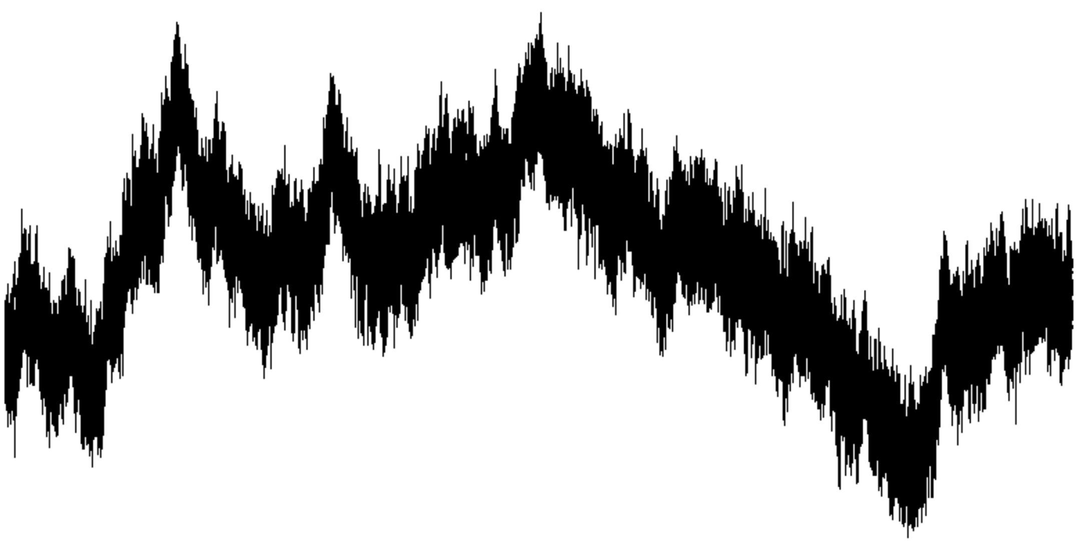
Consensus with stochastic disturbances

$$\dot{x}_i(t) \;=\; -\sum_{j\,\in\,\mathcal{N}_i} \Big(x_i(t) \;-\; x_j(t)\Big) \;\;+\;\; d_i(t)$$

$\downarrow$

white noise

- NETWORK AVERAGE
  - ⋆ **undergoes random walk**



connected network $\Rightarrow$ {
- **each $x_i(t)$ fluctuates around $\bar{x}(t)$**
- **deviation from average:** $\tilde{x}_i(t) \;:=\; x_i(t) \;-\; \bar{x}(t)$

Design of undirected consensus networks

dynamics: $\dot{x} \;=\; d \;+\; u$

control: $u \;=\; -\,F\,x$

objective: $J \;=\; \lim_{t\,\to\,\infty} \mathcal{E}\left(x^T(t)\,Q\,x(t) \;+\; u^T(t)\,R\,u(t)\right)$

SDP characterization:

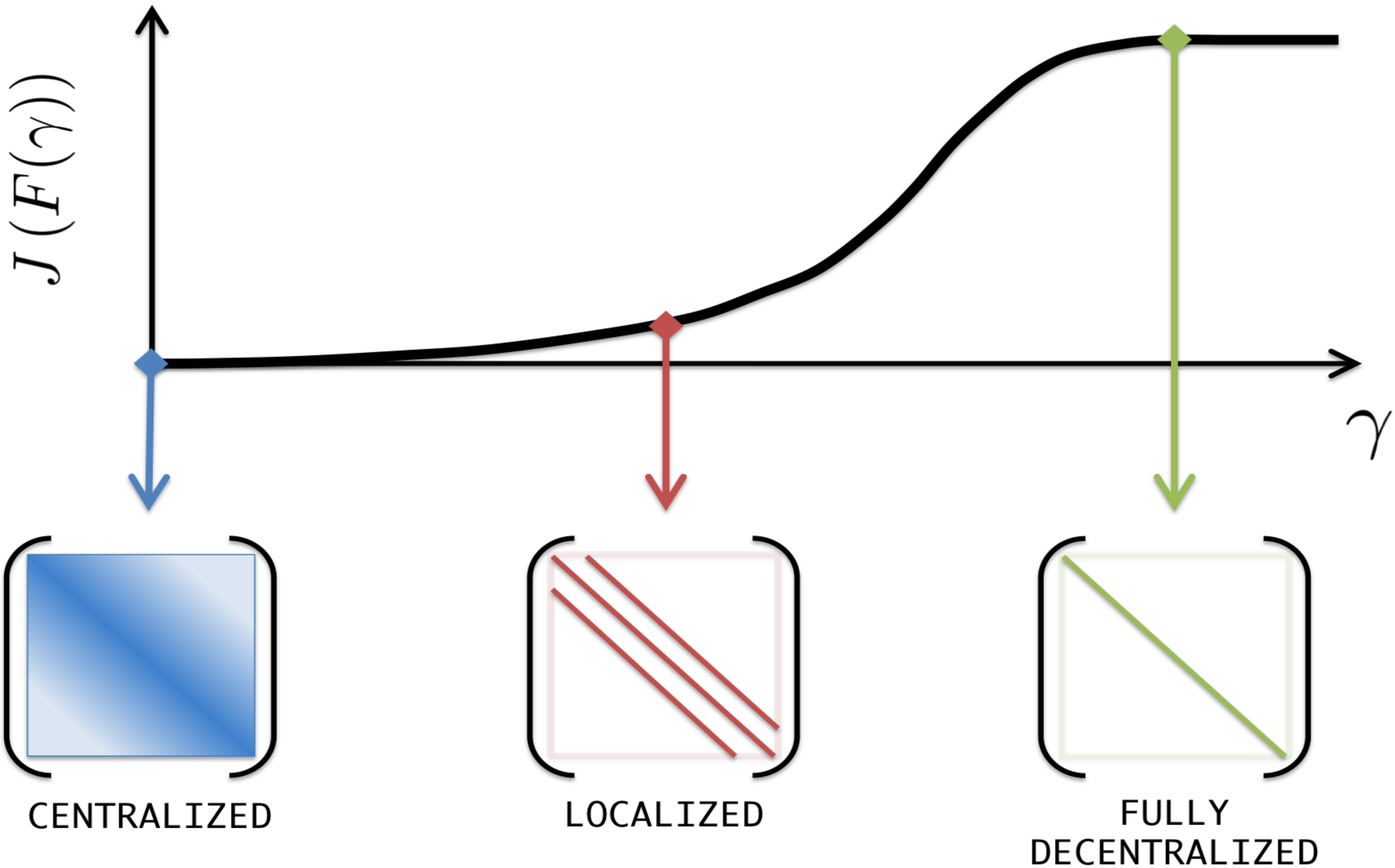
$$\begin{array}{ll} \text{minimize} & \operatorname{trace}\left(X \;+\; R\,F\right) \;+\; \gamma\,\mathbb{1}^T\,Y\,\mathbb{1} \\ \text{subject to} & \begin{bmatrix} X & Q^{1/2} \\ Q^{1/2} & F + \mathbb{1}\mathbb{1}^T/N \end{bmatrix} \succeq \; 0 \\ & -Y_{ij} \;\leq\; W_{ij}\,F_{ij} \;\leq\; Y_{ij} \\ & F\,\mathbb{1} \;=\; 0 \end{array}$$

Lin, Fardad, Jovanović, Allerton '12

Wu & Jovanović, ACC '14

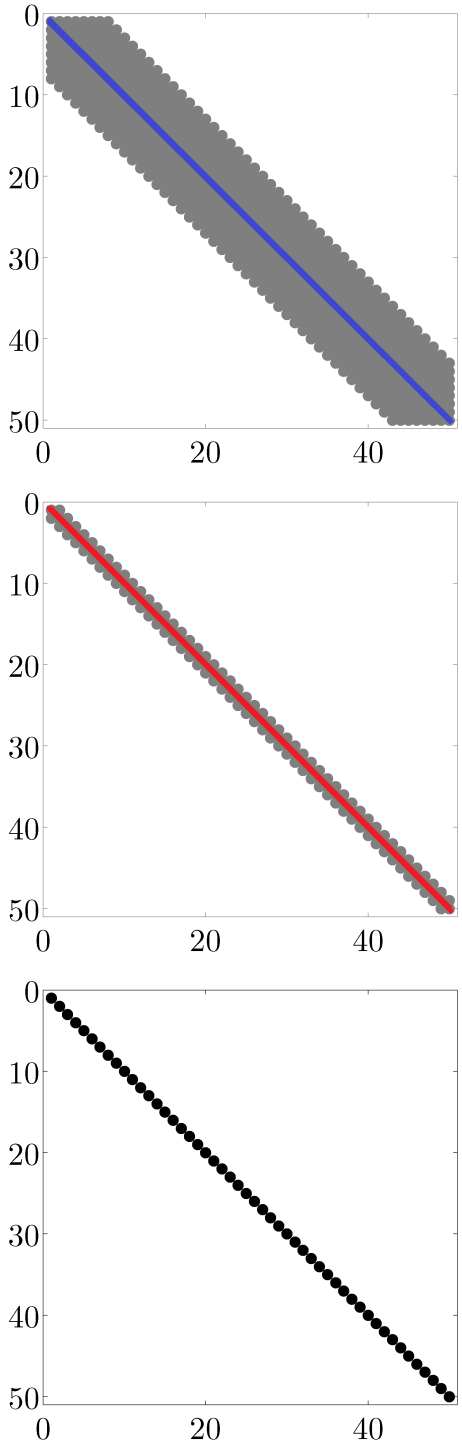
Parameterized family of feedback gains

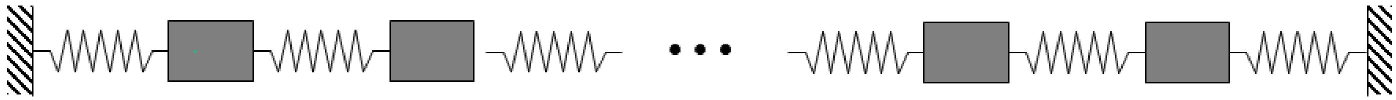
$$F(\gamma) \;:=\; \underset{F}{\operatorname{argmin}} \left(J(F) \;+\; \gamma\, g(F) \right)$$



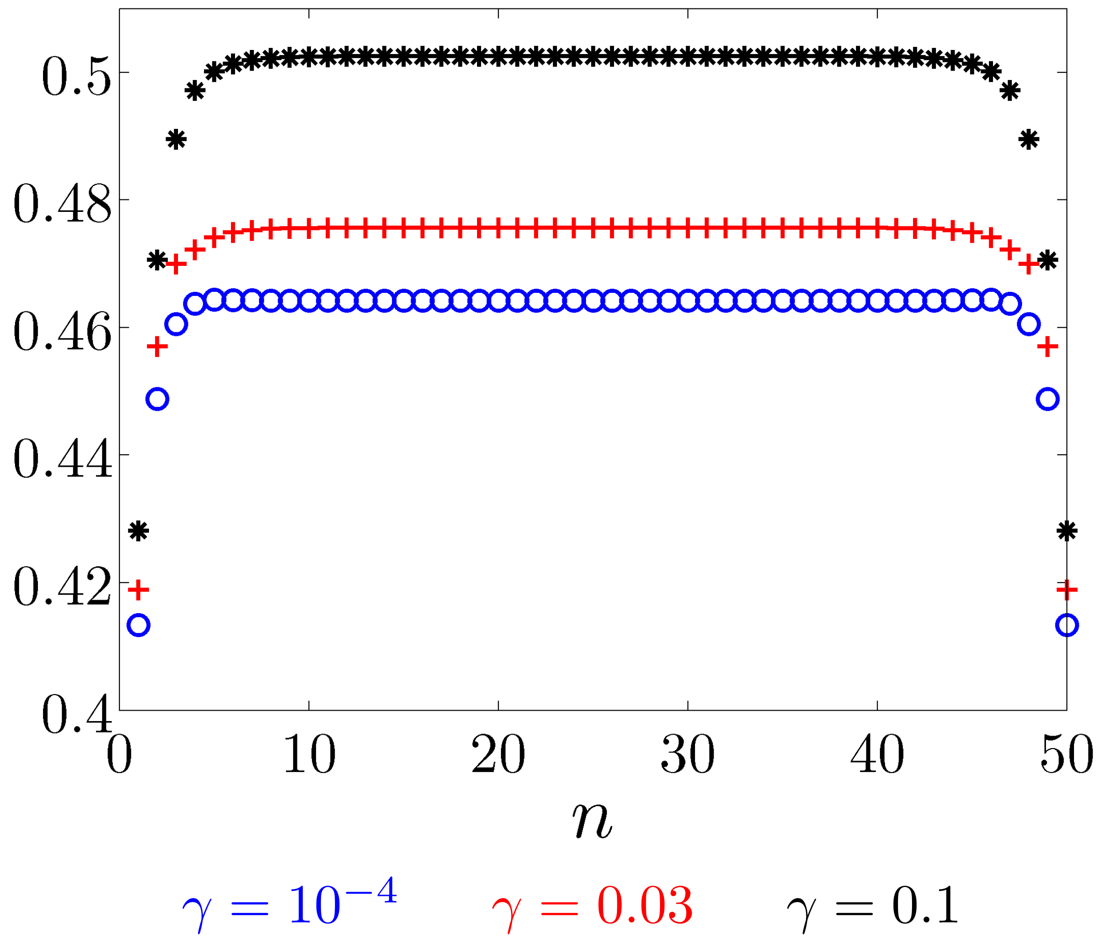
EXAMPLES

Mass-spring system



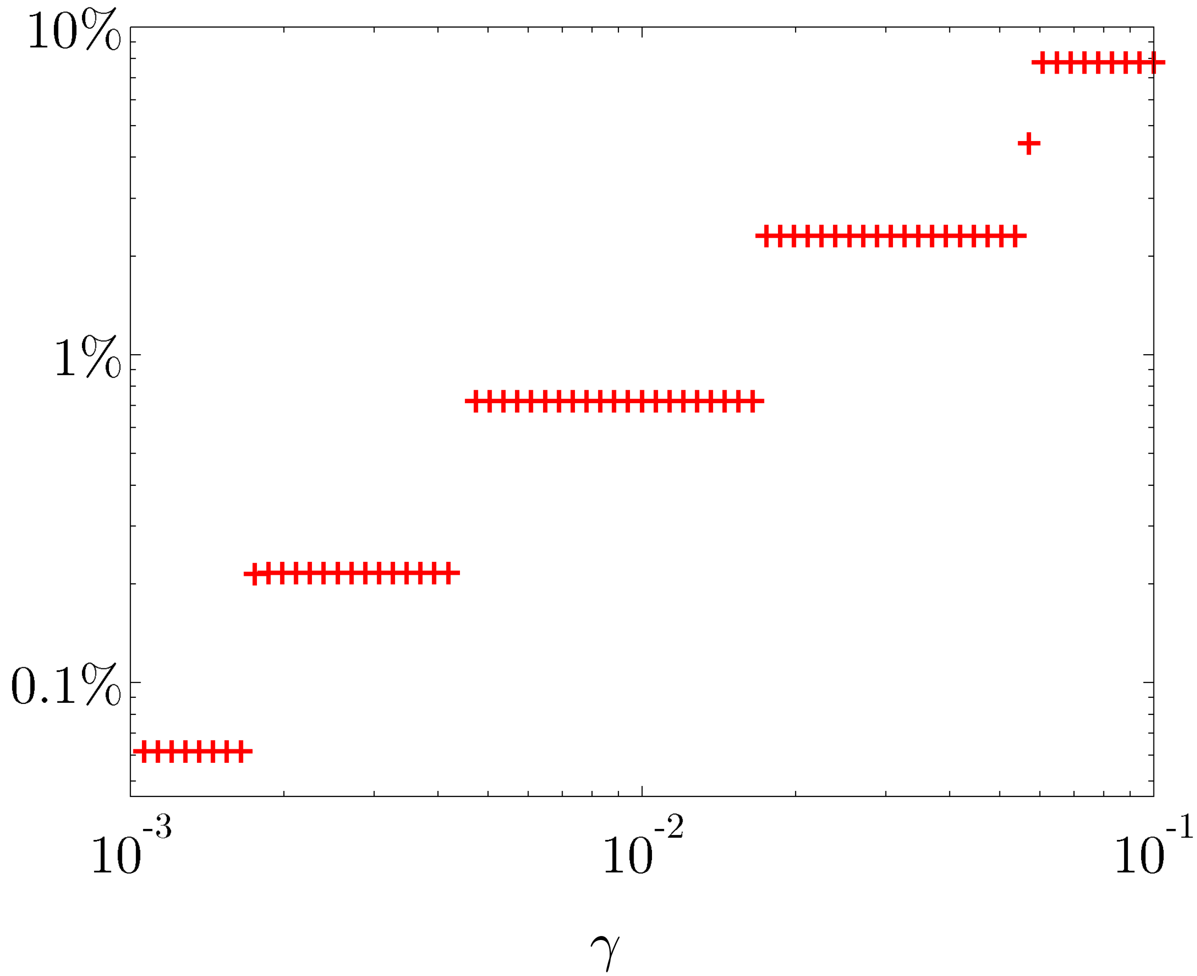


$\mathrm{diag}\,(F_v)$:



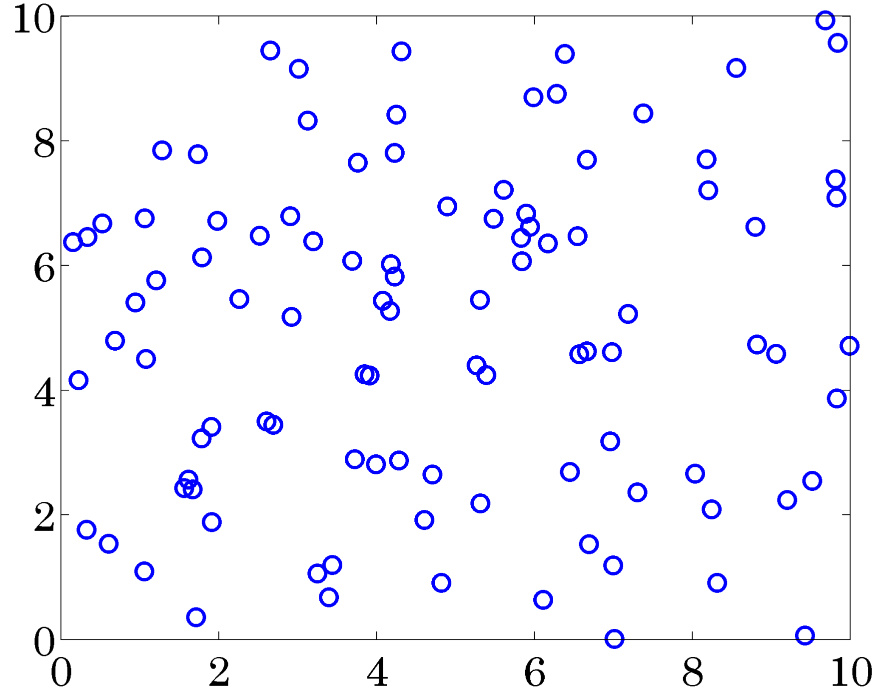
- **Performance comparison: sparse vs centralized**

$(J - J_c)/J_c$:



$\mathbf{card}(F)/\mathbf{card}(F_c)$	$(J - J_c)/J_c$
10%	0.75%
6%	2.4%
2%	7.8%

Network with 100 nodes

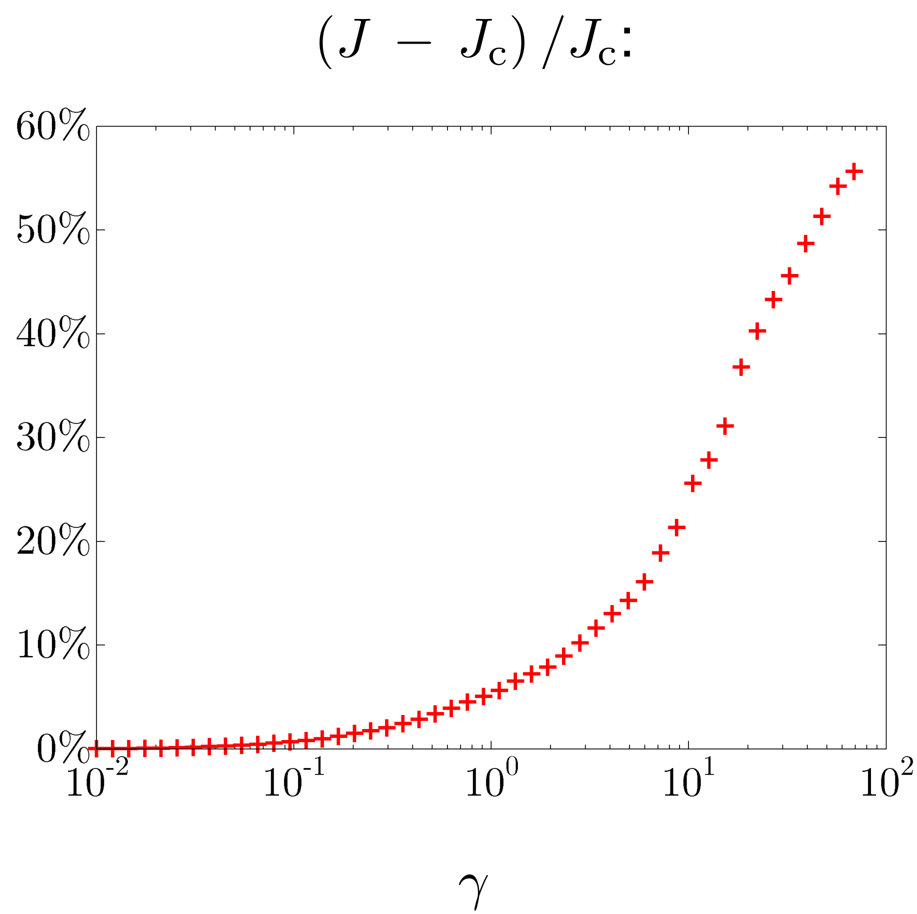


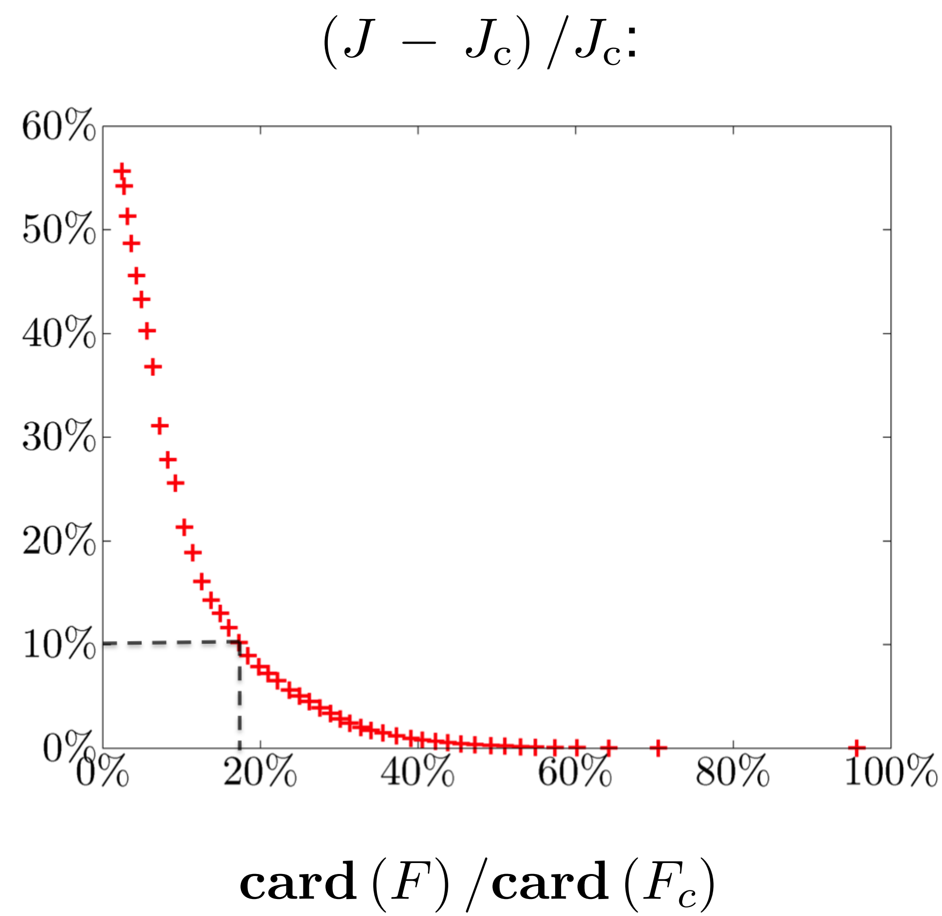
$$\begin{bmatrix} \dot{p}_i \\ \dot{v}_i \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} p_i \\ v_i \end{bmatrix}}_{\text{unstable dynamics}} + \underbrace{\sum_{j \neq i} \mathrm{e}^{-\alpha(i,j)} \begin{bmatrix} p_j \\ v_j \end{bmatrix}}_{\text{coupling}} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (d_i + u_i)$$

$\alpha(i,j)$: Euclidean distance between nodes i and j

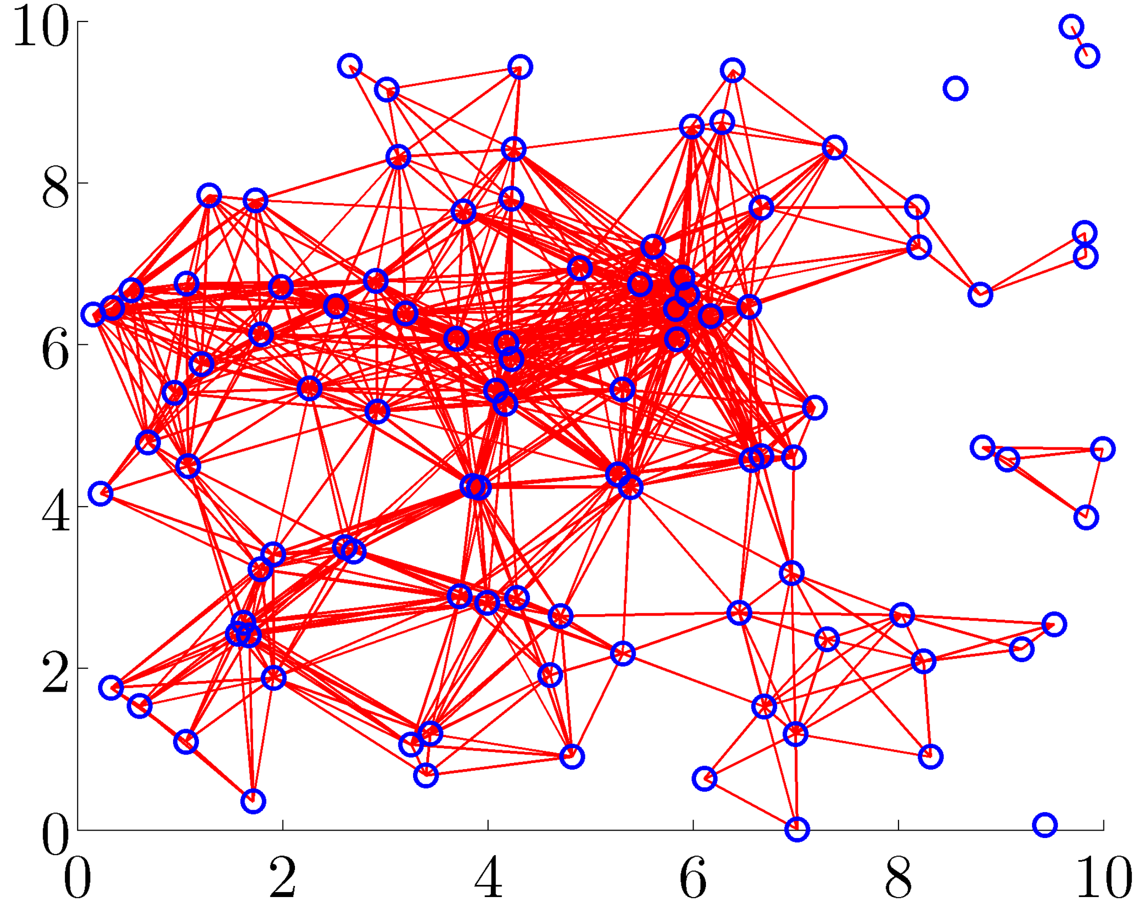
Motee & Jadbabaie, IEEE TAC '08

- **Performance comparison: sparse vs centralized**





identified communication graph:

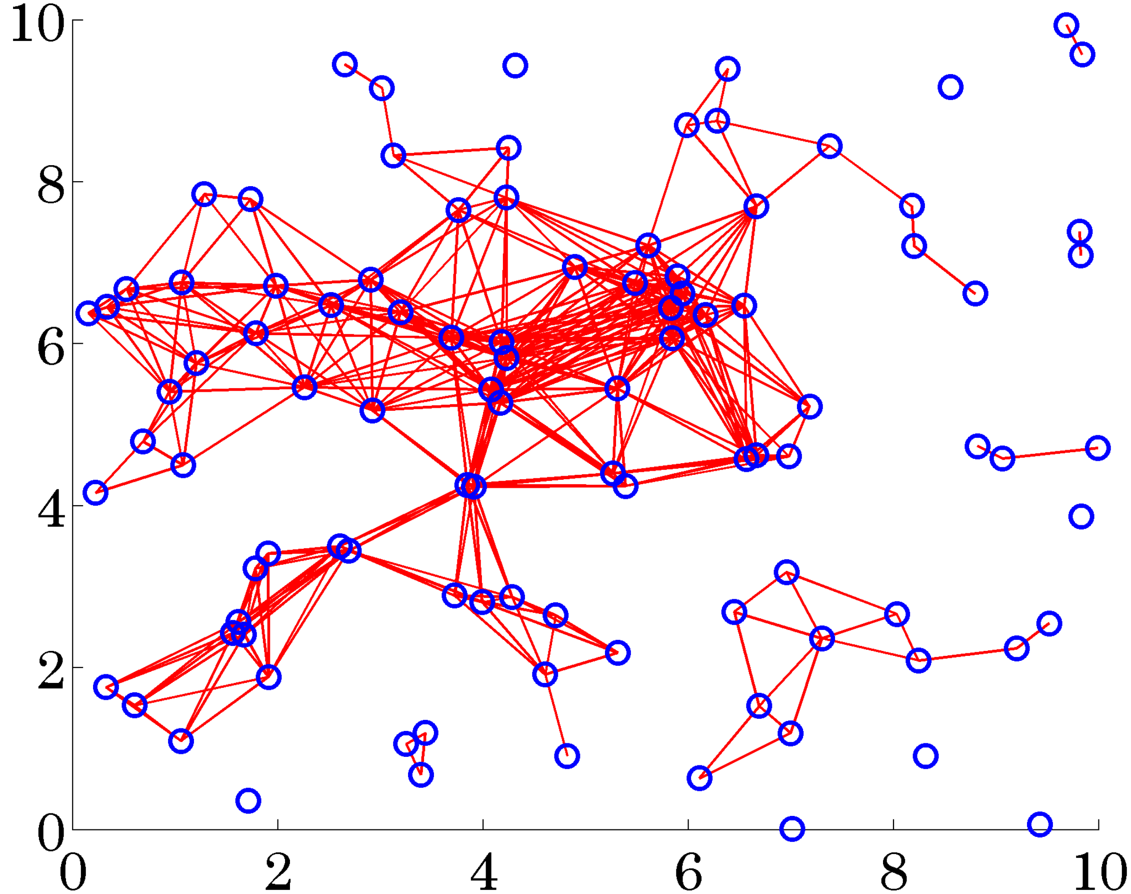


$$\gamma = 5$$

$$\textbf{card}\,(F)\,/\textbf{card}\,(F_{\rm c}) = 8.8\%$$

$$(J - J_{\rm c})\,/J_{\rm c} = 24.6\%$$

identified communication graph:

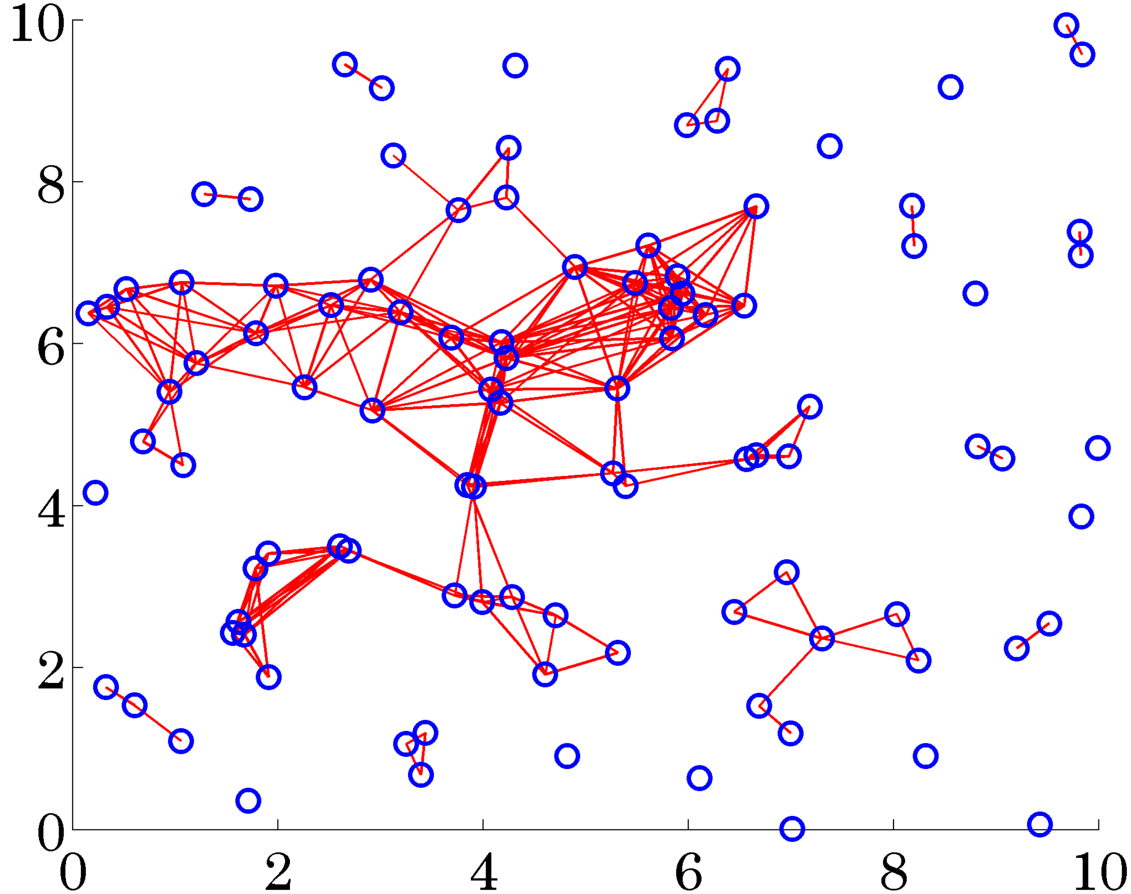


$$\gamma = 11$$

$$\mathbf{card}\,(F)\,/\,\mathbf{card}\,(F_c) = 5.1\%$$

$$(J - J_c)\,/\,J_c = 40.9\%$$

identified communication graph:

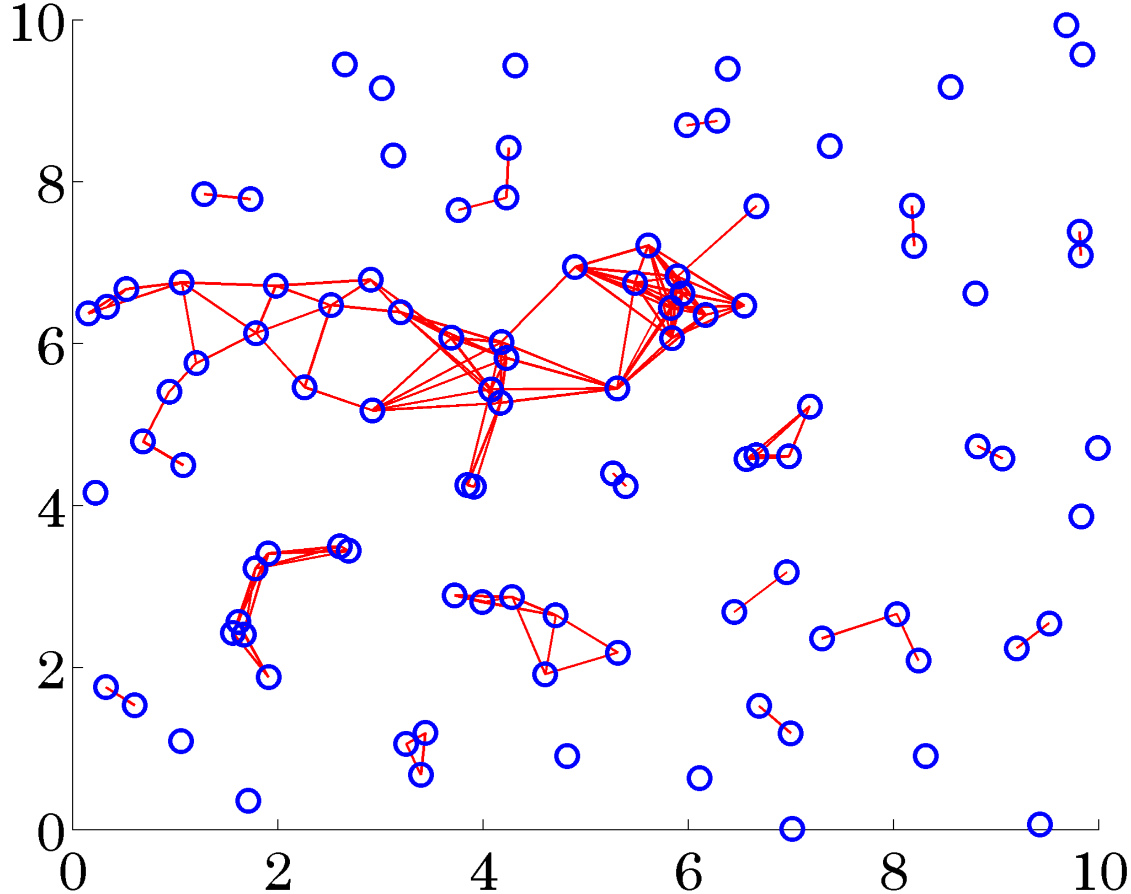


$$\gamma = 18$$

$$\mathbf{card}\,(F)\,/\mathbf{card}\,(F_{\rm c}) = 3.4\%$$

$$(J - J_{\rm c})\,/J_{\rm c} = 48.7\%$$

identified communication graph:

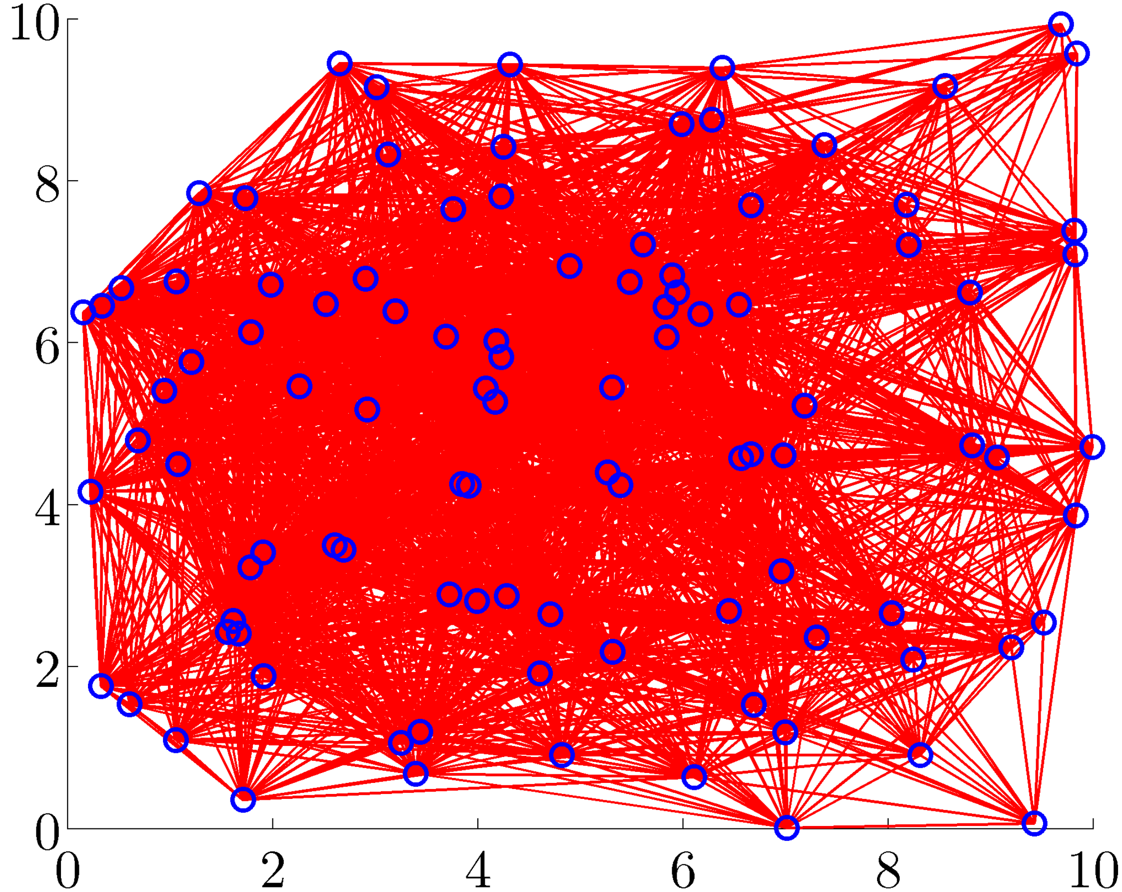


$$\gamma = 30$$

$$\mathbf{card}(F)/\mathbf{card}(F_c) = 2.4\%$$

$$(J - J_c)/J_c = 54.8\%$$

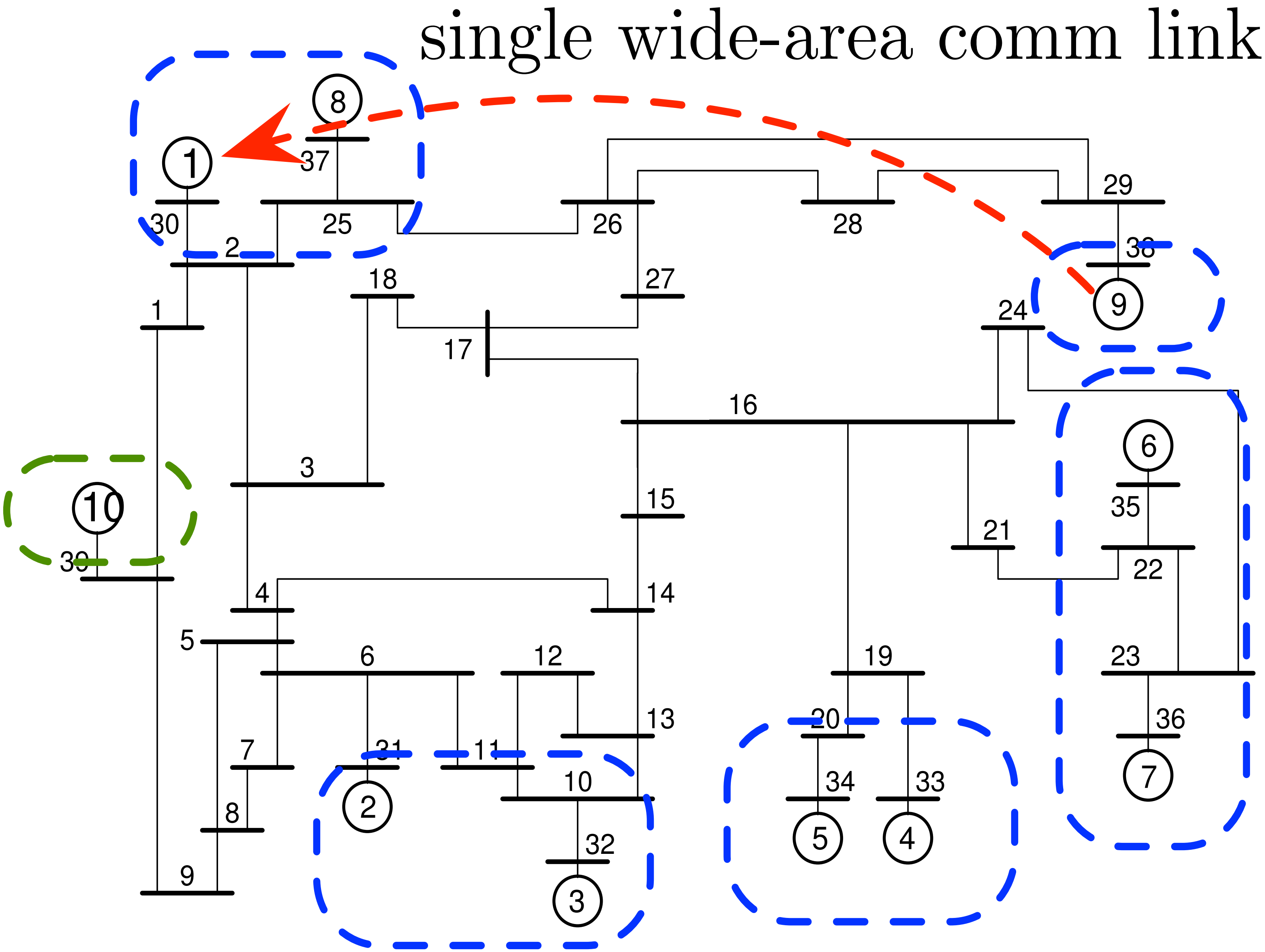
communication graph of a truncated centralized gain:



$\textbf{card}\,(F) \;=\; 7380\;\;(36.9\%)$

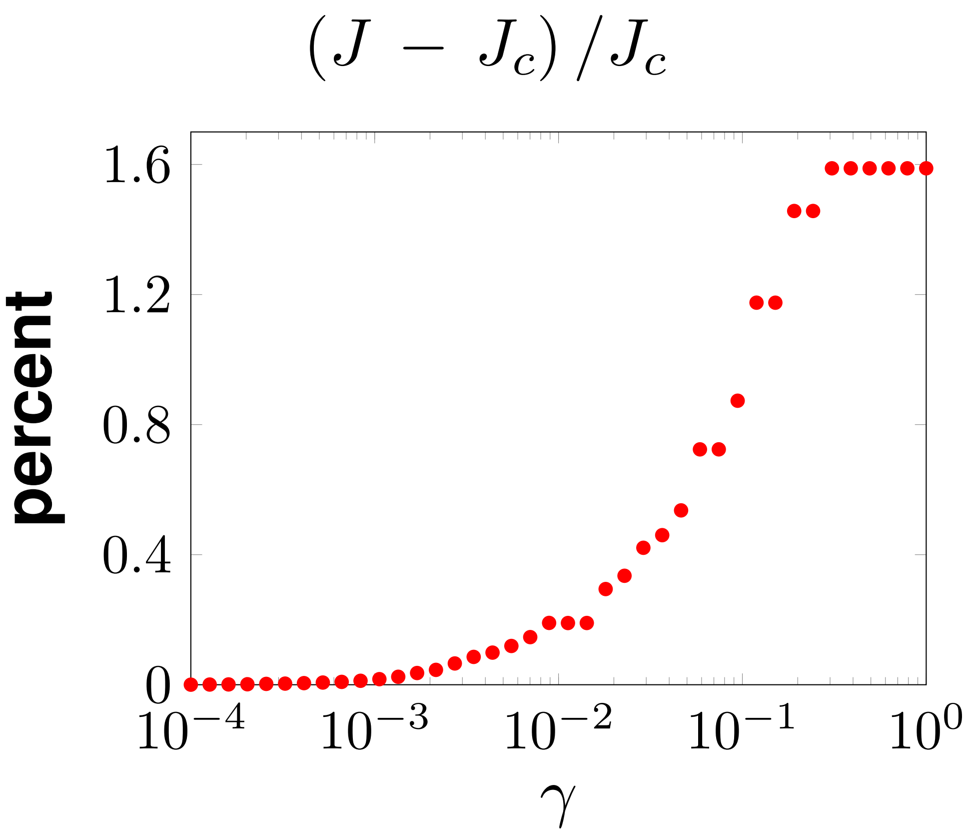
non-stabilizing

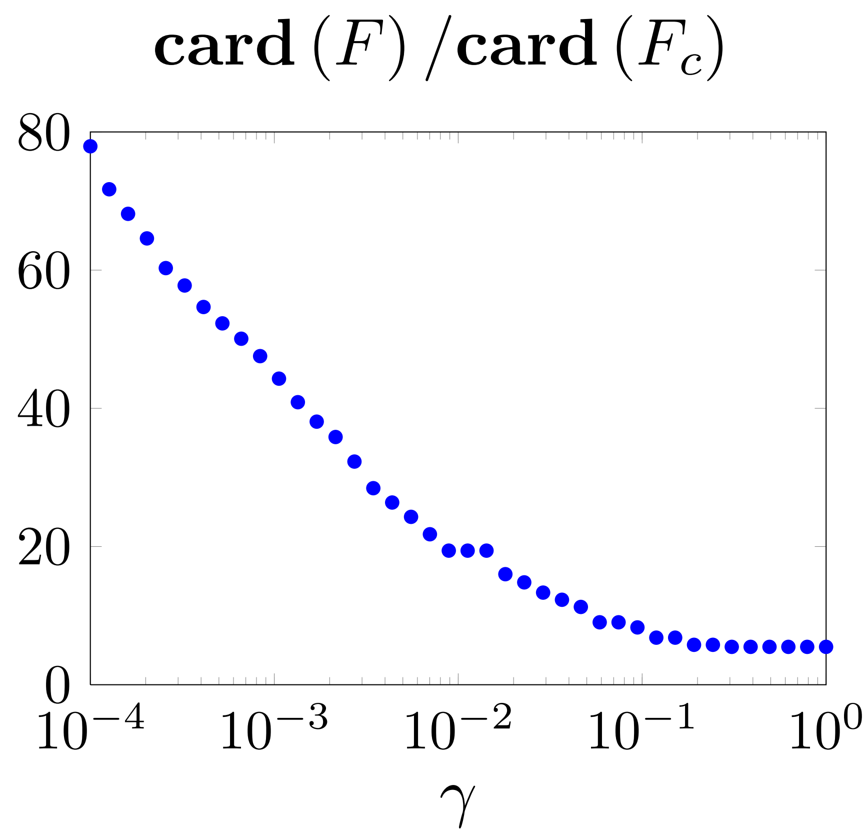
Wide area control of power networks



single long range interaction $\Rightarrow$ **nearly centralized performance**

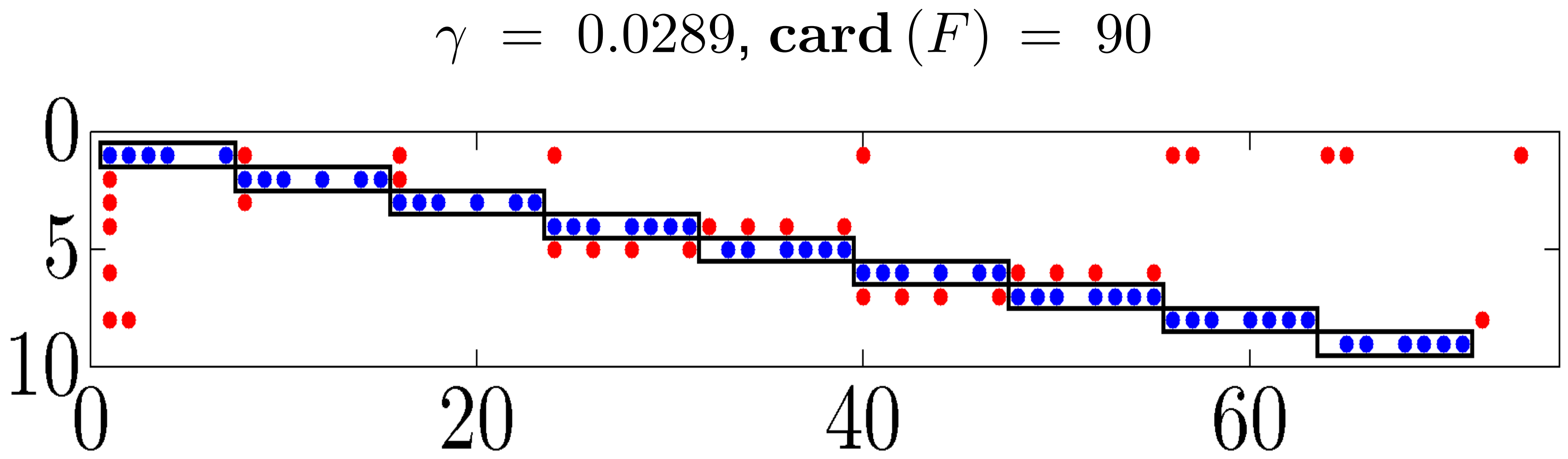
Performance vs sparsity

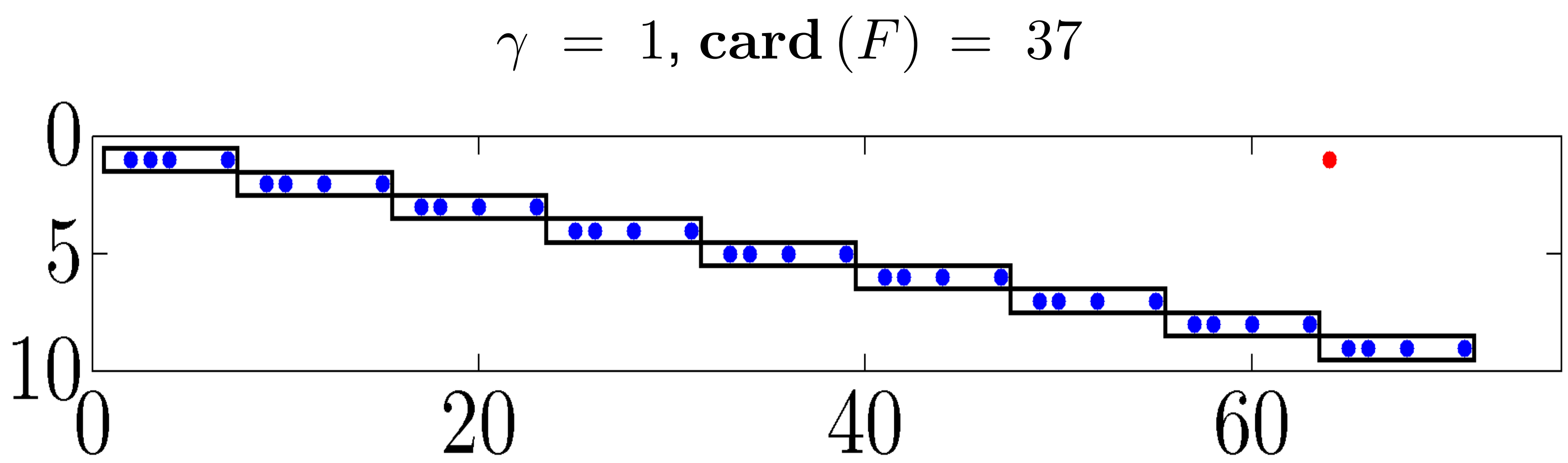




$$\gamma = 1 \xrightarrow{\text{relative to } F_c} \begin{cases} 1.6\,\% & \text{performance loss} \\ 5.5\,\% & \text{non-zero elements in } F \end{cases}$$

• Signal exchange network

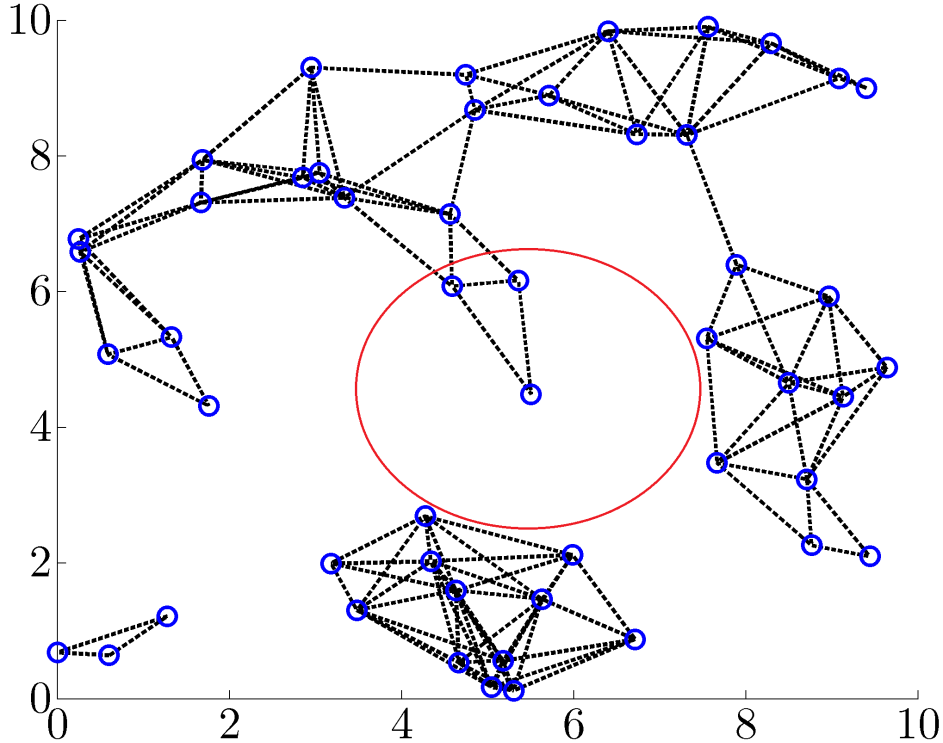




Dörfler, Jovanović, Chertkov, Bullo, IEEE TPWRS '14

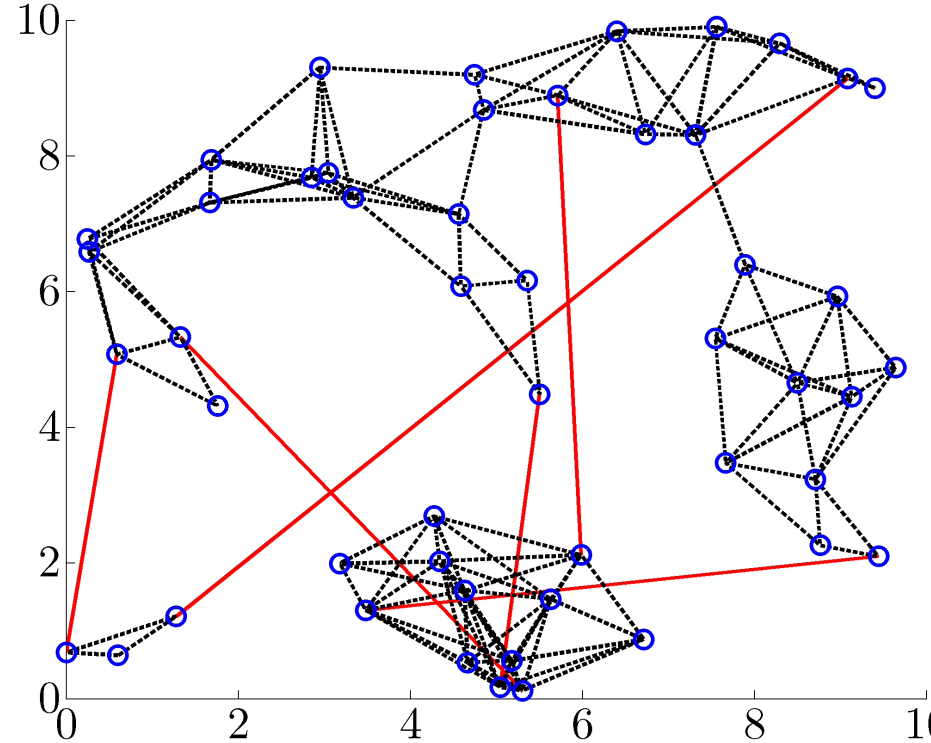
Sparsity-promoting consensus algorithm

local performance graph:



$$Q = Q_{\text{loc}} + \left(I - \frac{1}{N} \mathbb{1}\mathbb{1}^T \right)$$

identified communication graph:



$$\frac{J - J_c}{J_c} \approx 11\%$$

Lin, Fardad, Jovanović, Allerton '12

ALGORITHM

Alternating direction method of multipliers

$$\text{minimize} \quad J(F) \; + \; \gamma \, g(F)$$

- **Step 1: introduce additional variable/constraint**

$$\begin{array}{ll} \text{minimize} & J(F) \; + \; \gamma \, g(G) \\ \text{subject to} & F \; - \; G \; = \; 0 \end{array}$$

benefit: decouples J and g

- **Step 2: introduce augmented Lagrangian**

$$\mathcal{L}_\rho(F, G, \Lambda) \; = \; J(F) \; + \; \gamma \, g(G) \; + \; \text{trace}\left(\Lambda^T (F \; - \; G)\right) \; + \; \frac{\rho}{2} \, \|F \; - \; G\|_F^2$$

- **Step 3: use ADMM for augmented Lagrangian minimization**

$$\mathcal{L}_\rho(F, G, \Lambda) \;=\; J(F) \,+\, \gamma\, g(G) \,+\, \operatorname{trace}\big(\Lambda^T (F - G)\big) \,+\, \frac{\rho}{2}\, \|F - G\|_F^2$$

ADMM:

$$\begin{array}{lcl} F^{k+1} & := & \underset{F}{\operatorname{argmin}}\; \mathcal{L}_\rho(F, G^k, \Lambda^k) \\ G^{k+1} & := & \underset{G}{\operatorname{argmin}}\; \mathcal{L}_\rho(F^{k+1}, G, \Lambda^k) \\ \Lambda^{k+1} & := & \Lambda^k \,+\, \rho\,(F^{k+1} \,-\, G^{k+1}) \end{array}$$

MANY MODERN APPLICATIONS

- ⋆ distributed computing
- ⋆ distributed signal processing
- ⋆ image denoising
- ⋆ machine learning

Boyd et al., Foundations and Trends in Machine Learning ’11

- **Step 4: Polishing – back to structured optimal design**

$\star$ ADMM $\left\{ \begin{array}{l} \textbf{identifies sparsity patterns} \\ \textbf{provides good initial condition for structured design} \end{array} \right.$

$\star$ NECESSARY CONDITIONS FOR OPTIMALITY OF THE STRUCTURED PROBLEM

$$\begin{array}{rcl} (A - B_2 F)^T P + P(A - B_2 F) & = & -\left(Q + F^T R F\right) \\ (A - B_2 F) L + L (A - B_2 F)^T & = & -B_1 B_1^T \\ \left[\left(R F - B_2^T P\right) L\right] \circ I_{\mathcal{S}} & = & 0 \end{array}$$

Newton's method with conjugate gradient

$I_{\mathcal{S}}$ - **structural identity**

$$F = \begin{bmatrix} * & * & & \\ * & * & * & \\ & * & * & * \\ & & * & * \end{bmatrix} \quad \Rightarrow \quad I_{\mathcal{S}} = \begin{bmatrix} 1 & 1 & & \\ 1 & 1 & 1 & \\ & 1 & 1 & 1 \\ & & 1 & 1 \end{bmatrix}$$

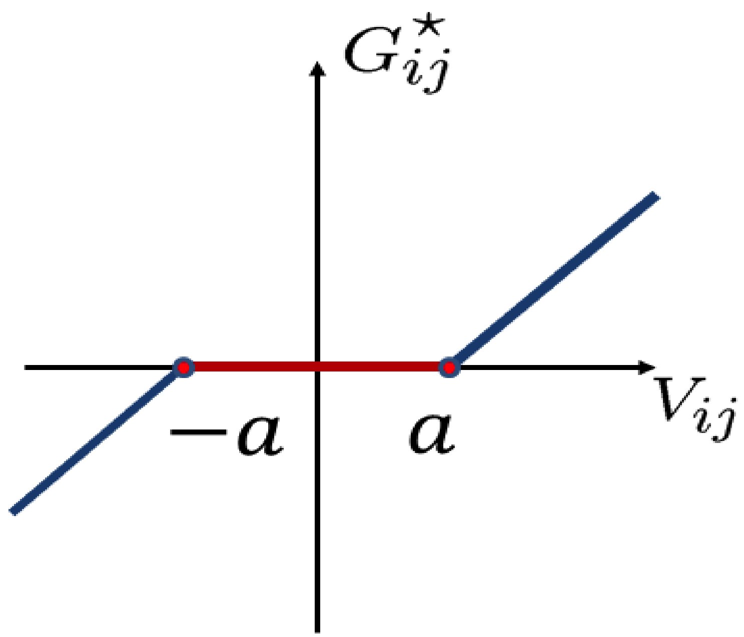
Solution to G-minimization problem

$$\underset{G_{ij}}{\text{minimize}} \quad \sum_{i,\,j} \left(\gamma \, W_{ij} \, |G_{ij}| \;+\; \frac{\rho}{2} \left(G_{ij} \;-\; V_{ij}^k \right)^2 \right)$$

$$V_{ij}^k \;:=\; F_{ij}^{k+1} \;+\; (1/\rho) \Lambda_{ij}^k$$

separability $\Rightarrow$ **element-wise analytical solution**

soft-thresholding



$$a \;=\; (\gamma/\rho) \, W_{ij}$$

Solution to F-minimization problem

$$\underset{F}{\text{minimize}} \quad J(F) \; + \; \frac{\rho}{2} \|F \; - \; U^k\|_F^2$$

$$U^k \; := \; G^k \; - \; (1/\rho)\Lambda^k$$

NECESSARY CONDITIONS FOR OPTIMALITY:

$$\begin{aligned} (A - B_2 F)L + L(A - B_2 F)^T &= -B_1 B_1^T \\ (A - B_2 F)^T P + P(A - B_2 F) &= -(Q + F^T R F) \\ FL + \rho(2R)^{-1} F &= R^{-1} B_2^T P L + \rho(2R)^{-1} U^k \end{aligned}$$

- ITERATIVE SCHEME

Given F_0 solve for $\{L_1, P_1\} \; \to \; F_1 \; \to \; \{L_2, P_2\} \; \to \; F_2 \; \cdots$

descent direction $+$ **line search** $\Rightarrow$ **convergence**

ALTERNATIVE FORMULATIONS?

Optimal control in discrete time

$$
\begin{aligned}
x_{t+1} &= A\, x_t \;+\; B_1\, d_t \;+\; B_2\, u_t \\
u_t &= -\,F\, x_t
\end{aligned}
$$

- **No structural constraints**

$$
\begin{array}{ll}
\underset{X,\,F}{\text{minimize}} & \text{trace}\left(X B_1 B_1^T\right) \\
\text{subject to} & X \;-\; (A \;-\; B_2 F)^T\, X\, (A \;-\; B_2 F) \;=\; Q \;+\; F^T R\, F \\
 & X \;\succ\; 0
\end{array}
$$

equivalent formulation:

$$
\begin{array}{ll}
\underset{X,\,Y,\,F,\,K}{\text{minimize}} & \text{trace}\left(X B_1 B_1^T\right) \\
\text{subject to} & X \;-\; (A \;-\; B_2 F)^T\, Y^{-1}\, (A \;-\; B_2 F) \;\succeq\; Q \;+\; K \\
 & X \;\succ\; 0, \quad K \;\succeq\; F^T R\, F \\
 & X\, Y \;=\; I
\end{array}
$$

• A POSSIBLE APPROACH

$$\begin{array}{ll} \underset{X,\,Y,\,F,\,K}{\text{minimize}} & \text{trace}\left(X B_1 B_1^T\right) \\ \text{subject to} & X \,-\, (A \,-\, B_2 F)^T\, Y^{-1}\, (A \,-\, B_2 F) \;\succeq\; Q \,+\, K \\ & K \;\succeq\; F^T R\, F, \quad X \;\succeq\; Y^{-1}, \quad Y \;\preceq\; X^{-1} \\ & X \;\succ\; 0, \quad Y \;\succ\; 0 \end{array}$$

CONVEX APPROXIMATION?

$$\begin{array}{ll} \underset{X,\,Y,\,F,\,K}{\text{minimize}} & \text{trace}\,(X\, Y_k \,+\, X_k\, Y) \,+\, \text{trace}\left(X B_1 B_1^T\right) \\ \text{subject to} & \begin{bmatrix} X - Q - K & (A - B_2 F)^T \\ A - B_2 F & Y \end{bmatrix} \;\succeq\; 0 \\ & \begin{bmatrix} K & F^T \\ F & R^{-1} \end{bmatrix} \;\succeq\; 0, \quad \begin{bmatrix} X & I \\ I & Y \end{bmatrix} \;\succeq\; 0 \end{array}$$

Fardad & Jovanović, ACC '14

Summary

- Sparsity-promoting optimal control
  - ⋆ Performance vs sparsity tradeoff

    Lin, Fardad, Jovanović, IEEE TAC '13

  - ⋆ Software

    www.umn.edu/∼mihailo/software/lqrsp/

- Related effort
  - ⋆ Leader selection in large dynamic networks

    Lin, Fardad, Jovanović, IEEE TAC '14

  - ⋆ Optimal synchronization of sparse oscillator networks

    Fardad, Lin, Jovanović, IEEE TAC '14

  - ⋆ Optimal dissemination of information in social networks

    Fardad, Zhang, Lin, Jovanović, CDC '12

  - ⋆ Sparse or infrequently changing (in time) control signals

    Jovanović & Lin, ECC '13

Acknowledgments

Makan Fardad
(Syracuse)

Fu Lin
(Argonne)

Neil Dhingra
(U of M)

Xiaofan Wu
(U of M)

SUPPORT:

NSF Award CMMI-09-27720

NSF Award ECCS-14-07958

DISCUSSIONS:

Stephen Boyd

Roland Glowinski