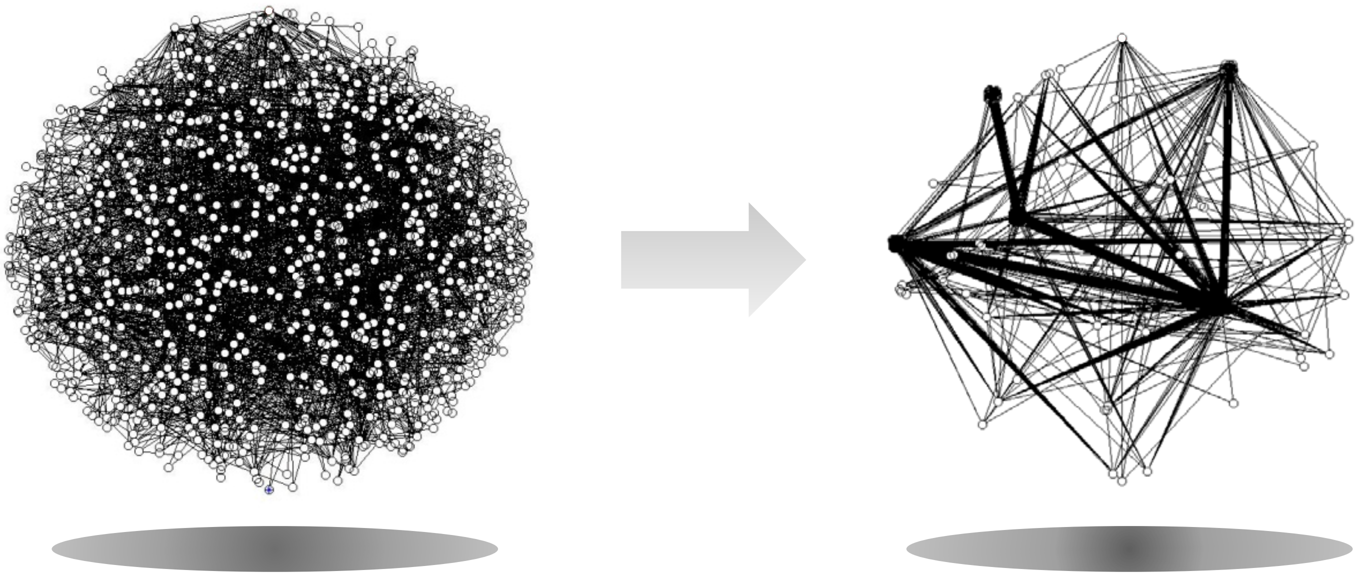
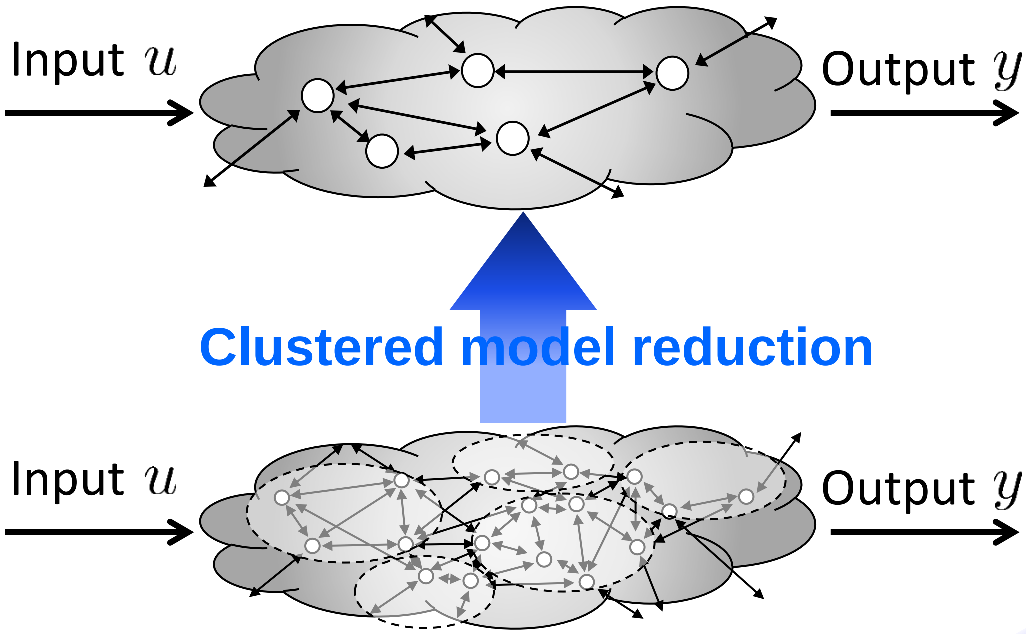
大規模ネットワークシステムのクラスタ低次元化



石崎　孝幸　（東京工業大学）

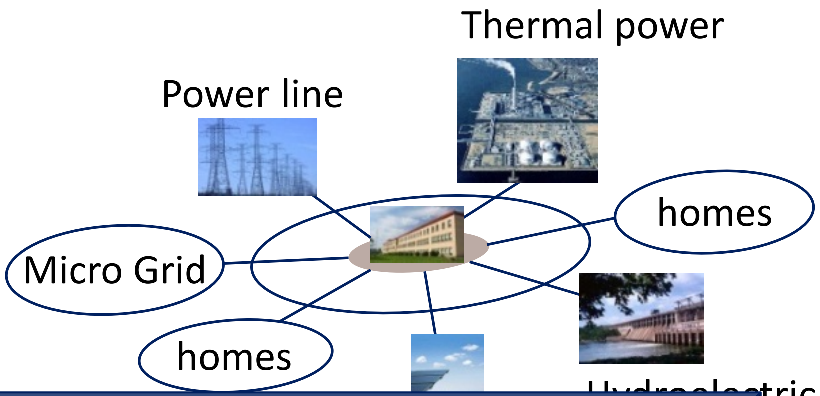
Outline

- Introduction: Why clustered model reduction?
- Clustered Model Reduction Theory
- Examples
- Conclusion



Large-Scale Network Systems

Power Network

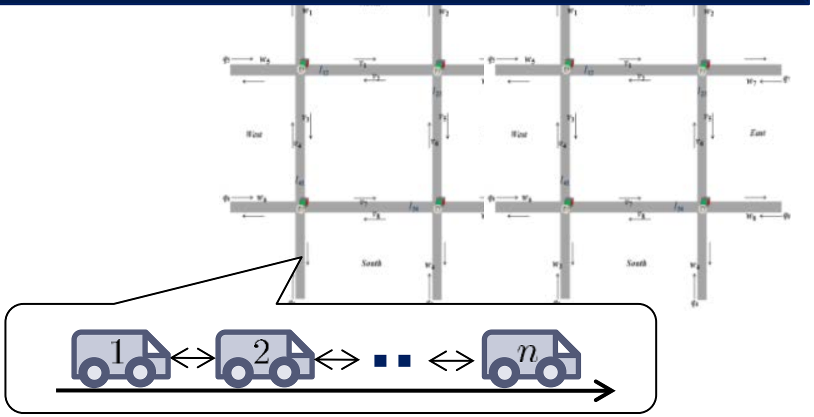


- Tokyo area: 20 million houses
  - rate of houses with PV will increase up to 50% by 2030 (PV2030)
  - 50% of total maximum power

Model reduction is one prospective approach

Traffic Network

- Center of Tokyo area: 5 million cars
  - Heavy traffic jam
  - average velocity 20km/h

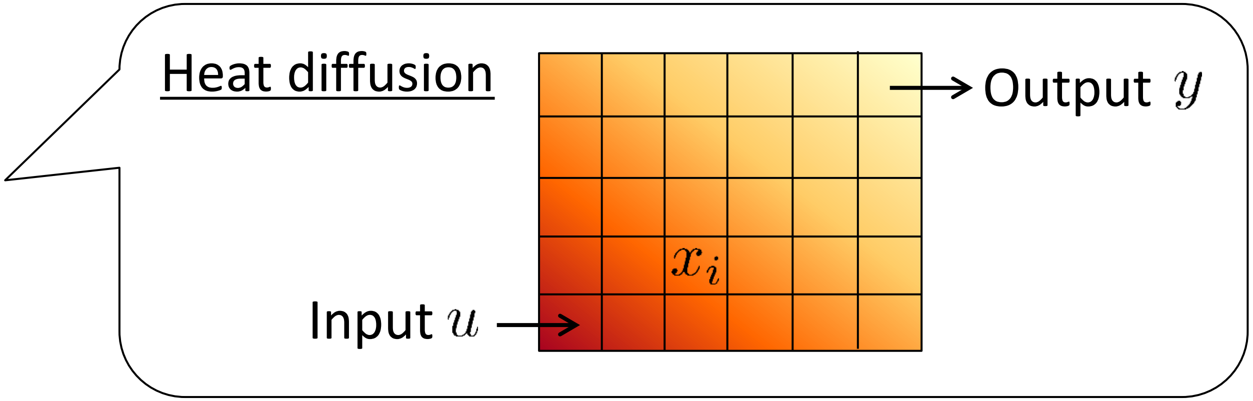


How should we improve?

Model Reduction Methods

$$\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

$$x \in \mathbb{R}^n$$



Heat diffusion

Output y

x_i

Input u

Main goal: Find P such that $\|y - \hat{y}\|$ is small enough

+ stability preservation, error bound derivation, low computational cost

Standard methods:

- Balanced truncation, Hankel norm approximation
  - error bound, stability preservation ☺ high computational cost ☹
- Krylov projection
  - lower computational cost ☺ possibly unstable model, no error bound ☹

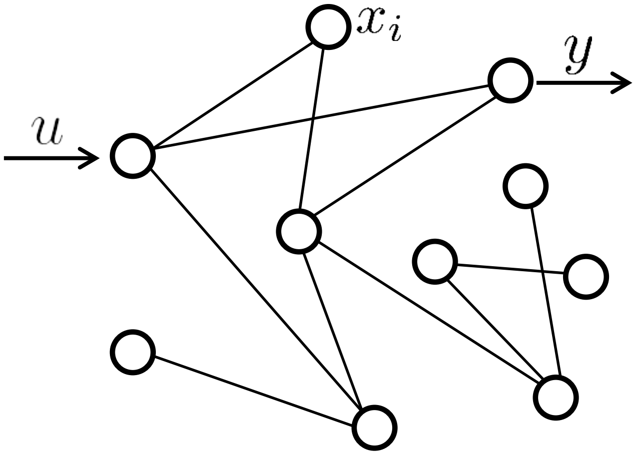
Drawback of Standard Methods

Network system

$$\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

Reduced model

$$\hat{\Sigma} : \begin{cases} \dot{\hat{x}} = PAP^{\dagger}\hat{x} + PBu \\ \hat{y} = CP^{\dagger}\hat{x} \end{cases}$$

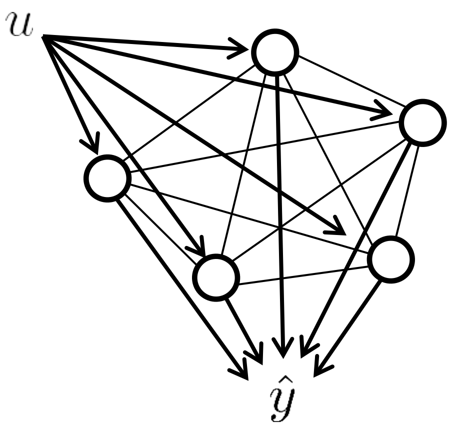


$$Px = \hat{x}$$

$$\longrightarrow$$

$$P \in \mathbb{R}^{\hat{n} \times n}, \ \hat{n} < n$$

No specific structure



Sparse ☺

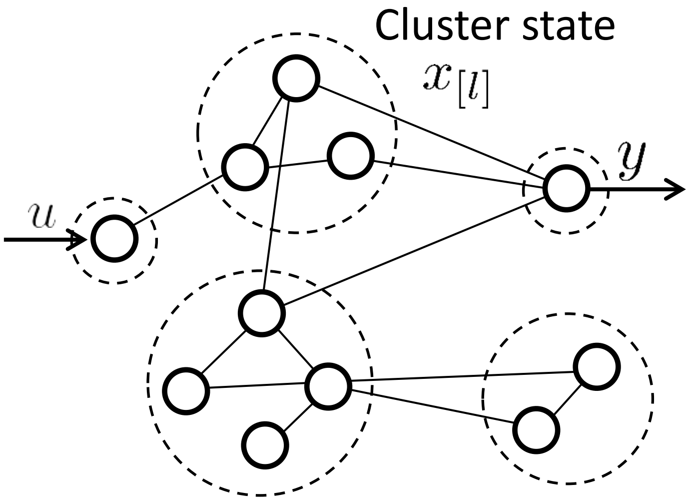
Dense ☹

Drawback: Network structure is lost through reduction

Clustered Model Reduction

Network system

$$\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$



Sparse ☺

$$Px = \hat{x}$$

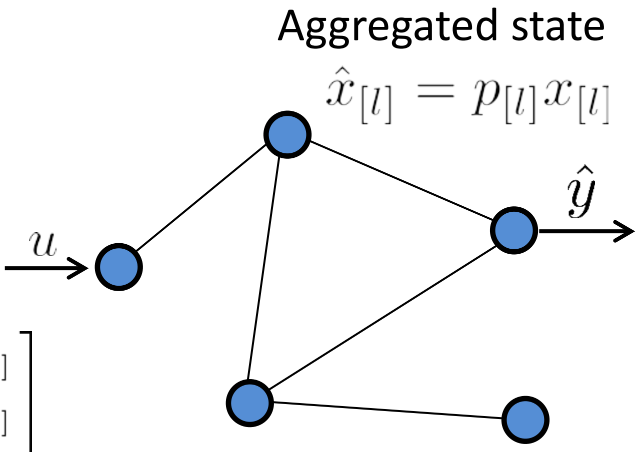
$$\longrightarrow$$

$p_{[l]}$: row vector

$$\begin{bmatrix} p_{[1]} & & \\ & p_{[2]} & \\ & & \ddots \end{bmatrix} \begin{bmatrix} x_{[1]} \\ x_{[2]} \\ \vdots \end{bmatrix} = \begin{bmatrix} \hat{x}_{[1]} \\ \hat{x}_{[2]} \\ \vdots \end{bmatrix}$$

Reduced model

$$\hat{\Sigma} : \begin{cases} \dot{\hat{x}} = PAP^{\dagger}\hat{x} + PBu \\ \hat{y} = CP^{\dagger}\hat{x} \end{cases}$$

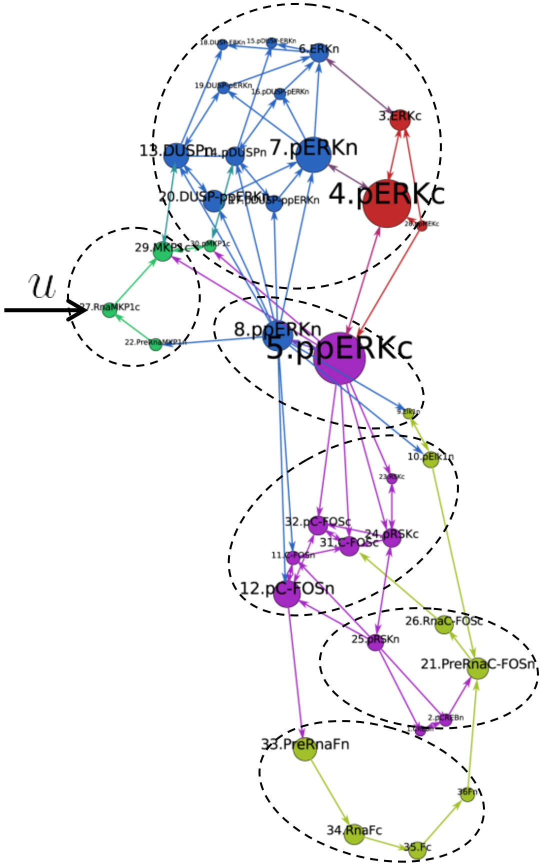


Sparse ☺

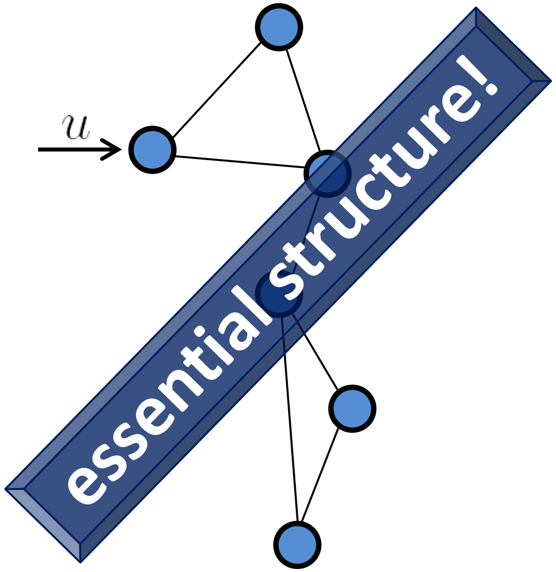
Preservation of network structure among clusters

Why Clustered Model Reduction?

<u>Gene Network</u> [Mochizuki et al. , J. Theoretical Biology (2010)]



Clustered model reduction

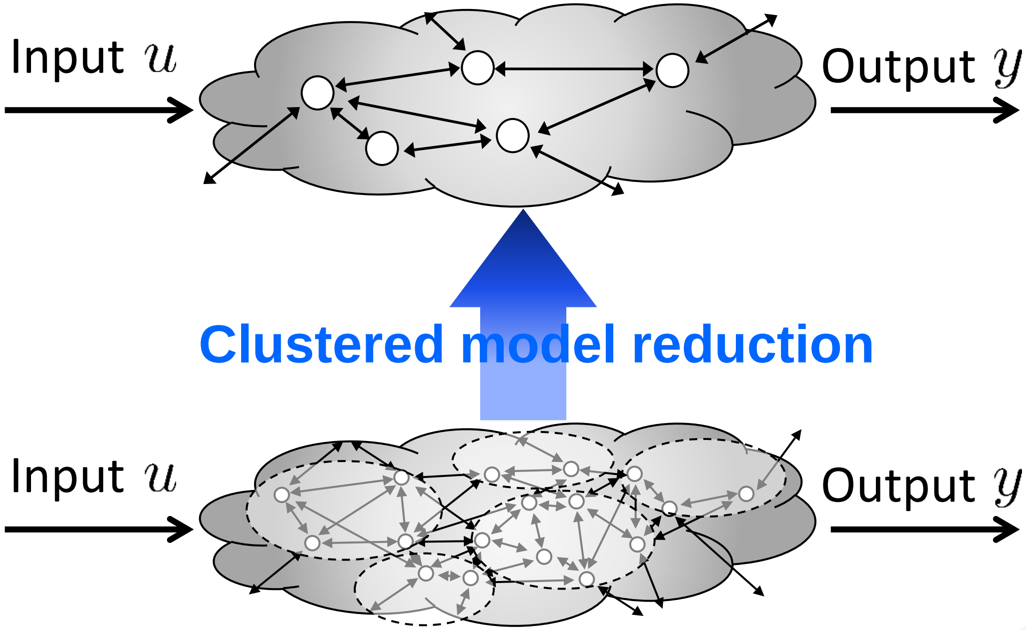


Extract essential principle
to show mechanism of functions

Other possible application: Hierarchical decentralized control

Outline

- Introduction: Why clustered model reduction?
- Clustered Model Reduction Theory
- Examples
- Conclusion



Typical System Norm

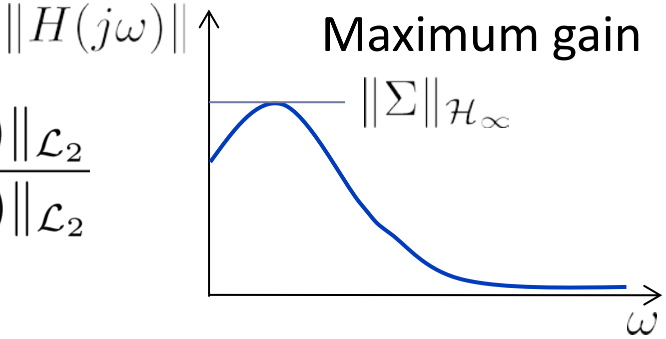
Stable system $\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$ solution $y(t) = \int_0^t h(t-\tau)u(\tau)d\tau$

impulse response $h(t) := Ce^{At}B$

Transfer function $H(s) := \mathscr{L}[h] = C(sI_n - A)^{-1}B$ ✓ $\mathscr{L}[*]$: Laplace transform

$\mathcal{H}_\infty$-norm

$$\|H\|_{\mathcal{H}_\infty} := \sup_{\omega\in\mathbb{R}} \|H(j\omega)\| = \sup_{u\neq 0} \frac{\|y(t)\|_{\mathcal{L}_2}}{\|u(t)\|_{\mathcal{L}_2}}$$



$\mathcal{H}_2$-norm $\|H\|_{\mathcal{H}_2} := \left(\int_{-\infty}^{\infty} \|H(j\omega)\|_F^2 \frac{d\omega}{2\pi}\right)^{\frac{1}{2}} = \|h\|_{\mathcal{L}_2}$ Energy of impulse response

✓ $\mathcal{L}_2$-norm of $f : \mathbb{R}_+ \to \mathbb{R}^{p\times m}$ $\|f\|_{\mathcal{L}_2} := \left(\int_0^\infty \|f(t)\|_F^2\, dt\right)^{\frac{1}{2}}$

$\|*\|_F$: Frobenius norm

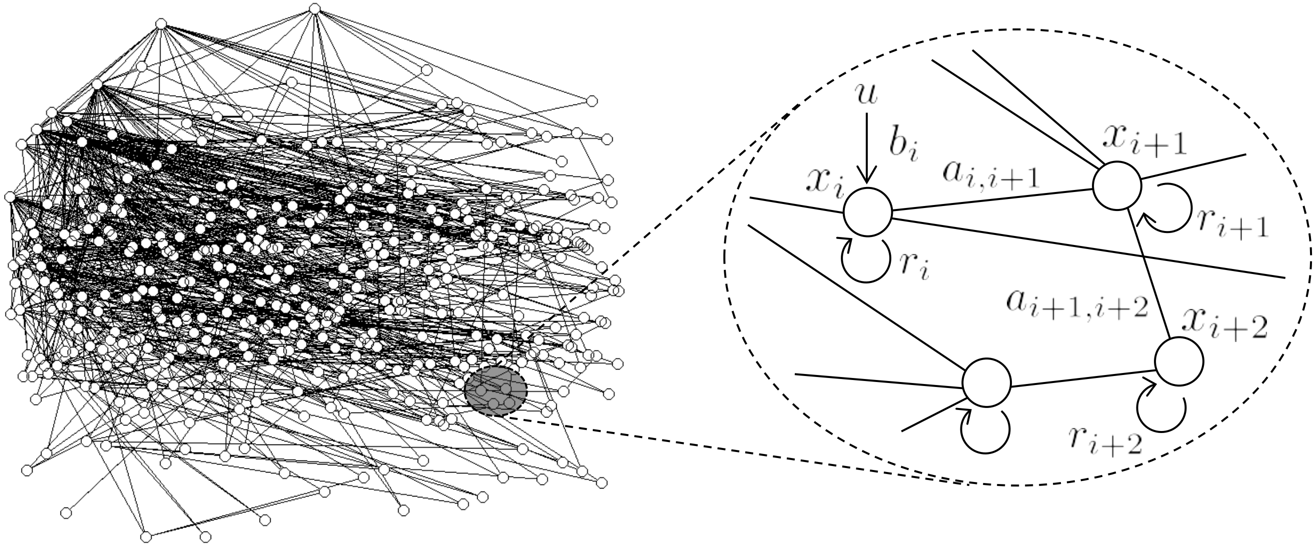
System Description

[Definition] **Bidirectional Network** (A, b)

$\dot{x} = Ax + Bu$ with $A = \{a_{i,j}\} \in \mathbb{R}^{n \times n}$ and $B = \{b_i\} \in \mathbb{R}^n$

is said to be *bidirectional network* (A, B) if A is symmetric and stable.

Reaction-diffusion systems: $\dot{x}_i = -r_i x_i + \sum_{j=1, j \neq i}^{n} a_{i,j} (x_j - x_i) + b_i u$



Clustered Model Reduction Problem

✓ $PP^{\mathsf{T}} = I_{\hat{n}}$

Bidirectional network

$$\dot{x} = Ax + Bu$$

Reduced model

$$\dot{\hat{x}} = PAP^{\mathsf{T}}\hat{x} + PBu$$

(i) Find a set of reducible clusters

(ii) Find suitable aggregation weights

Sparse ☺

Sparse ☺

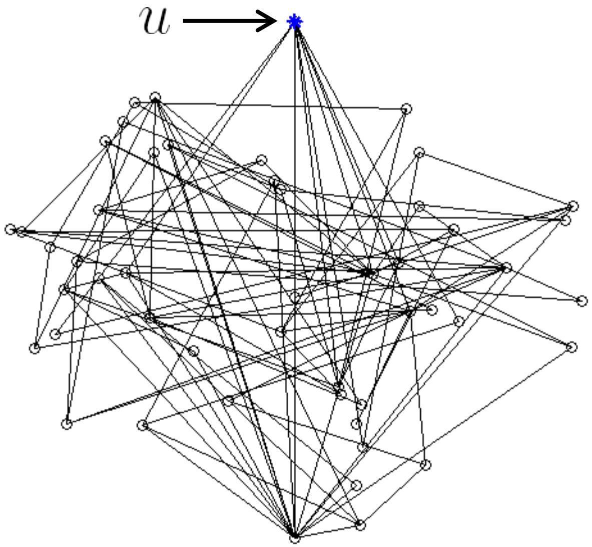
[Problem] Given $\epsilon \geq 0$, find a block-diagonal $P \in \mathbb{R}^{\hat{n} \times n}$ such that

$$\|g(s) - \hat{g}(s)\|_{\mathcal{H}_\infty} \leq \epsilon$$

where $g(s) := (sI_n - A)^{-1}B$ and $\hat{g}(s) := P^{\mathsf{T}}(sI_{\hat{n}} - PAP^{\mathsf{T}})^{-1}PB.$

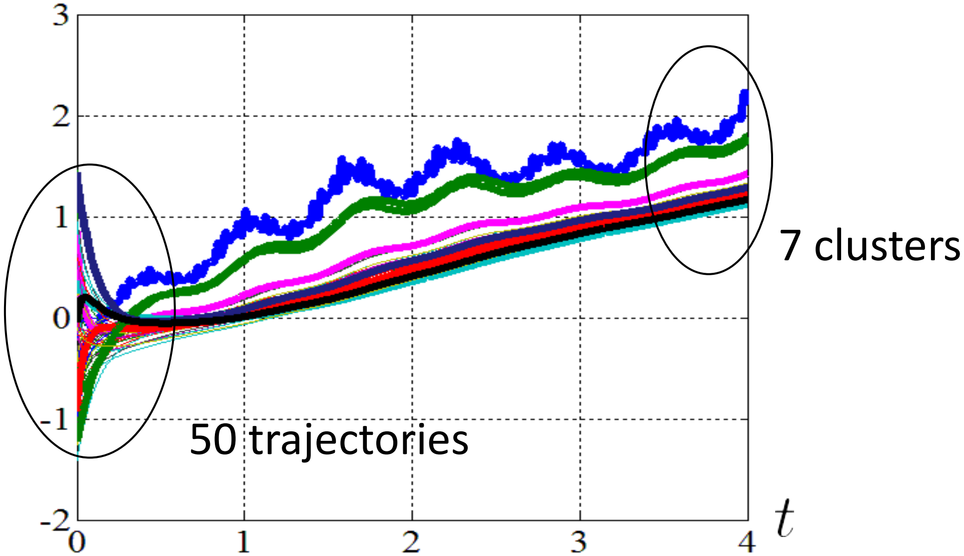
How to Formulate Reducibility?

<u>Bidirectional network</u> $\dot{x} = Ax + Bu$



50 nodes, nonzero $a_{i,j}$ is randomly chosen from $(0, 1]$

[State trajectory under random u]



$x \in \mathbb{R}^{50}$ can be aggregated into 7-dim. variable?

[Definition] **Reducible cluster**

Let $x(0) = 0$. A cluster $\mathcal{I}_{[l]} \subseteq \{1, \ldots, n\}$ is said to be ***reducible*** if

$$\forall i, j \in \mathcal{I}_{[l]}, \ \exists \rho_{i,j} \geq 0 \ \text{ s.t. } \ x_i(t) \equiv \rho_{i,j} x_j(t)$$ for any $u(t)$.

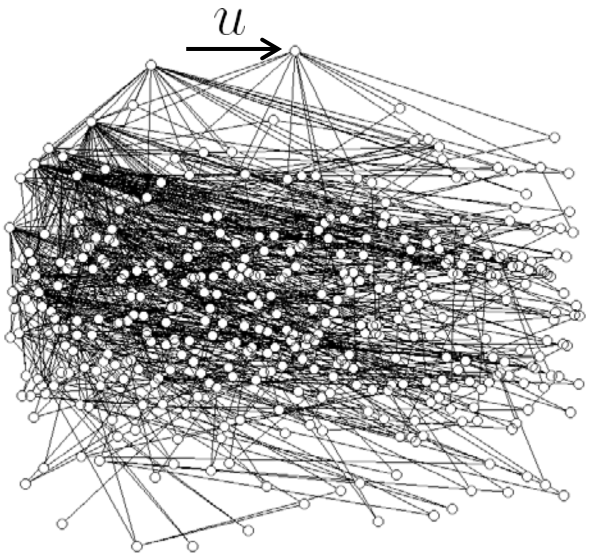
Positive Tridiagonalization

[Lemma] For every bidirectional network (A, B), there exists a unitary $H \in \mathbb{R}^{n \times n}$ such that $(\tilde{A}, \tilde{B}) = (H^{\mathsf{T}}AH, H^{\mathsf{T}}B)$ has the following structure.

Bidirectional network (A, B)

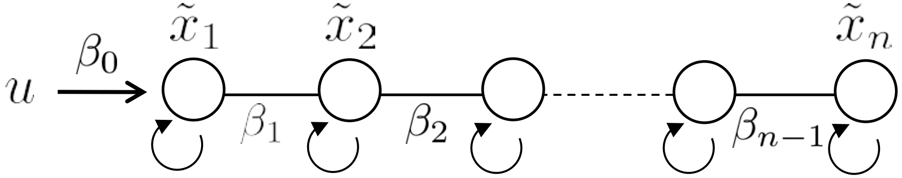
$$\dot{x} = Ax + Bu, \quad A = A^{\mathsf{T}}$$

(not necessarily positive)



$$\overset{x = H\tilde{x}}{\longleftrightarrow}$$

Positive tridiagonal realization $(\tilde{A}, \tilde{B})$



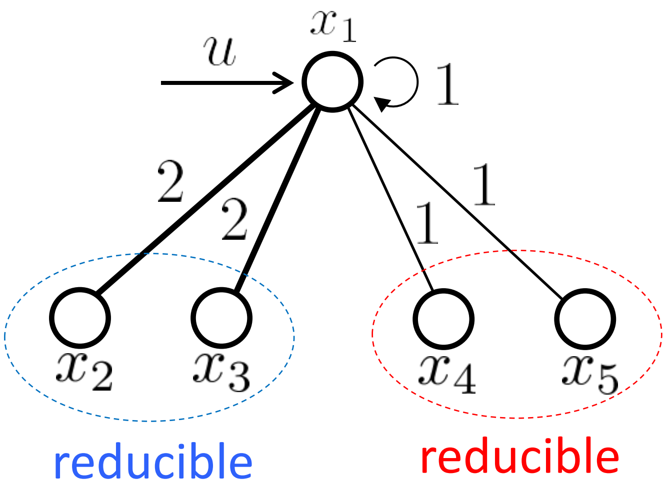
Metzler

$$\tilde{A} = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_{n-1} \\ & & & \beta_{n-1} & \alpha_n \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} \beta_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

✓ $\beta_i \geq 0$

Reducibility Characterization

Bidirectional network (A, B)



$$\begin{cases} (\tilde{A}, \tilde{B}) : \text{positive tridiagonal realization} \\ H \quad : \text{transformation matrix} \end{cases}$$

<u>Index matrix</u>

$$\Phi := H \mathrm{diag}(-\tilde{A}^{-1}\tilde{B})$$

$-\tilde{A}^{-1}\tilde{B}$: DC-gain ⇔ Maximal gain

<u>owing to positivity</u>

$$\Phi = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1.20 & -0.20 & 0 & 0 \\ \hline 0 & 1.20 & -0.20 & 0 & 0 \\ \hline 0 & 0.60 & 0.40 & 0 & 0 \\ \hline 0 & 0.60 & 0.40 & 0 & 0 \end{array}\right]$$

identical

identical

<u>Equivalent characterization of cluster reducibility</u>

θ-Reducible Cluster Construction

[Definition]   **θ-reducible Cluster**

A cluster $\mathcal{I}_{[l]}$ is said to be *θ-reducible* if

$$\forall j \in \mathcal{I}_{[l]},\ \exists i \in \mathcal{I}_{[l]},\ \rho_{i,j} \geq 0 \ \text{ s.t. } \|\mathrm{row}_i[\Phi] - \rho_{i,j}\mathrm{row}_j[\Phi]\|_{l_\infty} \leq \theta$$

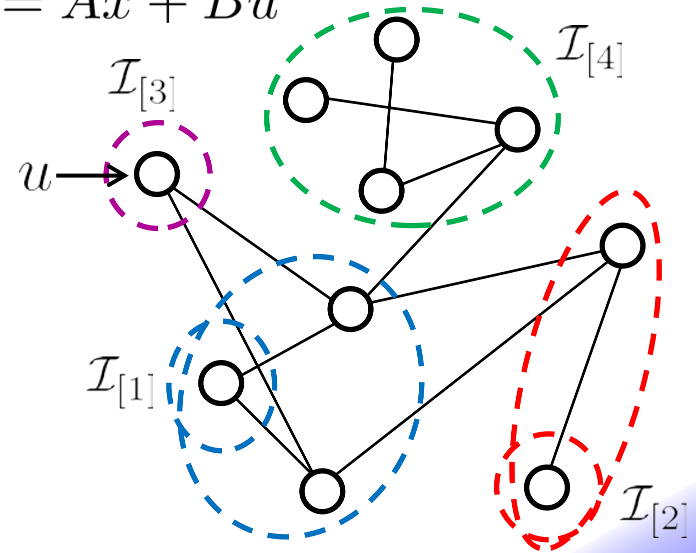
θ : coarseness parameter

Procedure for finding a cluster set

(i) Given $\dot{x} = Ax + Bu$, calculate the index matrix $\Phi = H\mathrm{diag}(-\tilde{A}^{-1}\tilde{B})$

(ii) For a fixed θ, find $\{\mathcal{I}_{[1]}, \ldots, \mathcal{I}_{[\hat{n}]}\}$ satisfying above definition of θ-reducibility

$\dot{x} = Ax + Bu$



Reducible Cluster Aggregation

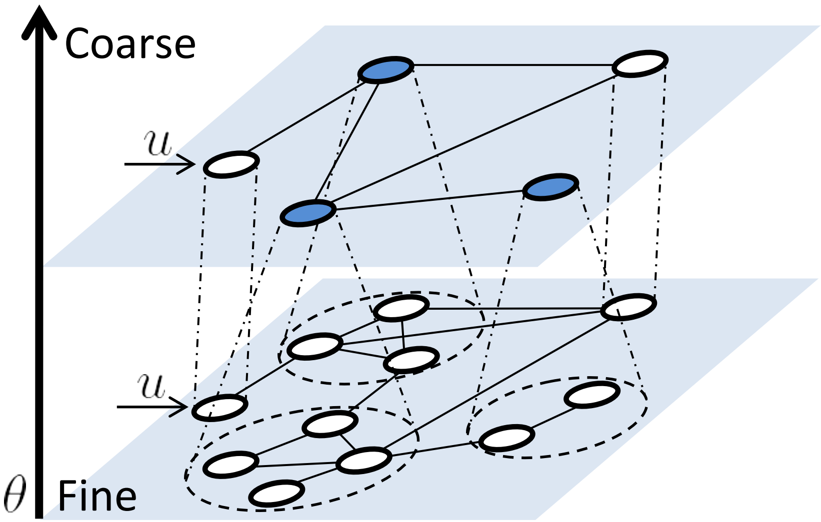
[Theorem] Define $P = \mathrm{Diag}\left(p_{[1]}, \ldots, p_{[\hat{n}]}\right)$ with $p_{[l]} = \frac{[\rho_{i,j}]_{j \in \mathcal{I}_{[l]}}}{\|[\rho_{i,j}]_{j \in \mathcal{I}_{[l]}}\|}$.

If all clusters are θ-reducible, then

$$\|g(s) - \hat{g}(s)\|_{\mathcal{H}_\infty} \leq \sqrt{\alpha}\,\|(PAP^{\mathsf{T}})^{-1}PA\|\theta$$

holds where $\alpha := \sum_{l=1}^{\hat{n}} |\mathcal{I}_{[l]}|(|\mathcal{I}_{[l]}| - 1)$.

linear dependence on θ



Reduced model (PAP^{T}, PB)

$$\hat{g}(s) = P^{\mathsf{T}}\left(sI_{\hat{n}} - PAP^{\mathsf{T}}\right)^{-1}PB$$

with $P = \mathrm{Diag}(p_{[1]}, \ldots, p_{[\hat{n}]}) \in \mathbb{R}^{\hat{n} \times n}$

Bidirectional network (A, B)

$$g\left(s\right) = \left(sI_n - A\right)^{-1} B$$

Remarks

- Clustered model reduction
  - preserves stability & network structure of original system
  - provides a priori $\mathcal{H}_\infty$-error bound
  - gives method to find a cluster set
  - requires low computational cost $O(n^3)$

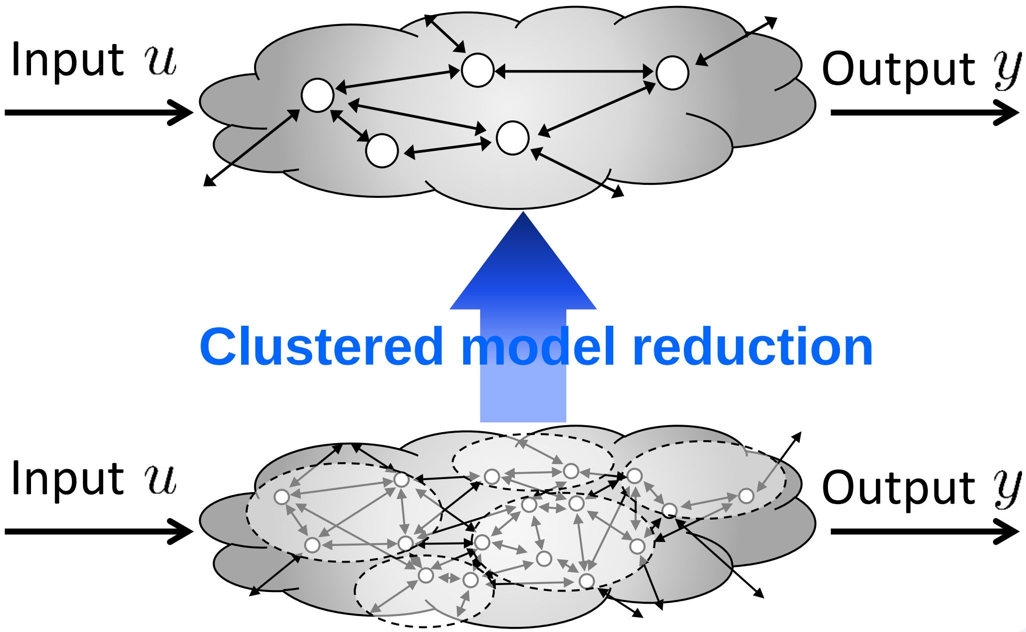
[T. Ishizaki et al. IEEE TAC (2014)]

- This method can be extended to
  - multi-input systems
  - positive directed networks
    - asymmetric A with nonnegative off-diagonal entries
    - a priori $\mathcal{H}_2$-error bound

[T. Ishizaki et al. ACC12, CDC12]

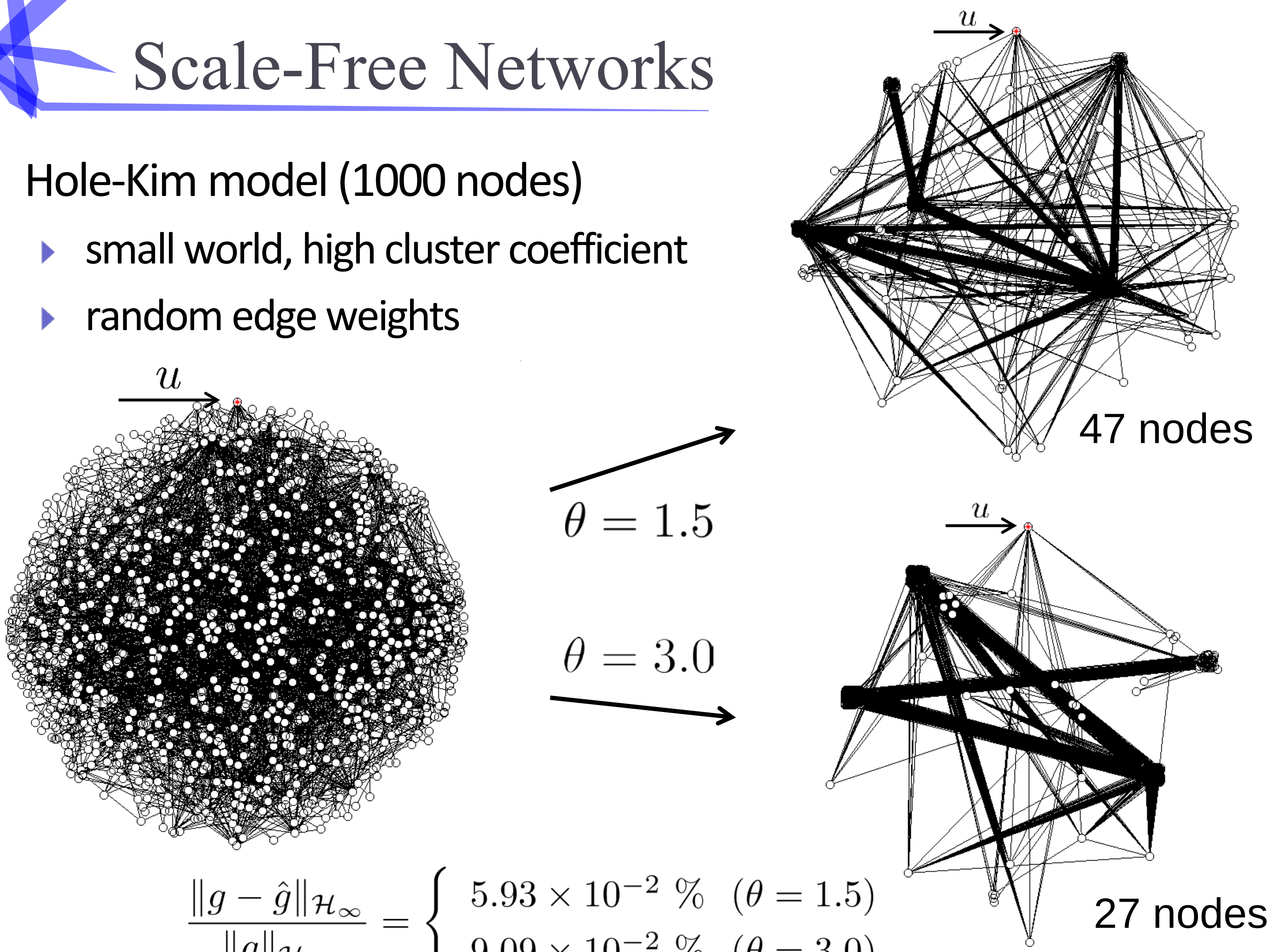
Outline

- Introduction: Why clustered model reduction?
- Clustered Model Reduction Theory
- Examples
- Conclusion



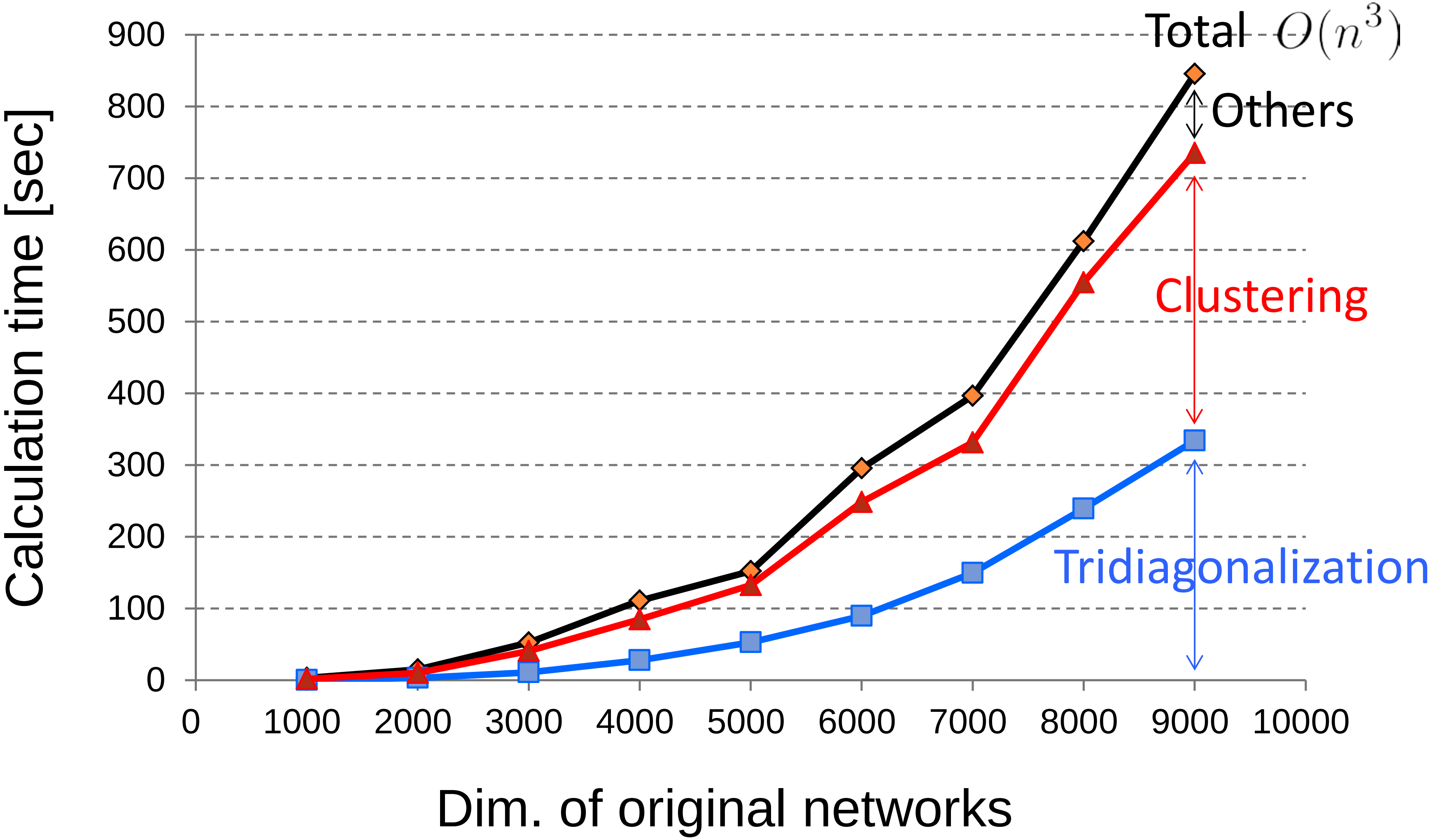
Scale-Free Networks

- Hole-Kim model (1000 nodes)
  - small world, high cluster coefficient
  - random edge weights



$$\frac{\|g-\hat{g}\|_{\mathcal{H}_\infty}}{\|g\|_{\mathcal{H}_\infty}} = \begin{cases} 5.93 \times 10^{-2}\ \% & (\theta = 1.5) \\ 9.09 \times 10^{-2}\ \% & (\theta = 3.0) \end{cases}$$

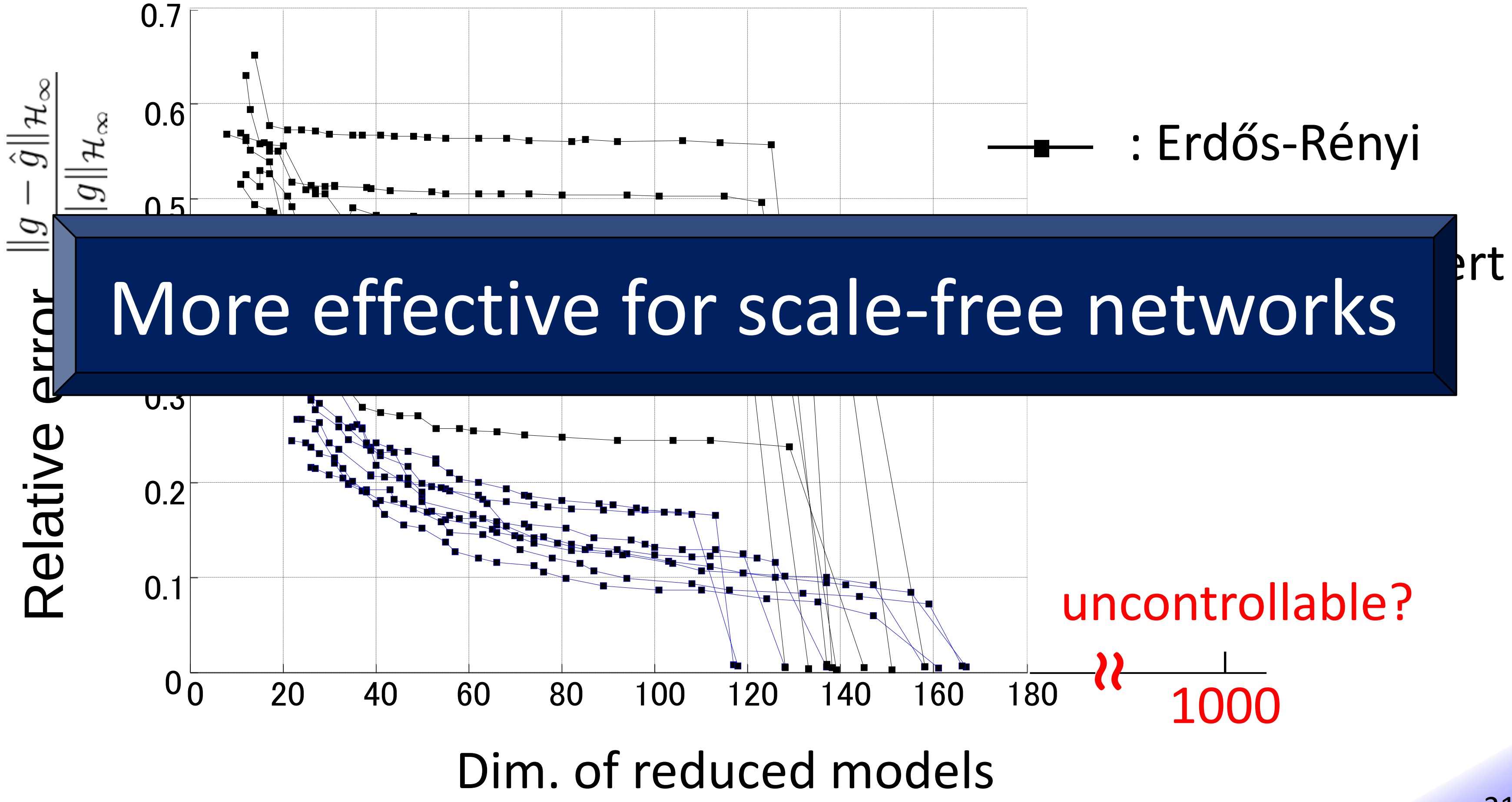
Computational Complexity



Matlab R2011b 64bit Intel® Core™ i7-2620M CPU @2.70GHz, RAM 16.0GB

Scale-Free vs Erdős-Rényi Networks

- 1000 nodes, around 2000 edges, two inputs
- Random edge weights



More effective for scale-free networks

Application to Chemical Reaction

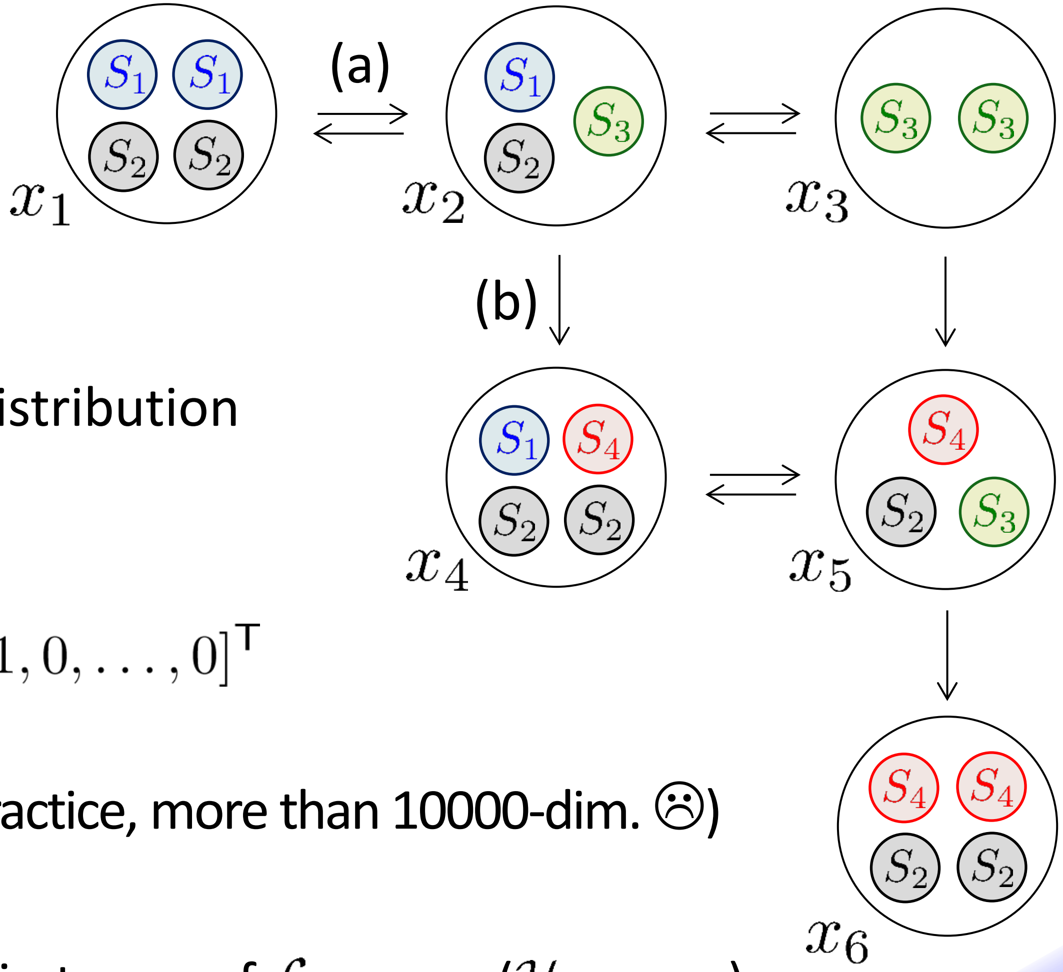
Michaelis-Menten system

$$S_1 + S_2 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_3 \qquad \text{(a)}$$

$$S_3 \xrightarrow{c_3} S_2 + S_4 \qquad \text{(b)}$$

$$(c_1 = 1,\ c_2 = 0.1,\ c_3 = 3)$$

of S_1 & S_2 are initially both $K = 2$



$x_i \in [0, 1]$: Probability of ith distribution

CME expression

$$\dot{x} = Ax, \quad x(0) = [1, 0, \ldots, 0]^{\mathsf{T}}$$

✓ $\frac{(K+1)(K+2)}{2}$-dim. (in practice, more than 10000-dim. ☹)

Approximate $e^{At}x(0)$ in terms of $\mathcal{L}_2$-norm ($\mathcal{H}_2$-norm)

Clustered Model Reduction of CME

$$\dot{x} = Ax, \quad x(0) = x_0 \xrightarrow{Px = \hat{x}} \dot{\hat{x}} = PAP^{\mathsf{T}}\hat{x}, \quad \hat{x}(0) = Px_0$$

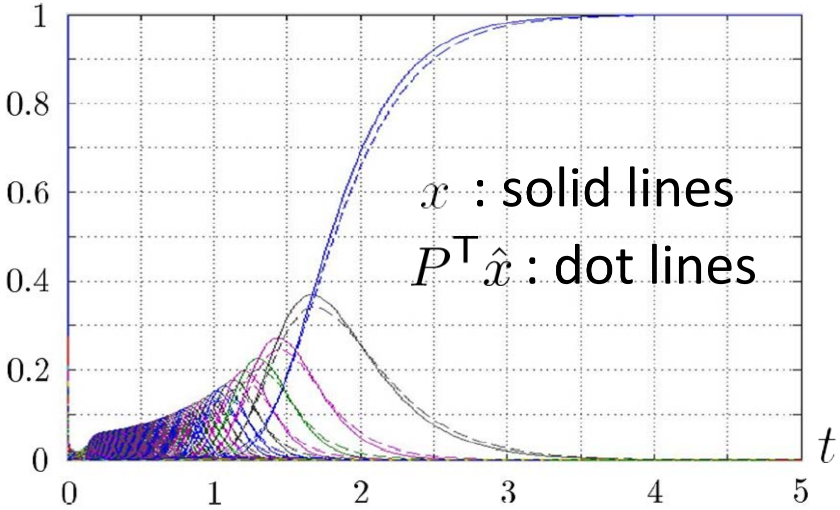
10011 th order

if $\theta = 5 \times 10^{-5}$

1077 th order

✓ **Preservation:** $\underbrace{P^{\mathsf{T}}\hat{x} \in [0,1]^n, \quad \sum_{i=1}^{n}[P^{\mathsf{T}}\hat{x}(t)]_i \equiv 1,}_{\text{property as probability}} \quad \underbrace{x(\infty) = P^{\mathsf{T}}\hat{x}(\infty)}_{\text{steady state}}$

[Trajectories of x and $P^{\mathsf{T}}\hat{x}$]

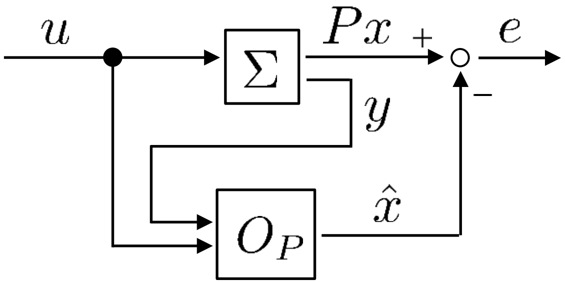


2.4% relative error

954 states are not aggregated

one HUGE cluster is made

Application to Average State Observation



Network system

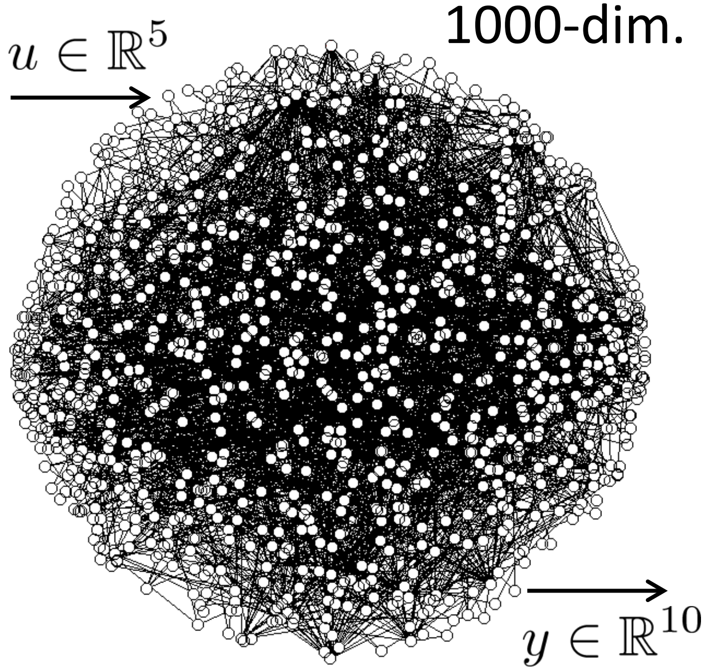
$$\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

Average state observer

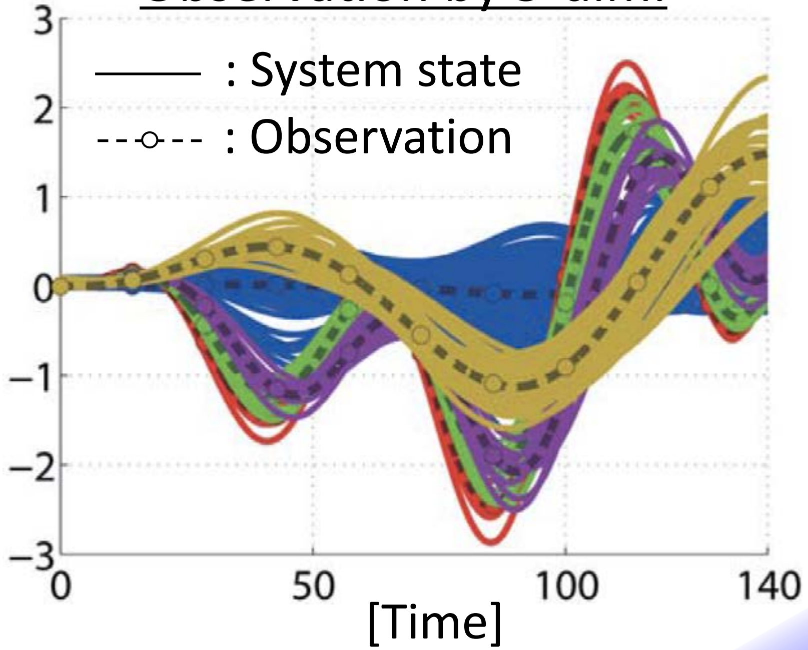
$$O_P : \begin{cases} \dot{\hat{x}} = PAP^{\mathsf{T}}\hat{x} + PBu + H(y - \hat{y}) \\ \hat{y} = CP^{\mathsf{T}}\hat{x} \end{cases}$$

Find $P = \mathrm{Diag}\left(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_L}\right)$ such that $\|Px - \hat{x}\|$ is small enough

1000-dim.



Observation by 5-dim.



Concluding Remarks

- Clustered model reduction
  - extract essential information on input-to-state mapping
  - application to scale-free networks, CMEs, average state observation

- Open problems:
  - approximation of input-to-output mapping
  - more sophisticated clustering algorithm
  - networks of high-dimensional subsystems
  - nonlinear systems
  - application to control system design

Acknowledgements

- Collaborators:
  - Kenji Kashima (Kyoto University)
  - Jun-ichi Imura (Tokyo Institute of Technology)
  - Antoine Girard (University of Grenoble)
  - Luonan Chen (Chinese Academy of Sciences)
  - Kazuyuki Aihara (The University of Tokyo)

Thank you for your attention!