$\mathcal{H}_2$-Clustering of Closed-loop Consensus Networks under a Class of LQR Design

Nan Xue[1] and Aranya Chakrabortty[2]

***Abstract*— Given any positive integer r, our objective is to develop a strategy for grouping the states of a n-node network into $r \leq n$ distinct non-overlapping groups. The criterion for this partitioning is defined as follows. First, a LQR controller is defined for the original n-node network. Then, a r-dimensional reduced-order network is created by imposing a projection matrix P on the n-node open-loop network, and a reduced-order r-dimensional LQR controller is constructed. The resulting controller is, thereafter, projected back to its original coordinates, and implemented in the n-node network. The problem, therefore, is to find a grouping strategy or P that will minimize the difference between the closed-loop transfer matrix of the original network with the full-order controller and that with the projected controller, in the sense of $\mathcal{H}_2$ norm. We derive an upper bound on this difference in terms of P, and, thereby propose a design for P using K-means that tightens the bound while guaranteeing numerical feasibility.**

***Index Terms*— Consensus, Clustering, Model Reduction.**

I. Introduction

Designing control systems for stability and performance of consensus networks has been a problem of perennial interest in network science and engineering. Over the past two decades, seminal papers such as [1]-[2], and references therein, have proposed various sets of control designs for stabilization of complex consensus networks, while those such as [3]-[4] have proposed different optimization and adaptation techniques to enhance their closed-loop performance. Translating these designs to networks with very large number of nodes in practice, however, is a rather challenging task. For example, majority of today's networks, ranging from power system networks to wireless networks to social or biological networks, consist of thousands to tens of thousands of nodes. Designing complicated controllers for such large networks is not only computationally taxing but also severely limited by scalability and tractability. Network engineers are, therefore, often interested in exploring model reduction techniques that can help them in designing simpler controllers by exploiting the inherent topological structure of the network. There exists an extensive literature, developed mostly in the 1980's for large-scale power system networks, on how network structure can be used for open-loop model reduction via *clustering* of network nodes. The fundamental tools used for such clustering include singular perturbation theory [5], Krylov projections [6] and etc. Recent papers such as [7] have presented clustering techniques using $\mathcal{H}_2$ norm optimization. All of these results, however, pertain to open-loop analysis, and not closed-loop control.

To bridge this gap, in this paper we develop a strategy for grouping the closed-loop states of a n-node consensus network into $r \leq n$ distinct non-overlapping groups, where $r > 0$ is a desired number of clusters, specified by the network administrator. The criterion for this partitioning is defined as follows. First, we define a LQR controller for the original n-node network under certain assumptions on Q and R. Note that this LQR controller only needs to be defined in theory for analyses, it does not need to be 'designed' in practice. The closed-loop transfer matrix (TM) of the network output from a disturbance input under this full-order LQR controller is defined as the benchmark TM. Then, a r-dimensional reduced-order network is created by imposing a projection matrix P on the n-node open-loop network, and a reduced-order r-dimensional LQR controller is constructed for this reduced-order system. The resulting controller is, thereafter, projected back to its original coordinates, and implemented in the n-node network. The problem, therefore, is to find a grouping strategy or P that will minimize the difference between the benchmark TM and the closed-loop transfer matrix of the original network with the projected controller in terms of their $\mathcal{H}_2$ norms. We derive an upper bound on this difference as a function of P, and, thereby propose a design for P using K-means that tightens the bound while guaranteeing numerical feasibility.

Clustering of networks, in general, is accepted as a NP-hard problem, and hence a lot of research has been devoted by numerical graph theorists on developing heuristic algorithms over different types of graphs. An extensive survey of these numerical clustering algorithms can be found in [8]. One of such widely used algorithms is K-means [9]. Technically speaking, any clustering algorithm, can be used for constructing P for our problem. For the sake of simplicity, however, in this paper we stick to only K-means.

Notation The following notations will be used throughout this paper: $\mathbf{1}_n$: a column vector of size n with all 1 entries, I_k: the unitary marix of size k, $\|M\|_F$: Frobenius norm of M, i.e. $\|M\|_F = \sqrt{tr(MM^*)}$, $|m|$: absolute value of m, $|\mathcal{S}|_c$: cardinality of a set $\mathcal{S}$, $diag(\mathbf{m})$: diagonal matrix with the vector $\mathbf{m}$ on its diagonal, $[M_{i,j}]$: a matrix whose (i,j) element is defined by $M_{i,j}$. A transfer matrix is defined as $g(s) = C(sI - A)^{-1}B + D$, with a realization form of $g(s) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Furthermore, the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norm of a stable transfer matrix $g(s)$ are defined by $\|g(s)\|_{\mathcal{H}_2} =$

N. Xue and A. Chakrabortty are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, 27695 USA, e-mail: nxue@ncsu.edu, achakra2@ncsu.edu

The work is supported partly by the US Department of Energy grant DE-OE0000654

$\sqrt{\int_{-\infty}^{\infty} tr[g^*(t)g(t)]\mathrm{d}t} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} tr[g^*(j\omega)g(j\omega)]\mathrm{d}\omega}$ and $\|g(s)\|_{\mathcal{H}_\infty} = sup_\omega \bar{\sigma}[g(j\omega)]$, where $\bar{\sigma}$ represents the largest singular value.

II. Problem Formulation

Consider a consensus network defined over an undirected graph $\mathcal{G}$ containing n nodes, each defined by a scalar state variable x_i and a scalar control input u_i, $i = 1, \ldots, n$. To formulate our problem, we define the following four representations of this model - namely, (1) the full-order open-loop model, (2) the full-order closed-loop model with a full-order LQR controller, (3) the reduced-order open-loop model for designing a reduced-order LQR controller, and finally, (4) the full-order closed-loop model with a projected LQR controller. To begin with, the full-order open-loop system is defined as

$$\dot{x}(t) = Ax(t) + Bu(t) + bd(t), \quad y(t) = Cx(t) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^n$ represent the vector of state and control variables. $d(t) \in \mathbb{R}$ is a scalar disturbance input entering any node in the network, while $b \in \mathbb{R}^n$ is a known non-zero vector. $A = A^T \in \mathbb{R}^{n\times n}$ is an edge-weighted graph Laplacian defined by $A(i,j) = \begin{cases} w_{i,j} & i \neq j \\ -\sum_i w_{i,j} & i = j \end{cases}$, where $w_{i,j}$ is the non-negative weight of the edge connecting the i^{th} and the j^{th} node, $B = I_n$ and $C = I_n$ (for full state feedback). By definition, $A \preceq 0$ contains a 0 eigenvalue spanned by $\mathbf{1}_n$. We also assume A to be diagonalizable. The LQR design for system (1) is then posed as finding a full-state feedback $u(t) = -Kx(t)$, where $K = K^T \succ 0$, to minimize the cost function

$$J = \int_0^\infty [x^T(t)Qx(t) + u^T(t)Ru(t)]\mathrm{d}t, \tag{2}$$

where $Q = \gamma I_n$, $R = \upsilon I_n$ and γ and υ are fixed positive scalars. To facilitate the analysis, we denote the transfer matrices (TMs) of the plant and the controller of (1) by

$$G(s) := (sI_n - A)^{-1}, \; K(s) := \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & K \end{array}\right] = K. \tag{3}$$

With this full-order LQR controller, we can write the closed-loop TM from disturbance d to output y as

$$g_c(s) := G(s)[I_n + K(s)G(s)]^{-1}b = G_c(s)b. \tag{4}$$

We next define a state aggregation strategy to repose the control problem using a reduced-order controller as follows.

Definition 2.1: Given an integer $r > 0$, define r non-empty, distinct, and non-overlapping sub-sets of the state variables $\{x_1, ..., x_n\}$, denoted as $\mathcal{S}_1, ..., \mathcal{S}_r$, such that $\mathcal{S}_1 \cup ... \cup \mathcal{S}_r = \{x_1, ..., x_n\}$. The aggregate state $\tilde{x} = [\tilde{x}_1, ..., \tilde{x}_r]^T \in \mathbb{R}^r$ is acquired by $\tilde{x} = Px$, where the projection matrix $P \in \mathbb{R}^{r\times n}$ is defined by

$$P(i,j) := \begin{cases} \frac{1}{\sqrt{|\mathcal{S}_i|_c}} & x_j \in \mathcal{S}_i \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

with $r \leq n$ and $PP^T = I_r$.

For example, P corresponding to a clustering strategy of $\mathcal{S}_1 = \{x_1, x_2, x_3\}$, $\mathcal{S}_2 = \{x_4, x_5\}$ and $\mathcal{S}_3 = \{x_6\}$ can be written as

$$P = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

We next apply the projection P on the open-loop model (1) to define an aggregate model:

$$\dot{\tilde{x}}(t) = \tilde{A}\tilde{x}(t) + \tilde{B}\tilde{u}(t) + \tilde{b}d(t), \quad \tilde{y}(t) = \tilde{C}\tilde{x}(t) \tag{7}$$

where $\tilde{A} = PAP^T \in \mathbb{R}^{r\times r}$, $\tilde{B} = PBP^T = I_r$, $\tilde{b} = Pb \in \mathbb{R}^r$, $\tilde{C} = PCP^T = I_r$. $\tilde{u}(t) \in \mathbb{R}^r$ is similarly projected by $\tilde{u}(t) = Pu(t)$. Next, we design a reduced-order LQR controller using $\tilde{u} = -\tilde{K}\tilde{x}$, $\tilde{Q} = PQP^T = \gamma I_r$ and $\tilde{R} = PRP^T = \upsilon I_r$, where $\tilde{K} = \tilde{K}^T \in \mathbb{R}^{r\times r} \succ 0$, and denote the reduced-order plant and controller, respectively, as

$$\tilde{G}(s) := (sI_r - PAP^T)^{-1}, \; \tilde{K}(s) := \tilde{K}. \tag{8}$$

Finally, we project the reduced-order controller to its original coordinates, and implement it in the full-order model using $u = P^T\tilde{u} = -P^T\tilde{K}Px$ in (1), which implies that in this case the effective feedback matrix is $P^T\tilde{K}P$. Therefore, the closed-loop system implemented with the projected controller can be written as

$$\hat{g}_c(s) := G(s)[I_n + P^T\tilde{K}(s)PG(s)]^{-1}b = \hat{G}_c(s)b. \tag{9}$$

Using these definitions, we next state our problem of $\mathcal{H}_2$-clustering for closed-loop consensus network as follows:

Problem Statement: Given system (1) and an integer $r > 0$, find a clustering strategy $\mathcal{S}_1, ..., \mathcal{S}_r$ such that the corresponding P matrix in (5) minimizes $\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2}$. However, this minimization is, unfortunately, non-convex. Our main contribution in this paper, therefore, is to derive an upper bound for $\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2}$ as a function of P, and thereafter design P using K-means to minimize the bound. Two advantages of our design are:

1) One only needs to design a r-dimensional LQR controller instead of n.
2) The recovered feedback matrix $P^T\tilde{K}P$ has same entries corresponding to the nodes assigned to a cluster. That means we only need to compute one control signal for every node inside a cluster, which simplifies the implementation of the controller, especially when $n \gg r$.

III. Recap of $\mathcal{H}_2$-Clustering for Open-loop Consensus Network

$\mathcal{H}_2$-clustering for a positive system, namely $A \preceq 0$ being a Metzler matrix, has recently been presented in [7]. Since a Laplacian matrix is also Metzler, we recall the main theorem from [7] here.

Theorem 3.1: Given $G(s)$ in (3) and $\tilde{G}(s)$ in (8), define $\Phi := \Phi^{\frac{1}{2}}\Phi^{\frac{T}{2}} = W^T\int_0^\infty e^{WAW^T\tau}e^{WA^TW^T\tau}\mathrm{d}\tau W$, where W is the complement of $\mathbf{1}_n^T/\sqrt{n}$. It follows

$$\|G(s) - P^T\tilde{G}(s)P\|_{\mathcal{H}_2} \leq \|\Xi(s)\|_{\mathcal{H}_\infty}\theta, \tag{10}$$

where $\Xi(s)$ is stable and

$$\Xi(s) := P^T(sI_r - PAP^T)^{-1}PA + I_n,$$
$$\theta := \|P^T P\Phi^{\frac{1}{2}} - \Phi^{\frac{1}{2}}\|_F.$$

Theorem 3.1 provides a metric for clustering an open-loop system, where $\|G(s) - P^T\tilde{G}(s)P\|_{\mathcal{H}_2}$ is upper bounded by a function of θ. θ represents the K-means error of $\Phi^{\frac{1}{2}}$. Therefore, P in (7) was found by providing $\Phi^{\frac{1}{2}}$ as an input to K-means algorithm. This approach, however, is not directly applicable for our problem as explained in the following.

Since $A = A^T$, a straightforward expression for $g_c(s) - \hat{g}_c(s)$ from (4) and (9) can be written as

$$g_c(s) - \hat{g}_c(s) = \hat{G}_c(s)[P^T\tilde{K}(s)P - K(s)]G_c(s)b, \quad (11)$$

which further yields the inequality

$$\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} \leq \|P^T\tilde{K}P - K\|_F\|\hat{G}_c(s)\|_{\mathcal{H}_\infty}\|g_c(s)\|_{\mathcal{H}_\infty}. \quad (12)$$

Since $g_c(s)$ and $\hat{G}_c(s)$ are stable or, equivalently, $\|g_c(s)\|_{\mathcal{H}_\infty}$ and $\|\hat{G}_c(s)\|_{\mathcal{H}_\infty}$ are both bounded, the $\mathcal{H}_2$ norm of $g_c(s) - \hat{g}_c(s)$ is linearly bounded by $\|P^T\tilde{K}P - K\|_F$. Therefore, by Theorem 3.1 a natural choice for P could have followed from minimizing $\|P^T\tilde{K}P - K\|_F$ with respect to P. However, different from (10), the difficulty of this problem lies in the fact that a direct projection P only applies from (1) to (7) but not from K to $\tilde{K}$. In other words, Theorem 3.1 cannot quantify the relationship between P and $\|P^T\tilde{K}P - K\|_F$. Hence, in the following section, we present an alternative approach for solving the problem by constructing a new bound on $\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2}$, which reveals the rule for selecting P.

IV. $\mathcal{H}_2$-Clustering for Closed-loop Consensus Network

To circumvent the problem of relating $(P^T\tilde{K}P - K)$ to P, we introduce an intermediate TM $\tilde{g}_c(s)$, which is defined as the TM from d to $P^T\tilde{y}$ as

$$\tilde{g}_c(s) := P^T\tilde{G}(s)[I_r + \tilde{K}(s)\tilde{G}(s)]^{-1}Pb = \tilde{G}_c(s)b, \quad (13)$$

and rewrite the LHS of (12) as

$$\begin{aligned}\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} &= \|g_c(s) - \tilde{g}_c(s) + \tilde{g}_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} \\ &\leq \|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2} + \|\tilde{g}_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2}. \quad (14)\end{aligned}$$

The LHS is bounded by the $\mathcal{H}_2$ norms of two error systems, namely, $g_c(s) - \tilde{g}_c(s)$, which is the error between full-order and reduced-order closed-loop systems, and $\tilde{g}_c(s) - \hat{g}_c(s)$, which is the error between implementing the same controller on reduced-order and full-order open-loop systems. In the following two subsections, we will detail the derivation of these two error terms separately, and then combine them together to form the algorithm for selecting P.

A. Upper Bound on $\|\tilde{g}_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2}$

Theorem 4.1: Given the aggregated model (7) and the LQR feedback controller $\tilde{K}(s)$ in (8), when the recovered controller $P^T\tilde{K}(s)P$ is applied to the open-loop system (1), then the closed-loop TM $\hat{g}_c(s)$ will satisfy

$$\|\tilde{g}_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} \leq \psi\theta, \quad (15)$$

where $\psi = \|\Psi_1(s)\|_{\mathcal{H}_\infty}\|\Psi_2(s)\|_{\mathcal{H}_\infty}$ and

$$\Psi_1(s) = P^T[I_r + \tilde{K}(s)\tilde{G}(s)]^{-T}P[I_n - \Xi(s)] - I_n,$$
$$\Psi_2(s) = [I_n + P^T\tilde{K}(s)PG(s)]^{-1}b,$$

are stable, with the same θ and $\Xi(s)$ as in Theorem 3.1.

Proof: (I): Define U as the complement of P, say $\begin{bmatrix}P^T & U^T\end{bmatrix}^T$ is unitary, then

$$\begin{aligned}\tilde{g}_c(s) - \hat{g}_c(s) &= P^T\tilde{G}(s)PP^T\underbrace{[I_r + \tilde{K}(s)\tilde{G}(s)]^{-1}}_{T_1^{-1}}Pb \\ &\quad - G(s)\underbrace{[I_n + P^T\tilde{K}(s)PG(s)]^{-1}}_{T_2^{-1}}b \\ &= P^T T_1^{-T}PP^T\tilde{G}(s)PT_2T_2^{-1}b - P^T T_1^{-T}PP^T T_1PG(s)T_2^{-1}b \\ &\quad - U^T UG(s)T_2^{-1}b \\ &= P^T T_1^{-T}P[P^T\tilde{G}(s)P - P^T PG(s)]T_2^{-1}b - U^T UG(s)T_2^{-1}b \\ &= P^T T_1^{-T}P[I_n - \Xi(s)]U^T UG(s)T_2^{-1}b - U^T UG(s)T_2^{-1}b \\ &= \{P^T T_1^{-T}P[I_n - \Xi(s)] - I_n\}U^T UG(s)T_2^{-1}b.\end{aligned}$$

Taking norm on both sides, and using $\|U^T UG(s)\|_{\mathcal{H}_2} = \|U^T U\Phi^{\frac{1}{2}}\|_F = \theta$ from Theorem 3.1 proves (15).

(II): The stability of $\Psi_1(s)$ and $\Psi_2(s)$ is equivalent to the stability of T_1^{-1} and T_2^{-1}. Since $\Xi(s)$ is stable by Theorem 3.1, T_1^{-1} and T_2^{-1} can be written as

$$T_1^{-1} = -\tilde{K}(sI_r - \tilde{A} + \tilde{K})^{-1} + I_r, \quad (16)$$
$$T_2^{-1} = -P^T\tilde{K}P(sI_n - A + P^T\tilde{K}P)^{-1} + I_n, \quad (17)$$

where $\tilde{A} - \tilde{K} \prec 0$. The negativeness of $A - P^T\tilde{K}P \prec 0$ holds if and only if any non-zero vector $v \in \mathbb{R}^n$ satisfies

$$v^T(A - P^T\tilde{K}P)v = v^T Av - v^T P^T\tilde{K}Pv < 0,$$

where $v^T Av \leq 0$ and $-v^T P^T\tilde{K}Pv \leq 0$. Note that $v^T Av = 0$ if and only if $v \in span\{\mathbf{1}_n\}$ while $-vP^T\tilde{K}Pv^T = 0$ if and only if $v^T \in span\{W^T\}$. Therefore, it follows that $A - P^T\tilde{K}P \prec 0$. ∎

Remark 4.2: For the interest of this paper, the controller $\tilde{K}(s)$ is limited to only LQR feedback. However, (15) holds, for any arbitrary controller $\tilde{K}(s)$ as long as $\tilde{K}(s) = \tilde{K}^T(s)$, and $\tilde{g}_c(s)$ and $\hat{g}_c(s)$ are internally stable.

B. Two Supporting Lemmas

Before delving into the LQR theory, we first give the definition of Cauchy matrix, and provide two lemmas which will help us in deriving the bound on $\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2}$.

Definition 4.3: Given $\Lambda = diag([\lambda_1, ..., \lambda_n])$ and $M = diag([\mu_1, ..., \mu_r])$, a Cauchy matrix $\mathcal{C}_{\Lambda,M}$ is defined by $[-\frac{1}{\lambda_i + \mu_j}]$, where $\lambda_i + \mu_j \neq 0$ for $i = 1, ..., n$, $j = 1, ..., r$.

The first lemma guides us in constructing the controllability Gramian:

Lemma 4.4: Consider matrices $\bar{A} \in \mathbb{R}^{n\times n}$, $\bar{b} \in \mathbb{R}^n$ with $\bar{A} \prec 0$. Let the eigenvalue decomposition of $\bar{A}$ be written as $\bar{A} = V\Lambda V^{-1}$, then the controllability Gramian $\bar{\Phi} := \int_0^\infty e^{\bar{A}\tau}\bar{b}\bar{b}^T e^{\bar{A}^T\tau}\mathrm{d}\tau$ can be found through $\bar{\Phi} = \mathcal{R}\mathcal{C}_{\Lambda,\Lambda}\mathcal{R}^T$, where the residue matrix $\mathcal{R} := V \cdot diag(V^{-1}\bar{b})$.

The proof follows from the basic definition of controllability Gramian and can be found in [10]. For the same matrices $\bar{A}$ and $\bar{b}$ where $\bar{A} \prec 0$, the next lemma states a general

expression for the $\mathcal{H}_2$ norm of the error between two stable systems.

Lemma 4.5: Given P as defined in (5), and stable TMs $g_1(s) = (sI_n - \bar{A})^{-1}\bar{b}$ and $g_2(s) = P^T(sI_r - P\bar{A}P^T)^{-1}P\bar{b}$, the $\mathcal{H}_2$ norm of the error system $g_1 - g_2$ is given by

$$\|g_1(s) - g_2(s)\|_{\mathcal{H}_2}^2 = trace(R_1\mathcal{C}_{\Lambda_1,\Lambda_1}R_1^T - R_1\mathcal{C}_{\Lambda_1,\Lambda_2}R_2^T - R_2\mathcal{C}_{\Lambda_2,\Lambda_1}R_1^T + R_2\mathcal{C}_{\Lambda_2,\Lambda_2}R_2^T), \quad (18)$$

where $R_1 = V_1 \cdot diag(V_1^{-1}\bar{b})$ and $R_2 = P^TV_2 \cdot diag(V_2^{-1}P\bar{b})$, with the eigenvalue decompositions $\bar{A} = V_1\Lambda_1V_1^{-1}$ and $P\bar{A}P^T = V_2\Lambda_2V_2^{-1}$.

Proof: See Appendix. ∎

C. Upper Bound on $\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2}$

Since $g_c(s)$ and $\tilde{g}_c(s)$ are both equipped with LQR controllers, we first state a few results from LQR theory that will help us derive the bound on $\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2}$. The algebraic Riccati equation (ARE) corresponding to (2) is

$$A^TX + XA + Q - XBR^{-1}B^TX = 0, \quad (19)$$

where $K = R^{-1}B^TX$ and $X \succ 0$. The Hamiltonian matrix corresponding to (19) can be written as

$$H = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} = \begin{bmatrix} A & -\frac{1}{\upsilon}I_n \\ -\gamma I_n & -A \end{bmatrix}. \quad (20)$$

Suppose the columns of the matrix $\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}_{2n\times n}$ span the stable invariant subspace of H, i.e.

$$H\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}\Lambda^- \quad (21)$$

where $\Lambda^- = diag(\begin{bmatrix}\lambda_1^- & \cdots & \lambda_n^-\end{bmatrix}) \prec 0$. The stabilizing solution of the ARE can then be found by $X = Z_2Z_1^{-1}$ [11]. The important property of Hamiltonian is that Λ^- and normalized Z_1, are the eigenvalue and right eigenspace of $A - BK$. Hence,

$$A - BK = Z_1\Lambda^-Z_1^{-1}. \quad (22)$$

For the reduced-order system (7), the corresponding LQR design is posed by the following ARE

$$PA^TP^T\tilde{X} + \tilde{X}PAP^T + PQP^T - \tilde{X}PBR^{-1}B^TP^T\tilde{X} = 0,$$

and the Hamiltonian matrix $\tilde{H} = \begin{bmatrix} P & \\ & P \end{bmatrix} H \begin{bmatrix} P^T & \\ & P^T \end{bmatrix}$ can be similarly written as

$$\tilde{H}\begin{bmatrix} \tilde{Z}_1 \\ \tilde{Z}_2 \end{bmatrix} = \begin{bmatrix} \tilde{Z}_1 \\ \tilde{Z}_2 \end{bmatrix}\tilde{\Lambda}^-, \quad (23)$$

where $\tilde{\Lambda}^- = diag(\begin{bmatrix}\tilde{\lambda}_1^- & \cdots & \tilde{\lambda}_r^-\end{bmatrix}) \prec 0$. Here we denote the eigenvalue decomposition of A and $\tilde{A}$ by

$$A = z\Lambda^Az^T,\ \tilde{A} = \tilde{z}\tilde{\Lambda}^A\tilde{z}^T, \quad (24)$$

where $\Lambda^A = diag([\lambda_1^A \cdots \lambda_n^A])$ and $\tilde{\Lambda}^A = diag([\tilde{\lambda}_1^A \cdots \tilde{\lambda}_r^A])$ are both semi-stable. Hence, we are able to write the Λ^- and Z_1 from H in terms of eigenvalues and eigenvectors of A explicitly as follows.

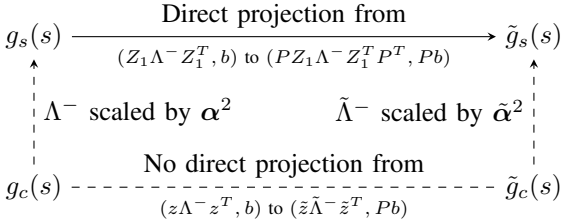

Fig. 1: Relations between $g_c(s)$, $\tilde{g}_c(s)$, $g_s(s)$ and $\tilde{g}_s(s)$

Proposition 4.6: For H and A defined above, the stable eigenvalues of H satisfy:

$$(\lambda_i^A)^2 - (\lambda_i^-)^2 = -\frac{\gamma}{\upsilon} \quad (25)$$

Furthermore $Z_1 = z\boldsymbol{\alpha}$ and $Z_2 = z\boldsymbol{\beta}$, with $\boldsymbol{\alpha}^2 + \boldsymbol{\beta}^2 = I_n$, where $\boldsymbol{\alpha} = diag(\begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix})$, $\boldsymbol{\beta} = diag(\begin{bmatrix}\beta_1 & \cdots & \beta_n\end{bmatrix})$, and for $i = 1, ..., n$

$$\alpha_i = \sqrt{\frac{\frac{1}{\upsilon^2}}{\frac{1}{\upsilon^2} + (\lambda_i^A - \lambda_i^-)^2}},\ \beta_i = \sqrt{\frac{(\lambda_i^A - \lambda_i^-)^2}{\frac{1}{\upsilon^2} + (\lambda_i^A - \lambda_i^-)^2}}. \quad (26)$$

Proposition 4.6 also applies to $(\tilde{H}, \tilde{A}, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$, where $\tilde{Z}_1 = \tilde{z}\tilde{\boldsymbol{\alpha}}$, $\tilde{Z}_2 = \tilde{z}\tilde{\boldsymbol{\beta}}$, by substituting $(\lambda_i^A, \lambda_i^-)$ in (26) with $(\tilde{\lambda}_j^A, \tilde{\lambda}_j^-)$ from $j = 1, ..., r$. With these notations, we can rewrite $g_c(s)$ and $\tilde{g}_c(s)$ by

$$\begin{cases} g_c(s) = (sI_n - z\Lambda^-z^T)^{-1}b \\ \tilde{g}_c(s) = P^T(sI_r - \tilde{z}\tilde{\Lambda}^-\tilde{z}^T)^{-1}Pb \end{cases}. \quad (27)$$

However, from (27), there is still no direct projection between $z\Lambda^-z^T$ and $\tilde{z}\tilde{\Lambda}^-\tilde{z}^T$. To quantify the error between $g_c(s)$ and $\tilde{g}_c(s)$, we next introduce two slack TMs $g_s(s)$ and $\tilde{g}_s(s)$ as

$$g_s(s) := (sI_n - Z_1\Lambda^-Z_1^T)^{-1}b = (sI_n - z\boldsymbol{\alpha}\Lambda^-\boldsymbol{\alpha}z^T)^{-1}b, \quad (28)$$

$$\tilde{g}_s(s) := P^T(sI_r - \tilde{Z}_1\tilde{\Lambda}^-\tilde{Z}_1^T)^{-1}Pb = P^T(sI_r - \tilde{z}\tilde{\boldsymbol{\alpha}}\tilde{\Lambda}^-\tilde{\boldsymbol{\alpha}}\tilde{z}^T)^{-1}Pb = P^T(sI_r - PZ_1\Lambda^-Z_1^TP^T)^{-1}Pb. \quad (29)$$

By definition, $g_s(s)$ and $\tilde{g}_s(s)$ are directly constructed from stable eigenvalues and eigenspace of H and $\tilde{H}$ respectively. Therefore, $\tilde{g}_s(s)$ can be projected from $g_s(s)$ by the projection matrix P, with the corresponding projection error given by Theorem 3.1. Furthermore, the eigenvalues of $g_s(s)$ are the scaled eigenvalues of $g_c(s)$, and similarly those of $\tilde{g}_s(s)$ are the scaled eigenvalues of $\tilde{g}_c(s)$. Using this property, we can approximate $\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2}$ by the next theorem.

Theorem 4.7: Given $g_c(s)$, $\tilde{g}_c(s)$, $g_s(s)$ and $\tilde{g}_s(s)$, with $\mathcal{R}_1 = z \cdot diag(z^Tb)$, $\mathcal{R}_2 = P^T\tilde{z} \cdot diag(\tilde{z}^TPb)$, $g_c(s) - \tilde{g}_c(s)$ satisfies

$$\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2} \le \sqrt{|\|\Xi_s(s)\|_{\mathcal{H}_\infty}^2\theta_s^2 - trace(\xi)|}, \quad (30)$$

where,

$$\theta_s := \|P^TP\Phi_s^{\frac{1}{2}} - \Phi_s^{\frac{1}{2}}\|_F,$$

$$\Phi_s := \Phi_s^{\frac{1}{2}}\Phi_s^{\frac{T}{2}} = \int_0^\infty e^{Z_1\Lambda^-Z_1^T\tau}bb^Te^{Z_1\Lambda^-Z_1^T\tau}\mathrm{d}\tau,$$

$$\Xi_s(s) := P^T(sI_r - PZ_1\Lambda^-Z_1^TP^T)^{-1}PZ_1\Lambda^-Z_1^T + I_n,$$

$$\xi := \mathcal{R}_1O_1\mathcal{R}_1^T - \mathcal{R}_1O_2\mathcal{R}_2^T - \mathcal{R}_2O_2^T\mathcal{R}_1^T + \mathcal{R}_2O_3\mathcal{R}_2^T$$

and O_1, O_2, O_3 denote

$$O_1 = \left[-\frac{\beta_i^2 \lambda_i^- + \beta_j^2 \lambda_j^-}{(\alpha_i^2 \lambda_i^- + \alpha_j^2 \lambda_j^-)(\lambda_i^- + \lambda_j^-)} \right],$$
$$O_2 = \left[-\frac{\beta_i^2 \lambda_i^- + \tilde{\beta}_j^2 \tilde{\lambda}_j^-}{(\alpha_i^2 \lambda_i^- + \tilde{\alpha}_j^2 \tilde{\lambda}_j^-)(\lambda_i^- + \tilde{\lambda}_j^-)} \right],$$
$$O_3 = \left[-\frac{\tilde{\beta}_i^2 \tilde{\lambda}_i^- + \tilde{\beta}_j^2 \tilde{\lambda}_j^-}{(\tilde{\alpha}_i^2 \tilde{\lambda}_i^- + \tilde{\alpha}_j^2 \tilde{\lambda}_j^-)(\tilde{\lambda}_i^- + \tilde{\lambda}_j^-)} \right].$$

Proof: The proof follows directly from Lemma 4.5,

$$\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2}^2 = \|g_s(s) - \tilde{g}_s(s)\|_{\mathcal{H}_2}^2 - trace(\xi), \quad (31)$$

where

$$\xi = \mathcal{R}_1(\mathcal{C}_{\boldsymbol{\alpha}\Lambda\boldsymbol{\alpha},\boldsymbol{\alpha}\Lambda\boldsymbol{\alpha}} - \mathcal{C}_{\Lambda,\Lambda})\mathcal{R}_1^T - \mathcal{R}_1(\mathcal{C}_{\boldsymbol{\alpha}\Lambda\boldsymbol{\alpha},\tilde{\boldsymbol{\alpha}}\tilde{\Lambda}\tilde{\boldsymbol{\alpha}}} - \mathcal{C}_{\Lambda,\tilde{\Lambda}})\mathcal{R}_2^T$$
$$- \mathcal{R}_2(\mathcal{C}_{\tilde{\boldsymbol{\alpha}}\tilde{\Lambda}\tilde{\boldsymbol{\alpha}},\boldsymbol{\alpha}\Lambda\boldsymbol{\alpha}} - \mathcal{C}_{\tilde{\Lambda},\Lambda})\mathcal{R}_1^T + \mathcal{R}_2(\mathcal{C}_{\tilde{\boldsymbol{\alpha}}\tilde{\Lambda}\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\alpha}}\tilde{\Lambda}\tilde{\boldsymbol{\alpha}}} - \mathcal{C}_{\tilde{\Lambda},\tilde{\Lambda}})\mathcal{R}_2^T.$$

Applying Theorem 3.1 to $\|g_s(s) - \tilde{g}_s(s)\|_{\mathcal{H}_2}$, and then expanding the RHS of (31) yields (30). ∎

From Theorem 4.7, the error ξ is governed by $\boldsymbol{\alpha}$, $\tilde{\boldsymbol{\alpha}}$, $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$, which, as expressed in (26), are functions of $(\lambda_i^A - \lambda_i^-)$ or $(\tilde{\lambda}_i^A - \tilde{\lambda}_i^-)$. However, it shows in (25) that the square value of an open-loop eigenvalue λ_i^A is increased by $\frac{\gamma}{\upsilon}$ after applying this particular LQR feedback, which means $(\lambda_i^A - \lambda_i^-)$ and $(\tilde{\lambda}_i^A - \tilde{\lambda}_i^-)$ can be at most $\sqrt{\frac{\gamma}{\upsilon}}$. And if for any i λ_i^A has a large negative value, this value would be very close to λ_i^-. Therefore, we can push the entries in O_1, O_2 and O_3 to 0 by choosing γ and υ such that $\frac{1}{\upsilon} \gg \frac{\gamma}{\upsilon}$, that is by choosing $\gamma \ll 1$, which guarantees $trace(\xi) \approx 0$.

D. Design of Clustering Matrix P

Combining (15) and (30), we get our final result:

$$\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} \leq \sqrt{|\|\Xi_s(s)\|_{\mathcal{H}_\infty}^2 \theta_s^2 - trace(\xi)|} + \psi\theta. \quad (32)$$

We choose $\gamma \ll 1$ such that $trace(\xi) \approx 0$, in which case (32) can be written as,

$$\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} \leq \|\Xi_s(s)\|_{\mathcal{H}_\infty} \theta_s + \psi\theta. \quad (33)$$

The magnitude of the RHS of (33) depends on θ and θ_s. Hence, we can put weights on $\Phi^{\frac{1}{2}}$ and $\Phi_s^{\frac{1}{2}}$, and form

$$\Phi_{inp} := \left[(1-\rho)\Phi_s^{\frac{1}{2}} \quad \rho\Phi^{\frac{1}{2}} \right], \ 0 \leq \rho \leq 1 \quad (34)$$

as the input to a K-means algorithm. Since we are interested in closed-loop clustering, ideally one should choose $\rho = 0$. The trade-off is that the resulting P has no correlation with the structure of open-loop network, which, depending on situations, may or may not be favorable for the control design. Hence, in general, it is advisable to include both $\Phi_s^{\frac{1}{2}}$ and $\Phi^{\frac{1}{2}}$ in Φ_{inp}. The corresponding K-means error $\theta_{inp} := \|P^T P \Phi_{inp} - \Phi_{inp}\|_F$ will satisfy

$$\theta_{inp} = \sqrt{(1-\rho)^2\theta_s^2 + \rho^2\theta^2}. \quad (35)$$

Thus by minimizing θ_{inp}, θ and θ_s will both be reduced, but in different proportions. For a standard K-means algorithm, $\Phi_s^{\frac{1}{2}}$ should be provided along with the number of clusters r. Hence, r and ρ can be both tuned with respect to the refineness of the clustering results. The clustering algorithm is summarized in Algorithm 1. Note that when $\rho = 1$, Algorithm 1 is equivalent to the open-loop clustering.

Algorithm 1: Algorithm for finding $\mathcal{H}_2$ closed-loop clustering matrix P

Input : A, b, Q (γ), R (υ), r and ρ
1 Compute eigenvalue Λ^A and eigenspace z from A;
2 Compute Φ from Theorem 3.1 and Lemma 4.4;
3 Compute closed-loop eigenvalue Λ^-;
4 Calculate $\boldsymbol{\alpha}$ from (26);
5 Compute Φ_s from Theorem 4.7 and Lemma 4.4;
6 Find $\Phi^{\frac{1}{2}}$ and $\Phi_s^{\frac{1}{2}}$ through LDL decomposition, and form Φ_{inp} with ρ from (34);
7 Input Φ_{inp} with r to K-means algorithm;
8 With clustering labels from the K-means results, construct P based on Definition 2.1.
Output: P

V. Simulation Results

We verify our results by simulating a 51-node consensus network, whose structure is shown in Fig. 4. We assume the disturbance to enter from node 4, i.e., b is the 4^{th} column of I_{51}. As we have discussed in Proposition 4.6, the value of $\frac{\gamma}{\upsilon}$ will decide the eigenvalues of the closed-loop system. Hence, considering a fixed ratio of $\frac{\gamma}{\upsilon} = 100$, we choose four pairs of LQR specifications based on the value of $\frac{1}{\upsilon}$ in a descending order as $\{\gamma = 10^{-3}, \upsilon = 10^{-5}\}$, $\{\gamma = 1, \upsilon = 10^{-2}\}$, $\{\gamma = 10, \upsilon = 0.1\}$ and $\{\gamma = 100, \upsilon = 1\}$. The corresponding approximation errors $trace(\xi)$ are shown in Fig. 2. As r, the specified number of clusters, increases to n $trace(\xi)$ for all the cases move to 0. But it is also clear that for this fixed value of $\frac{\gamma}{\upsilon}$, as γ decreases, $trace(\xi)$ becomes sufficiently small, which makes $\|g_s(s) - \tilde{g}_s(s)\|_{\mathcal{H}_2}$ a good approximation for $\|g_c(s) - \tilde{g}_c(s)\|_{\mathcal{H}_2}$. Therefore, in the remaining simulations, we apply $\{\gamma = 10^{-3}, \upsilon = 10^{-5}\}$ as the design parameter for LQR.

To cluster this 51-node network, the weight ρ is provided as an input to Algorithm 1. Fig. 3 shows the ratio $\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2} / \|g_c(s)\|_{\mathcal{H}_2}$ for different values of ρ with respect to the resulting P matrices. When $\rho = 0$, i.e., the input to the K-means is only $\Phi_s^{\frac{1}{2}}$, the resulting clustering strategy outperforms all the other scenarios in approaching the closed-loop $\mathcal{H}_2$ performance of $g_c(s)$. This means that one can achieve $\mathcal{H}_2$ performance close to that of full-order LQR by only designing a much smaller dimensional controller. In contrast, the worst-case scenario given by $\rho = 1$ shows that by applying a LQR controller designed from a model clustered only by open-loop structure, $\|g_c(s) - \hat{g}_c(s)\|_{\mathcal{H}_2}$ is significantly larger than 0 for a small r ($r \leq 20$).

To illustrate the difference of our closed-loop clustering strategy from the open-loop clustering of the system, we draw the network graph for $\rho = 0$ and $\rho = 1$ with $r = 5$

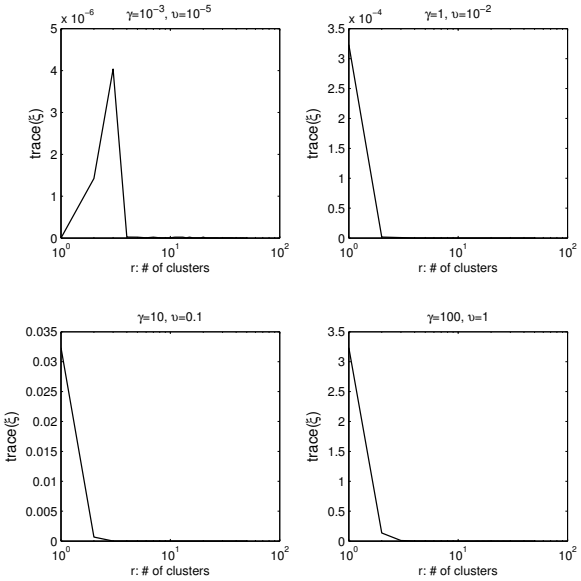

Fig. 2: Values of $trace(\xi)$ with respect to four LQR specifications for a 51-node network

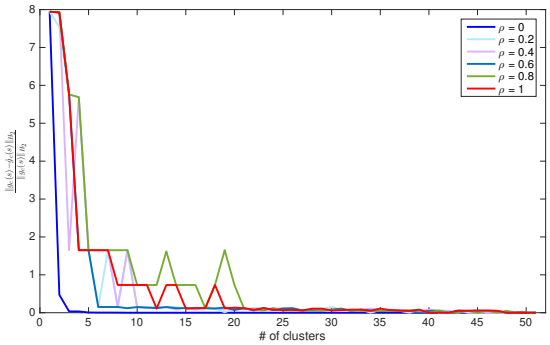

Fig. 3: Ratios of $\frac{\|g_c(s)-\hat{g}_c(s)\|_{\mathcal{H}_2}}{\|g_c(s)\|_{\mathcal{H}_2}}$ with respect to choices of ρ for a 51-node network

clusters as shown in Fig. 4. The natural clustering of this network following from coherency [5] is defined by the nodes set $\{1-10\}$, $\{11-19\}$, $\{20-28\}$, $\{29-39\}$ and $\{40-51\}$. The clusters for $\rho = 1$ coincide exactly with these sets. The sets for $\rho = 0$, however, are very different as clearly visible from Fig. 4. This shows that the similarity of nodes in our problem is not based on coherency, but is rather related to the controllability subspace. It should be noted, however, that the A matrix in (1) does not need to have any natural clustering for our results to be applicable.

VI. Conclusion

In this paper we proposed an algorithm to find a clustering matrix for a consensus network such that the LQR controller designed from the clustered model can approximate the full-order LQR in closed-loop performance. Simulations show that the controller resulting from the clustering when projected on the full-order system, significantly outperforms the controller designed from the open-loop clustering.

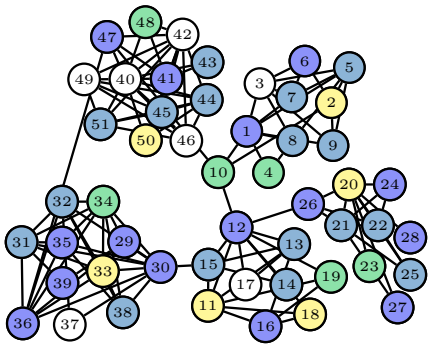

Fig. 4: Clustering of a 51-node network with $r = 5$ and $\rho = 0$ (marked by colors).

APPENDIX

Proof of Lemma 4.5:

$$g_1(s) - g_2(s) = \underbrace{\begin{bmatrix} I & -P^T \end{bmatrix}}_{C_e} (sI - \underbrace{\begin{bmatrix} \bar{A} & \\ & P\bar{A}P^T \end{bmatrix}}_{A_e})^{-1} \underbrace{\begin{bmatrix} \bar{b} \\ P\bar{b} \end{bmatrix}}_{b_e}.$$

The following eigenvalue decomposition is easily found,

$$A_e = V_e \Lambda_e V_e^{-1} = \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \begin{bmatrix} V_1^{-1} & \\ & V_2^{-1} \end{bmatrix}. \quad (36)$$

Hence, $\|g_1(s) - g_2(s)\|_{\mathcal{H}_2}^2 = trace(C_e \Phi_e C_e^T)$, where $\Phi_e := \int_0^\infty e^{A_e \tau} b_e b_e^T e^{A_e^T \tau} \mathrm{d}\tau$ and according to Lemma 4.4,

$$C_e \Phi_e C_e^T = \begin{bmatrix} \mathcal{R}_1 & -\mathcal{R}_2 \end{bmatrix} \begin{bmatrix} \mathcal{C}_{\Lambda_1,\Lambda_1} & \mathcal{C}_{\Lambda_1,\Lambda_2} \\ \mathcal{C}_{\Lambda_2,\Lambda_1} & \mathcal{C}_{\Lambda_2,\Lambda_2} \end{bmatrix} \begin{bmatrix} \mathcal{R}_1^T \\ -\mathcal{R}_2^T \end{bmatrix}^T. \quad (37)$$

Expanding (37) yields $C_e \Phi_e C_e^T = \mathcal{R}_1 \mathcal{C}_{\Lambda_1,\Lambda_1} \mathcal{R}_1^T - \mathcal{R}_1 \mathcal{C}_{\Lambda_1,\Lambda_2} \mathcal{R}_2^T - \mathcal{R}_2 \mathcal{C}_{\Lambda_2,\Lambda_1} \mathcal{R}_1^T + \mathcal{R}_2 \mathcal{C}_{\Lambda_2,\Lambda_2} \mathcal{R}_2^T$, which completes the proof. ■

References

[1] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," *in Proceedings of the IEEE*, vol.95(1), pp.215-233, Jan 2007.

[2] A. Jadbabaie and J. Lin, "Coordination of Groups of Mobile Autonomous Agents Using Nearest Neighbor Rules," *IEEE Transactions on Automatic Control*, vol. 48(6), pp. 988-1001, 2003.

[3] X. Wu and M. R. Jovanovic, "Sparsity-Promoting Optimal Control of Consensus and Synchronization Networks," *in American Control Conference (ACC)*, pp. 2936-2941, June 2014.

[4] H. Bai, M. Arcak and J. Wen, "Adaptive Motion Coordination: Using Velocity Feedback to Track a Reference Velocity," *Automatica*, vol. 45(4), pp. 1020-1025, April 2009.

[5] J. Chow and P. Kokotovic, "Time Scale Modeling of Sparse Dynamic Networks," *IEEE Transactions on Automatic Control*, vol. 30(8), pp. 714-722, 1985.

[6] D. Chaniotis and M. A.Pai, "Model Reduction in Power Systems Using Krylov Subspace Methods," *IEEE Transactions on Power Systems*, vol.20(2), pp. 888-894, May 2005.

[7] T. Ishizaki, K. Kashima, A. Girard, J. Imura, L. Chen and K. Aihara, "Clustering-Based $\mathcal{H}_2$-State Aggregation of Positive Networks and Its Application to Reduction of Chemical Master Equations," 51*st IEEE Conference on Decision and Control (CDC)*, 2012.

[8] S. E. Schaeffer, "Graph Clustering," *Computer Science Review*, vol. 1, pp. 27-64, 2007.

[9] S. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28(2), pp. 129-137, Mar 1982.

[10] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, SIAM, 2005.

[11] K. Zhou and J. C. Doyle, *Essentials of Robust Control*, vol. 180, Prentice Hall Upper Saddle River, NJ, 1998.