This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

Model Reduction of Multi-Agent Systems Using Dissimilarity-Based Clustering

Xiaodong Cheng, *Student Member,* Yu Kawano, *Member,* and Jacquelien M.A. Scherpen, *Senior Member*

***Abstract*—This technical note investigates a model reduction scheme for large-scale multi-agent systems. The studied system is composed of identical linear subsystems interconnected by undirected weighted networks. To reduce the network complexity, a notion of nodal dissimilarity is established on the $\mathcal{H}_2$-norms of transfer function deviations, and a new graph clustering algorithm is proposed to aggregate the pairs of nodes with smaller dissimilarities. The simplified system is verified to preserve an interconnection structure and the synchronization property. Moreover, a computable bound of the approximation error between the full-order and reduced-order models is provided, and the feasibility of the proposed approach is demonstrated by network examples.**

***Index Terms*—Model reduction; Multi-agent systems; Graph clustering; Synchronization.**

I. Introduction

In recent decades, multi-agent systems (or network systems) have received increasing attention from the system and control field [1], [2]. However, multi-agent systems with complex interconnection structures are often modeled by high-dimensional differential equitations, which complicate the analysis, online simulation, controller design, etc. Thus, it is of clear importance to find a less complex model to approximate the input-output characteristics of a full-order model. Meanwhile, due to the need for applications, e.g., distributed controller designs and sensor allocations, the reduced-order model is required to retain a network topology. This technical note aims to lower the complexity of networks by reducing the number of agents.

Conventional model reduction techniques, e.g., balanced truncation and Krylov subspace methods, can produce reduced-order models in systematic ways. However, direct applications of these standard methods to multi-agent systems may lose the network interpretation. In recent years, graph clustering has been employed in the model reduction of network systems (see e.g. [3]–[11]), mainly because this approach potentially preserves the spatial structure of networks and shows an insightful physical interpretation of the reduction process. An early result in [3] interprets the clustering-based approach in the Petrov-Galerkin framework, while it does not discuss how to select clusters, which is the most crucial issue in the clustering-based model reduction, as it determines the quality of the approximation. An *almost equitable partition* (AEP) is suggested by [4], [5] to be treated as a clustering candidate. However, for general graphs, finding AEPs itself is fairly difficult, which causes a major limitation in practical applications. [6] combines graph clustering with balanced truncation approach, where diagonal generalized Gramians are used to identify the importance of edges, and then the nodes connected by less important edges are aggregated. This result is restricted to the networks with tree topologies. Another pioneering approach is proposed in [9], [10] and the references therein. The definition of *reducibility* is introduced, which is characterized by the uncontrollability of clusters. Merging more reducible clusters leads to a reduced-order model that still maintains the network structural information.

In this technical note, we investigate the model reduction problem of multi-agent systems also based on network clustering, where reducing the complexity of underlying networks is of particular interest. Related to this work, our preliminary results on networked single-integrators and double-integrators can be found in [11]–[14]. To characterize a broader class of networks, we consider, in this paper, systems that are composed of identical higher-order linear subsystems interconnecting through a general undirected graph. The notion of *dissimilarity* from [11] is extended to characterize pairwise distances among agents. Specifically, this paper interprets the behaviors of agents as the transfer matrices from external control inputs to the outputs of individual agents, and the generalize dissimilarity between two agents to the $\mathcal{H}_2$-norm of the transfer matrix deviations. In contrast to [9], [10], where the clustering selection requires a prescribed error bound that relies on the positivity of the network system, the proposed framework utilizes a pairwise notion, the vertex dissimilarity, such that a dissimilarity matrix is established. It is an extension and generalization of the concept in conventional clustering problems in data mining, see e.g. [15], [16], where static data objects are classified. Owing to the consistency, many existing clustering algorithms in computer graphics can be adapted to efficiently reduce the complexity of dynamical network systems. Furthermore, the pairwise dissimilarities allow for an easy modification to only aggregate adjacent nodes as in [6]. Finally, the proposed method shows that the simplified model retains the network structure and preserves the synchronization property of the network.

The remainder of this technical note is organized as follows. The models of multi-agent systems and the form of reduced-order models are presented in Section II. In Section III, the cluster selection algorithm is provided based on the concept of dissimilarity, and the $\mathcal{H}_2$ error bound is given. Section IV illustrates the proposed method by simulation examples, and Section V summarizes this technical note.

Notation and terminology: The symbol $\mathbb{R}$ denotes the set of

Xiaodong Cheng, Yu Kawano, and Jacquelien M.A. Scherpen are with Jan C. Willems Center for Systems and Control, Engineering and Technology Institute Groningen, Faculty of Science and Engineering, University of Groningen, Nijenborgh 4, 9747 AG Groningen, the Netherlands. {x.cheng,y.kawano,j.m.a.scherpen}@rug.nl

This work of X. Cheng was supported by Chinese Scholarship Council (CSC) and the work of Y. Kawano was partly supported by JST CREST Grant Number JPMJCR15K2, Japan.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

real numbers, whereas I_n and $\mathbf{1}_n$ represent the identity matrix of size n and all-ones vector of n entries, respectively. The subscript n is omitted when no confusion arises. Moreover, $\mathbf{e}_i$ is the i-th column vector of I_n, and $\mathbf{e}_{ij} := \mathbf{e}_i - \mathbf{e}_j$. The cardinality of set $\mathcal{S}$ is denoted by $|\mathcal{S}|$. The Kronecker product of matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is denoted by $A\otimes B \in \mathbb{R}^{mp\times nq}$. The $\mathcal{H}_\infty$-norm and $\mathcal{H}_2$-norm of the transfer function of a linear system $\mathbf{\Sigma}$ are denoted by $\|\mathbf{\Sigma}\|_{\mathcal{H}_\infty}$ and $\|\mathbf{\Sigma}\|_{\mathcal{H}_2}$, respectively.

A real square matrix A is called *generalized negative definite* if its symmetric part $A_s = \frac{1}{2}(A + A^T)$ is strictly negative definite [17]. If A is generalized negative definite, then A is also Hurwitz.

II. Multi-agent system & Reduced Model

A. Multi-agent systems

Consider a multi-agent system consisting of identical the agent dynamics as

$$\begin{cases} \dot{x}_i = Ax_i + Bv_i, \\ y_i = Cx_i, \end{cases} \tag{1}$$

where $x_i \in \mathbb{R}^{\bar{n}}$, $v_i, y_i \in \mathbb{R}^{\bar{m}}$ are the state, control input and measured output of agent i, respectively. A diffusive coupling rule is applied such that

$$m_i v_i = -\sum_{j=1, j\neq i}^{n} w_{ij}\left(y_i - y_j\right) + \sum_{j=1}^{p} f_{ij}u_j, \tag{2}$$

where $m_i \in \mathbb{R} > 0$ is the inertia of node i, and $u_j \in \mathbb{R}$ with $j = \{1, 2, \cdots, p\}$ are external control signals. Furthermore, $f_{ij} \in \mathbb{R}$ represents the amplification of u_j acting on node i, and w_{ij} stands for the intensity of the coupling between nodes i and j. By (1) and (2), we establish a compact model describing the dynamics of the overall network. Let $F \in \mathbb{R}^{n\times p}$ be the collection of f_{ij} and denote inertia matrix $M := \mathrm{diag}\left(m_1, m_2, \cdots, m_n\right) \in \mathbb{R}^{n\times n}$. We then obtain

$$\mathbf{\Sigma}: \begin{cases} (M\otimes I_{\bar{n}})\dot{x} = (M\otimes A - L\otimes BC)\,x + (F\otimes B)u, \\ y = (I\otimes C)x. \end{cases} \tag{3}$$

with a combined state vector $x^T := \left[x_1^T, x_2^T, \cdots, x_n^T\right] \in \mathbb{R}^{n\bar{n}}$, external control inputs $u^T := \left[u_1^T, u_2^T, \cdots, u_p^T\right] \in \mathbb{R}^{p\bar{m}}$, and external measurements $y^T := \left[y_1^T, y_2^T, \cdots, y_n^T\right] \in \mathbb{R}^{n\bar{m}}$. In the model, $L \in \mathbb{R}^{n\times n}$ is the *Laplacian matrix* of the underlying graph, whose (i, j) entry is given by

$$L_{ij} = \begin{cases} \sum_{j=1, j\neq i}^{n} w_{ij}, & i = j \\ -w_{ij}, & \text{otherwise.} \end{cases} \tag{4}$$

The Laplacian matrix L indicates the interconnection topology and the edge weights of $\mathcal{G}$. Assume the multi-agent system is evolving over a connected, weighted undirected graph, then $L^T = L \succcurlyeq 0$ and $\mathtt{ker}(L) = \mathtt{span}(\mathbf{1}_n)$.

B. Reduced Model

In graph theory, clustering is an important tool to simplify the topology of a complex graph and capture its essential structure. This idea is applied to dynamical networks in this section. Before proceeding, relevant concepts are recalled from [4], [18] as follows.

Definition 1. *Let $\mathcal{G}$ be a connected graph with a nonempty node set $\mathcal{V}$. Then, a **graph clustering** partitions $\mathcal{V}$ into r nonempty disjoint subsets $\{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_r\}$ covering all the elements in $\mathcal{V}$. Here, $\mathcal{C}_i$ is called a **cluster** of $\mathcal{G}$.*

Definition 2. *Consider a graph clustering $\{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_r\}$ of $\mathcal{G}$ with node set $\mathcal{V}$. If $|\mathcal{V}| = n$, its characteristic matrix $P \in \mathbb{R}^{n\times r}$ is defined as*

$$P := \left[p(\mathcal{C}_1), p(\mathcal{C}_2), \cdots, p(\mathcal{C}_r)\right], \tag{5}$$

where $p(\mathcal{C}_i) \in \mathbb{R}^n$ is the characteristic vector of the cluster $\mathcal{C}_i$ such that the k-th element of $p(\mathcal{C}_i)$ is 1 when $k \in \mathcal{C}_i$ and 0 otherwise.

If n nodes are partitioned into r clusters, the reduced-order model can be formed as

$$\hat{\mathbf{\Sigma}}: \begin{cases} (\hat{M}\otimes I_{\bar{n}})\dot{z} = \left(\hat{M}\otimes A - \hat{L}\otimes B\right)z + (\hat{F}\otimes B)u, \\ \hat{y} = (P\otimes C)z, \end{cases} \tag{6}$$

where $\hat{M} := P^T M P$, $\hat{L} := P^T L P$ and $\hat{F} = P^T F$ with P corresponding characteristic matrix of the clustering. The reduced state z presents the dynamics of clusters, and $\hat{y} = (P\otimes I)z$ provides an approximation of the original outputs y.

Remark 1. *From Definition 2, P is a binary matrix, which satisfies $P\mathbf{1}_r = \mathbf{1}_n$ and $\mathbf{1}_n^T P = [|\mathcal{C}_1|, |\mathcal{C}_2|, \cdots, |\mathcal{C}_r|]$. The specific structure of P guarantees that $\hat{M}$ is diagonal positive definite and $\hat{L}$ is a Laplacian matrix [3], [12], [19]. Hence, the reduced model $\hat{\mathbf{\Sigma}}$ is again in the form of system (3) and can be interpreted as a multi-agent system with less agents. Furthermore, due to $\mathtt{ker}(\hat{L}) = \mathtt{span}(\mathbf{1}_r)$, $\hat{L}$ characterizes a connected reduced graph with r nodes.*

C. Synchronization Preservation

Synchronization is an important property in the context of multi-agent systems. With $u = 0$, the system $\mathbf{\Sigma}$ in (3) *synchronizes* if

$$\lim_{t\to\infty}\left[x_i(t) - x_j(t)\right] = 0,\ \forall\ i, j = \{1, 2, \cdots, n\}. \tag{7}$$

Note that $M^{-1}L$ have only real eigenvalues, which are denoted by $\lambda_1 \geq \cdots \geq \lambda_{n-1} > \lambda_n = 0$. Based on the eigenvalues, the following lemma then provides a sufficient condition for the synchronization of $\mathbf{\Sigma}$.

Lemma 1. *The multi-agent system $\mathbf{\Sigma}$ synchronizes if $A - \lambda_1 BC$ and $A - \lambda_{n-1}BC$ are **generalized negative definite**.*

Proof. Denote $\Phi_i := A - \lambda_i BC$. For any $\lambda_1 \geq \lambda_i \geq \lambda_{n-1}$, there exists a pair of constants c_1, $c_2 \geq 0$ with $c_1 + c_2 = 1$ such that $\Phi_i = c_1\Phi_1 + c_2\Phi_{n-1}$. Observe that

$$\frac{1}{2}\left(\Phi_i + \Phi_i^T\right) = \frac{c_1}{2}\left(\Phi_1 + \Phi_1^T\right) + \frac{c_2}{2}\left(\Phi_{n-1} + \Phi_{n-1}^T\right) \prec 0.$$

Thus, Φ_i is generalized negative definite, which implies that Φ_i is Hurwitz for all $i = 1, 2, \cdots, n-1$. The synchronization of $\mathbf{\Sigma}$ then follows from e.g. [2], [20]. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

Note that the agent system (1) is allowed to be unstable as the synchronization condition in Lemma 1 does not require A to be Hurwitz. However, to avoid the trajectories of agents converging to infinity, we still exclude the agent system (1) from having poles in the open right-half plane. Based on Lemma 1, the following theorem shows that the cluster-based model reduction method preserves the synchronization property in the reduced-order multi-agent system.

Theorem 1. *Consider the original system $\mathbf{\Sigma}$ and the corresponding reduced-order model $\hat{\mathbf{\Sigma}}$ resulting from graph clustering. The eigenvalues of $M^{-1}L$ **interlace** those of $\hat{M}^{-1}\hat{L}$. Moreover, if $\mathbf{\Sigma}$ satisfies the synchronization condition in Lemma 1, then $\hat{\mathbf{\Sigma}}$ also **synchronizes**, and their impulse responses, denoted by $\xi(t)$ and $\hat{\xi}(t)$, converge to the same trajectories as*

$$\lim_{t\to\infty}\xi(t)=\lim_{t\to\infty}\hat{\xi}(t)=\sigma_M^{-1}\mathbf{1}\mathbf{1}^TF\otimes Ce^{At}B, \tag{8}$$

where $\sigma_M=\mathbf{1}^TM\mathbf{1}$.

Proof. It follows from [18] that the eigenvalues of matrix $\hat{M}^{-1/2}\hat{L}\hat{M}^{-1/2}$ *interlace* those of $M^{-1/2}LM^{-1/2}$, since there exists a matrix $S:=M^{1/2}P\hat{M}^{-1/2}\in\mathbb{R}^{n\times r}$ with $S^TS=I$ such that

$$S^T\left(M^{-1/2}LM^{-1/2}\right)S=\hat{M}^{-1/2}\hat{L}\hat{M}^{-1/2}. \tag{9}$$

As $M^{-1}L$ and $\hat{M}^{-1}\hat{L}$ are similar to $M^{-1/2}LM^{-1/2}$ and $\hat{M}^{-1/2}\hat{L}\hat{M}^{-1/2}$, respectively, we obtain that the eigenvalues of $\hat{M}^{-1}\hat{L}$ also interlace those of $M^{-1}L$, i.e.,

$$\lambda_1\geq\hat{\lambda}_i\geq\lambda_{n-1},\forall\ i=1,2,\cdots,r-1, \tag{10}$$

where $\hat{\lambda}_i$ are the i-th largest eigenvalue of $\hat{M}^{-1}\hat{L}$. Moreover, $\mathbf{\Sigma}$ satisfies the synchronization condition in Lemma 1, i.e., $A-\lambda_1BC$ and $A-\lambda_{n-1}BC$ are generalized negative definite, which then leads to the generalized negative definiteness of $A-\hat{\lambda}_iBC$, $\forall\ i=1,2,\cdots,r-1$ due to (10). Thus, system $\hat{\mathbf{\Sigma}}$ also synchronizes by Lemma 1.

Next, we prove that the impulse responses of $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ converge to the same value. The proof of the synchronization of $\mathbf{\Sigma}$ follows from e.g., [2]. Consider the eigenvalue decomposition $M^{-1}L=\mathcal{U}\Lambda_o\mathcal{U}^{-1}$, where $\mathcal{U}\in\mathbb{R}^{n\times n}$ is nonsingular, and

$$\Lambda_o=\begin{bmatrix}0 & \\ & \bar{\Lambda}_o\end{bmatrix}\quad\text{with}\quad\bar{\Lambda}_o=\operatorname{diag}(\lambda_1,\cdots,\lambda_{n-1}). \tag{11}$$

The matrices $\mathcal{U}$ and $\mathcal{U}^{-1}$ are partitioned as

$$\mathcal{U}=\begin{bmatrix}\mathcal{U}_1 & \mathcal{U}_2\end{bmatrix},\ \mathcal{U}^{-1}=\begin{bmatrix}\mathcal{V}_1^T & \mathcal{V}_2^T\end{bmatrix}^T, \tag{12}$$

where $\mathcal{V}_1^T,\mathcal{U}_1\in\mathbb{R}^{n\times1}$ are the left and right eigenvectors corresponding to the zero eigenvalue, respectively. Here, $\mathcal{U}_1$ is a unit vector, and we have

$$(M^{-1}L)^T\mathcal{V}_1^T=0,\ (M^{-1}L)\mathcal{U}_1=0\text{ and }\mathcal{V}_1\mathcal{U}_1=1, \tag{13}$$

which yields

$$\mathcal{V}_1=\sqrt{n}\sigma_M^{-1}\mathbf{1}^TM,\ \text{and}\ \mathcal{U}_1=\frac{\mathbf{1}}{\sqrt{n}}. \tag{14}$$

Note that

$$\begin{aligned}e^{\left[I\otimes A-(M^{-1}L)\otimes BC\right]t}&=(\mathcal{U}\otimes I)e^{(I\otimes A-\Lambda_o\otimes BC)t}(\mathcal{U}^{-1}\otimes I)\\&=\mathcal{U}_1\mathcal{V}_1\otimes e^{At}+\mathcal{U}_2\mathcal{V}_2\otimes e^{\left(I_{n-1}\otimes A-\bar{\Lambda}_o\otimes BC\right)t},\end{aligned}$$

where $I_{n-1}\otimes A-\bar{\Lambda}\otimes BC$ is Hurwitz by Lemma 1. Therefore, the impulse response of the original system $\mathbf{\Sigma}$ converges as

$$\begin{aligned}\xi(t)=&(I\otimes C)\left[e^{\left(I\otimes A-M^{-1}L\otimes BC\right)t}\right](M^{-1}F\otimes B)\\&\to\mathcal{U}_1\mathcal{V}_1M^{-1}F\otimes Ce^{At}B,\ \text{as}\ t\to\infty,\\&=\sigma_M^{-1}\mathbf{1}\mathbf{1}^TF\otimes Ce^{At}B.\end{aligned} \tag{15}$$

Similarly, the impulse response of the reduced-order system $\hat{\mathbf{\Sigma}}$ is given by

$$\begin{aligned}\hat{\xi}(t)\to&(P\otimes C)\left(\sigma_M^{-1}\mathbf{1}_r\mathbf{1}_r^T\hat{F}\otimes e^{At}B\right),\ \text{as}\ t\to\infty,\\&=\sigma_M^{-1}P\mathbf{1}_r\mathbf{1}_r^TP^TF\otimes Ce^{At}B\\&=\sigma_M^{-1}\mathbf{1}_n\mathbf{1}_n^TF\otimes Ce^{At}B.\end{aligned}$$

To obtain the above result, the equations $P\mathbf{1}_r=\mathbf{1}_n$ and $\mathbf{1}_r^TP^TMP\mathbf{1}_r=\mathbf{1}_n^TM\mathbf{1}_n=\sigma_M^{-1}$ are used. That completes the proof. □

III. Approximation of Network Systems

A clustering-based model reduction framework for multi-agent systems is proposed in this section. Denote the transfer matrices of system $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ by

$$\eta(s)=(I\otimes C)\left[M\otimes(sI-A)+L\otimes BC\right]^{-1}(F\otimes B), \tag{16a}$$

$$\hat{\eta}(s)=(P\otimes C)\left[\hat{M}\otimes(sI-A)+\hat{L}\otimes BC\right]^{-1}(\hat{F}\otimes B). \tag{16b}$$

Then the transfer matrices from the external inputs to the outputs of individual subsystem are expressed as

$$\eta_i(s):=(\mathbf{e}_i^T\otimes I)\eta(s),\ \hat{\eta}_i(s):=(\mathbf{e}_i^T\otimes I)\hat{\eta}(s). \tag{17}$$

As a natural outcome of Theorem 1, the following corollary implies that the approximation error between $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ is always bounded, even if $\mathbf{\Sigma}$ is not asymptotically stable.

Corollary 1. *Consider the multi-agent system $\mathbf{\Sigma}$ and the reduced model $\hat{\mathbf{\Sigma}}$ resulting from an arbitrary clustering. Then, $\|\eta(s)-\hat{\eta}(s)\|_{\mathcal{H}_2}$ is always bounded.*

Proof. Denote $\Xi_{ij}(t):=(\mathbf{e}_i^T\otimes I)\xi(t)-(\mathbf{e}_j^T\otimes I)\hat{\xi}(t)$, where $\xi(t)$ and $\hat{\xi}(t)$ are the impulse responses of the system $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$, respectively. Then, by the definition of $\mathcal{H}_2$-norm in [23], we obtain

$$\|\eta_i(s)-\hat{\eta}_j(s)\|_{\mathcal{H}_2}^2=\int_0^\infty\mathtt{tr}\left[\Xi_{ij}^T(t)\Xi_{ij}(t)\right]dt, \tag{18}$$

Note that $\xi(t)$ and $\hat{\xi}(t)$ are bounded smooth functions of t. It follows from $\mathbf{e}_{ij}^T\mathbf{1}=0$ that

$$\lim_{t\to\infty}\Xi_{ij}^T(t)=\sigma_M^{-1}\mathbf{e}_{ij}^T\mathbf{1}\mathbf{1}^TF\otimes\left(\lim_{t\to\infty}Ce^{At}B\right)=0. \tag{19}$$

Thus, for bounded initial conditions $\xi_i(0)$ and $\hat{\xi}_j(0)$, the integral in (18) is bounded, i.e., $\|\eta_i(s)-\hat{\eta}_j(s)\|_{\mathcal{H}_2}^2<\infty$. Consequently, $\|\eta(s)-\hat{\eta}(s)\|_{\mathcal{H}_2}^2$ is finite. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

A. Vertex Dissimilarity

The transfer matrix $\eta_i(s)$ defined in (17) represents the mapping from the external control signals u to the outputs x_i, which can be interpreted as the behavior of the i-th agent. Thus, we define the dissimilarity of two nodes as the differences in their behaviors.

Definition 3. *Consider the multi-agent system $\mathbf{\Sigma}$ in (3), the **dissimilarity** of nodes i and j is defined by*

$$\mathcal{D}_{ij} := \|\eta_i(s) - \eta_j(s)\|_{\mathcal{H}_2}. \tag{20}$$

Particularly, if $\mathcal{D}_{ij} = 0$, nodes i and j are 0-dissimilar.

The dissimilarity matrix (or distance matrix), defined by $\mathcal{D} := [\mathcal{D}_{ij}]$, is nonnegative, symmetric, and with zero diagonal elements. The concept of dissimilarity matrix is commonly used in signal processing, as it describes a pairwise distance between two observations. Conventionally, the dissimilarity is characterized by the Euclidean distance, see e.g., [21], [22]. However, Definition 3 extends the domain of this concept to the norm of the difference between nodal dynamics. The idea to measure the similarity of transfer functions for clustering of dynamical networks can be also seen in [8], [9]. When each cluster only has two nodes, the notion of the dissimilarity in (20) coincides with that of the cluster reducibility in [8], [9].

An efficient computation of the $\mathcal{H}_2$-norm in (20) requires the controllability Gramian of $\mathbf{\Sigma}$, which however may not exist when $\mathbf{\Sigma}$ is not asymptotically stable [11]. Inspired by [9], we extract out the asymptotically stable parts from $\mathbf{\Sigma}$ by a specific transformation and employ the controllability Gramian of the asymptotically stable system, we develop an efficient way for the computation of the pairwise dissimilarities.

Theorem 2. *Consider the multi-agent system $\mathbf{\Sigma}$ in (3) that synchronizes. Denote*

$$\mathcal{S}_n := \frac{1}{n}\left[\mathbf{1}_{n-1}\mathbf{1}_{n-1}^T - nI_{n-1}, \mathbf{1}_{n-1}\right] \in \mathbb{R}^{(n-1)\times n}. \tag{21}$$

and $\bar{\mathcal{P}} \in \mathbb{R}^{\bar{n}(n-1)\times\bar{n}(n-1)}$ as the unique solution of the Lyapunov equation

$$\bar{A}\bar{\mathcal{P}} + \bar{\mathcal{P}}\bar{A} + \bar{\mathcal{B}}\bar{\mathcal{B}}^T = 0, \tag{22}$$

where

$$\begin{aligned}\bar{\mathcal{A}} :&= I_{n-1}\otimes A - (\mathcal{S}\mathcal{S}^T)^{-1}\mathcal{S}M^{-1}L\mathcal{S}^T \otimes BC,\\ \bar{\mathcal{B}} :&= (\mathcal{S}\mathcal{S}^T)^{-1}\mathcal{S}M^{-1}F\otimes B.\end{aligned} \tag{23}$$

*Then, the (i,j) entry of the **dissimilarity matrix** is computed as follows:*

- $\mathcal{D}_{ii} = 0$, *if* $i \in \{1,2,\cdots,n\}$;
- $\mathcal{D}_{ni}^2 = \mathcal{D}_{in}^2 = \mathtt{tr}\left[(\mathbf{e}_i^T\otimes C)\bar{\mathcal{P}}(\mathbf{e}_i\otimes C)\right]$, *if* $i \in \{1,2,\cdots,n-1\}$;
- $\mathcal{D}_{ij}^2 = \mathcal{D}_{ji}^2 = \mathtt{tr}\left[(\mathbf{e}_{ij}^T\otimes C)\bar{\mathcal{P}}(\mathbf{e}_{ij}\otimes C)\right]$, *if* $i,j \in \{1,2,\cdots,n-1\}$.

where $\mathbf{e}_i$, $\mathbf{e}_{ij} \in \mathbb{R}^{n-1}$.

Proof. Consider the following transformation matrices

$$\mathcal{T}_n := \begin{bmatrix}\mathcal{S}_n^T & \frac{1}{n}\mathbf{1}_n\end{bmatrix}, \mathcal{T}_n^{-1} = \begin{bmatrix}(\mathcal{S}_n\mathcal{S}_n^T)^{-1}\mathcal{S}_n\\ \mathbf{1}_n^T\end{bmatrix}, \tag{24}$$

and define new state variables

$$\delta := (\mathcal{T}_n^{-1}\otimes I_{\bar{n}})x = \left(\begin{bmatrix}-I_{n-1} & \mathbf{1}_{n-1}\\ \mathbf{1}_{n-1}^T & 1\end{bmatrix}\otimes I_{\bar{n}}\right)x := \begin{bmatrix}\delta_d\\ \delta_a\end{bmatrix},$$

where

$$\begin{aligned}\delta_d &= ([-I_{n-1}, \mathbf{1}_{n-1}]\otimes I_{\bar{n}})\,x \in \mathbb{R}^{m(n-1)},\\ \delta_a &= (\mathbf{1}_n^T\otimes I_{\bar{n}})x \in \mathbb{R}^m.\end{aligned} \tag{25}$$

Note that $(\mathbf{e}_i^T\otimes I_{\bar{n}})\delta_d \in \mathbb{R}^{\bar{n}}$ represents the error between the states of the i-th and the n-th agents, while $\delta_a \in \mathbb{R}^{\bar{n}}$ indicates the average of all the agent states.

We then substitute $x = (\mathcal{T}_n\otimes I_{\bar{n}})\cdot\delta$ to the network model $\mathbf{\Sigma}$ in (3) and multiply $\mathcal{T}_n^{-1}M^{-1}\otimes I_{\bar{n}}$ from the left side. It then leads to an equivalent representation of $\mathbf{\Sigma}$ as

$$\dot{\delta} = (I_n\otimes A - \tilde{L}\otimes BC)\delta + (\tilde{F}\otimes B)u. \tag{26}$$

where

$$\begin{aligned}\tilde{L} &= \begin{bmatrix}(\mathcal{S}_n\mathcal{S}_n^T)^{-1}\mathcal{S}_nM^{-1}L\mathcal{S}_n^T & 0\\ \mathbf{1}^TM^{-1}L\mathcal{S}_n^T & 0\end{bmatrix},\\ \tilde{F} &= \begin{bmatrix}(\mathcal{S}_n\mathcal{S}_n^T)^{-1}\mathcal{S}_nM^{-1}F\\ \mathbf{1}^TM^{-1}F\end{bmatrix}.\end{aligned} \tag{27}$$

It is not hard to see that $\bar{L}$ and $M^{-1}L$ share all the nonzero eigenvalues, and the synchronization of $\mathbf{\Sigma}$ implies that matrix $\bar{A}$ in (23) is Hurwitz [2]. Now, consider an output $y = H\delta$, and denote $\eta_d(s)$ as the transfer function of the system $(\bar{A}, \bar{B}, H)$, whose controllability Gramian $\bar{\mathcal{P}}$ is given by the unique solution of the Lyapunov equation in (22). Thus, it follows from [23] that

$$\|\eta_d(s)\|_{\mathcal{H}_2} = \sqrt{\mathtt{tr}(H\bar{\mathcal{P}}H^T)}, \tag{28}$$

We obtain the pairwise dissimilarities $\mathcal{D}_{ii}$, $\mathcal{D}_{ni}$ and D_{ij}, if H in (28) is replaced by $\mathbf{e}_j^T\otimes C$, $\mathbf{e}_i^T\otimes C$ or $H = \mathbf{e}_{ij}^T\otimes C$, respectively. □

B. Cluster Selection & Error Analysis

The appropriate selection of clusters is crucial for the approximation precision of network reduction. The main contribution of this paper comes from a novel clustering procedure. Compared to the existing results in e.g. [4], [6], [9], our idea is a generalization of the conventional clustering in signal processing [15]. But instead of classifying a large number of static data and measuring their differences by the Euclidean norms, we generalize the method for dynamical systems, where the domain of dissimilarity is extended to Definition 3. With the new notation of dissimilarity, the model reduction problem of networks is connected to the conventional data clustering problems.

Note that the value of $\mathcal{D}_{ij}$ indicates the dissimilarity of agent i and j in terms of transfer functions. Intuitively, clustering the agents with higher similarity can potentially deliver a reduced-order model with smaller approximation error. Based on this idea, standard clustering schemes in signal processing can be adapted to generate a suitable partition of the network, e.g., the iterative clustering in [12] and the hierarchical clustering in [11]. In this paper, a more efficient algorithm is proposed in Algorithm 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

Algorithm 1 Network Clustering Algorithm

1: Compute matrix $\mathcal{D}$ using Theorem 2.
2: Place each node into its own singleton cluster, i.e.,

$$\mathcal{C}_k \leftarrow \{k\},\ \forall\ 1 \le k \le n.$$

3: Find two clusters $\mathcal{C}_\mu$ and $\mathcal{C}_\nu$ such that

$$(\mu, \nu) := \arg\min \left(\max_{i,j \in \mathcal{C}_\mu \cup \mathcal{C}_\nu} \mathcal{D}_{ij} \right). \tag{29}$$

4: Merge clusters $\mathcal{C}_\mu$ and $\mathcal{C}_\nu$ into a single cluster.
5: If there are more than r clusters, repeat the steps 3 and 4. Otherwise, compute $P \in \mathbb{R}^{n\times r}$ and generate

$$\hat{M} \leftarrow P^T M P, \hat{L} \leftarrow P^T L P, \hat{F} \leftarrow P^T F.$$

The proposed algorithm is implicitly based on pairwise dissimilarities of the agents and minimizes within-cluster variances. The variance within a cluster is evaluated by the maximal dissimilarity between all pairs of nodes in the cluster. Note that the formation of clusters in Algorithm 1 does not focus on manipulating any individual edges. Even if two nodes are not adjacent, they can be placed into the same cluster when they have very similar behaviors.

Remark 2. *It should be emphasized that Algorithm 1 can be easily adapted to aggregate adjacent nodes only. To this end, we first introduce the definition of adjacent clusters as follows: Two clusters $\mathcal{C}_\mu$ and $\mathcal{C}_\nu$ are adjacent if there exist $i \in \mathcal{C}_\mu$ and $j \in \mathcal{C}_\nu$ such that nodes i and j are connected directly. Then, we modify step 3 in Algorithm 1 where we can find two adjacent clusters instead such that (29) holds.*

Now the approximation error between the original and reduced multi-agent systems is analyzed using the dissimilarities in (3). For simplicity, we denote

$$\mathcal{S}_r := \frac{1}{r}\left[\mathbf{1}_{r-1}\mathbf{1}_{r-1}^T - rI_{r-1}, \mathbf{1}_{r-1}\right] \in \mathbb{R}^{(r-1)\times r}, \tag{30}$$

and

$$\bar{P} := P\hat{M}^{-1}\mathcal{S}_r^T, \bar{M} := \bar{P}^T M \bar{P}, \bar{L} := \bar{P}^T L \bar{P}. \tag{31}$$

Theorem 3. *Consider the multi-agent system $\mathbf{\Sigma}$ in (3). Suppose $\mathbf{\Sigma}$ synchronizes (i.e., it satisfies Lemma 1). If the graph clustering is given by $\{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_r\}$, then the approximation error between $\mathbf{\Sigma}$ and the clustered system $\hat{\mathbf{\Sigma}}$ is bounded by*

$$\|\eta(s) - \hat{\eta}(s)\|_{\mathcal{H}_2} < \gamma_s \cdot \sum_{k=1}^{r} |\mathcal{C}_k| \cdot \max_{i,j\in\mathcal{C}_k} \mathcal{D}_{ij}, \tag{32}$$

where γ_s is a positive scalar satisfying

$$\begin{bmatrix} \bar{\mathcal{X}} & \bar{P}^T L \otimes BC & -\bar{P}^T \otimes C^T \\ L\bar{P} \otimes C^T B^T & -\gamma_s I & I \\ -\bar{P} \otimes C & I & -\gamma_s I \end{bmatrix} \prec 0, \tag{33}$$

with $\bar{\mathcal{X}} := \bar{M} \otimes (A^T + A) - \bar{L} \otimes (C^T B^T + BC)$. Particularly, if A in (1) is generalized negative definite, i.e., $A + A^T \prec 0$, then γ_s is characterized by

$$\begin{bmatrix} \mathcal{X} & L \otimes BC & -I \otimes C^T \\ L \otimes C^T B^T & -\gamma_s I & I \\ -I \otimes C & I & -\gamma_s I \end{bmatrix} \prec 0, \tag{34}$$

with $\mathcal{X} := M \otimes (A^T + A) - L \otimes (C^T B^T + BC)$.

Proof. The proof can be found in the Appendix. □

Specifically, when A is generalized negative definite, γ_a is obtained by an *a priori* calculation, i.e., its value is only determined by the original system $\mathbf{\Sigma}$. Besides, from (32), we can see that the proposed clustering algorithm is effective, as Algorithm 1 aims to minimize the maximal within-cluster dissimilarity of each cluster such that the sum term in (32) would be smaller. Consequently, the error bound of $\|\eta(s) - \hat{\eta}(s)\|_{\mathcal{H}_2}$ will potentially be lower.

IV. Illustrative Example

A. Path Network

To illustrate the feasibility of our method proposed in Section III, we use the example in [6] for comparison. The thermal model of interconnected rooms in a building is considered, where the network is described by a path graph with 6 nodes, see Fig. 1a, and each room is an agent as in (1) with

$$\begin{aligned} A &= R_i^{-1}\begin{bmatrix} C_1^{-1} & C_1^{-1} \\ C_2^{-1} & C_2^{-1} \end{bmatrix} + R_o^{-1}\begin{bmatrix} C_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}, \\ B &= \begin{bmatrix} C_1^{-1} & 0 \end{bmatrix}^T,\ C = \begin{bmatrix} 1 & 0 \end{bmatrix}. \end{aligned} \tag{35}$$

The meaning of the parameters can be found in [6], which provides their values as $C_1 = 4.35 \cdot 10^4$, $C_2 = 9.24 \cdot 10^6$, $R_i = 2.0 \cdot 10^{-3}$, and $R_o = 2.3 \cdot 10^{-2}$. Moreover, the inertia and Laplacian matrices are given by

$$M = I_6, L = R_w^{-1}\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \tag{36}$$

where $R_w = 1.6 \cdot 10^{-2}$ represents the nominal thermal resistance between two adjacent rooms. The input matrix $F = \begin{bmatrix} \mathbf{e}_3 & R_o^{-1}\mathbf{1}_6 \end{bmatrix}$ indicates the distribution of external inputs.

By Theorem 2, the dissimilarity matrix is computed as $\mathcal{D} =$

$$\begin{bmatrix} 0 & 0.0095 & 0.1332 & 0.0094 & 0.0011 & 0.0028 \\ 0.0095 & 0 & 0.1268 & 0.0004 & 0.0099 & 0.0114 \\ 0.1332 & 0.1268 & 0 & 0.1268 & 0.1332 & 0.1337 \\ 0.0094 & 0.0004 & 0.1268 & 0 & 0.0098 & 0.0112 \\ 0.0011 & 0.0099 & 0.1332 & 0.0098 & 0 & 0.0019 \\ 0.0028 & 0.0114 & 0.1337 & 0.0112 & 0.0019 & 0 \end{bmatrix} \cdot 10^{-3}$$

Then, Algorithm 1 generate a graph clustering: $\{\{1, 5, 6\}, \{2, 4\}, \{3\}\}$, see Fig. 1b, which is different from the result in [6], see Fig. 1c. Taking the output of the third agent as the external output of the whole system, we calculate the approximation error as $\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_{\mathcal{H}_\infty} = 9.4171 \cdot 10^{-5}$, while it is $8.4663 \cdot 10^{-4}$ in [6]. Thus, our method produces a more accurate approximation of the network system. Next, using Remark 2, we adapt Algorithm 1 to only cluster adjacent agents. It then yields an identical clustering result as in [6].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

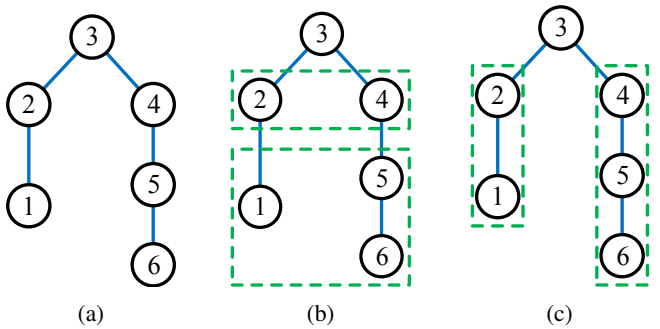

Fig. 1: The clustering of a network that is composed of 6 interconnected rooms. (a) The cluster selection generated by Algorithm 1. (b) The clustering result in [6] and modified Algorithm 1 in Remark 2.

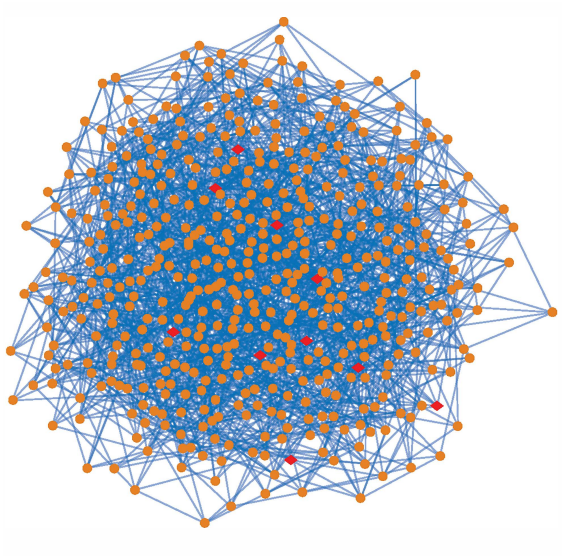

Fig. 2: Watts-Strogatz network with 500 nodes and 2000 edges

B. Small-World Network

Next, the efficiency of the proposed approach is verified by a large-scale small-world network example. This simulation is implemented with Matlab 2016a in the environment of a 64-bit operating system, which is equipped with Intel Core i5-3470 CPU @ 3.20GHz, RAM 8.00 GB.

The agents are oscillators with coefficients

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}, B = C = I_2. \tag{37}$$

The Laplacian matrix L representing the underlying network is created by the Watts-Strogatz model [24], which is a random graph generator producing graphs with small-world properties. In this example, the original network contains 500 nodes and 2000 edges, as shown in Fig. 2. In (3), the diagonal entries of the inertia matrix M are chosen randomly from the range 1 to 10, and $F \in \mathbb{R}^{500\times 10}$ is a binary matrix, whose (i, j) entry is 1 if the the j-th input affects the i-th node, and 0 otherwise. Here, we randomly assign 10 nodes to be controlled. In Fig. 2, the controlled nodes are labeled by diamond markers.

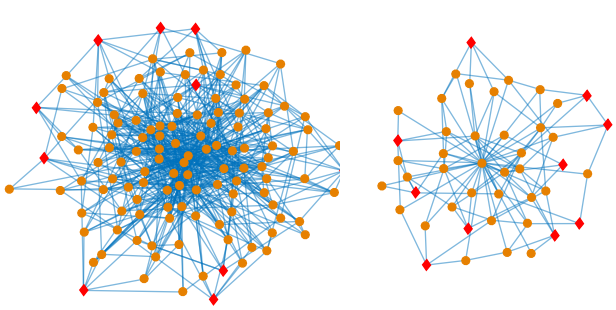

Fig. 3: Reduced Watts-Strogatz networks with 125 clusters (left) and 45 clusters (right)

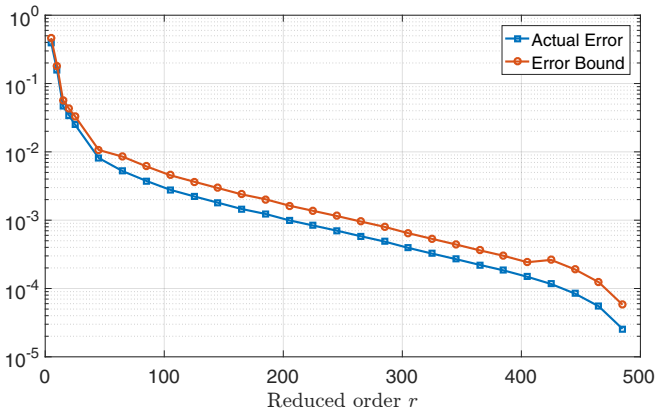

Fig. 4: Approximation error versus reduced order r

TABLE I: The exact approximation errors and the error bounds

Clusters	Actual Error	Error Bound	γ_s
$r = 45$	0.0081	0.107	1.3223
$r = 125$	0.0022	0.0037	1.6348
$r = 225$	$8.4 \cdot 10^{-4}$	$13.7 \cdot 10^{-4}$	1.6348
$r = 425$	$1.2 \cdot 10^{-4}$	$2.6 \cdot 10^{-4}$	2.2598

In the simulation, the original multi-agent system has a dimension of 1000, and we use Algorithm 1 to reduce the number of agents to 5. The approximation errors of the reduced models with different dimensions are shown in Fig. 4, which compares the actual approximation errors and the associated bounds in terms of $\mathcal{H}_2$-norms. From Fig. 4, the exact errors and the error bounds of $\|\mathbf{\Sigma}-\hat{\mathbf{\Sigma}}\|_{\mathcal{H}_2}$ show negative relations with the reduced dimension r. In particular, when $r < 40$, the approximation errors rapidly decrease as r increases. Table I list the actual approximation errors and the error bounds at different reduced order r. The reduced networks with 125 nodes and 45 nodes are shown in Fig. 3. The time for computation of the dissimilarity matrix is approximately 85 seconds, while it only takes 0.004s, on average, for Algorithm 1 to find a suitable clustering.

In conclusion, this simulation example demonstrates that hierarchical clustering algorithm is feasible and effective in model reduction of large-scale multi-agent systems.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

V. Conclusion

In this paper, we propose a general framework of structure preserving model reduction for multi-agent systems. The proposed method builds a connection to the conventional data clustering. The pairwise Euclidean distance in statistical clustering is generalized to the behavior dissimilarity in our framework, which is measured by the norm of transfer function variance. Based on the dissimilarity matrix, which is known as distance matrix in statistical analysis, we are able to adapt the well-developed algorithms for the statistical clustering to solve the model reduction problem. Therefore, the proposed method is a novel extension and generalization of the conventional clustering analysis. Moreover, to generate an appropriate graph clustering, an efficient clustering algorithm is proposed, which can be also easily adapted to only aggregate adjacent agents.

Appendix: Proof of Theorem 3

The approximation error can be evaluated by the following transfer function.

$$\eta_{\mathsf{e}}(s) := \eta(s) - \hat{\eta}(s) = \mathcal{C}_{\mathsf{e}}(sI - \mathcal{A}_{\mathsf{e}})^{-1}\mathcal{B}_{\mathsf{e}}, \tag{38}$$

with

$$\begin{aligned} \mathcal{A}_{\mathsf{e}} &= I_{n+r} \otimes A - \begin{bmatrix} M^{-1}L & \\ & \hat{M}^{-1}\hat{L} \end{bmatrix} \otimes BC, \\ \mathcal{B}_{\mathsf{e}} &= \begin{bmatrix} M^{-1}F \\ \hat{M}^{-1}\hat{F} \end{bmatrix} \otimes B,\ \mathcal{C}_{\mathsf{e}} = \begin{bmatrix} I_n & -P \end{bmatrix} \otimes I_{\bar{n}}. \end{aligned} \tag{39}$$

Inspired by [8], [9], we rewrite the error system into a cascade form. Consider the following nonsingular matrices

$$\mathcal{T} = \begin{bmatrix} 0 & I_n \\ I_r & \Pi \end{bmatrix} \otimes I_{\bar{n}},\ \mathcal{T}^{-1} = \begin{bmatrix} -\Pi & I_r \\ I_n & 0 \end{bmatrix} \otimes I_{\bar{n}}. \tag{40}$$

with $\Pi = \hat{M}^{-1}P^T M \in \mathbb{R}^{r\times n}$. It then follows that

$$\begin{aligned} sI - \mathcal{T}^{-1}\mathcal{A}_{\mathsf{e}}\mathcal{T} &= \begin{bmatrix} \hat{\Omega}(s) & -\Pi M^{-1}L(I - P\Pi) \otimes BC \\ 0 & \Omega(s) \end{bmatrix}, \\ \mathcal{T}^{-1}\mathcal{B}_{\mathsf{e}} &= \begin{bmatrix} 0 \\ M^{-1}F \end{bmatrix} \otimes B, \mathcal{C}_{\mathsf{e}}\mathcal{T} = \begin{bmatrix} -P & I - P\Pi \end{bmatrix} \otimes I, \end{aligned}$$

where $\Omega(s)$ and $\hat{\Omega}(s)$ are transfer functions defined by

$$\begin{aligned} \Omega(s) &:= sI_n - I_n \otimes A + M^{-1}L \otimes BC, \\ \hat{\Omega}(s) &:= sI_r - I_r \otimes A + \hat{M}^{-1}\hat{L} \otimes BC. \end{aligned} \tag{41}$$

Thus, applying a transformation using (40) to (38) leads to

$$\begin{aligned} \eta_{\mathsf{e}}(s) &= \mathcal{C}_{\mathsf{e}}\mathcal{T}(sI - \mathcal{T}^{-1}\mathcal{A}_{\mathsf{e}}\mathcal{T})^{-1}\mathcal{T}^{-1}\mathcal{B}_{\mathsf{e}} \\ &= -(P \otimes C)\hat{\Omega}^{-1}(s)\left[\Pi M^{-1}L(I - P\Pi) \otimes BC\right]\eta(s) \\ &\qquad + \left[(I - P\Pi) \otimes C\right]\eta(s) \\ &= \left[I \otimes C - (P \otimes C)\hat{\Omega}^{-1}(s)(\Pi M^{-1}L \otimes BC)\right] \\ &\qquad \cdot \left[(I - P\Pi) \otimes I\right]\eta(s), \end{aligned} \tag{42}$$

where $\eta(s) = \Omega^{-1}(s)(M^{-1}F \otimes B)$ and

$$\begin{aligned} &\Pi M^{-1}L(I - P\Pi) \otimes BC \\ &\qquad = (\Pi M^{-1}L \otimes BC)\left[(I - P\Pi) \otimes I\right], \end{aligned}$$

are used for derivation of the last expression in (42).

Denote the following two transfer functions

$$\eta_{\mathsf{e}}^a(s) := I \otimes C - (P \otimes C)\hat{\Omega}^{-1}(s)(\hat{M}^{-1}P^T L \otimes BC), \tag{43a}$$

$$\eta_{\mathsf{e}}^b(s) := \left[(I - P\Pi) \otimes I\right]\eta(s). \tag{43b}$$

such that $\eta_{\mathsf{e}}(s) := \eta_{\mathsf{e}}^a(s) \cdot \eta_{\mathsf{e}}^b(s)$. Thus, the approximation error between the original and reduced-order multi-agent systems, $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$, is bounded as

$$\|\eta_{\mathsf{e}}(s)\|_{\mathcal{H}_2} \le \|\eta_{\mathsf{e}}^a(s)\|_{\mathcal{H}_\infty}\|\eta_{\mathsf{e}}^b(s)\|_{\mathcal{H}_2}. \tag{44}$$

In the rest of the proof, we will show the boundedness of $\|\eta_{\mathsf{e}}^a(s)\|_{\mathcal{H}_\infty}$ and $\|\eta_{\mathsf{e}}^b(s)\|_{\mathcal{H}_2}$ and analyze their upper bounds respectively.

First, we discuss the transfer function $\eta_{\mathsf{e}}^a(s)$ in (43a), which is associated with a linear system with coefficient matrices

$$\begin{aligned} A_a &:= I_r \otimes A - \hat{M}^{-1}\hat{L} \otimes BC, \\ B_a &:= \hat{M}^{-1}P^T L \otimes BC, \\ C_a &:= -P \otimes C, \text{and } D_a := I_r \otimes C. \end{aligned} \tag{45}$$

Using the matrix $\mathcal{S}_r$ in (30), which satisfies $\mathcal{S}_r\mathbf{1}_r = 0$, we define transformation matrices

$$\mathcal{T}_r := \begin{bmatrix} \hat{M}^{-1}\mathcal{S}_r^T & \frac{1}{r}\mathbf{1}_r \end{bmatrix}, \mathcal{T}_r^{-1} = \begin{bmatrix} (\mathcal{S}_r\hat{M}^{-1}\mathcal{S}_r^T)^{-1}\mathcal{S}_r \\ \sigma_M^{-1}\mathbf{1}_r^T\hat{M} \end{bmatrix},$$

where $\sigma_M = \mathbf{1}_r^T\hat{M}\mathbf{1}_r = \mathbf{1}_n^T M\mathbf{1}_n$. It then leads to

$$\mathcal{T}_r^{-1}A_a\mathcal{T}_r = \begin{bmatrix} A_s & 0 \\ 0 & A \end{bmatrix}, \mathcal{T}_r^{-1}B_a = \begin{bmatrix} B_s \\ 0 \end{bmatrix}, C_a\mathcal{T}_r = \begin{bmatrix} C_s & * \end{bmatrix}$$

with

$$\begin{aligned} A_s &:= I_{r-1} \otimes A - \bar{M}^{-1}\bar{L} \otimes BC, \\ B_s &:= \bar{M}^{-1}\bar{P}^T L \otimes BC, \\ C_s &:= -\bar{P} \otimes C. \end{aligned} \tag{46}$$

The matrices $\bar{M}$, $\bar{L}$ and $\bar{P}$ are defined in (31).

Clearly, the above transformation splits $\eta_{\mathsf{e}}^a(s)$ as $\eta_{\mathsf{e}}^a(s) = \text{blkdiag}(C, g_s(s))$ with a static gain C and the system

$$g_s(s) := C_s(sI - A_s)^{-1}B_s + I_{r-1} \otimes C. \tag{47}$$

Observe that

$$\begin{aligned} &\bar{M}^{-1}\bar{L} = (\bar{P}^T M\bar{P})^{-1}\bar{P}^T L\bar{P} \\ &= (\mathcal{S}_r^T\hat{M}^{-1}P^T MP\hat{M}^{-1}\mathcal{S}_r^T)^{-1}\mathcal{S}_r^T\hat{M}^{-1}P^T LP\hat{M}^{-1}\mathcal{S}_r^T \\ &= (\mathcal{S}_r\hat{M}^{-1}\mathcal{S}_r^T)^{-1}\mathcal{S}_r(\hat{M}^{-1}\hat{L})\mathcal{S}_r^T. \end{aligned}$$

Thus, $\bar{M}^{-1}\bar{L}$ shares all the nonzero eigenvalues with $\hat{M}^{-1}\hat{L}$. Moreover, Theorem 1 implies that $\hat{\mathbf{\Sigma}}$ also synchronizes as the original multi-agent system $\hat{\mathbf{\Sigma}}$ synchronizes. Then, it follows from [2] that A_s is Hurwitz. Consequently, the transfer function $g_s(s)$ is shown to be asymptotically stable, and hence,

$$\|\eta_{\mathsf{e}}^a(s)\|_{\mathcal{H}_\infty} \le \max\{\|C\|_2, \|g_s(s)\|_{\mathcal{H}_\infty}\} \le \|g_s(s)\|_{\mathcal{H}_\infty}. \tag{48}$$

We use the bounded real lemma (see e.g. [25]) to characterize the $\mathcal{H}_\infty$-norm of $g_s(s)$ in (47). There exists a positive scalar γ_s such that $\|g_s(s)\|_{\mathcal{H}_\infty} < \gamma_s$, if the following inequality holds for a matrix $K \succ 0$.

$$\begin{bmatrix} A_s^T K + KA_s & KB_s & C_s^T \\ B_s^T K & -\gamma_s I & I \\ C_s & I & -\gamma_s I \end{bmatrix} \prec 0. \tag{49}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2853578, IEEE Transactions on Automatic Control

Let $K = \bar{M} \otimes I$ in (49), it then yields the LMI in (33), which is feasible as $\bar{\mathcal{X}}$ is negative definite.

In the special case that $A + A^T \prec 0$, $\mathcal{X}$ is also negative definite so that the LMI in (34) is feasible. Observe that (33) can be obtained from (34) by pre-multiplying and post-multiplying the matrix $\text{blkdiag}(\bar{P}^T, I, I)$ and its transpose, respectively. Thus, γ_s is a solution of (33) if it satisfies (34).

Next, the $\mathcal{H}_2$-norm of transfer function (43b) is discussed. Without loss of generality, let

$$P = \text{blkdiag}\left(\mathbf{1}_{|\mathcal{C}_1|}, \mathbf{1}_{|\mathcal{C}_2|}, \cdots, \mathbf{1}_{|\mathcal{C}_r|}\right),$$
$$M = \text{blkdiag}\left(M_1, M_2, \cdots, M_r\right),$$

with $M_i \in \mathbb{R}^{|\mathcal{C}_i| \times |\mathcal{C}_i|}$. Denote $\hat{m}_i = \mathbf{1}^T M_i \mathbf{1}$, then $\hat{M} = \text{diag}(\hat{m}_1, \hat{m}_2, \cdots, \hat{m}_r)$. Define a vector of transfer functions

$$\left(\eta^{\mathcal{C}_i}\right)^T := \left[\left(\eta_1^{\mathcal{C}_i}\right)^T, \left(\eta_2^{\mathcal{C}_i}\right)^T, \cdots, \left(\eta_{|\mathcal{C}_i|}^{\mathcal{C}_i}\right)^T\right], \tag{50}$$

where $\eta_k^{\mathcal{C}_i}$ represents the behavior of the k-th node in the cluster $\mathcal{C}_i$. We can also write the expression $[(I - P\Pi) \otimes I]\,\eta(s)$ into a block diagonal form whose i-th block diagonal entry is given by

$$\left(\left[I_{|\mathcal{C}_i|} - \mathbf{1}_{|\mathcal{C}_i|}\mathbf{1}_{|\mathcal{C}_i|}^T \hat{m}^{-1} M_i\right] \otimes I\right) \eta^{\mathcal{C}_i} = \begin{bmatrix} \sum_{k \in \mathcal{C}_i, k \neq 1} \hat{m}_i^{-1} m_k \left(\eta_1^{\mathcal{C}_i} - \eta_k^{\mathcal{C}_i}\right) \\ \vdots \\ \sum_{k \in \mathcal{C}_i, k \neq |\mathcal{C}_i|} \hat{m}_i^{-1} m_k \left(\eta_{|\mathcal{C}_i|}^{\mathcal{C}_i} - \eta_k^{\mathcal{C}_i}\right) \end{bmatrix}. \tag{51}$$

It is noted that $\|\eta_j^{\mathcal{C}_i} - \eta_k^{\mathcal{C}_i}\|_{\mathcal{H}_2} \leq \|\eta_{\max}^{\mathcal{C}_i}\|_{\mathcal{H}_2}$, where $\eta_{\max}^{\mathcal{C}_i}$ refers to the biggest divergence of node behaviors within the cluster $\mathcal{C}_i$. It then leads to

$$\left\| \sum_{k \in \mathcal{C}_i, k \neq 1} \hat{m}_i^{-1} m_k \left(\eta_1^{\mathcal{C}_i} - \eta_k^{\mathcal{C}_i}\right) \right\|_{\mathcal{H}_2} \leq \|\eta_{\max}^{\mathcal{C}_i}\|_{\mathcal{H}_2}. \tag{52}$$

Therefore, we obtain

$$\begin{aligned} \|\eta_{\mathsf{e}}^b(s)\|_{\mathcal{H}_2} &\leq \sum_{k=1}^{r} \left\| \left(I_{|\mathcal{C}_k|} - \hat{m}^{-1}\mathbf{1}_{|\mathcal{C}_k|}\mathbf{1}_{|\mathcal{C}_k|}^T M_k\right) \eta^{\mathcal{C}_k} \right\|_{\mathcal{H}_2} \\ &\leq \sum_{k=1}^{r} |\mathcal{C}_k| \cdot \|\eta_{\max}^{\mathcal{C}_k}\|_{\mathcal{H}_2} = \sum_{k=1}^{r} |\mathcal{C}_k| \cdot \max_{i,j \in \mathcal{C}_k} \mathcal{D}_{ij}. \end{aligned}$$

That completes the proof. □

REFERENCES

[1] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[2] Z. Li, Z. Duan, G. Chen, and L. Huang, "Consensus of multiagent systems and synchronization of complex networks: a unified viewpoint," *IEEE Transactions on Circuits and Systems I*, vol. 57, no. 1, pp. 213–224, 2010.

[3] A. J. van der Schaft, "On model reduction of physical network systems," in *Proceedings of 21st International Symposium on Mathematical Theory of Networks and Systems*, 2014, pp. 1419–1425.

[4] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel, "Projection-based model reduction of multi-agent systems using graph partitions," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 2, pp. 145–154, 2014.

[5] H.-J. Jongsma, P. Mlinarić, S. Grundel, P. Benner, and H. L. Trentelman, "Model reduction of linear multi-agent systems by clustering with H_2 and H_∞ error bounds," *Mathematics of Control, Signals, and Systems*, vol. 30, no. 1, p. 6, Apr 2018.

[6] B. Besselink, H. Sandberg, and K. H. Johansson, "Clustering-based model reduction of networked passive systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2958–2973, 2016.

[7] P. Mlinarić, S. Grundel, and P. Benner, "Efficient model order reduction for multi-agent systems using QR decomposition-based clustering," in *Proceedings of 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 4794–4799.

[8] T. Ishizaki, K. Kashima, J. I. Imura, and K. Aihara, "Model reduction and clusterization of large-scale bidirectional networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 48–63, 2014.

[9] T. Ishizaki, K. Kashima, A. Girard, J.-i. Imura, L. Chen, and K. Aihara, "Clustered model reduction of positive directed networks," *Automatica*, vol. 59, pp. 238–247, 2015.

[10] T. Ishizaki, R. Ku, and J.-i. Imura, "Clustered model reduction of networked dissipative systems," in *Proceedings of 2016 American Control Conference*. IEEE, 2016, pp. 3662–3667.

[11] X. Cheng, Y. Kawano, and J. M. A. Scherpen, "Reduction of second-order network systems with structure preservation," *IEEE Transactions on Automatic Control*, vol. 62, pp. 5026 – 5038, 2017.

[12] ——, "Graph structure-preserving model reduction of linear network systems," in *Proceedings of 2016 European Control Conference*, 2016, pp. 1970–1975.

[13] X. Cheng and J. M. A. Scherpen, "Introducing network gramians to undirected network systems for structure-preserving model reduction," in *Proceedings of 55th IEEE Conference on Decision and Control*, 2016, pp. 5756–5761.

[14] X. Cheng, J. M. A. Scherpen, and Y. Kawano, "Model reduction of second-order network systems using graph clustering," in *Proceedings of 55th IEEE Conference on Decision and Control*, 2016, pp. 7471–7476.

[15] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[16] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. CRC press, 2013.

[17] G.-R. Duan and R. J. Patton, "A note on hurwitz stability of matrices," *Automatica*, vol. 34, no. 4, pp. 509–511, 1998.

[18] C. Godsil and G. F. Royle, *Algebraic Graph Theory*. Springer Science & Business Media, 2013, vol. 207.

[19] N. Monshizadeh and A. J. van der Schaft, "Structure-preserving model reduction of physical network systems by clustering," in *Proceedings of 53rd IEEE Conference on Decision and Control*, 2014, pp. 4434–4440.

[20] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel, "Stability and synchronization preserving model reduction of multi-agent systems," *Systems & Control Letters*, vol. 62, no. 1, pp. 1–10, 2013.

[21] Y. Xu, S. M. Salapaka, and C. L. Beck, "Aggregation of graph models and markov chains by deterministic annealing," *IEEE Transactions on Automatic Control*, vol. 59, no. 10, pp. 2807–2812, Oct 2014.

[22] J. Anderson and A. Papachristodoulou, "A decomposition technique for nonlinear dynamical system analysis," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1516–1521, June 2012.

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

[24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[25] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.