AIAA Guidance, Navigation, and Control Conference
08 - 11 August 2011, Portland, Oregon

AIAA 2011-6398

Initial-Condition Estimation in Network Synchronization Processes: Algebraic and Graphical Characterizations of the Estimator

Mengran Xue*, Enoch Yeung♯, Anurag Rai†,
Sandip Roy*, Yan Wan♢, Sean Warnick†

July 14, 2011

Abstract

We pursue a graph-theoretic analysis of state *inference* in network synchronization (or diffusive) processes. Precisely, we study estimation of a non-random initial condition of a canonical synchronization dynamics defined on a graph, from noisy observations at a single network node. By characterizing the maximum-likelihood estimation of the initial condition and the associated Cramer-Rao bound, we identify graph properties (e.g., symmetries, interconnection strengths, spectral measures) that determine 1) whether or not estimation is possible and 2) the quality of the estimate.

1 Introduction

Synchronization processes are common in networks: for instance, groups of fireflies come to flash in unison, reservoir water levels reach equilibrium, autonomous vehicles are designed to move in unison, generators in power systems align in phase, and players in a financial market reach consensus on prices [1–5]. The dynamics of synchronizing networks have been exhaustively studied in the physical and biological sciences, and are well-known to be deeply tied to the topology of the underlying network. Recently, control engineers have also revisited the study of synchronization, with a particular interest in *designing* a network's topology or its nodal controllers to shape an associated synchronization dynamics [6–8].

As the analysis and especially the design of synchronization becomes increasingly important, a rich class of new problems are arising in network synchronization that we believe require the joint attention of physicists and control engineers. These include problems regarding state inference, network-model identification, and fairness in design,

*This work was supported by grants NSF-0528882, NSF-0725589, NSF-0901137, NSF-1058124, and NASA NNA-06CN26A. Correspondence should be sent to mxue@eecs.wsu.edu . Affiliations: ∗ Department of Electrical Engineering, Washington State University; ♯ Department of Computing and Mathematical Sciences, California Institute of Technology; † Information and Decision Algorithms Laboratories, Brigham Young University; ♢ Department of Electrical Engineering, University of North Texas.

among many others. Addressing these tasks for synchronization processes brings forth a core new challenge: the central role played by the network topology should be identified and exploited (in analogy with the analysis of synchronization in the physical sciences), while systematic tools for estimation and controller design (coming from control community) are also brought to bear. Here, we pursue a topology-based study of one new problem in network synchronization, namely the inference of the network's initial state from noisy temporal observations at one network node. Although literature on inference of synchronization processes is quite sparse, it is worth noting a couple related efforts. First, this study complements our previous work on parameter inference (specifically, mode estimation) in synchronization dynamics, see [9]. Our initial-condition estimation study is also closely related to the study of state estimation in synchronization processes given in [10], but with the core difference that we relate estimator structure/performance to the network's topology rather than presenting an estimation algorithm.

The particular network state inference problem pursued here has numerous potential applications in both natural and engineered synchronization processes. These applications are often ones where the initial condition and/or state dynamics must be evaluated from partial and noisy observations. allows *a posteriori* evaluation. For both these applications, an explicit characterization of estimator structure and performance in terms of the underlying network structure is of central interest for several reasons, including for sensor (observation) design, bounding of estimator performance, and network design.

In this article, we develop graph-theoretic characterizations of initial-condition inference in network synchronization as follows. First we pose the initial-condition inference problem as a non-random parameter estimation problem (Section 3). We invoke classical results on inference to obtain algebraic expressions for the optimal estimators and their performance (Sections 3.1). Using these algebraic characterizations and applying various algebraic graph-theory and control theoretic constructs, we then obtain several characterizations of the estimators and their performance in terms of the network's topological structure (Sections 3.2 and Section 3.3).

Although our focus in this article is specifically on inferring the initial state of synchronization processes, we view this work as a component in an encompassing study of network-dynamics inference. The case that we pursue here serves as a prototype for various network dynamics, and hence potentially provides a starting point for dynamics-estimation in such diverse applications as virus-spreading control, sensor networking, and air traffic management. More broadly, we recall that the characterization of observability/controllability in networks and the estimation of state information from local observations plays a crucial role in several decentralized controls methods, e.g., [11]; our efforts here can thus possibly give insight into decentralized control methods, albeit (for now) in a limited class of models.

2 Problem Formulation

We are concerned with initial-condition inference for a network synchronization process, i.e. for a *synchronizing* or *diffusive* dynamics defined on a graph. In this section, let us review the classical model for diffusive network dynamics (see [2, 9] for background) that we use here, and formally introduce the initial-condition inference problem.

Most broadly, we consider a linear time-invariant (LTI) dynamics specified by a weighted and directed graph Γ. Precisely, let us consider a graph $\Gamma = (V,\ E{:}W)$, where the **vertex set** V contains m elements labeled $1, \ldots, m$, the **edge set** E contains q **edges** or ordered pairs of distinct vertices, and each edge (i, j) in E has associated with it a positive weight w_{ij} as given in the **weight set** W. To describe the diffusive network dynamics, we find it convenient to specify an $m \times m$ **diffusion matrix** $\mathcal{L}$ from the graph Γ, as follows:

- We set $\mathcal{L}_{ij}$ equal to $-w_{ji}$, for each ordered pair $(i, j) \in E$.
- We set $\mathcal{L}_{ij}$, $i \neq j$, equal to 0 otherwise.
- We choose $\mathcal{L}_{ii} = -\sum_{j=1, j\neq i}^{n} \mathcal{L}_{ij}$. That is, we choose the diagonal entries so that each row sums to 0.

Now let us specify a network's (discrete-time) dynamics in terms of the diffusion matrix specified above. Specifically, let us consider a network with m components or nodes labelled $1, \ldots, m$, which correspond to the m vertices in the graph Γ. We associate with each component a state $x_i[k]$ that evolves in discrete time (for $k = 0, 1, 2, \ldots$). To specify the network dynamics, let us define a **network state** $\mathbf{x}[k] = \begin{bmatrix} x_1[k] & \cdots & x_m[k] \end{bmatrix}^T$. We consider the following evolution of the network state:

$$\mathbf{x}[k+1] = \mathbf{x}[k] - \delta \mathcal{L} \mathbf{x}[k], \tag{1}$$

where $\delta \in R$ scales the magnitude of the interactions among the network components specified in $\mathcal{L}$ (and, for instance, may represent a time-step in a discretization of a continuous-time process). We note that the dynamics in (1) describes a process of state-equalization through *balancing* or *flow* between each component and its graphical neighbors, and hence can be viewed as a diffusive dynamics. The edge-weight w_{ij} in the graph captures the strength of impact of node i's current state on node j's next state, or in other words the extent of diffusing coupling from i to j.

Our primary focus in this article is to infer the unknown initial state $\mathbf{x}[0]$ of the dynamics above from a sequence of noisy measurements at a single component $j \in \{1, \ldots, m\}$, and specifically to relate the estimator and its performance to the structure of the graph Γ. Specifically, we consider inference from a sequence of observations $y[0], y[1], ...y[n]$, where the observation signal at time step k ($k = 0, 1, \cdots, n$) is given by

$$y[k] = e_j^T \mathbf{x}[k] + g[k], \tag{2}$$

where e_j is a standard basis vector with the unity entry in the j^{th} component, and $g[k]$ is a sample of a scalar Gaussian white noise process with zero mean and variance σ^2. We refer to the component j where the observation is being made as the observation location/node.

In the rest of the paper we will discuss the scenario of initial state estimation, namely estimating a *non-random initial condition* (Section 3), for the diffusive dynamics. We will review the classical algebraic formulae for the estimator and its performance, and then motivate and provide a family of graph-theoretic characterizations.

Holistically, we find it convenient to refer to the state dynamics and observation model described above as a **diffusive network model**, and to refer to the inference problem as

the **initial condition estimation problem**. While many of our graph-theoretic results are for arbitrary diffusive networks, we will also at times limit ourselves to the case that the edge weights satisfy $w_{ij} = w_{ji}$ for each i, j (i.e. the graph Γ is undirected), and so the diffusion matrix is symmetric. The network matrix $\mathcal{L}$ in this case is well-known to be a **Laplacian matrix** (see [12, 13] for many applications of the Laplacian), and so we refer to the state and observation model as a **Laplacian network model** in this case.

3 Non-Random Initial Condition Estimation

In this section, we will consider the initial condition estimation problem where the initial condition $\mathbf{x}[0]$ is non-random, i.e. $\mathbf{x}[0]$ is fixed and unknown with no *a priori* probability distribution. In estimation theory, *maximum likelihood estimation* (MLE) is a classical and popular method to estimate unknown deterministic parameters, which also guarantees that the variance of the estimator achieves the minimum variance specified by the *Cramer-Rao bound* (see [9, 14]). Here given a sequence of the observations with Gaussian white noise added, we will use the MLE method to provide estimates for the non-random initial condition $\mathbf{x}[0]$.

To fully address this non-random estimation case, we will first in Section 3.1 build the maximum likelihood estimator and derive its performance (i.e., the estimator's covariance matrix). Our key effort, to relate the estimator and its performance to the underlying graph structure, is developed in the later subsections (Section 3.2, Section 3.3). We will illustrate the relationship in the following two aspects. First, in Section 3.2, we will give graph-theoretic conditions for whether or not the initial condition can be estimated at all [1]. Then, in the case where the initial condition can be estimated, we will tie the estimator's structure and performance to the underlying graph structure (Section 3.3). Specifically, in Section 3.3.1 we will first study the estimator's asymptotic structure in general. Also, we will use the slow-coherency theory to broadly identify graph structures that, while permitting estimation, are weakly connected and have poor estimator performance (Section 3.3.2). Finally, we will give some graph-theoretic characterizations of estimator performance for more general graph structures and upon changes to the graph (Section 3.3.3).

3.1 Algebraic Expressions for the Estimator and its Performance

Here, we present algebraic expressions for the ML estimator and its error covariance, for the non-random initial-condition estimation problem. The algebraic characterization serves as a foundation for the graph-theoretic characterizations that we seek in this article. We recall that the ML estimator is *efficient* in the sense that it achieves minimum variance among unbiased estimators, and so our development yields graph-theoretic characterizations of the best possible estimate of the initial condition.

To begin, let us note that each observation $y[k]$ can be viewed as a linear function of the initial condition $\mathbf{x}[0]$, corrupted by an additive zero-mean (independent) Gaussian noise sample. Thus, the initial condition estimation problem resolves to that of estimating

[1] The estimability concept also turns out to connect with the control-theoretic notion of *observability*.

a non-random parameter from a sequence of independent Gaussian random variables whose means are specified by that parameter. This problem of non-random estimation from Gaussian observations has been classically solved in generality [15, 16], and we only need apply the result to obtain the estimator and its covariance for our problem. To present the estimator, we find it convenient to define some further notation. Specifically, let us define a matrix:

$$Q = \begin{bmatrix} e_j^T \\ e_j^T G \\ \vdots \\ e_j^T G^n \end{bmatrix}, \tag{3}$$

where $G := \mathbf{I} - \delta\mathcal{L}$ is the state transition matrix for discrete-time dynamics and j identifies the observation location. Let us also introduce the notation $\widehat{\mathbf{x}}[0]$ to represent the ML estimate of the initial condition $\mathbf{x}[0]$.

Now, applying the standard condition, we observe that finite-variance estimation is possible (i.e., a maximum-likelihood estimate exists) if and only if the matrix Q has full column rank. Under this condition, the estimator is given by

$$\widehat{\mathbf{x}}[0] = (Q^T Q)^{-1} Q^T \begin{bmatrix} y[0] & \cdots & y[n] \end{bmatrix}^T. \tag{4}$$

Thus, we see that the estimate can be computed as a linear function of the observations, with the mapping given by the **estimator matrix** $(Q^T Q)^{-1} Q^T$.

The ML estimator of a non-random parameter vector achieves minimum variance among unbiased estimators, for any function of the parameters in the vector. Thus, the error covariance matrix of the ML estimate not only measures the performance of this estimate, but bounds the performance of all unbiased estimators of the initial condition (see the literature on the Cramer-Rao bound, e.g. [9, 14], for details). Based on this understanding, we view the **error covariance matrix** $C = E[(\widehat{\mathbf{x}}[0] - \mathbf{x}[0])(\widehat{\mathbf{x}}[0] - \mathbf{x}[0])^T]$ of the initial-condition estimate as a key construct in studying the performance of the estimator. With just a little effort, this error covariance can be shown to be

$$C = E[(\widehat{\mathbf{x}}[0] - \mathbf{x}[0])(\widehat{\mathbf{x}}[0] - \mathbf{x}[0])^T] = \sigma^2 (Q^T Q)^{-1}. \tag{5}$$

We also note that several particular performance measures can be defined from the error covariance matrix. Let us list several relevant measures, and briefly describe the motivations for their use:

- The trace of the error covariance, i.e. the sum of its diagonal entries, is commonly used as a performance measure for numerous non-random estimation tasks including optimal sensor design problems [17,18]. We note that the trace, which we denote by $tr(C)$, captures the total expected squared error in the parameter estimates [18]; in our case, $tr(C)$ captures the total expected squared error in estimating the entries in the initial condition vector. In Section 3.3.2, we will use this measure to describe estimator performance for slow coherency graph structures.

- The determinant of the error covariance matrix, $det(C)$, is also widely used as a performance measure in non-random estimation problems because 1) it captures the volume of the error ellipsoid around the true parameter value and 2) it measures

mutual information between unknown parameters and observations in estimation problems [18]. Based on either interpretation, $det(C)$ is important for us as a measure of the ability of the observations to pin down the initial condition. In Section 3.3.3, we will derive explicit graph- eigenvalue- based expression for $det(C)$ for arbitrary graphs.

- Often, we may be interested in the squared error in the estimate of a particular linear combination of the initial condition, i.e. for our estimate of $\mathbf{w}^T\mathbf{x}(0)$ for some vector $\mathbf{w}$. The best estimate is seen to be $\mathbf{w}^T\widehat{\mathbf{x}}[0]$ in that case, and the corresponding estimation error is $E[(\mathbf{w}^T(\widehat{\mathbf{x}}[0] - \mathbf{x}[0]))^2] = \mathbf{w}^T C\mathbf{w}$. For instance, we may be interested in the estimate quality for the average initial condition (corresponding to $\mathbf{w} = \mathbf{1}$) or for a particular entry in the initial condition vector (corresponding to $\mathbf{w}$ that is an basis vector). Of particular importance, we may be interested in the minimum and maximum possible squared errors among unitary linear combinations of the initial condition. These can be shown to equal the minimum and maximum eigenvalues of C, respectively.

As we characterize the various performance measures, we will also briefly discuss their particular applications in synchronization.

We have thus given explicit algebraic conditions for whether or not estimation of the non-random initial condition is possible, and for the estimate and its error covariance when estimation is possible. From the expressions above as well as our intuition regarding synchronization processes, we recognize that the Laplacian matrix, and hence the underlying graph structure, plays a critical role in whether or not estimation is possible and in the form/performance of the estimator. Characterizing this relationship is potentially valuable for several reasons, including for permitting estimator design without full knowledge of the network structure, designing the network structure to permit (or prevent) estimation, and facilitating sensor placement for controller design. With these goals in mind, we examine the relationship between graph structure and estimation in the following section.

Remark: Control theorists will notice that the matrix Q is the observability matrix and that the condition for whether or not finite-variance estimation can be achieved is equivalent to the condition for observability of the initial condition. Thus, our ensuing characterization of whether or not estimation is possible is also a characterization of the dynamical system's observability.

3.2 Graphical Conditions for Maximum Likelihood Estimation

In this subsection, we will develop conditions for MLE of the non-random initial condition that are phrased in terms of the network structure. While many of our results will apply to general diffusive networks, we will also develop some specialized graph-theoretic results for the symmetric (Laplacian network) case.

In order to develop the graphical results, we find it convenient to invoke an eigenvalue-based condition for MLE, that follows immediately from the classical spectral test for observability of linear systems [19] . Here is the foundational eigenvalue-based lemma:

Lemma 1. *Consider a diffusive network with graph Γ, and call the diffusion matrix $\mathcal{L}$. Say that we place our observer at the j^{th} node in the corresponding network. Then an ML*

estimate for the non-random initial condition exists if and only if every right eigenvector of $\mathcal{L}$ has a nonzero j^{th} entry.

Proof. From the classical modal condition for observability, the LTI system is observable if and only if $e_j^T \mathbf{v} \neq 0$ for each right eigenvector $\mathbf{v}$ of G [19]. Since $G = \mathbf{I} - \delta\mathcal{L}$, the vector $\mathbf{v}$ is also the right eigenvector of $\mathcal{L}$. Thus, we immediately find that the system is observable if and only if the j^{th} component of each eigenvector of $\mathcal{L}$ is non-zero. Invoking the equivalence of observability and existence of an MLE, we see that ML estimate is possible if and only if every right eigenvector of $\mathcal{L}$ has a nonzero j^{th} entry. □

Let us briefly discuss the special case that the diffusive matrix $\mathcal{L}$ is Laplacian. In this case, since $\mathcal{L}$ is symmetric, every eigenvalue of $\mathcal{L}$ is simple, i.e., all Jordan blocks in the spectral factorization of $\mathcal{L}$ have size 1. Thus, for symmetric case, multiple independent eigenvectors are associated with any repeated eigenvalues, and we can always construct a corresponding eigenvector with zero j^{th} entry. Therefore, MLE is possible only if the Laplacian $\mathcal{L}$ has no repeated eigenvalue.

Before we give explicit graphical conditions for the existence of estimators as well as the structure/performance, we relate the eigenvector components (and hence the possibility for MLE) to the eigenvalues of the Laplacian $\mathcal{L}$ and certain related matrices in the Laplacian network case. This characterization of the eigenvector component gives an interesting structural interpretation to estimability and estimation performance, and also serves as a stepping-stone toward other graphical results. We will present the result in the case that the Laplacian $\mathcal{L}$ has m distinct eigenvalues, noting that MLE is necessarily impossible otherwise. We find it convenient to introduce some further notation before presenting the result. First, we will use $\widehat{\mathcal{L}}(j)$ to denote the $(m-1)$ by $(m-1)$ *grounded Laplacian* matrix formed by deleting the j^{th} row and the j^{th} column of $\mathcal{L}$. We will also use $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_m$ to denote the m distinct eigenvalues of $\mathcal{L}$ and use $0 < \mu_1 \leq \mu_2 \leq \ldots \leq \mu_{m-1}$ to denote the $(m-1)$ eigenvalues of $\widehat{\mathcal{L}}(j)$. Now let us present an explicit expression of the eigenvector components in terms of the eigenvalues when the network has a Laplacian $\mathcal{L}$.

Theorem 1. *Consider a Laplacian network for which $\mathcal{L}$ has distinct eigenvalues, and say we place our observer at the j^{th} node in the network. Then the j^{th} component in the eigenvector $\mathbf{v}_i$ of $\mathcal{L}$, $\mathbf{v}_{i,j}$, can be computed as*

$$\mathbf{v}_{i,j} = \sqrt{\frac{\Pi_{z=1}^{m-1}(\mu_z - \lambda_i)}{\Pi_{z=1,z\neq i}^{m}(\lambda_z - \lambda_i)}}, \tag{6}$$

where $i = 1, 2, \ldots, m$.

Proof. Our method of proof is similar to that used for Theorem 1 in [20], and we refer the reader to [20] for a more detailed presentation. To develop the result, we consider the following continuous LTI system:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathcal{L}\mathbf{x}(t) + e_j u(t), \\ y(t) &= e_j^T \mathbf{x}(t), \end{aligned} \tag{7}$$

where $\mathbf{x}(t)$ is the state vector of length m and $y(t)$ is the scalar output [2]. We will compare two expressions for the impulse response of the system to obtain an expression for the eigenvector component. First, by solving the system based on the Jordan form of $\mathcal{L}$, we obtain that the impulse response is

$$y_i(t) = \sum_{i=1}^{m} \mathbf{v}_{i,j}^2 e^{\lambda_i t}, t \geq 0. \tag{8}$$

Meanwhile, we can find an alternate expression for the system dynamics:

$$\begin{aligned} \dot{y}(t) &= \mathcal{L}_{jj} y(t) + \mathcal{L}_j \mathbf{x}_a(t) + u(t), & (9) \\ \dot{\mathbf{x}}_a(t) &= \widehat{\mathcal{L}}(j)\mathbf{x}_a(t) + \mathcal{L}_j^T y(t), & (10) \end{aligned}$$

where $\mathbf{x}_a = (x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_m)$, $\mathcal{L}_{jj}$ is the j^{th} diagonal entry of $\mathcal{L}$, and row vector $\mathcal{L}_j = \big(\mathcal{L}_{j,1}, \cdots, \mathcal{L}_{j,(j-1)}, \mathcal{L}_{j,(j+1)}, \cdots, \mathcal{L}_{j,m}\big)$. Thus the system can be viewed in the feedback form with a first-order dynamics (9) in the forward path and an order-$(m-1)$ dynamics (10) in the feedback path. We see that the forward path transfer function is

$$H_a(s) = \frac{1}{s - \mathcal{L}_{jj}}, \tag{11}$$

and the feedback path transfer function is

$$H_f(s) = \frac{r(s)}{\Pi_{z=1}^{m-1}(s - \mu_z)}, \tag{12}$$

where $r(s)$ is a polynomial of degree less than $m-1$. From (11) and (12), we see that the transfer function from the to the input u to the output y is

$$H(s) = \frac{H_a(s)}{1 - H_a(s)H_f(s)} = \frac{\Pi_{z=1}^{m-1}(s - \mu_z)}{(s - \mathcal{L}_{jj})\Pi_{z=1}^{m-1}(s - \mu_z) - r(s)}. \tag{13}$$

Note that the system of $H(s)$ has m poles that are the m distinct eigenvalues of $\mathcal{L}$. We thus obtain

$$H(s) = \frac{\Pi_{z=1}^{m-1}(s - \mu_z)}{\Pi_{i=1}^{m}(s - \lambda_i)} = \sum_{i=1}^{m} \frac{A_i}{s - \lambda_i}, \tag{14}$$

where $A_i = \frac{\Pi_{z=1}^{m-1}(\mu_z - \lambda_i)}{\Pi_{z=1, z\neq i}^{m}(\lambda_z - \lambda_i)}$.
Hence the impulse response is

$$y_i(t) = \sum_{i=1}^{m} A_i e^{\lambda_i t}, t \geq 0, \tag{15}$$

where A_i are us defined above. Comparing the two expressions for the impulse response in (8) and (15), we thus obtain the result in the theorem. □

[2]Notice that this dynamics is *not* the diffusive network dynamics, but simply a construct to characterize the eigenvector component.

We stress here that Theorem 1 is not a rephrase of the result in [20], since here we give an explicit expression of individual eigenvector components instead of the differences between eigenvector components in [20]. Also, the LTI system 7 that we consider in the proof is different from that from [20]. Such an analytical expression is powerful in many aspects, and we will address this again in the later sections.

Let us now invoke the eigenstructural results to obtain several graphical conditions for MLE.

As a most basic result, we first formalize that non-random initial condition estimation is not possible in disconnected graphs. This basic result regarding disconnected graphs serves as a foundation to investigate MLE performance in networks comprising weakly connected subgraphs in Section 3.3.

Lemma 2. *Consider the initial condition estimation problem for a diffusive network model. If there is at least one vertex for which there does not exist a directed path to the vertex where the observer is located, then MLE of the initial condition is not possible.*

Proof. Let $\Gamma = (V, E)$ denote the graph of the diffusive network and $j \in V$ denote the vertex where the observer is located. Let us partition Γ into two subgraphs $\Gamma_A = (V_A, E_A)$ and $\Gamma_B = (V_B, E_B)$ such that: V_B contains all vertices which do not have directed paths to vertex j, and V_A contains all other vertices in V including vertex j. Let $\mathcal{L}$, $\mathcal{L}_A$, $\mathcal{L}_B$ denote the diffusion matrices of the networks associated with graphs Γ, Γ_A, Γ_B respectively. Since no directed path exists from any vertex in V_B to vertex j, there are no edges directed to any vertex in V_A from vertices in V_B. Without loss of generality, we can order the vertices in V so that $\mathcal{L}$ becomes $\mathcal{L} = \left[\begin{smallmatrix} \mathcal{L}_A & \mathbf{0} \\ \mathcal{L}_{AB} & \mathcal{L}_B \end{smallmatrix}\right]$, where $\mathcal{L}_{AB}$ denotes the directed edge weights from vertices in Γ_A to vertices in Γ_B. Let $\mathbf{v}_B$ denote an eigenvector of the diffusion matrix $\mathcal{L}_B$. We see that $\left[\begin{smallmatrix} \mathbf{0} \\ \mathbf{v}_B \end{smallmatrix}\right]$ is an eigenvector of the diffusion matrix $\mathcal{L}$. Thus by Lemma 1, MLE is not possible. In fact, estimation is not possible from any vertex in V_A. □

While the previous lemma illustrated that estimation is impossible for disconnected graphs, the next set of results show how, even in connected graphs, certain internal structures can make MLE impossible at specific locations. In particular, the following set of results show how internal symmetries in a network can cause multiple subsets of nodes to identically impact neighbouring dynamics, thus making their own dynamics indistinguishable to the observer.

One form of symmetry arises when two nodes connect to the same neighbours via identical edge weights. Our first symmetry based result shows that estimating the initial condition of the network from anywhere besides these two nodes is impossible. In particular, observers located anywhere else in the network will find dynamics initiated at the two nodes to be indistinguishable, since they identically impact the surrounding dynamics in the network. Here is the result:

Lemma 3. *Consider the non-random initial condition estimation problem for a diffusive network model. Let $\Gamma = (V, E)$ be a graph and $\mathcal{L}$ the associated diffusion matrix. Suppose there are two vertices $r, s \in V$ such that the weights of edges (r, s) and (s, r) are equal (including possibly zero i.e. there is no edge in either direction). Suppose for every vertex $q \neq s$ connected to r, q is also connected to s and that the edge weights (s, q) and (r, q) are equal. Finally, suppose the total edge weight coming into r and coming into s are*

identical. If the observer is located at any vertex $j \in V, j \neq r, s$, then the maximum likelihood estimate for the non-random initial condition does not exist.

Proof. We will take advantage of the special network structure to identify an eigenvector whose j^{th} entry is nil. Let e_i denote the i^{th} standard basis vector. We claim that because of the symmetry in the edge weights between the shared neighbourhood of r and s, the vector $\mathbf{v} = e_r - e_s$ is an eigenvector of the diffusion matrix $\mathcal{L}$ associated with the graph Γ.

To show this, let the i^{th} row of $\mathcal{L}$ be denoted by l_i^T. For $i \neq r, s$ we get $l_i^T\mathbf{v} = 0$ for vertices connected to r, s because the symmetry leads to cancellation. Note that if the vertex j is not connected to r, s the equation $l_i^T\mathbf{v} = 0$ still holds since the r^{th} and s^{th} components of l_i are nil.

Now for $i = r$ we can write $l_r^T\mathbf{v} = (\mathcal{L}_{rr} - \mathcal{L}_{rs})$ and for $j = s$ we get $l_s^T\mathbf{v} = (\mathcal{L}_{sr} - \mathcal{L}_{ss}) = -(\mathcal{L}_{rr} - \mathcal{L}_{rs})$, as the weights of the edges (r, s) and (s, r) are equal and sum of the incoming edges from all q to r and s and are equal. Thus, $\mathbf{v}$ is an eigenvector. Since there is a zero in the j^{th} position of $\mathbf{v}$, we can apply Lemma 1 and the result follows. □

Let us make several observations about this result. First we note the symmetry-based result does not hold for diffusive networks if the edge weights are not identical from r and s to their neighbours. Thus, strategic perturbations to a diffusive network with such a symmetry may resolve estimation failure. This lemma also suggests that nodes added to a network to enhance estimator performance should have edge weights designed to avoid symmetry.

Our next result generalizes the neighbour-symmetry case above to larger sets of vertices (more than just 2) exhibiting similar connective symmetry. Broadly speaking, the next theorem shows that measurement outside of a Laplacian subnetwork whose nodes are connected to all exterior nodes in a uniform fashion does not permit MLE.

Theorem 2. *Consider a diffusive network associated with the graph $\Gamma = (V, E)$. Suppose that Γ has a set of vertices A with the following properties: 1) the diffusive matrix associated with the induced subgraph on A is Laplacian and the subgraph contains at least 2 vertices; 2) the sum of all edge weights from vertices outside A to each vertex in A is identical; 3) if there exists a directed edge from a vertex in A to a vertex outside of A, then there are directed edges with identical edge weight from every vertex in A to that vertex outside of A. Then ML estimation is not possible if the observation node j corresponds to any vertex outside A.*

Proof. For convenience and without loss of generality, let us order the vertices as those inside A followed by those outside of A (which we call set B) and partition the diffusive matrix $\mathcal{L}$ accordingly. Then $\mathcal{L}$ can be written as follows:

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_A & 0 \\ 0 & \mathcal{L}_B \end{bmatrix} + \begin{bmatrix} -\alpha I & R \\ S & D \end{bmatrix} \tag{16}$$

where $\mathcal{L}_A$ is a Laplacian matrix, $\mathcal{L}_B$ is a diffusion matrix describing the network dynamics of vertices outside of A, each row in R has identical sum α, each row in S has identical entries, and D is diagonal. Let $\mathbf{v}$ be an eigenvector of $\mathcal{L}_A$ orthogonal to $\mathbf{1}$. We immediately see that $\hat{\mathbf{v}} = \begin{bmatrix} \mathbf{v} \\ \mathbf{0} \end{bmatrix}$ is an eigenvector of $\mathcal{L}$ with eigenvalue $\lambda_{\mathcal{L}} = \lambda_A - \alpha$ since $\mathbf{v}$ is orthogonal to each row of S. Applying Lemma 1, the proof is complete. □

Before we give further discussion regarding Theorem 2, let us present a simple example (Fig. 1) to help illustrate which graphs satisfy the premises of Theorem 2. In this example, we claim that the premises hold when vertices 1 and 2 are considered as the set of vertices A. In particular, the subgraph associated with A has two vertices and symmetric edge weights, hence Premise 1 is met. Meanwhile, the sum of the edge weights from outside A to each vertex in A is identical (specifically, 4.9, see also the matrix R), so Premise 2 is met. Finally, the edge weights from Vertices 1 and 2 to Vertex 3 are identical (they have value 1.8) and Vertices 1 and 2 are not connected to Vertex 4 or Vertex 5 (as can also be seen in the matrix S), so Premise 3 is met. Hence, Theorem 2 holds and estimation is not possible from outside the set of vertices A.

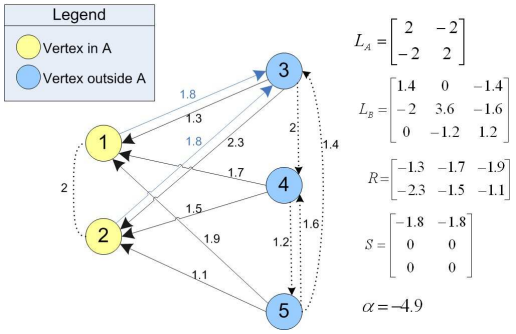

Figure 1: An example graph that meets Premises 1) - 3) in Theorem 2. We note that the matrices on the right specify parts of the diffusive matrix and clarify that the theorem holds.

Now let us make several observations. First, we note that since any Laplacian network is diffusive, the result holds for Laplacian networks as well; in this case, Premise 3 of the theorem statement, i.e., the equality in interaction strengths, implies the other premises and hence directly yields unestimability. The theorem also encompasses the case that there are no directed edges from vertices outside of A to vertices in A. Broader results can be obtained even in the case that the network induced on A is not Laplacian. However, this result is less intuitive in terms of graph structure, thus we do not give a detailed discussion. We note that this result can also be generalized to show that adding more observers in the same subnetwork does not improve our ability to estimate the initial state. Regardless of the number of observers, the edge weight symmetry destroys our ability to estimate the non-random initial condition.

Let us discuss one particular use of the above result. While it may often be difficult to find a subgraph satisfying the premises of Theorem 2 in an existing network, it is conceivable that edges can be *designed* to satisfy the premises 1) - 3) in Theorem 2 for a set of vertices A. This design capability may be useful when state information needs to be hidden or secured in an engineered synchronization process.

The above lemma and theorem show that networks whose global dynamics have localized eigenvectors are subject to estimation failure. In some sense, such dynamics are like those of a disconnected network which, as we have seen above, also cannot be estimated. Moreover, a connection strategy of this type allows for full localization of a particular mode of the network dynamics.

Next, we present a theorem concerned with whether or not MLE is possible, when a

component (node) is added to a diffusive network. We note that such results regarding estimation upon modification of the network may be valuable for a couple reasons: 1) to permit assembly of networks with desirable estimation properties and 2) to allow characterization of estimation in the common circumstance that a network is altered. Here is the result:

Theorem 3. *Suppose a new vertex is connected to only one vertex of a graph; and consider estimation in the diffusive network associated with the new graph. MLE is possible from observation at the new network node if and only if MLE was possible in the original diffusive network from observation at the node to which the new node is connected.*

Proof. Denote the diffusion matrix from the original graph as $\mathcal{L}$ and the diffusion matrix from the graph with the added vertex as $\mathcal{L}'$. Let m denote the vertex in the original graph and $(m+1)$ denote the added vertex. Suppose $(m+1)$ is connected to m via edges $(m, m+1)$ and $(m+1, m)$ with weights $-b$ and $-a$ respectively. Without loss of generality, we can assume the vertex m corresponds to the last row and column of $\mathcal{L}$. The diffusion matrix for the graph with the added vertex (upon appropriate ordering of the vertices) can then be written as

$$\mathcal{L}' = \begin{bmatrix} \mathcal{L} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & a & -a \\ 0 & \cdots & -b & b \end{bmatrix}.$$

To prove sufficiency, suppose that MLE is not possible in the original network when the observer is located at vertex m. Then by Lemma 1, there exists an eigenvector $\mathbf{v}$ of $\mathcal{L}$ with m^{th} entry $v_m = 0$. It is easy to see that $\mathbf{v}' = \begin{bmatrix} \mathbf{v}^T & 0 \end{bmatrix}^T$ is also an eigenvector of $\mathcal{L}'$; by Lemma 1, MLE from vertex $m+1$ is impossible in the new network.

To prove necessity, suppose that MLE is not possible from node $(m+1)$ in the network with the added node. By Lemma 1, $\mathcal{L}'$ has an eigenvector $\mathbf{v}' = \begin{bmatrix} \mathbf{v}^T & 0 \end{bmatrix}^T$ for some non-zero vector $\mathbf{v}$. Let λ denote the eigenvalue associated with $\mathbf{v}'$. Again, let v_m denote the m^{th} entry of vector $\mathbf{v}$ and let v'_{m+1} denote the $(m+1)^{th}$ entry of vector v'. Since the eigenvector equation $\mathcal{L}'\mathbf{v}' = \lambda\mathbf{v}'$ holds, we have that $av_m - av'_{m+1} = \lambda v_m$ and $-bv_m - bv'_{m+1} = \lambda v'_{m+1}$. Therefore, v_m is zero and since $\mathbf{v}'$ is an eigenvector of $\mathcal{L}'$, $\mathbf{v}$ is also an eigenvector of $\mathcal{L}$ as well. Recalling Lemma 1, we obtain the result. □

We note that the above theorem immediately implies that a diffusive network associated with a connected line graph can be estimated from observations at the ends.

So far we have developed several graph-based conditions under which MLE is not possible. We stress that these conditions are by no means exhaustive (and also are not necessarily easy to find in a graph) however they do capture several typical connection structures that may be found or designed in networks that prevent estimation. Let us conclude the discussion of unestimability by giving detailed graph-theoretic conditions for the canonical case of an unweighted and undirected line graph. This example serves to illustrate that quite detailed characterizations of circumstances preventing estimation are possible in some simple examples, and also to illustrate that estimation may be impossible even when the graph-theoretic conditions above do not hold.

Examples:

Precisely, let us consider an unweighted and undirected line graph of m vertices, i.e. a graph Γ such that two vertices i and j are connected by an edge and that edge has weight 1, if and only if $|i-j|=1$. We shall consider non-random initial condition estimation in the associated Laplacian network (henceforth called the Laplacian line network) from observations at various nodes j. We obtain the following results regarding unestimability. These lemmas follow immediately from simple constructions of an eigenvector followed by application of the spectral test. The first result is a direct result of symmetry when the number of vertices is odd:

Lemma 4. *For a Laplacian line network with $m=2j+1$ nodes, $j\in\mathbb{N}$, estimation from the $(j+1)^{th}$ node is not possible.*

Next, we use the following idea: embedded within line graphs are smaller line graphs that exhibit their own internal symmetry. Thus, large Laplacian line networks contain levels of symmetry that play a role in judicious observation placement. The following lemma demonstrates how a Laplacian line network with size divisible by 3 can be regarded as a composition of 3-node line graphs, each with a center spot that is a poor location for state estimation.

Lemma 5. *Consider a Laplacian line network with $m=3s$ nodes, $s\in\mathbb{N}$. MLE is not possible from observations at nodes $3j+2$, $j=0,1,2,...,s-1$.*

We note that, for both these simple line graph results, MLE is impossible even though the premises for Lemma 3 and Theorem 2 do not hold.

So far, we have presented several results on graphical conditions on non-estimability for MLE when the observer is placed at some special vertex in the underlying graph. These graphical results are mostly developed based on Lemma 1: a right eigenvector with a zero entry at the observation location is constructed for those cases with special graph structures, which conflicts to the estimability condition in Lemma 1. We note that, these results can not be applied to all cases (e.g., Laplacian cases with more general graph structures). As for the general Laplacian case, although Lemma 1 indicates that the Laplacian matrix with non-repeated eigenvalues is not sufficient for the existence of an MLE, the above results do not provide any further intuition on Laplacian graph structures. Nevertheless, we recall that in the early part of Section 3.2, Theorem 1 builds an analytical relationship (Equation 6) between eigenvector components of the original Laplacian matrix and eigenvalues of both the reduced grounded Laplacian matrix and the original one. From both Lemma 1 and Theorem 1, we thus can conclude a necessary and sufficient condition for the existence of an MLE in the following theorem.

Theorem 4. *Consider a Laplacian network where the corresponding Laplacian matrix $\mathcal{L}$ has m distinct eigenvalues: $0=\lambda_1<\cdots<\lambda_m$, and say that the observer is placed at the j^{th} node in the network. We form a grounded Laplacian matrix $\hat{\mathcal{L}}(j)$ by deleting the j^{th} row and the j^{th} column of $\mathcal{L}$, and use $0<\mu_1\leq\cdots\leq\mu_{m-1}$ to denote the $m-1$ eigenvalues of $\hat{\mathcal{L}}(j)$. Then, there exists an MLE at the j^{th} node if and only if $\lambda_i\neq\mu_z$, for all $i=1,\cdots,m$ and $z=1,\cdots,m-1$.*

The proof of the theorem can be obtained directly from Equation 6 and Lemma 1. Moreover, the *Cauchy interlacing theorem* (see [21]) tells us that $\lambda_i \leq \mu_i \leq \lambda_{i+1}$, for $i = 1, \cdots, m-1$. In fact, if the ML estimate exists (i.e., Q^TQ is invertible), then we have $\lambda_i < \mu_i < \lambda_{i+1}$ for $i = 1, 2, \cdots, m-1$ (or $0 = \lambda_1 < \mu_1 < \lambda_2 < \cdots < \lambda_{m-1} < \mu_{m-1} < \lambda_m$). Otherwise, if there exists a μ_i which is equal to either λ_i or λ_{i+1}, Q^TQ will lose rank and hence is not invertible.

Since Theorem 4 provides a necessary and sufficient condition for the existence of an MLE for a general Laplacian case, it is valuable to relate the condition to the network topology and hence underlying graph structure of the network. To capture the relationship from a graphical viewpoint, let us develop a new concept which we call a *sub-Laplacian network* first. To do so, let us consider an observer located at the j^{th} node in the original Laplacian network. We simply form the sub-Laplacian network by deleting this observation node and all its connections to the other nodes from the network, or in other words enforce that this node 1) has state value fixed at zero and 2) has no impact on the other nodes' dynamics. Our definition of the sub-Laplacian network is such that the grounded Laplacian matrix $\hat{\mathcal{L}}(j)$ is the state matrix for this sub-Laplacian network's dynamics. Thus, we see that Theorem 4 can equivalently be phrased in terms of the dynamics (and hence the graph structures) of the original Laplacian network and the sub-Laplacian network. That is, if the sub-Laplacian network has a certain special structure, or the node j connects to the sub-Laplacian network in a special way, the eigenvalue equivalence condition in Theorem 4 does not hold and hence an MLE does not exist for measurements at node j. The results for non-estimability obtained in the theorems/lemmas prior to Theorem 4 capture special cases where the Laplacian and sub-Laplacian network structure lead to non-estimability. We note that these above graphical results on non-estimability can also be easily proved for the Laplacian case based on Theorem 4 (as an alternative to the first-principles arguments that we gave before); we omit the details.

3.3 Estimator Structure and Performance

We have thus far developed various conditions for whether or not MLE is possible. In the case that estimation is possible, the structure of the estimator as well as its performance should also be dependent on the diffusive network's topological structure. In this subsection, we refine the graph-theoretic analysis of non-random initial condition MLE, by giving graph-theoretic characterizations of the estimator and its performance. We begin by characterizing the asymptotic structure of the estimator.

3.3.1 Asymptotic structure of the estimator

As a first effort on characterizing estimator structure/performance, we describe how the number of samples affects the estimator matrix structure, and clarify the form of the estimator matrix's later columns (the ones that incorporate observation data from large time-steps, henceforth called the estimator's asymptotic structure). The results that we obtain are valid for any graph structure, as long as MLE is possible. Thus, our results provide a quite general characterization of asymptotic estimator structure, which arises generally from the diffusive characteristic of the model. As such, these results are also

useful in building estimators for networks where the state transition matrix of the diffusive network is unknown or only partially known.

We begin with a couple of technical lemmas. These characterize the final column in Q^T and an eigenvector of Q^TQ. Finally, Theorem 5 elucidates the contribution of all the later measurements to the initial condition estimate $\widehat{\mathbf{x}}[0]$; the later measurements asymptotically have an equal contribution to the estimate that is inversely proportional to the number of samples n. Thus, we see that the estimator averages the measured data asymptotically.

Lemma 6. *Consider a diffusive network for which MLE is possible if the observer is located at vertex j. As the sample size n grows large, the final column of Q^T converges to* $\mathbf{w}$, *where* $\mathbf{w}$ *is the strictly positive left eigenvector of $\mathcal{L}$ associated with the non-repeated 0 eigenvalue.*

Proof. Consider the eigenvalue decomposition $G = V\Lambda W$. Since estimation is possible, there is a path from every node to the observation vertex. It then follows automatically from non-negative matrix theory that $\mathcal{L}(G)$ has a single zero (unity) eigenvalue and right eigenvector $\mathbf{1}$ and left eigenvector $\mathbf{w} > 0$ [22]. Using the fact that the other eigenvalues of G are in the unit circle, we see that $G^n \to \mathbf{1}\mathbf{w}^T$ as n becomes large. We thus see that the last column of Q^T is $(G^T)^n e_j = \mathbf{w}$. □

Lemma 7. *Q^TQ has an eigenvector that approaches* $\mathbf{w}$ *as the number of time steps, n, becomes large.*

Proof. Using the asymptotic structure of Q^TQ proved in Lemma 6, we decompose Q as $Q = \mathbf{1}_{n+1} \times \mathbf{w}^T + \Phi$, where $\mathbf{1_{n+1}}$ is a *n+1* dimensional column vector of ones and Φ^T is a $O(1)$ matrix such that its final columns approach a zero vector exponentially. Then

$$\begin{aligned} Q^TQ &= (\Phi^T + \mathbf{w}\mathbf{1}_{n+1}^T)(\mathbf{1}_{n+1}\mathbf{w}^T + \Phi) \\ &= (n+1)\mathbf{w}\mathbf{w}^T + \mathbf{w}\mathbf{1}_{n+1}^T\Phi + \Phi^T\mathbf{1}_{n+1}\mathbf{w}^T + \Phi^T\Phi. \end{aligned}$$

The entries of $\mathbf{1}_{n+1}^T\Phi$, $\Phi^T\mathbf{1}(n+1)$ and $\Phi^T\Phi$ do not increase substantially as n increases since the entries of each column of Φ decreases exponentially and its rows approach $\mathbf{0}^T$. Also, $\mathbf{w}$ is fixed for all n. Thus the second, third and fourth terms (together called ϵ hereafter) are matrices with nearly fixed $O(1)$ entries. Hence, the matrix $\frac{1}{n+1}Q^TQ$ can be thought of as the matrix $\mathbf{w}\mathbf{w}^T$ with an $O\left(\frac{1}{n}\right)$ perturbation added. Since the matrix $\mathbf{w}\mathbf{w}^T$ has distinct eigenvalue 1 and eigenvector $\mathbf{w}$, the perturbation of order $\frac{1}{n}$ implies that Q^TQ has an eigenvalue within order 1 of $(n+1)$ and an eigenvector within order $\frac{1}{n}$ of $\mathbf{w}$. Therefore as n gets large, the eigenvector approaches $\mathbf{w}$. □

Theorem 5. *Given the estimator $(Q^TQ)^{-1}Q^T$, as the sample size n grows large, each row of the estimator converges to a (in general different) number. Moreover, the limit point of each of the estimator matrix rows is inversely proportional to n.*

Proof. We note that the symmetric matrix $(Q^TQ)^{-1}$ has an eigenvalue near $\frac{1}{n}$ with corresponding eigenvector within $O(\frac{1}{n})$ of $\mathbf{w}$, while its remaining eigenvalues are $O(1)$ and the corresponding eigenvectors $\hat{\mathbf{v}}$ are nearly orthogonal to $\mathbf{w}$ (specifically, $\hat{\mathbf{v}}^T\mathbf{w}$ is $O(\frac{1}{n})$). Thus, we immediately see that $(Q^TQ)^{-1}\mathbf{w}$ is on the order of $\frac{1}{n}$, we can write

$(Q^TQ)^{-1}\mathbf{w} = \frac{1}{(n+1)\mathbf{w}^T\mathbf{w}}\mathbf{w} - \frac{1}{(n+1)\mathbf{w}^T\mathbf{w}}(Q^TQ)^{-1}\epsilon\mathbf{w}$. And so the entries in the final column of the estimator matrix are inversely proportional to n for large n. Since the columns of Q^T approach $\mathbf{w}$ for sufficiently large n, each row of $(Q^TQ)^{-1}Q^T$ also approaches to the particular value in its final column. Thus, the result is proved. □

3.3.2 Estimator Performance for Slow-Coherent Graph Structures

Many modern large-scale dynamical networks, including electric power networks and various biological networks, turn out to comprise multiple sub-networks that are strongly linked internally but only weakly coupled to each other. Diffusive or synchronizing dynamics in networks with such structure have been extensively studied under the label of slow-coherency theory, and are well-known to display 1) global but synchronous slow responses and 2) highly localized fast dynamics (see [23] for an overview). Here, we study MLE for networks with weakly-connected subnetworks, or in other words ones that display slow-coherent dynamics. Intuitively we might expect that measurement in one subnetwork would not permit high-quality MLE since the fast dynamics of the other subnetwork are almost unobservable upon measurement. More precisely, we might expect that the slow and local-fast dynamics can be estimated well, while the remaining aspects of the dynamics are difficult to estimate. The following theorem makes this intuition precise.

Theorem 6. *Consider the initial condition estimation problem for a Laplacian network model. Suppose that the network's graph can be partitioned into two subgraphs that are connected via small edge weights, i.e. ones that are at most $\epsilon << 1$. Now consider the covariance matrix $\sigma^2(Q^TQ)^{-1}$ of the estimate, partitioned in accordance with the network's partition. Then the entries in each block of the matrix are at least of the following order: $\begin{bmatrix} O(1) & O(\frac{1}{\epsilon}) \\ O(\frac{1}{\epsilon}) & O(\frac{1}{\epsilon^2}) \end{bmatrix}$.*

Proof. Let A and B represent the two subgraphs, and say that they have m_A and m_B vertices respectively ($m_A + m_B = m$). As per the theorem statement, the edge weights between the vertices in A and B are at most ϵ for some $\epsilon << 1$. For convenience, in this proof we will specify the dimension of each ones vector $\mathbf{1}$ (or zeros vector $\mathbf{0}$) with a subscript, e.g. $\mathbf{1}_n$ is a n dimensional column vector with all unity entries.

We will take the following approach to characterize MLE for the weakly linked network: we will use the classical spectral characterization of the state transition matrix for such systems to infer the structure and spectrum of Q^TQ, and hence characterize its inverse. We can write the state transition matrix G as $G = \begin{bmatrix} G_A & 0 \\ 0 & G_B \end{bmatrix} + \Phi$ where Φ is a Laplacian perturbation matrix of order ϵ. It is well known that G has an eigenvalue decomposition with the following structure:

$$G = \begin{bmatrix} S & F_A & F_B \end{bmatrix} \begin{bmatrix} D_0 & 0 & 0 \\ 0 & D_A & 0 \\ 0 & 0 & D_B \end{bmatrix} \begin{bmatrix} S^T \\ F_A^T \\ F_B^T \end{bmatrix},$$

where $F_A = \begin{bmatrix} \bar{W} \\ [\epsilon_1] \end{bmatrix} \in \mathbb{R}^{m\times(m_A-1)}$, $F_B = \begin{bmatrix} [\epsilon_2] \\ \bar{V} \end{bmatrix} \in \mathbb{R}^{m\times(m_B-1)}$, $S = \begin{bmatrix} \frac{1}{\sqrt{m}}\mathbf{1}_{m_A} & \sqrt{\frac{m_B}{mm_A}}\mathbf{1}_{m_A} \\ \frac{1}{\sqrt{m}}\mathbf{1}_{m_B} & -\sqrt{\frac{m_A}{mm_B}}\mathbf{1}_{m_B} \end{bmatrix} +$ $[\epsilon_3] \in \mathbb{R}^{m\times 2}$, $D_0 = \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix}$, and μ_2 is $O(\epsilon)$ away from 1, D_A and D_B are diagonal matri-

ces with diagonal entries $O(1)$ away from 1 (and in the unit circle), $[\epsilon_1]$, $[\epsilon_2]$, and $[\epsilon_3]$ are matrices of $O(\epsilon)$, and $\bar{W}, \bar{V}$ are matrices of $O(1)$. Note the first column of $[\epsilon_3]$ is zero since Φ is Laplacian. Using this special structure, we will characterize each row of Q and thus determine the order of the entries in Q^TQ. This will in turn allow us to demonstrate that the covariance matrix $(Q^TQ)^{-1}$ has blocks of the order given in the theorem statement.

Now for an integer k,

$$\begin{aligned}
G^k &= \begin{bmatrix} S & F_A & F_B \end{bmatrix} \begin{bmatrix} D_0^k & 0 & 0 \\ 0 & D_A^k & 0 \\ 0 & 0 & D_B^k \end{bmatrix} \begin{bmatrix} S^T \\ F_A^T \\ F_B^T \end{bmatrix} \\
&= SD_0^kS^T + \begin{bmatrix} \bar{W} & [\epsilon_1] \\ [\epsilon_2] & \bar{V} \end{bmatrix} \begin{bmatrix} D_A^k & 0 \\ 0 & D_B^k \end{bmatrix} \begin{bmatrix} \bar{W}^T & [\epsilon_2]^T \\ [\epsilon_1]^T & \bar{V}^T \end{bmatrix} \\
&= \left(\begin{bmatrix} \frac{1}{\sqrt{m}}\mathbf{1}_{m_A} & \sqrt{\frac{m_B}{mm_A}}\mathbf{1}_{m_A} \\ \frac{1}{\sqrt{m}}\mathbf{1}_{m_B} & -\sqrt{\frac{m_A}{mm_B}}\mathbf{1}_{m_B} \end{bmatrix} + [\epsilon_3] \right) \begin{bmatrix} 1^k & 0 \\ 0 & \mu^k \end{bmatrix} \left(\begin{bmatrix} \frac{1}{\sqrt{m}}\mathbf{1}_{m_A}^T & \frac{1}{\sqrt{m}}\mathbf{1}_{m_B}^T \\ \sqrt{\frac{m_B}{mm_A}}\mathbf{1}_{m_A}^T & -\sqrt{\frac{m_A}{mm_B}}\mathbf{1}_{m_B}^T \end{bmatrix} + [\epsilon_3]^T \right) \\
&+ \begin{bmatrix} \bar{W}D_A^k\bar{W}^T + [\epsilon_1]D_B^k[\epsilon_2]^T & \bar{W}D_A^k[\epsilon_1]^T + [\epsilon_1]D_B^k\bar{V}^T \\ [\epsilon_2]D_A^k\bar{W}^T + \bar{V}D_B^k[\epsilon_2]^T & \bar{V}D_B^k\bar{V}^T + [\epsilon_2]D_A^k[\epsilon_1]^T \end{bmatrix}.
\end{aligned}$$

From this form, we immediately find that i^{th} row of Q can be written as

$$e_j^TG^i = \left[\, {}^{c_1(i)\mathbf{1}_{m_A}^T + k(i)} \mid {}^{c_2(i)\mathbf{1}_{m_B}^T} \, \right] + [\epsilon_4],$$

where $c_1(i)$ and $c_2(i)$ are scalars of order 1 for each i, $k(i)$ is a vector whose norm is decreased exponentially with i (at a rate that has order 1) and has order 1, and $[\epsilon_4]$ is a vector of order ϵ. Noticing that $Q^TQ = \sum_{i=0}^{n} (e_j^TG^i)^T(e_j^TG^i)$, we immediately recover that Q^TQ can approximated by the matrix:

$$\begin{bmatrix} Q_{AA} + \alpha_{AA}\mathbf{1}\mathbf{1}^T & \epsilon Q_{AB} + \alpha_{AB}Z\mathbf{1}^T \\ \epsilon Q_{AB}^T + \alpha_{AB}\mathbf{1}Z^T & \epsilon^2 Q_{BB} + \epsilon_{BB}\mathbf{1}\mathbf{1}^T \end{bmatrix},$$

where the matrices Q_{AA}, Q_{AB}, Q_{BB}, and vector Z are all of order 1, and all other perturbations are of lower order. To characterize the error covariance matrix $(Q^TQ)^{-1}$, let us apply the block-matrix inverse formula. Specifically, assuming that each partition has at least two vertices, we immediately obtain the following:

1. $(Q_{AA} + \alpha_{AA}\mathbf{1}\mathbf{1}^T)^{-1}$ is of order 1;
2. $(\epsilon^2 Q_{BB} + \alpha_{BB}\mathbf{1}\mathbf{1}^T - (\epsilon Q_{AB}^T + \alpha_{AB}\mathbf{1}Z^T)(Q_{AA} + \alpha_{AA}\mathbf{1}\mathbf{1}^T)^{-1}(\epsilon Q_{AB}^T + \alpha_{AB}\mathbf{1}Z^T))^{-1}$ is at least of order $\frac{1}{\epsilon^2}$.

From the above two facts and the block matrix inversion formula, we immediately infer the result of the theorem. □

3.3.3 Characterizations of Performance: General Case

We conclude our study of the structure and performance of Laplacian network estimators by pursuing spectral and graphical characterizations of estimator performance measures

for some more general classes of graphs (ones that may not have weak-link structures). Let us begin with a characterization of the performance measure $det(C)$ for Laplacian networks, which we recall captures the volume of the error ellipsoid around the true initial conditions and also is a measure of information content. Specifically, we give a formula for the determinant in terms of the eigenvalues of the Laplacian and associated grounded Laplacian matrices. This expression constitutes an interesting representation of estimator performance in terms of graph eigenvalues, and also serves as a starting point for graph-theoretic characterizations since much is known about matrix spectra. Subsequently, we also study performance in estimating linear combinations of the initial condition, again giving spectral and graphical characterizations.

To obtain the result, we progress as follows. We know that the eigen-decomposition of Laplacian $\mathcal{L}$ is $\mathcal{L} = V\Lambda V^{-1}$, where each column of V is an eigenvector of $\mathcal{L}$. Since $G = I - \delta\mathcal{L}$, the eigenvalues of G are $\eta_i = 1 - \delta\lambda_i$, $i = 1, \cdots, m$. We note that $\eta_1 = 1$ while the other eigenvalues lie within $(-1, 1)$ for small δ. We shall again limit our analysis to the case that the eigenvalues are distinct. In this notation, $G = V\Lambda_G V^{-1}$, where $\Lambda_G = diag\{\eta_i\}_{i=1}^m$. Then the matrix Q can be written as

$$\begin{aligned} Q &= \begin{bmatrix} e_j^T \\ e_j^T G \\ \vdots \\ e_j^T G^n \end{bmatrix} = \begin{bmatrix} e_j^T \\ e_j^T V\Lambda_G V^{-1} \\ \vdots \\ e_j^T V\Lambda_G^n V^{-1} \end{bmatrix} = \begin{bmatrix} e_j^T V \\ e_j^T V\Lambda_G \\ \vdots \\ e_j^T V\Lambda_G^n \end{bmatrix} V^{-1} \\ &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \eta_1 & \eta_2 & \cdots & \eta_m \\ \vdots & \vdots & & \vdots \\ \eta_1^n & \eta_2^n & \cdots & \eta_m^n \end{bmatrix} \begin{bmatrix} \mathbf{v}_{1,j} & & & \\ & \mathbf{v}_{2,j} & & \\ & & \ddots & \\ & & & \mathbf{v}_{m,j} \end{bmatrix} V^{-1} \end{aligned} \tag{17}$$

We will characterize the determinant of Q^TQ and hence of $(Q^TQ)^{-1}$ when n is large, i.e., sufficient data is available. Before we derive this expression, we find it convenient to limit ourselves to the case where the number of time steps is a multiple of m, i.e. $n = \bar{n}m - 1$ for some $\bar{n}$. In this case, Q has $\bar{n}m$ rows. Since we are only concerned with the asymptotic characteristics of Q^TQ and hence $det(Q^TQ)$, limiting ourselves to this case is sufficient. Then the matrix Q becomes

$$Q = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \eta_1 & \eta_2 & \cdots & \eta_m \\ \vdots & \vdots & & \vdots \\ \eta_1^{\bar{n}m-1} & \eta_2^{\bar{n}m-1} & \cdots & \eta_m^{\bar{n}m-1} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{1,j} & & & \\ & \mathbf{v}_{2,j} & & \\ & & \ddots & \\ & & & \mathbf{v}_{m,j} \end{bmatrix} V^{-1}. \tag{18}$$

Now let us define a m by m block matrix as

$$M_k = \begin{bmatrix} \eta_1^{(k-1)m} & \eta_2^{(k-1)m} & \cdots & \eta_m^{(k-1)m} \\ \eta_1^{(k-1)m+1} & \eta_2^{(k-1)m+1} & \cdots & \eta_m^{(k-1)m+1} \\ \vdots & \vdots & & \vdots \\ \eta_1^{km-1} & \eta_2^{km-1} & \cdots & \eta_m^{km-1} \end{bmatrix},$$

where $k = 1, 2, \cdots, n$. Note that M_1 is in fact a *Vandermonde matrix* (see [24]). The determinant of M_1 is

$$\begin{aligned} det(M_1) &= \Pi_{1\le i<j\le m}(\eta_j - \eta_i) \\ &= \delta^{\frac{m(m-1)}{2}}\Pi_{1\le i<j\le m}(\lambda_i - \lambda_j) \end{aligned} \tag{19}$$

We also define $M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{bmatrix}$. Then Q can be written as

$$Q = M \begin{bmatrix} \mathbf{v}_{1,j} & & & \\ & \mathbf{v}_{2,j} & & \\ & & \ddots & \\ & & & \mathbf{v}_{m,j} \end{bmatrix} V^{-1},$$

and $det(Q^T Q)$ can be written as

$$det(Q^T Q) = det(M^T M) \Pi_{i=1}^{m} \mathbf{v}_{i,j}^2.$$

Next, we will analyze the determinant of $M^T M$. Since

$$M_k = \begin{bmatrix} \eta_1^{(k-1)m} & & & \\ & \eta_2^{(k-1)m} & & \\ & & \ddots & \\ & & & \eta_m^{(k-1)m} \end{bmatrix} M_1$$

for each k, we obtain

$$\begin{aligned} M^T M &= \sum_{k=1}^{n} M_k^T M_k \\ &= M_1^T \begin{bmatrix} \sum_{k=1}^{n} \eta_1^{2m(k-1)} & & & \\ & \sum_{k=1}^{n} \eta_2^{2m(k-1)} & & \\ & & \ddots & \\ & & & \sum_{k=1}^{n} \eta_m^{2m(k-1)} \end{bmatrix} M_1 \\ &= M_1^T \begin{bmatrix} n & & & \\ & \frac{1-\eta_2^{2mn}}{1-\eta_2^{2m}} & & \\ & & \ddots & \\ & & & \frac{1-\eta_m^{2mn}}{1-\eta_m^{2m}} \end{bmatrix} M_1. \end{aligned} \tag{20}$$

Finally, let us consider the expression for $det(Q^T Q)$. Combining this expression with (19) and Theorem 1, we have

$$\begin{aligned} det(Q^T Q) &= det(M^T M) \prod_{i=1}^{m} \mathbf{v}_{i,j}^2 \\ &= (det(M_1))^2 \left(n \prod_{i=2}^{m} \frac{1-\eta_i^{2mn}}{1-\eta_i^{2m}} \right) \prod_{i=1}^{m} \mathbf{v}_{i,j}^2 \\ &= n\delta^{m(m-1)} \prod_{1 \le i < j \le m} (\lambda_i - \lambda_j)^2 \prod_{i=2}^{m} \frac{1-\eta_i^{2mn}}{1-\eta_i^{2m}} \prod_{i=1}^{m} \frac{\prod_{z=1}^{m-1} (\mu_z - \lambda_i)}{\prod_{z=1, z \ne i}^{m} (\lambda_z - \lambda_i)} \\ &= n\delta^{m(m-1)} \left(\prod_{i=2}^{m} \frac{1-\eta_i^{2mn}}{1-\eta_i^{2m}} \right) \left(\prod_{i=1}^{m} \prod_{z=1}^{m-1} |\mu_z - \lambda_i| \right) \\ &\to n\delta^{m(m-1)} \left(\prod_{i=2}^{m} \frac{1}{1-\eta_i^{2m}} \right) \left(\prod_{i=1}^{m} \prod_{z=1}^{m-1} |\mu_z - \lambda_i| \right), \end{aligned} \tag{21}$$

as n becomes large.

We now can present the explicit expression for the determinant of the error covariance matrix.

Theorem 7. *Consider the Laplacian network with Laplacian matrix $\mathcal{L}$. For sufficiently large n, the determinant of the error covariance matrix approaches*

$$det(C) = \frac{\sigma^{2m}\Pi_{i=2}^{m}(1-\eta_i^{2m})}{n\delta^{m(m-1)}\Pi_{i=1}^{m}\Pi_{z=1}^{m-1}|\mu_z - \lambda_i|}. \tag{22}$$

Proof. We note that the error covariance matrix C is $C = \sigma^2(Q^TQ)^{-1}$. Then the result in the theorem automatically follows (21). □

From the above theorem, we see that when n is increased (i.e., we have more data), the performance of the estimator improves. Meanwhile, when the scalar δ is decreased (i.e., the interactions among the network components are weak), the performance becomes worse. We also see that the performance of the estimator can be improved by spreading out the eigenvalues λ_i and placing the eigenvalues μ_z as far from the λ_i as possible (through design of the network topology and observation location). As a very first result in this direction, it is easy to show that $det(C)$ is lower-bounded by an exponentially growing function of n, i.e. estimator quality is poor for networks with many nodes regardless of the interconnection structure. We leave it to future work to further characterize the determinant measure.

Finally, let us characterize performance measures defined from estimates of linear combinations of the initial condition. In particular, we recall that the error in the MMSE estimate of $\mathbf{w}^T\mathbf{x}(0)$ is given by $E[(\mathbf{w}^T(\widehat{\mathbf{x}}[0] - \mathbf{x}[0]))^2] = \mathbf{w}^TC\mathbf{w}$. Two particularly interesting performance measures defined for such linear-combination estimates are the maximum and minimum error variances for unitary linear combinations, namely $\sigma_{min} = \min_{\mathbf{w}\, s.t.\, ||\mathbf{w}||_2=1} \mathbf{w}^TC\mathbf{w}$ and $\sigma_{max} = \max_{\mathbf{w}\, s.t.\, ||\mathbf{w}||_2=1} \mathbf{w}^TC\mathbf{w}$. These maximum and minimum error variances are a measure of how easy or difficult it may be to obtain initial-condition statistics, and the vectors achieving the minimum and maximum indicate which statistics are easy or hard to obtain. We note in particular that estimates of and lower bounds on σ_{max} provide an indication of how difficult estimation of some initial-condition statistic may be, and hence are an indication of the level of *security* of the full state; meanwhile, estimates of and upper bounds on σ_{min} indicate how easy estimation of some initial condition statistic may be, and hence indicate the vulnerability of the network to inference of some initial-condition statistic. Here, we provide several characterizations of these performance measures, in terms of properties of the network matrix G (and hence in terms of its associated graph).

Let us begin by characterizing the minimum error variance σ_{min}. Based on our asymptotic characterization of the minimum eigenvalue and corresponding eigenvector of the error covariance matrix C, we can also characterize σ_{min} in the asymptotic case. The results are in the following theorem:

Theorem 8. *As the number of observations n is increased, the minimum possible error variance for an initial-condition statistic's estimate, or σ_{min}, approaches $\frac{\sigma^2}{n}$. Furthermore, the vector* $\mathbf{w}$ *that achieves the minimum approaches the left eigenvector of G associated with its unity eigenvalue. Thus, the initial-condition statistic that is easiest to*

approximate in the limit of large n is the synchronization value, i.e. the weighted linear combination of the agents' initial states that is achieved by each agent asymptotically.

The result given in the above theorem is not surprising, given the asymptotic characteristic of synchronization processes. After sufficient time has passed, the observation will simply be a noisy measurement of the synchronization value, with any other information about the initial condition suppressed in the measurement. Thus, it is no surprise that the synchronization value becomes the initial-condition statistic that is easiest to estimate, with the estimator performance replicating that of scalar estimation from a set of independent noisy measurements.

In the next two theorems, we provide lower bounds on the maximum possible error variance when an initial-condition statistic is estimated. We phrase these bounds in terms of the spectrum of the network matrix G. After the theorems are presented, we will discuss how the spectral representations of the bounds can be translated to graph-theoretic representations. For convenience of presentation, we assume that the variance of the scalar Gaussian white noise process $g[k]$ in Equation 2 is unit, i.e. $\sigma^2 = 1$. Here is the first result, which shows that estimation will necessarily be poor if G 1) has an eigenvalue far inside the unit circle or 2) has an eigenvector whose entry at the measurement component is small:

Theorem 9. *Consider the initial-condition estimation problem with measurement made a node j for an arbitrary duration and unit variance of noise $g[k]$ (i.e., $\sigma^2 = 1$ in Equation 2), assume that estimation is possible, and assume that the eigenvalues of G are real and simple. Consider the maximum error variance σ_{max}, in estimating a unitary linear statistic of the network's initial condition. The maximum error variance is bounded as follows: $\sigma_{max} \geq \max_i \frac{1-\lambda_i^2}{v_{ij}^2}$, where λ_i is the i^{th} eigenvalue of matrix G, and v_{ij} is the j^{th} entry of the i^{th} right eigenvector $\mathbf{v}_i$ of G.*

Proof. By definition, we have that $\sigma_{max} = \max_{\mathbf{w}\, s.t. ||\mathbf{w}||=1} \mathbf{w}^T (Q^T Q)^{-1} \mathbf{w}$. However, since $(Q^T Q)^{-1}$, the Courant-Fisher theorem holds, and so σ_{max} can be rewritten as: $\sigma_{max} = \lambda_{max}((Q^T Q)^{-1})$, i.e. it is equal to the largest eigenvalue of $(Q^T Q)^{-1}$. Using the relationship between eigenvalues of a matrix and its inverse and then the Courant-Fisher theorem, we then recover that $\sigma_{max} = \frac{1}{\lambda_{min}(Q^T Q)} = \frac{1}{\min_{\mathbf{w}\, s.t. ||\mathbf{w}||=1} \mathbf{w}^T (Q^T Q) \mathbf{w}}$. For convenience, noting that we have assumed the eigenvalues of G to be simple, let us express the vector $\mathbf{w}$ in the expression above in terms of the right eigenvectors of G, i.e. as $\mathbf{w} = \sum_{i=1}^m \alpha_i \mathbf{v}_i$, where $\mathbf{v}_i$ are the right eigenvectors of G normalized to unit length. Substituting, we obtain the following expression: $\sigma_{max} = \frac{1}{\min_{\alpha_1,\ldots,\alpha_m} (\sum_{i=1}^m \alpha_i Q \mathbf{v}_i)^T (\sum_{i=1}^m \alpha_i Q \mathbf{v}_i)}$, where $\alpha_1^2 + \cdots + \alpha_m^2$ is constrained to equal 1. Invoking Equation 3 for Q and using the fact that the $\mathbf{v}_i$ are eigenvectors, we readily obtain: $\sigma_{max} = \frac{1}{\min_{\alpha_1,\cdots,\alpha_m} (\sum_i \alpha_i v_{ij})^2 + (\sum_i \alpha_i \lambda_i v_{ij})^2 + \cdots + (\sum_i \alpha_i \lambda_i^n v_{ij})^2}$, where the condition on the α_i remains, v_{ij} is the jth entry of $\mathbf{v}_i$, and measurements have been made from time 0 to n. Noting that performance only improves with increasing data, we then see that $\sigma_{max} \geq \frac{1}{\min_{\alpha_1,\ldots,\alpha_m} (\sum_i \alpha_i v_{ij})^2 + (\sum_i \alpha_i \lambda_i v_{ij})^2 + \cdots}$, where the denominator summation is an infinite sum in this case. Finally, we use the fact that the

quantity on the right in the previous expression is lower-bounded when any particular set of $\alpha_1, \ldots, \alpha_m$ (subject to the constraint) is used in place of the minimum. Thus, choosing $\alpha_i = 1$ and $\alpha_r = 0$ for $r \neq i$, we obtain that $\sigma_{max} \geq \frac{1}{v_{ij}^2(1+\lambda_i^2+\lambda_i^4+\cdots)}$ for any i. Evaluating the infinite sum and choosing the best bound with respect to i, we obtain the result of the theorem. □

Let us briefly interpret the result given in the theorem. Noting that the eigenvector entries v_{ij} are always less than 1 in magnitude, we see that σ_{max} is always lower-bounded by $1 - \lambda_i^2$. Thus, if even one eigenvalue of G is far from the unit circle, we see that some linear combination of the initial condition becomes difficult to estimate. Similarly, if any eigenvector has a small-magnitude component at the measurement vertex, the estimation of some initial-condition statistic is necessarily poor. Let us recall that numerous results in the algebraic-graph-theory community relate eigenvalues and eigenvector components to a corresponding graph's topological structure: these relationships permit us to translate the above spectral bound on σ_{max} to graphical bounds. We kindly ask the reader to see e.g. [12] for the details of such relationships.

Next, let us give a second lower bound on the maximum possible error variance in estimating unitary statistics, which clarifies that not only the eigenvalue locations but the distances between the eigenvalues modulate estimation performance. Here is the result:

Theorem 10. *Consider the initial-condition estimation problem with measurement made at node j for an arbitrary duration and unit variance of noise $g[k]$ (i.e., $\sigma^2 = 1$ in Equation 2), and assume that estimation is possible. Also, assume that the eigenvalues of G are real and simple. Consider the maximum error variance σ_{max}, in estimating a unitary linear statistic of the network's initial condition. The maximum error variance is bounded as follows:* $\sigma_{max} \geq \max_{i,r} \frac{(1-\lambda_i^2)^3(v_{ij}^2 + v_{rj}^2)}{v_{ij}^2 v_{rj}^2 (\lambda_i - \lambda_r)^2}$.

Proof. From the proof of the previous theorem, we recall that the following inequality holds: $\sigma_{max} \geq \frac{1}{\min_{\alpha_1,\ldots,\alpha_n}(\sum_i \alpha_i v_{ij})^2 + (\sum_i \alpha_i \lambda_i v_{ij})^2 + \cdots}$, where the α_i are subject to the constraint $\alpha_1^2 + \cdots + \alpha_n^2 = 1$. To continue, we evaluate the argument of the minimization expression for $\alpha_i = \frac{v_{rj}}{(v_{ij}^2+v_{rj}^2)^{1/2}}$, $\alpha_r = \frac{v_{ij}}{(v_{ij}^2+v_{rj}^2)^{1/2}}$, and $\alpha_q = 0$ otherwise, for each possible i and r. The result of the theorem follows with some algebra. □

Again, let us take a moment to interpret the above theorem. One key insight gained from the theorem is that, if G has two eigenvalues that are bounded away from unity are moved close to each other, then the maximum error variance becomes increasingly large. The increasing difficulty in estimation as two eigenvalues approach each other is not surprising, since the dynamics become unestimable if two simple eigenvalues of G are collocated. This theorem further clarifies that the lower bound on maximum error variance is inversely proportional to the square of the difference between the eigenvalues: thus, estimation rapidly becomes difficult as two eigenvalues are made close to each other.

References

[1] S. Strogatz, "Exploring complex networks," *Nature*, 410(2001) pp. 268-276.

[2] C. Wu, L. Chua, "Application of Kronecker Products to the Analysis of Systems with Uniform Linear Coupling," *IEEE Transactions on Circuits and Systems*, 42(10)(1995), pp. 777-780.

[3] X. Li, G. Chen, "Synchronization and desynchronization of complex dynamical networks: an engineering viewpoint," *IEEE Trans. Circuits Syst.* I 50(11)(2003) pp. 1381-1390.

[4] W. A. Hashlamoun, M. A. Hassouneh, and E. H. Abed, "New results on modal participation factors: revealing a previously unknown dichotomy," *IEEE Transactions on Automatic Control*, VOL. 54, NO. 7, July 2009.

[5] E. Scholtz, P. Sonthikorn, G. C. Verghese, and B. C. Lesieutre, "Observers for swing-state estimation of power systems," *North American Power Symposium*, October 2002.

[6] S. Roy, A. Saberi, K. Herlugson, "A control-theoretic perspective on the design of distributed agreement protocols," *International Journal of Robust and Nonlinear Control*, 17(2007), pp. 1034-1066.

[7] Y. Wan, S. Roy, X. Wang, A. Saberi, T. Yang, M. Xue, B. Malek, "On the structure of graph edge designs that optimize the algebraic connectivity," *Proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, Dec.9-11, 2008.

[8] A. Y. Pogromsky, "Controlled synchronization of dynamical systems based on passivity technique," *Proceedings of the 14th IFAC World Congress*, pp. 473-478, Beijing, China, 1999.

[9] Y. Wan, S. Roy, "On inference of network time constants from impulse response data: graph-theoretic Cramer-Rao bounds," in *Proceeding of the 48th IEEE Conference on Decision and Control*, Shanghai, China, December 14-16, 2009.

[10] A. Maybhate, R. E. Amritkar, D. R. Kulkarni, "Estimation of initial conditions and secure communication," *arXiv.org e-Print archive*, arXiv:nlin/0110039v1.

[11] J. P. Corfmat and A. S. Morse, "Stabilization with deentralized feedback control," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, Feb. 1976.

[12] F. R. K. Chung, *Spectral Graph Theory*, American Mathematical Society Press: Providence, RI, 1997.

[13] R. Merris, "Laplacian matrices of graphs," *Linear Algebra and Appl.*, 197-198 (1994), pp. 143-176.

[14] L. T. McWhorter and L.L. Scharf, "Cramer-Rao bounds for deterministic modal analysis," *IEEE Trans. Signal Proc.*, SP-41, pp. 1847-1862, May 1993.

[15] A. Papoulis, S. U. Pillai, *Probability, random variables, and stochastic processes*, TaTa McGraw-Hill: New Delhi, 2002.

[16] A. Wiesel, Y. C. Eldar, and A. Beck, "Maximum likelihood estimation in linear models with a Gaussian model matrix," *IEEE Signal Processing Letters*, vol. 13, NO. 5, May 2006.

[17] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, no.2, February 2009.

[18] V. V. Fedorov, *Theory of Optimal Experiments*, Academic Press: New York, 1972.

[19] W. J. Rugh, *Linear system theory*, Prentice Hall: New Jersey, 1996.

[20] S. Roy and Y. Wan, "An explicit formula for differences between laplacian-eigenvector components," *Linear Algebra and Its Applications*, vol. 433, no. 7, pp. 1329-1335, Dec. 2010.

[21] W. Haemers, "Interlacing eigenvalues and graphs", *Linear Algebra and its Applications*, 226-228(1995), 593-616.

[22] A. Berman, R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Academic Press: New York, 1979.

[23] J. H. Chow, *Time-scale modeling of dynamic networks with applications to power systems*, Springer-Verlag: Boston, May 1983.

[24] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.