Robustness in Consensus Networks

T. Sarkar, M. Roozbehani, M. A. Dahleh

Abstract—We consider the problem of robustness in large consensus networks that occur in many areas such as distributed optimization. Robustness, in this context, is the scaling of performance measures, e.g.: $H_2$-norm, as a function of network dimension. We provide a formal framework to quantify the relation between such performance scaling and the convergence speed of the network. Specifically, we provide upper and lower bounds for the convergence speed in terms of robustness and discuss how these bounds scale with the network topology. The main contribution of this work is that we obtain tight bounds, that hold regardless of network topology. The work here also encompasses some results in convergence time analysis in previous literature.

I. INTRODUCTION

The goal of this paper to develop a general framework to assess robustness in large consensus networks. Robustness quantifies how some performance measure scales with the network dimension. Large networks have become ubiquitous in many socioeconomic and engineering areas, e.g.: finance, economics, transportation etc. Therefore, to guarantee graceful performance scaling as a function of network dimension and its interconnections, there is a need to develop a general framework for robustness analysis. We use the ideas introduced in [1] and extend them to consensus networks. It is well known that convergence speeds of consensus networks are limited by the proximity of the second largest eigenvalue to unity. However, the precise relation between robustness and convergence speed is still unclear. It should, however, be clear intuitively that graceful performance scaling implies “fast” convergence. In this work our goal is to study this notion more precisely.

Many performance measures to capture dimension dependent scaling have been studied in [2], [3], [4]. Specifically, in [2], $H_2$–norm based performance measures are studied. $H_{\infty}$ based performance scaling, known as harmonic instability, is studied in [3] for the case of one-dimensional transportation networks. A generalized framework for robustness of stable networks is provided in [1]. The focus there has been limited to stable networks. The case of robustness in consensus networks is studied in [5], [6], [7], [8]. The work in [6], [7], [8] is limited to the case of undirected networks. A class of performance measures called schur convex systemic (SCS) are introduced there. These performance measures have the property that they are schur convex in the eigenvalues of the network. $H_2$-norm is used as a performance measure for analyzing directed networks in [5]; however they do not show any relation between convergence speed and network robustness. Additionally, the analysis there is based on a non-unique reduction of the network, which is different from the framework we develop here.

The convergence speeds for stochastic matrices has been widely studied (See [9], [10], [11], [12] and references therein). Specifically, [11] studies with convergence speeds in consensus networks that are characterized by nodal degrees. The work in [11], [12], [13] also deal with the case when the network topology is time varying. A primary motivation of previous literature has been to analyze, for e.g., if a distributed consensus dynamics converges geometrically. Let us denote by $d_i$ as the distance to consensus, where $d_i < 1/16$ implies convergence. Further assume that $d_i \leq \alpha$, then it is clear that if $\alpha < 1$ the convergence is geometric. When $\alpha$ is a function of the network dimension then so does the convergence time, i.e., $t^*$ at which $d_t < 1/16$. We are specifically interested in the case when $\alpha$ varies with network dimension. The network architectures for which such dependence is known are limited, and the method of analyses do not extend to general structures.

In this work we seek to obtain a relation between the aforementioned convergence time and dimension dependent performance scaling in networks. In previous literature such scaling has been attributed to network interconnections and studies there have resulted in more “robust” network design. We believe that an understanding of robustness in the context of consensus networks will help in areas such as identification of links that slow convergence. Deriving from results in [1], we demonstrate that there exists a network projection which is equivalent to the consensus network. We will then show that the robustness of this projected network is closely related to the convergence time of its consensus counterpart. Such an alternate formulation is needed because a consensus network is marginally stable, and most performance measures require network stability. The main contribution of this work is that we provide a general set of tools that help us connect the convergence time of a consensus time to the “convergence” time (to origin) of a stable network, which can then be measured by a concrete algebraic quantity – which we refer to as robustness. Further, the framework studied here is not limited by network architecture as in previous cases, and can be extended to general time varying or random networks.

The paper is organized as follows – in section II we introduce some mathematical notation. Section III develops the model and reviews some results in consensus networks and network robustness. In Section IV we state and prove our main result. Specifically, we show how convergence time can be interpreted as a measure of robustness. We present some examples in Section V that demonstrate the tightness of our bounds. Finally we conclude in Section VI.

II. MATHEMATICAL NOTATION

Matrix Theory: A vector $v \in \mathbb{R}^{n \times 1}$ is of the form $[v_1, \ldots, v_n]^T$, where $v_i$ denotes the $i$th element, unless specified otherwise. The vector $1$ is the all 1s vector of appropriate dimension; to specify the dimension we sometimes refer to it as $1_n$, where it is a $n \times 1$ vector. Similarly, we will refer to a $n \times n$ matrix $A$ as $A_n$ when we want to specify its dimension. For a matrix, $A$, we denote by $\rho(A)$ its spectral...
radius, and by \(\sigma_q(A)\) the \(i^{th}\) largest singular value of \(A\). \(I\) is the identity matrix of appropriate dimension. The \(\mathcal{L}_p\) norm of a matrix, \(A\), is given by \(\|A\|_p = \sup_{v\neq 0} \frac{|Av|_p}{|v|_p}\) and more generally,

\[
\|A\|_{p \rightarrow q} = \sup_{|v|_p \leq 1} \|Av\|_q
\]

Finally, \(Q_n = I_n - 1_n, 1_n^T / n\) is the projection matrix. A network is Schur stable (or simply stable) if \(p(A) < 1\).

**Order Notation:** For functions, \(f(\cdot), g(\cdot)\), we have \(f(n) = O(g(n))\), when there exist constants \(C, n_0\) such that \(f(n) \leq Cg(n)\) for all \(n \in \mathbb{N} \geq n_0\). Further, if \(f(n) = O(g(n))\), then \(g(n) = \Omega(f(n))\). For functions \(g(\cdot), h(\cdot)\), we have \(g(n) = \Theta(h(n))\) when there exist constants \(C_1, C_2, n_1\) such that \(C_1 h(n) \leq g(n) \leq C_2 h(n)\) for all \(n \in \mathbb{N} > n_1\). Finally, for functions \(h_1(\cdot), h_2(\cdot)\), we have \(h_1(n) = o(h_2(n))\) when \(\lim_{n \rightarrow \infty} |h_1(n)/h_2(n)| = 0\).

**Graph Theory:** A graph is the tuple \(G = (\mathcal{V}_G, \mathcal{E}_G, w_G)\), where \(\mathcal{V}_G = \{v_1, v_2, \ldots, v_n\}\) represents the set of nodes and \(\mathcal{E}_G \subseteq \mathcal{V}_G \times \mathcal{V}_G\) represents the set of edges or communication links. An edge or link from node \(i\) to node \(j\) is denoted by \(e(i,j) = (v_i, v_j) \in \mathcal{E}_G\), and \(w_G : \mathcal{E}_G \rightarrow \mathbb{R}\). Denote by \(A_G\) the incidence matrix of \(G\). A graph, \(G\), is symmetric or undirected if \(w_G(v_i, v_j) = w_G(v_j, v_i)\) for all \(1 \leq i, j \leq |\mathcal{V}_G|\). \(G\) is induced by a matrix, \(A_{n \times n}\), if \(w_G = \{1, \ldots, n\}\), and \((i,j) \in \mathcal{E}_G\). Further, \(|A_{ij}| \neq 0\), and \(w_G(i,j) = |A_{ij}|\).

**Miscellaneous:** Denote by \(P\) the family of polynomials.

### III. PRELIMINARIES

We follow a similar model framework to [1]. Consider the following discrete time LTI dynamics:

\[
x(k+1) = A x(k) + \omega \delta(0,k), \quad k \in \{0,1,2,\ldots\}
\]

Here \(x(k) = [x_1(k), \ldots, x_n(k)]^T\) is the vector of state variables. \(A\) is the \(n \times n\) state transition matrix. \(\delta(0,k)\) is the Kronecker delta function, with \(\delta(0,0) = 1\) and \(\delta(0,k) = 0 \forall k \neq 0\) and \(\omega = [\omega_1, \ldots, \omega_n]^T\) is exogenous to the system. We further assume that \(x(0) = 0\).

We impose the following assumptions on \(A\) and \(\omega\) (if random) throughout the paper explicitly unless otherwise stated.

**Assumption 1:** \(A\) is a primitive (aperiodic and irreducible) row stochastic matrix.

**Assumption 2:** \(\omega\) is an \(n \times 1\) random vector with \(E[\omega^T] = I_{n \times n}\) and \(E[\omega] = 0\).

**Definition 1:** A signal, \(w(k)\), is a shock if \(w(k) = 0\) for all \(k > 0\) but \(w(0) \neq 0\).

Then \(w(k) = \omega\delta(0,k)\) in Eq. (1) is a shock.

**Definition 2:** A consensus network, \(\mathcal{N}(A; G)\), is a graph \(\mathcal{G} = (\mathcal{V}_G, \mathcal{E}_G, w_G)\), where \(\mathcal{V}_G = \{1,2,\ldots,n\}\), and for each node, \(i \in \mathcal{V}_G\), there is an associated dynamical behavior, \(i \rightarrow x_i(\cdot)\), given by Eq. (1). Further, \(w_G(i,j) = |A_{ij}|\) for all \(1 \leq i, j \leq n\). The graph \(\mathcal{G}\) is sometimes referred to as the network topology.

**Remark 1:** The focus of this paper will be on “large” consensus networks. Following the discussion in [1], we will denote a large network \(A\) by a sequence of networks \(\{A_n\}_{n=1}^{\infty}\) that have a “fixed” topology but growing network dimension.

**Remark 2:** Based on Definition 2 it is obvious that a network matrix \(A\) uniquely defines the consensus network \(\mathcal{N}(A; G)\). So for shorthand, we will refer to the network \(\mathcal{N}(A; G)\) by its network matrix \(A\).

Throughout this paper, we use some commonly encountered network topologies. For these topologies, the network matrices are of the form \(A_n = D_n^{-1} A_{n_0}\), where \(D_n = \text{Diag}(d_1, d_2, \ldots, d_n)\). \(A_{n_0}\) is the incidence matrix and \(d_i\) is the degree of node \(i\). Further, \(i \rightarrow j\) denotes a directed edge from \(i\) to \(j\), i.e., \([A_{n}]_{ij} = 1\) but \([A_{n}]_{ji} = 0\). The network topologies considered in this work are shown in Fig. 1.

![Fig. 1: Different Network Topologies](image-url)

Now, under Assumption 1, we have the following result on the steady state dynamics of a consensus network.

**Proposition 1 (Perron-Frobenius Theorem):** For a stochastic matrix \(A\) we have that \(\lim_{k \rightarrow \infty} A^k = 1 \pi^T\), where \(\pi^T A = \pi^T\), i.e., \(\pi\) is the invariant distribution. Additionally, \(\pi_i > 0\) for all \(i \in \{1,2,\ldots,n\}\).

It is important to note that Assumption 1 is a stronger condition than ergodicity of \(A\), i.e., there exists \(\pi\) such that \(\pi^T A = \pi\). For example, consider the directed cycle network \(DC_n\) in Fig. 1. Then it is easy to see that \(\pi_{DC_n} = (1/n)1\). Further, \(DC_n^k \neq DC_n^{k+1}\) for all \(k > 0\), and in fact \(DC_n^{m \times n} = I\) for all \(m \in \mathbb{N}\). Thus \(\lim_{k \rightarrow \infty} DC_n^k\) does not exist.

**Remark 3:** Throughout the paper we assume that Proposition 1 holds for all stochastic matrices considered unless otherwise stated.

For the class of consensus networks we consider, to quantify how “fast” they approach the invariant distribution we define the notion of convergence time.

**Definition 3:** We define \(\epsilon\)-convergence time, \(t_\epsilon(A)\), for consensus network \(A\), as

\[
t_\epsilon(A) = \inf \{k | \|A^k \pi - \pi^T \|_\infty < \epsilon\}
\]

where \(\pi\) is the invariant distribution of \(A\). We say that consensus has been reached for all \(k > t(\epsilon)\).

Further for the consensus network in Eq. (1) with an input shock we have that

\[
x(k) = A^k \omega
\]

\[
\lim_{k \rightarrow \infty} x(k) = 1(\pi^T \omega)
\]

This means that as \(t \rightarrow \infty, x_i(k) \rightarrow x_j(k)\) for all \(i, j \in \{1,2,\ldots,n\}\). Now, consider \(x^Q(\cdot)\)

\[
x^Q(k) = Q_n A^k \omega
\]

where \(x^Q(\cdot)\) is the projection of \(x(\cdot)\) onto the subspace orthogonal to \(1_n\), i.e., \(Q\). Then it follows that

\[
\lim_{k \rightarrow \infty} \|x^Q(k)\|_\infty = 0
\]
It follows from Proposition 3 in [1] that this convergence is exponential, i.e.,
\[
\lim_{k \to \infty} \|x^Q(k)\|^{1/k} = \alpha < 1
\]
The goal of this work is to show that there exists a stable large network whose robustness is equivalent to the 1/2–convergence time of the consensus network. For completeness we define robustness for a large network (See [1]).

**Definition 4:** A large network, \(A_n\), is (asymptotically) robust if we have:

- Network matrix, \(A_n\), is stable for each \(n\)
- \(\text{trace}(P(A_n)) = O(p(n))\)

Here \(p(\cdot) \in \mathcal{P}\) and \(A_n P(A_n) A_n^T + I = P(A_n)\).

**Definition 5:** The pair \(\|\cdot\|, (P(A_n))\) is a robustness measure for the network matrix \(A_n\) if \(\|\cdot\| \in \mathcal{C}\) where
\[
\mathcal{C} = \{\|\cdot\| : (p(n))^{-1}\text{trace}(P(A_n)) \leq \|A\|, \|A\| \leq (p(n))\text{trace}(P(A_n))\}
\]
and \(p(\cdot) \in \mathcal{P}\).

Examples of norms that can be considered as part of robustness measures include \(L_p\)-induced matrix norms, Schatten norms etc. Consequently, we call a large network, \(A_n\), fragile if it lacks robustness in the sense of super-polynomial or exponential scaling of \(\text{trace}(P(A_n))\).

**Remark 4:** Definition 4 uses the “controllability” gramian (assuming \(B = I\), \(P(A_n)\)), i.e.,
\[
A_n P(A_n) A_n^T + I = P(A_n)
\]
In [1], [14], robustness measure was instead based on the “observability” gramian (assuming \(C = I\)), i.e.,
\[
A_n^T P^O(A_n) A_n + I = P^O(A_n)
\]
From the perspective of asymptotic robustness, the type of gramian used is not important as \(\text{trace}(P(A_n)) = \text{trace}(P^O(A_n))\). With some more effort similar bounds (as shown in this paper) can be obtained for \(P^O(A_n)\).

**IV. CONVERGENCE TIME AS ROBUSTNESS**

In this section we will describe the relationship between the convergence time of a consensus network and the robustness of its stable, projected counterpart. We first characterize this stable network, then we formalize the relationship that we seek between robustness and convergence time, and finally we show how the two are connected. Such an analysis is similar to the one introduced in [1] for consensus networks. However, here we provide a formal justification of why such a projection is appropriate.

**A. Projected Network Dynamics**

Consider the consensus network dynamics in Eq. (1), we define the projected network dynamics (for the consensus network) as follows

**Definition 6:** The projected network dynamical behavior, \(\text{Proj}_Q(A)\), for the network matrix \(A\) is given by,
\[
x_{\Pi}(k) = \Pi A x_{\Pi}(k-1) + w(k)
\]
where \(\Pi\) is the projection operator and \(\Pi A\) is the projected network matrix.

In this work we are interested specifically in \(\text{Proj}_Q(A)\) where \(Q = I - 11T/n\). We show next the stability of the network dynamics \(\text{Proj}_Q(A)\).

**Proposition 2:** The projected network matrix \(QA\) is stable.

**Proof:** We first show that \((QA)^k = QA^k\). For \(k = 1\) this is trivially true. Now, assume that this holds for some \(k = k_0\), then we have for \((QA)^{k+1}\) that
\[
(QA)^{k+1} = (QA)^k QA = QA^k QA
\]
Now, since \(A^{k_0}\) is also row stochastic we have that \(A^{k_0}Q = A^{k_0} - 11T/n\) (since \(A^{k_0} = 1\)). Then, \(QA^{k_0}Q = QA^{k_0}\) since \(Q\) is orthogonal to \(11T/n\) by definition. Thus we have \((QA)^{k+1} = QA^{k+1}\) and our assertion is true.

From Theorem 4.9 in [10] we have that
\[
\|A^k - 1\pi^T\|_\infty \leq 2C_{\alpha}^k \implies A^k = 1\pi^T + W
\]
where \(\|W\|_\infty \leq 2C_{\alpha}^k\) and \(\alpha \in (0,1)\), then it follows that
\[
\lim_{k \to \infty} \|QA^k\|^{1/k} = \alpha < 1
\]
Since \(QA^k = (QA)^k\) we have \(\rho(QA) < 1\) and \(QA\) is stable.

The proof of Proposition 2 shows that \((QA)^k = QA^k\). Then under the same input shock, \(x_Q(k)\) (generated by \(\text{Proj}_Q(A)\)) is identical to \(x^Q(k)\), that is the projection of \(x(k)\) onto \(Q\) generated in Eq. 1, i.e., the projection of network output is equivalent to output of the network projection. This commutativity between the projection and network operators is what helps us connect the notions of robustness and convergence time.

Since the projected network dynamical behavior, \(\text{Proj}_Q(A)\), is stable, its gramian exists and is unique. This is given by the solution to the matrix equation
\[
(QA) X (QA)^T + I = X
\]
The solution will be denoted as \(X = P(QA)\). In this discussion we showed how for every consensus network that we consider, there exists a stable projected network. In the following sections we will show that there exists a strong relationship between the \(\epsilon\)-convergence time of a consensus network and the robustness of its projected network.

**B. Convergence Time in Consensus Networks**

We defined the \(\epsilon\)-convergence time of a consensus network to measure the speed of convergence to the invariant distribution. However, it seems that this speed depends on \(\epsilon\). Here we show that the choice of \(\epsilon\) is not important as long as it is less than a certain threshold.

**Proposition 3:** For a consensus network, \(A_n\),
\[
t_{A_n}(1/2) = \Theta(p(n))
\]
then
\[
t_{A_n}(\epsilon) = \Theta(p(n))
\]
for all \(\epsilon < 1\) (uniformly).

**Proof:** We first consider the case of \(\epsilon < 1/2\). If \(t_A(1/2) = \Theta(p(n))\) this means that there exists \(k = \Theta(p(n))\) such that
\[
\|A^k - 1\pi^T\|_\infty \leq (1/2)
\]
Now it follows from \(A^k 1\pi^T = 1\pi^T\) and submultiplicativity that
\[
\|A^{k+\tau} - 1\pi^T\|_\infty \leq \|A^k - 1\pi^T\|_\infty
\]
Let \( a = \inf \{ t \mid t \in \mathbb{N} \text{ and } t \geq \log (1/2\epsilon) \} \), then we have that
\[
\| A^{k_0} - \pi^T\|_\infty \leq (1/2)^a \leq \epsilon
\]
Thus \( t_A(\epsilon) = O(p(n)) \) but \( t_A(1/2) = \Theta(p(n)) \) and since \( \epsilon \leq 1/2 \implies t_A(\epsilon) = \Theta(p(n)) \) \( t_A(\epsilon) \) is non-increasing because \( \| QA^k \|_\infty \) is non-increasing in \( k \).

Next assume that \( t_{A_n}(1/2) = \Theta(p(n)) \) but \( t_{A_n}(\epsilon) \geq o(p(n)) \), for some \( 1/2 < \epsilon < 1 \). Then pick any \( a > \log (1/\epsilon) \), then \( t_{A_n}(\epsilon^a) = o(ap(n)) \implies t_{A_n}(1/2) = o(p(n)) \) because \( t_A(\cdot) \) is non-increasing. But this is a contradiction. ■

Proposition 3 states that as long as \( \epsilon \) does not depend on the network dimension the convergence times are equal in an order sense. For example if \( \epsilon = 3/4 \), then it is the same as \( \epsilon = 1/2 \) (in an order sense). However, in the definition of \( \epsilon \)-convergence time, it is possible that \( t_A(\epsilon) > 0 \) for \( \epsilon > 1 \) and therefore it is important to consider an “appropriate” class of \( \epsilon \) where the equivalence of Proposition 3 holds. For example, consider the following stochastic network, \( B_n \), and \( \alpha = 1 - 1/(n-1) \)
\[
B_n = \alpha_n I + (1 - \alpha_n)11^T/n
\]
The invariant distribution of \( B_n = 1/n \), and further since it is a positive matrix, therefore Perron Frobenius theorem holds. Then it follows that,

**Proposition 4:** For the network, \( B_n \), we have
\[
t_{B_n}(2-4/n) = 1
\]
but
\[
t_{B_n}(\epsilon^{-1}) = \Theta(n)
\]

*Proof:* We first show that \( B_n^{k_0} = \alpha^{k_0}_n I + (1 - \alpha^{k_0}_n)11^T/n \).
For the base case \( k = 1 \), it is trivial. We assume that this holds for \( k = k_0 \), then

\[
B_n^{k_0+1} = (\alpha^{k_0}_n I + (1 - \alpha^{k_0}_n)11^T/n)(\alpha I + (1 - \alpha)11^T/n)
\]

Thus our inductive hypothesis is true for \( k_0 + 1 \), and hence for all \( k \in \mathbb{N} \).

Now, observe that
\[
\| B^{k_0} - 11^T/n \|_\infty = \alpha^{k_0} \| I - 11^T/n \|_\infty = 2\alpha^{k_0}(1 - 1/n)
\]
Then for \( k = 1 \), we get that \( \| B - 11^T/n \|_\infty = 2(1 - 2/n) \). It is easy to verify that for \( k = 2(n-1) \) we see that
\[
\| B^{k} - 11^T/n \|_\infty < e^{-1}\]
but for \( k = n - 1 \) we have that
\[
\| B^{k} - 11^T/n \|_\infty > e^{-1}\]
Thus, \( t_{B_n}(\epsilon^{-1}) = \Theta(n). \) ■

We have shown that \( t_A(\epsilon) \) are essentially equal (in an order sense) for every \( \epsilon \) in a certain class; further notice that the class of \( \{ \epsilon < 1 \} \) (uniformly) is not very restrictive given that \( \| Q \|_\infty = 2, \| A \|_\infty = 1 \) for all \( k \). Following this whenever we say convergence time of \( A \), we will mean \( t_A(1/2) \).

Next, we relate convergence time to the dynamics of the projected network. This will be the first step towards making connections between the robustness of the projected network and convergence time of the original consensus network.

**Theorem 1:** For a consensus network, \( A \), we have that
\[
t_A(1/2) = \Theta(p(n))
\]
if and only if for some \( k = \Theta(p(n)) \) we have
\[
\| QA^k \|_\infty \leq 1/2
\]

*Proof:* If \( t_A(1/2) = \Theta(p(n)) \), then from Proposition 3 it follows that \( t_A(1/4) = \Theta(p(n)) \), i.e., there exists \( k = \Theta(p(n)) \) such that
\[
\| A^{k} - 11^T \|_\infty \leq 1/4
\]
Then \( QA^k = Q(A^{k} - 11^T) \) (since \( Q11^T = 0 \)) and we have
\[
\| Q(A^{k} - 11^T) \|_\infty \leq \| Q \|_\infty \| (A^{k} - 11^T) \|_\infty \leq 1/2
\]
The last inequality follows from the fact \( \| Q \|_\infty = 2 \).

Next assume that \( QA^k \leq 1/2 \) for some \( k = \Theta(p(n)) \), and since \( QA^{2k} = (QA^2)^k \) from Proposition 2, it follows that \( QA^k \leq 1/4 \) for \( k = \Theta(p(n)) \). Then we have that for some \( W_k \), in the space orthogonal to \( 11^T \) and \( \| W \|_\infty \leq 1/4 \), that
\[
A^k = 1\delta_k + W_k
\]
Now, we will show that \( \delta_k \) and \( \pi \) do not vary too much in norm.
\[
\lim_{\alpha \to \infty} A^{k_0} = 1\delta_k + \pi W_k
\]
\[
1\pi = 1\delta_k + \pi W_k
\]
\[
\| 1\pi - 1\delta_k \|_\infty \leq (1/4)\| 1\pi \|_\infty
\]
Then we have
\[
\| A^k - 1\pi \|_\infty \leq \| 1\delta_k - 1\pi \|_\infty + \| W_k \|_\infty \leq 1/2
\]
This means that \( t_A(1/2) = O(p(n)). \) ■

We will show that the convergence time of the consensus network depends on robustness of the projected network and some variational properties of the consensus network, i.e.,

\[
\text{VAR}_1(A)||P(QA)||_{(1)} \leq t_A(1/2) \leq \text{VAR}_2(A)||P(QA)||_{(2)}
\]
Here \( (||A||_{(i)}, P(QA)) \) is some appropriately chosen robustness measure and \( \text{VAR}_i(A) \) is a measure of the “variance” in \( A \). It will turn out that in some non-trivial cases robustness meets convergence time with equality (in an order sense). We describe these quantities more precisely in the following discussion.

**C. Convergence Time as a Robustness Measure**

In the preceding discussion, we related the convergence time to the norm properties of the projected network dynamics (specifically \( ||QA^k||_\infty \)). It can be shown that similar bounds can also be obtained for the projection \( Q_k = I - 11^T \).
In fact, there is some hope that these are stronger results compared to those given by \( Q \). However for the sake of completeness we limit ourselves to results for \( Q \).

**Assumption 3:**
\[
\sup_{ij} |a_{ij} - (1/n) \sum_{i=1}^n a_{ij}| > 0
\]
i.e., all rows are not identical and consensus has not been reached.

This assumption is not restrictive as a consensus network with identical rows would imply that \( t_A(\epsilon) = 1 \) for all \( \epsilon \) which is not a very interesting case. For the purpose of our main theorem we define a variance parameter.
Definition 7:

\[ \text{Var}_0(QA) = \sup |a_{i,j} - (1/n) \sum a_{i,j}| \]

Theorem 2: Let \( t_A(1/2) = t^* = \Theta(p(n)) \), then we have that

\[ (1/n)\text{trace}(P(QA)) \leq 1 + \text{Var}_0(A)\Theta(p(n)) \]

\[ 1 + (1/n)\Theta(p(n)) \leq \|P(QA)\|_2 \leq 1 + \left( \frac{\pi_{\max}}{\pi_{\min}} \right) \Theta(p(n)) \]

where \( (\pi_{\max}, \pi_{\min}) = (\max_i \pi_i, \min_i \pi_i) \).

The proof of this can be found in Section VII-B.

\[ \text{Var}_0(A) \] parameter measures the furthest any element in the stochastic matrix is away from its column average. If all elements are very close to the columnar average, then this implies consensus from preceding results. In such a case the average is the stationary value. Another point to note is that the bounds are not tight. However, they show that the grammian measures the convergence time reasonably well.

In previous literature, there exist bounds on convergence times of the form

\[ t_A(1/2) \leq C(1 - |\lambda_2|)^{-1} \]

Where \( C \) can be a function of network size. General lower bounds are obtained in terms of bottleneck ratio, which is a discrete optimization problem (See Chapter 7 [10]). Further to obtain tighter and/or lower bounds on convergence times, special structure was required. In this work we find explicitly characterize the effect of network dimension in terms of robustness and “variance”. Moreover, our bounds do not require any special structural constraints. Our result can be summarized as follows

\[ \text{trace}(P(QA)) = O(t_A(1/2)) \]

\[ n\text{Var}_0(A) = \Omega(t_A(1/2)) \]

V. Examples

We will present a few examples remarking on the tightness of our bounds. First we define a class of derived networks that will be useful in our discussion.

Definition 8: For any network \( A \), its lazy version is given by

\[ A_n^L = (1/2)(A_n + I) \]

The lazy version is useful because first, \( A_n^L \) is aperiodic and irreducible even if \( A_n \) is not and second, it preserves the topological and spectral properties of \( A_n \).

Fig. 2: Lazy versions of Network Topologies

A. Tightness of bounds

Consider the star network, \( S_n \), and the cycle network, \( C_n \), in Fig. 1. We consider the lazy versions of these networks denoted by \( S_n^L \), \( C_n^L \) respectively. Then,

Proposition 5: For the lazy random walk networks, we have \( \text{trace}(C_n^L) = \Theta(n^3), \|P(C_n^L)\|_2 = \Theta(n^2), \) and \( \text{trace}(S_n^L) = \Theta(n), \|P(S_n^L)\|_2 = \Theta(1) \).

Quite surprisingly it turns out that the star network exhibits no scaling with network dimension. The centrality in star topology lends fast “mixing” property. The cycle network has poor “mixing” properties and as a result the robustness measure scales quadratically with network dimension. Further these robustness measures are identical to convergence times for these networks (See [10]).

Proof: Star Network

\[ S_n^L = \begin{bmatrix}
0.5 & 0 & \cdots & 0 \\
0.5 & 0.5 & \cdots & 0 \\
0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0
\end{bmatrix} \]

Now, the projection of \( S_n^L \) on \( Q \) is

\[ QS_n^L = \begin{bmatrix}
0 & 1 - \frac{1}{2(n-1)} & \cdots & 0 \\
0 & -\frac{1}{2(n-1)} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & -\frac{1}{2(n-1)} & \cdots & \frac{1}{2} - \frac{1}{2(n-1)}
\end{bmatrix} \]

Now,

\[ \|P(QS_n^L)\|_2 \leq 1 + ||Q S_n^L||_2^{1/2} ||P(QS_n^L)||_2 \]

This follows from

\[ AP(A)A^T + I = P(A) \implies ||A^T P(A)A||_2 + 1 \geq ||P(A)||_2 \]

From the structure of \( QS_n^L \) it follows that \( ||QS_n^L||_2 = 0.5 \) and then it follows that \( ||P(QS_n^L)||_2 \leq 1/(1 - ||QS_n^L||_2) \). Further, \( ||P(QS_n^L)||_2 \geq ||Q||_2 = 1 \) and our claim follows.

Cycle Network

\[ C_n^L = \begin{bmatrix}
0.5 & 0.25 & \cdots & 0 & 0.25 \\
0.25 & 0.5 & \cdots & 0 & 0 \\
0 & 0.25 & \cdots & 0 & 0.25 \\
0 & 0 & \cdots & 0.5 & 0.25 \\
0.25 & 0 & \cdots & 0 & 0.5
\end{bmatrix} \]

\[ A_n^L = 1/2I_n \times n + (1/4)(DL_n + DL_n^T) \]

Here \( DL_n \) is the matrix with 1s on the off-diagonal. This is a circulant matrix (and symmetric). From [15], we have that

\[ \rho_2(A_n^L) = \cos^2(\pi/n) \]

Additionally, the eigenvalue corresponding to \( \rho_2(A_n^L) \) is positive \( \implies ||P_1(A_n^L)||_2 = \Theta(1 - \cos^2(\pi/n))^{-1} = \Theta((\sin^2(\pi/n))^{-1}) = \Theta(n^2) \). The argument for trace follows from symmetry.

The proof of the preceding propositions show that \( t_{SL}(1/2) = \Theta(1) \) and \( t_{CL}(1/2) = \Theta(n^2) \) (this follows from Section 5.3 in [10]). To show that some of the bounds in Theorem 2 are tightly achieved we provide a plot of \( ||P(QA)||_2 \) with the number of nodes in the network (Fig. 3). We observe that this ratio remains a constant for the two topologies. This is a general trend we observe for common examples. However, we have been unable to tighten our
Next, we provide upper bounds on convergence times for a class of networks discussed in [12].

**Definition 9:** A consensus network, $A_n$, is an undirected flocking network if

$$A_n = D_n^{-1} A_n$$

where $D_n$ is the matrix with degree of each node on the diagonal and $A_n$ is the incidence matrix. Further, $[A_n]_{ij} = [A_n]_{ji}$.

We will now provide a bound for the convergence time of a connected undirected flocking network. This result also readily follows from the Theorem 1 in [12].

**Proposition 6:** Any undirected flocking consensus network, $A_n$, is robust. Further $\| P(Q A_n) \| = O(n^2)$.

**Proof:** An undirected flocking matrix, as in [12], has a symmetric adjacency matrix. Then, we know that $A_n = D_n^{-1} A_n$ can be symmetrized to

$$S_n = D_n^{1/2} A_n D_n^{1/2} = D_n^{1/2} A_n D_n^{1/2}$$

Next note that $P^{'}(Q A_n) = (Q A_n)^T P^{'}(Q A_n)(Q A_n) + I$ can be alternately written $P = A_n^T P A_n + Q$, i.e., $P = P(Q A_n)$. Note that this is different from our traditional definition but in terms of the trace they are identical, as discussed before. Then we have,

$$P = A_n^T P A_n + Q$$

$$P = A_n^T D_n^{1/2} D_n^{-1/2} D_n^{-1/2} A_n + Q$$

$$D_n^{-1/2} P D_n^{-1/2} = S_n^T D_n^{-1/2} P D_n^{-1/2} S_n = D_n^{-1/2} Q D_n^{-1/2}$$

Here $Q D$ is orthogonal to $D_n^{1/2} [1, 1, \ldots, 1]^T$, which is the eigenvector of $S_n$ corresponding to the eigenvalue 1. Since $S_n$ is symmetric, the second greatest eigenvalue (in magnitude) will be in an orthogonal subspace, i.e., in $Q D$.

Since $S_n$ is symmetric, the singular values of $S_n$ are the eigenvalues of $S_n$ (albeit some permutation). From [16], we know that for a flocking matrix, $\lambda_2(S_n) \leq 1 - 1/n$, where $\lambda_2(A)$ is the second largest eigenvalue (in magnitude) of $A$.

Then, we get that $\sigma_1(D_{PD}) = O(n)$ further $D_n^{1/2} P D_n^{1/2} = P(Q A_n) \implies \sigma_1(P(Q A_n)) = O(n^2)$.

Note from Theorem 1 in [12], we have that the convergence time is $O(n^3)$. Further, since it follows from Proposition 6 that $n \| P(Q A_n) \| = O(n^3)$, our bound in Theorem 2 is tight from below.

**VI. CONCLUSIONS**

In this work we showed that convergence time for consensus networks can be represented as robustness in its projected counterpart. This gave us a general framework, one that is not limited by the type of consensus network, to estimate convergence times. The bounds obtained here do not assume knowledge of the invariant distribution. We believe that such a relation will help in analyzing cases when the network is time varying or random. For example, in [12] it was shown that there exist time varying networks where convergence is exponential. Such a result may follow from our framework in the following way — since it is clear that exponential convergence time implies lack of robustness, then all we have to do is study the scaling of $\mathcal{H}_2$—norm with network size. Another future direction may be to study network projection when the invariant distribution is known, i.e., when $Q$ is replaced by $Q\pi = I - \pi T$ where $\pi$ is the invariant distribution. For such projections, we believe that the bounds in Theorem 2 can be made tighter.
VII. APPENDIX

A. Results

For strong bounds on convergence times in terms of the gramian, we need projection operator $Q_x = I - \pi T$. We briefly show that all the properties of the projection $QA$ can be summarized for $Q_x A$

- $Q_x A^k = (Q_x A)^k$. This follows because
  $$Q_x A = AQ_x$$

- $Q_x A$ is Schur stable. This follows from Theorem 4.9 in [10].

\[ \lim_{k \to \infty} ||Q_x A^k||^{1/k} < 1 \]

- Theorem 1 holds by definition.
- $Q_x A^{m+n} = Q_x A^m Q_x A^n$. This follows because of the projection property

$$Q_x^2 = Q_x$$

B. Proof of Theorem 2

First, we show a lower bound on the convergence time.

$$P(QA) = I + \sum_{k=1}^{\infty} Q(A^k)TQ$$

$$\text{trace}(P(QA)) = n + \sum_{k=1}^{\infty} \sum_{i,j=1}^{n} |a_{ij}^{(k)} - (1/n) \sum_{i=1}^{n} a_{ij}|^2$$

$$\leq n(1 + \sup_{k=1}^{\infty} \sum_{i,j=1}^{n} |a_{ij}^{(k)} - (1/n) \sum_{i=1}^{n} a_{ij}|^2)$$

$$\leq n(1 + 2||QA||_{\text{max}} \times \sup_{k=1}^{\infty} \sum_{i,j=1}^{n} |a_{ij}^{(k)} - (1/n) \sum_{i=1}^{n} a_{ij}|)$$

Let $t^*$ be when $||QA^k||_\infty \leq 1/2$. By definition $t^* = \Theta(p(n))$, then it follows from the monotonicity of $||QA^k||_\infty$ in $k$ and the fact that $||QA^{m+n}||_\infty \leq ||QA^n||_\infty ||QA^m||_\infty$, $r_{mt^*,i} = \sum_{k=m+t^*}^{\infty} \sum_{j=1}^{n} |a_{ij}^{(k)} - (1/n) \sum_{i=1}^{n} a_{ij}|$, $r_{mt^*} = \sup_{i} r_{mt^*,i} \leq (1/2)^m t^*$

Since we have,

$$\text{trace}(P(QA)) \leq n(1 + 2||QA||_{\text{max}} \sum_{m=0}^{\infty} r_{mt^*})$$

Then it follows that

$$(1/n)\text{trace}(P(Q_x A)) = 1 + ||QA||_{\text{max}} O(p(n))$$

Consider the two scenarios - when $||QA||_{\text{max}} = o(1/n)$ and $||QA||_{\text{max}} = \Omega(1/n)$. In the first case, when $||QA||_{\text{max}} = o(1/n)$, notice that $||QA||_\infty = o(1)$ and we are done by definition, i.e., consensus has been reached. The interesting case is when $||QA||_{\text{max}} = \Omega(1/n)$, we start “away” from consensus. Indeed, a similar bound could have been obtained by taking $Q_x$ instead of $Q$. It turns out that

$$\text{trace}(P(Q_x A)) \geq \text{trace}(P(QA))$$

Thus the lower bound is tighter for $Q_x$.

For the other direction, we design a construction. First observe that, we need a lower bound $\sum_{j=1}^{n} |a_{ij} - (1/n) \sum_{i=1}^{n} a_{ij}|^2$. Although we have a lower bound for the $l_1$ by design ($> 1/4$), it is not that trivial to obtain such a bound for the $l_2$ case. Consider again the gramian, for any $k$ we have

$$P(QA) \geq \sum_{i=0}^{k} QA^i(A^T)^i Q$$

Now, the diagonal term of $[QA^i(A^T)^i Q]_{mm} = \sum_{j=1}^{n} |a_{ij} - (1/n) \sum_{i=1}^{n} a_{ij}|^2$

$$\sigma_1(P(QA)) \geq 1 + \sum_{k=1}^{t^*} \sum_{j=1}^{n} |a_{ij} - (1/n) \sum_{i=1}^{n} a_{ij}|^2$$

$$\geq 1 + \sum_{k=1}^{t^*} (1/4)n^{-1}$$

$$\geq 1 + \Theta(p(n))n^{-1}$$

Here $t^* + 1$ is the first time when $||QA||_\infty \leq 1/4$. Then the $l_2$-norm of the row with $l_2$-norm $> 1/4$ will be at least $(1/4)n^{-1}$ (By Jensen’s Inequality).

For an upper bound on $||P(QA)||_2$, we see that

$$P(QA) = \sum_{k=0}^{\infty} QA^k(A^T)^k Q$$

$$||P(QA)||_2 \leq \sum_{k=0}^{\infty} ||QA^k(A^T)^k Q||_2$$

$$||P(QA)||_2 \leq \sum_{k=0}^{\infty} ||QA^k(A^T)^k Q||_\infty$$

$$||P(QA)||_2 \leq \sum_{k=0}^{\infty} ||QA^k||_\infty ||(A^k)^T Q||_\infty$$

$$||P(QA)||_2 \leq 2\sum_{k=0}^{\infty} ||QA^k||_\infty ||(A^k)^T||_\infty$$

Now $||(A^k)^T||_\infty \leq (\pi_{\text{max}}/\pi_{\text{min}})$, this follows from the fact that

$$\pi^T A^k = \pi^T$$

$$||\pi^T A^k||_\infty \leq \pi_{\text{max}}$$

$$||\pi_{\text{min}}||_\infty ||(A^k)^T||_\infty \leq \pi_{\text{max}}$$

Then our claim follows.