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.*: $\mathcal{H}_2$–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], $\mathcal{H}_2$–norm based performance measures are studied. $\mathcal{H}_\infty$ 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. $\mathcal{H}_2$–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

T. Sarkar, M. Roozbehani, M. A. Dahleh are with the EECS Department, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: `tsarkar, mardavij,dahleh@mit.edu`).

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_t as the distance to consensus, where $d_t < 1/16$ implies convergence. Further assume that $d_t \leq \alpha^t$, then it is clear that if $\alpha < 1$ the convergence is geometric. When α 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 α 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 ith 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 $\mathbf{1}_n$, 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 $\sigma_i(A)$ the ith 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 $||A||_p = \sup_v ||Av||_p/||v||_p$ and more generally,

$$||A||_{p \to q} = \sup_{||v||_p \leq 1} ||Av||_q$$

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

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 $n \in \mathbb{N} > n_0$. 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 $\lim_{n \to \infty} |h_1(n)/h_2(n)| = 0$.

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

Miscellaneous: Denote by $\mathcal{P}$ is the family of polynomials.

III. PRELIMINARIES

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

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

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 ω (if random) throughout the paper unless explicitly stated otherwise.

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

Assumption 2: ω is an $n \times 1$ random vector with $\mathbb{E}[\omega\omega^T] = I_{n \times n}$ and $\mathbb{E}[\omega] = \mathbf{0}$.

Definition 1: A signal, $w(k)$, is a shock if $w(k) = \mathbf{0}$ for all $k > 0$ but $w(0) \neq \mathbf{0}$.
Then $w(k) = \omega\delta(0,k)$ in Eq. (1) is a shock.

Definition 2: A consensus network, $\mathcal{N}(A; \mathcal{G})$, is a graph $\mathcal{G} = (\mathcal{V}_\mathcal{G}, \mathcal{E}_\mathcal{G}, w_\mathcal{G})$, where $\mathcal{V}_\mathcal{G} = \{1, 2, \ldots, n\}$, and for each node, $i \in V_\mathcal{G}$, there is an associated dynamical behavior, $i \to x_i(\cdot)$, given by Eq. (1). Further, $w_\mathcal{G}(i,j) = [A]_{ij}$ 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 $\mathcal{N}(A; \mathcal{G})$. So for shorthand, we will refer to the network $\mathcal{N}(A; \mathcal{G})$ 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 $A_n = D_n^{-1} \mathbb{A}_n$, where $D_n = \text{Diag}(d_1, d_2, \ldots, d_n)$, $\mathbb{A}_n$ is the incidence matrix and d_i is the degree of node i. Further, $i \to j$ denotes a directed edge from i to j, *i.e.*, $[\mathbb{A}_n]_{ij} = 1$ but $[\mathbb{A}_n]_{ji} = 0$. The network topologies considered in this work are shown in Fig. 1.
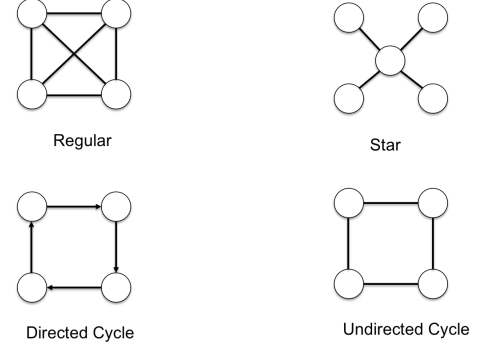

Fig. 1: Different Network Topologies

Now, based on our assumptions, 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 $\lim_{k \to \infty} A^k = \mathbf{1}\pi^T$, where $\pi^T A = \pi^T$, *i.e.*, π 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 π 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 $\pi_{DC_n} = (1/n)\mathbf{1}$. 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 $\lim_{k \to \infty} DC_n^k$ does not exist.

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 ϵ–convergence time, $t_A(\epsilon)$, for consensus network A, as

$$t_A(\epsilon) = \inf\{k| \, ||A^k - \mathbf{1}\pi^T||_\infty < \epsilon\}$$

where π 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

$$x(k) = A^k \omega$$
$$\lim_{k \to \infty} x(k) = \mathbf{1}(\pi^T \omega)$$

This means that as $t \to \infty$, $x_i(k) \to x_j(k)$ 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 $\mathbf{1}_n$, *i.e.*, Q. Then it follows that

$$\lim_{k \to \infty} ||x^Q(k)||_\infty = 0$$

It follows from Proposition 3 in [1] that this convergence is

exponential, *i.e.*,

$$\lim_{k \to \infty} ||(x^Q(k))||_\infty^{1/k} = \alpha < 1$$

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
- $||P(A_n)||_2 = O(p(n))$

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

Definition 5: The pair $(|| \cdot ||, P(A_n))$ is a robustness measure for the network matrix A_n if $|| \cdot || \in \mathcal{C}$ where

$$\mathcal{C} = \{|| \cdot || \mid (p(n))^{-1}||A||_2 \le ||A|| \le (p(n))||A||_2\}$$

and $p(\cdot) \in \mathcal{P}$

Examples of norms that can be considered as part of robustness measures include $\mathcal{L}_p$–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 $||P(A_n)||_2$.

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, $\text{Proj}_\Pi(A)$, for the network matrix A is given by,

$$x_\Pi(k) = \Pi A x_\Pi(k-1) + w(k)$$

where Π is the projection operator and ΠA is the projected network matrix.

In this work we are interested specifically in $\text{Proj}_Q(A)$ where $Q = I - 11^T/n$. 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 $(QA)^{k_0+1}$ that

$$(QA)^{k_0+1} = (QA)^{k_0} QA$$
$$= QA^{k_0} QA$$

Now, since A^{k_0} is also row stochastic we have that $A^{k_0}Q = A^{k_0} - 11^T/n$ (since $A^{k_0}1 = 1$). Then, $QA^{k_0}Q = QA^{k_0}$ since Q is orthogonal to $11^T/n$ by definition. Thus we have $(QA)^{k_0+1} = QA^{k_0+1}$ and our assertion is true.

From Theorem 4.9 in [10] we have that

$$||A^k - 1\pi^T||_\infty \le 2C\alpha^k \implies A^k = 1\pi^T + W$$

where $||W||_\infty \le 2C\alpha^k$ and $\alpha \in (0,1)$, then it follows that

$$\lim_{k \to \infty} ||QA^k||_\infty^{1/k} = \alpha < 1$$

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

$$(QA)^T X(QA) + I = X$$

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 ϵ–convergence time of a consensus network and the robustness of its projected network.

B. Convergence Time in Consensus Networks

We defined the ϵ–convergence time of a consensus network to measure the speed of convergence to the invariant distribution. However, it seems that this speed depends on ϵ. Here we show that the choice of ϵ 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

$$||A^k - 1\pi^T||_\infty \le (1/2)$$

Now it follows from $A^\tau 1\pi^T = 1\pi^T$ and submultiplicativity that

$$||A^{k+\tau} - 1\pi^T||_\infty \le ||A^k - 1\pi^T||_\infty$$

Let $a = \inf\{t \mid t \in \mathbb{N} \text{ and } t \ge \log(1/2\epsilon)\}$, then we have that

$$||A^{ka} - 1\pi^T||_\infty \le (1/2)^a$$
$$\le \epsilon$$

Thus $t_A(\epsilon) = O(p(n))$ but $t_A(1/2) = \Theta(p(n))$ and since $\epsilon \le 1/2 \implies t_A(\epsilon) = \Theta(p(n))$ ($t_A(\cdot)$ is non-increasing because $||QA^k||_\infty$ is non-increasing in k).

Next assume that $t_{A_n}(1/2) = \Theta(p(n))$ but $t_{A_n}(\epsilon) = o(p(n))$ for some $1/2 < \epsilon < 1$. Then pick any $a > \frac{\log(2)}{\log(1/\epsilon)}$, 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 ϵ 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

ϵ–convergence time, it is possible that $t_A(\epsilon) > 0$ for $\epsilon > 1$ and therefore it is important to consider an "appropriate" class of ϵ where the equivalence of Proposition 3 holds. For example, consider the following stochastic network, B_n, and $\alpha = 1 - 1/(n-1)$

$$B_n = \alpha_n I + (1 - \alpha_n)\mathbf{1}\mathbf{1}^T/n$$

The invariant distribution of $B_n = \mathbf{1}/n$, 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

$$t_{B_n}(e^{-1}) = \Theta(n)$$

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

$$B_n^{k_0+1} = (\alpha^{k_0} I + (1 - \alpha^{k_0})\mathbf{1}\mathbf{1}^T/n)(\alpha I + (1-\alpha)\mathbf{1}\mathbf{1}^T/n)$$
$$= (\alpha^{k_0+1} I + (1 - \alpha^{k_0+1})\mathbf{1}\mathbf{1}^T/n)$$

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

Now, observe that

$$||B^k - 11^T/n||_\infty = \alpha^k||I - 11^T/n||_\infty$$
$$= 2\alpha^k(1 - 1/n)$$

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, $t_{B_n}(e^{-1}) = \Theta(n)$. ∎

We have shown that $t_A(\epsilon)$ are essentially equal (in an order sense) for every ϵ in a certain class; further notice that the class of $\{\epsilon < 1 \text{ (uniformly)}\}$ is not very restrictive given that $||Q||_\infty = 2, ||A^k||_\infty = 1$ 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

$$||QA^k||_\infty \le 1/2$$

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

$$||A^k - 1\pi^T||_\infty \le 1/4$$

Then $QA^k = Q(A^k - 1\pi^T)$ (since $Q1\pi^T = 0$) and we have

$$||Q(A^k - 1\pi^T)||_\infty \le ||Q||_\infty ||(A^k - 1\pi^T)||_\infty$$
$$\le 1/2$$

The last inequality follows from the fact $||Q||_\infty = 2$.

Next assume that $QA^\tau \le 1/2$ for some $\tau = \Theta(p(n))$, and since $QA^{2\tau} = (QA^\tau)^2$ from Proposition 2, it follows that $QA^k \le 1/4$ for $k = \Theta(p(n))$. Then we have that for some

W_k, in the space orthogonal to $\mathbf{1}\mathbf{1}^T$ and $||W_k||_\infty \le 1/4$, that

$$A^k = 1\delta_k^T + W_k$$
$$\implies A^{k+\alpha} = A^\alpha 1\delta_k^T + A^\alpha W_k$$

Now, we will show that δ_k and π do not vary too much in norm.

$$\lim_{\alpha \to \infty} A^{k+\alpha} = 1\delta_k^T + 1\pi^T W_k$$
$$1\pi^T = 1\delta_k^T + 1\pi^T W_k$$
$$||1\pi^T - 1\delta_k^T||_\infty \le (1/4)||1\pi^T||_\infty$$
$$= (1/4)$$

Then we have $||A^k - 1\pi^T||_\infty \le ||1\delta_k - 1\pi^T||_\infty + ||W_k||_\infty \le (1/2)$. 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.*,

$$\text{VAR}_1(A)||P(QA)||_{(1)} \le t_A(1/2) \le \text{VAR}_2(A)||P(QA)||_{(2)}$$

Here $(||\cdot||_{(i)}, P(QA))$ 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 $||QA^k||_\infty$). We now state some properties of $||QA^k||_\infty$.

Property 1: For any stochastic matrix, A, we have
- $||QA^k||_\infty$ is non-increasing in k.
- $[QA^k]_{ij} = (a_{ij}^{(k)} - \frac{1}{n}\sum_{l=1}^n a_{lj}^{(k)})$

where $[A^k]_{ij} = a_{ij}^{(k)}$

Assumption 3:

$$\sup_{ij} |a_{ij} - (1/n)\sum_{l=1}^n a_{lj}| > 0$$

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 ϵ which is not a very interesting case.

Theorem 2: Let $t_A(1/2) = t^*$, then we have that

$$\text{trace}(P(QA)) \ge t^*(1/16n)||QA||_{2\to\infty}^{-2} + (n - 1)$$
$$\text{trace}(P(QA)) \le n(1 + 2||QA||_{\max}t^*)$$

Further if $t_A(1/2) = \Theta(p(n))$ then

$$\text{trace}(P(QA)) = O(C(n)p(n) + n) \tag{2}$$
$$\text{trace}(P(QA)) = \Omega(p(n)/B(n) + n) \tag{3}$$
$$||P(QA)||_2 = \Omega(p(n))D(n) + 1 \tag{4}$$

Here

$$B(n) = n \sup_i \sum_{j=1}^n (a_{ij} - (1/n) \sum_{l=1}^n a_{lj})^2$$

$$C(n) = n \sup_{i,j} |a_{ij} - (1/n) \sum_{l=1}^n a_{lj}|$$

$$D(n) = \inf_j \frac{\sup_i (a_{ij} - (1/n) \sum_{l=1}^n a_{lj})^2}{\sum_i^n (a_{ij} - (1/n) \sum_{l=1}^n a_{lj})^2}$$

The proof of this can be found in Section **??**.

$C(n)$ measures the distance of each column from the subspace $\mathbf{1}^T$. For example, if $A = 1\pi^T$ (*i.e.*, consensus was reached at $k \geq 1$), then $C(n) = 0$ since $a_{ij} = a_{kj}$ for each j. Specifically, $C(n)$ is low when each column of A is "close" to the subspace $\mathbf{1}^T$. Similarly, $B(n)$ measures the maximum distance of a row of A to the invariant distribution, *i.e.*, distance of the worst r_i (the ith row) to π. It might seem counter-intuitive that $B(n)$ is in the denominator, but what it shows is that the convergence might be faster if there is only one dominating term, *i.e.*, $B(n) \approx n$ but slower when there are many nodes away from the invariant distribution. $D(n)$ is similar to $B(n)$ in the sense that if there are many nodes away from from the invariant distribution then $D(n) = (1/n)$, if there is one dominating term then $D(n) = 1$. In some cases Eq. (4) gives us tighter upper bounds than Eq. (3).

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

$$\frac{\text{trace}(P(QA))}{\text{Var}_1(A)} = O(t_A(1/2))$$

$$\frac{\text{trace}(P(QA))}{\text{Var}_2(A)} = \Omega(t_A(1/2))$$

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 7: 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.

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,
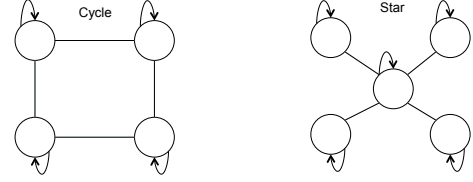

Fig. 2: Lazy versions of Network Topologies

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

$$S_n^L = \begin{bmatrix} 0.5 & \frac{1}{2(n-1)} & \cdots & \frac{1}{2(n-1)} \\ 0.5 & 0.5 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0.5 & 0 & \cdots & 0.5 \end{bmatrix}$$

Now, the projection of S_n^L on Q is

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

Now,

$$||P(QS_n^L)||_2 \leq 1 + ||QS_n^L||_2^2 ||P(QS_n^L)||_2$$

This follows from

$$A^T P(A)A + I = P(A) \implies ||A^T P(A)A||_2 + 1 \geq ||P(A)||_2$$

From the structure of QS_n^L it follows that $||QS_n^L||_2 = 0.5$ and then it follows that $||P_\Pi(S_n^L)||_2 \leq 1/(1 - ||QS_n^L||_2^2)$. Further, $||P(QS_n^L)||_2 \geq ||\Pi||_2 = 1$ and our claim follows.

Cycle Network

$$C_n^L = \begin{bmatrix} 0.5 & 0.25 & \cdots & 0 & 0.25 \\ 0.25 & 0.5 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0.5 & 0.25 \\ 0.25 & 0 & \cdots & 0 & 0.5 \end{bmatrix}$$

$$A_n^L = 1/2 I_{n \times n} + (1/4)(DL_n + DL_n^T)$$

Here DL_n is the matrix with 1s on the off-diagonal. This is a circulant matrix (and symmetric). From [14], we have that

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

Additionally, the eigenvalue corresponding to $\rho_2(A_n^L)$ is positive $\implies ||P_\Pi(A_n)||_2 = \Theta((1 - \cos^2(\pi/n))^{-1}) =$

$\Theta((\sin^2(\pi/n))^{-1}) = \Theta(n^2)$. The argument for trace follows from symmetry. ∎

The proof of the preceding propositions show that $t_{S_n^L}(1/2) = \Theta(1)$ and $t_{C_n^L}(1/2) = \Theta(n^2)$ (this follows from Section 5.3 in [10]). We see that from Theorem 2, the bounds in Eq. (2) and Eq. (4) are achieved tightly for both lazy networks. However, Eq. (3) is still loose by a factor of n. This is a general trend we observe for common examples. However, we have been unable to tighten the Eq. (3) for general consensus topologies.

Next, we provide upper bounds on convergence times for a class of networks discussed in [12].

Definition 8: A consensus network, A_n, is an undirected flocking network if

$$A_n = D_n^{-1} \mathbb{A}_n$$

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

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 $(1/n)\text{trace}(P(QA_n)) = O(n^3)$.

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

$$S_n = D_n^{-1/2} \mathbb{A}_n D_n^{-1/2} = D_n^{1/2} A_n D_n^{-1/2}$$

Next note that $P(QA_n) = (QA_n)^T P(QA_n)(QA_n) + I$ can be alternately written $P = A_n^T P A_n + Q$, *i.e.*, $P = P(QA_n)$. Then we have,

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

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 [15], 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 $\sigma_1(P) = O(n^2)$, further $D_n^{1/2} P D_n^{1/2} = P(QA_n) \implies \sigma_1(P(QA_n)) = O(n^3)$, *i.e.*, $(1/n)\text{trace}(P(QA_n)) = O(n^3)$. ∎

Note from Theorem 1 in [12], we have that the convergence time is $O(n^3)$. Further, since it follows from Proposition 6 that $(1/n)\text{trace}(P(QA_n)) = O(n^3)$, 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 $Q_\pi = I - 1\pi^T$ where π is the invariant distribution. For such projections, we believe that the upper bound in Theorem 2 (Eq. (3)) can be made tighter.

References

[1] T. Sarkar, M. Roozbehani, and M. A. Dahleh, "Robustness scaling in large networks," in *2016 American Control Conference (ACC)*, July 2016, pp. 197–202.

[2] Q. Huang, Y. Yuan, J. Gonalves, and M. A. Dahleh, "H2 norm based network volatility measures," in *2014 American Control Conference*, June 2014, pp. 3310–3315.

[3] I. Herman, D. Martinec, Z. Hurak, and M. Sebek, "Nonzero bound on fiedler eigenvalue causes exponential growth of h-infinity norm of vehicular platoon," *CoRR*, vol. abs/1409.4605, 2014. [Online]. Available: http://arxiv.org/abs/1409.4605

[4] D. Acemoglu, V. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, vol. 80, no. 5, pp. 1977–2016, 2012.

[5] G. F. Young, L. Scardovi, and N. E. Leonard, "Robustness of noisy consensus dynamics with directed communication," in *American Control Conference (ACC), 2010*. IEEE, 2010, pp. 6312–6317.

[6] M. Siami and N. Motee, "Systemic measures for performance and robustness of large-scale interconnected dynamical networks," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 5119–5124.

[7] ——, "Fundamental limits and tradeoffs on disturbance propagation in linear dynamical networks," *IEEE Transactions on Automatic Control*, vol. 61, no. 12, pp. 4055–4062, 2016.

[8] ——, "Growing linear consensus networks endowed by spectral systemic performance measures," *arXiv preprint arXiv:1706.07832*, 2017.

[9] I. C. Ipsen and T. M. Selee, "Ergodicity coefficients defined by vector norms," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 1, pp. 153–200, 2011.

[10] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Soc., 2009.

[11] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and control," *arXiv preprint math/0612682*, 2006.

[12] ——, "Degree fluctuations and the convergence time of consensus algorithms," *Automatic Control, IEEE Transactions on*, vol. 58, no. 10, pp. 2626–2631, 2013.

[13] ——, "On the nonexistence of quadratic lyapunov functions for consensus algorithms," *IEEE Transactions on Automatic Control*, vol. 53, no. 11, pp. 2642–2645, 2008.

[14] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006. [Online]. Available: http://dx.doi.org/10.1561/0100000006

[15] M. Cao, D. A. Spielman, and A. S. Morse, "A lower bound on convergence of a distributed network consensus algorithm," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*. IEEE, 2005, pp. 2356–2361.