Degree Fluctuations and the Convergence Time of Consensus Algorithms

Alex Olshevsky and John N. Tsitsiklis, *Fellow, IEEE.*

Abstract—We consider a consensus algorithm in which every node in a sequence of undirected, B-connected graphs assigns equal weight to each of its neighbors. Under the assumption that the degree of each node is fixed (except for times when the node has no connections to other nodes), we show that consensus is achieved within a given accuracy ϵ on n nodes in time $B+4n^3 B\ln(2n/\epsilon)$. Because there is a direct relation between consensus algorithms in time-varying environments and inhomogeneous random walks, our result also translates into a general statement on such random walks. Moreover, we give a simple proof of a result of Cao, Spielman, and Morse that the worst case convergence time becomes exponentially large in the number of nodes n under slight relaxation of the degree constancy assumption.

Index Terms—consensus protocols, distributed control, Markov chains.

I. INTRODUCTION

Consensus algorithms are a class of iterative update schemes that are commonly used as building blocks for the design of distributed control laws. Their main advantage is robustness in the presence of time-varying environments and unexpected communication link failures. Consensus algorithms have attracted significant interest in a variety of contexts such as distributed optimization [22], [19] coverage control [13], and many other contexts involving networks in which central control is absent and communication capabilities are time-varying.

While the convergence properties of consensus algorithms in time-varying environments are well understood, much less is known about the corresponding convergence times. An inspection of the classical convergence proofs ([4], [14]) leads to convergence time upper bounds that grow exponentially with the number of nodes. It is then natural to look for conditions under which the convergence time only grows polynomially, and this is the subject of this paper.

In our main result, we show that a consensus algorithm in which every node assigns equal weight to each of its neighbors in a sequence of undirected graphs has polynomial convergence time if the degree of any given node is constant in time (except possibly during the times when the node has no connections to other nodes).

Alex Olshevsky is with the Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Email: `aolshev2@illinois.edu`

John N. Tsitsiklis is with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. Email: `jnt@mit.edu`.

A. Model, notation, and background

In this subsection, we define our notation, the model of interest, and some background on consensus algorithms.

We will consider only undirected graphs in this paper; this will often be stated explicitly, but when unstated every graph should be understood to be undirected by default. Given a graph G, we will use $N_i(G)$ to denote the set of neighbors of node i, and $d_i(G)$ to denote the cardinality of $N_i(G)$. Given a sequence of graphs $G(0), G(1), \ldots, G(k-1)$, we will use the simpler notation $N_i(t), d_i(t)$ in place of $N_i(G(t))$, $d_i(G(t))$, and we will make a similar simplification for other variables of interest.

We are interested in analyzing a consensus algorithm in which a node assigns equal weight to each one of its neighbors. We consider n nodes and assume that at each discrete time t, node i stores a real number $x_i(t)$. We let $x(t) = (x_1(t), \ldots, x_n(t))^T$. For any given sequence of graphs $G(0), G(1), G(2), \ldots$, all on the node set $\{1, \ldots, n\}$, and any initial vector $x(0)$, the algorithm is described by the update equation

$$x_i(t+1) = \frac{1}{d_i(t)} \sum_{j \in N_i(t)} x_j(t), \qquad i = 1, \ldots, n, \quad (1)$$

which can also be written in the form

$$x(t+1) = A(t)x(t), \qquad (2)$$

for a suitably defined sequence of matrices $A(0), A(1), \ldots, A(t-1)$. The graphs $G(t)$, which appear in the above update rule through $d_i(t)$ and $N_i(t)$, correspond to information flow among the agents; the edge (i, j) is present in $G(t)$ if and only if agent i uses the value $x_j(t)$ of agent j in its update at time t. To reflect the fact that every agent always has access to its own information, we assume that every graph $G(t)$ contains all the self-loops (i, i); as a consequence, $d_i(t) \geq 1$ for all i, t. Note that we have $[A(t)]_{ij} > 0$ if and only if (i, j) is an edge in $G(t)$.

We will say that the graph sequence $G(t)$ is B-connected if, for every $k \geq 0$, the graph obtained by taking the union of the edge sets of $G(kB), G(kB+1), \ldots, G((k+1)B-1)$ is connected. It is well known ([22], [14]) that if the graph sequence is B-connected for some positive integer B, then every component of $x(t)$ converges to a common value. In this paper, we focus on the convergence rate of this process in some natural settings. To quantify the progress of the algorithm towards consensus, we will use the function $S(x) = \max_i x_i - \min_i x_i$. For any $\epsilon > 0$, a sequence of stochastic matrices

$A(0), A(1), \ldots, A(k-1)$ *results in ϵ-consensus* if

$$S(A(k-1) \cdots A(1)A(0)x(0)) \leq \epsilon S(x(0))$$

for all initial vectors $x(0)$; alternatively, a sequence of graphs $G(0), G(1), \ldots$ achieves ϵ-consensus if the sequence of matrices $A(t)$ defined by Equations (1) and (2) achieves ϵ-consensus.

As mentioned previously, we will focus on graph sequences in which every graph $G(t)$ is undirected. There are a number of reasons to be especially interested in undirected graphs within the context of consensus. For example, $G(t)$ is undirected if: (i) $G(t)$ contains all the edges between agents that are physically within some distance of each other; (ii) $G(t)$ contains all the edges between agents that have line-of-sight views of each other; (iii) $G(t)$ contains the edges corresponding to pairs of agents that can send messages to each other using a protocol that relies on acknowledgments.

It is an immediate consequence of existing convergence proofs ([4], [14]) that any sequence of $C n^{nB} \ln(1/\epsilon)$ undirected B-connected graphs, with self-loops at every node, results in ϵ-consensus. Here, C is a constant that does not depend on the problem parameters n, B, and ϵ. We are interested in simple conditions on the graph sequence under which the undesirable $O(n^{nB})$ scaling becomes polynomial in n and B.

B. Our results

Our contributions are as follows. First, in Section II, we prove our main result.

Theorem 1. *Consider a sequence $G(0), G(1), \ldots, G(k-1)$ of B-connected undirected graphs with self-loops at each node. Suppose that for each i there exists some d_i such that $d_i(t) \in \{1, d_i\}$ for all t (note that $d_i(t) = 1$ means node i has no links to any other node). If the length k of the graph sequence is at least $B + 4n^3 B \ln \frac{2n}{\epsilon}$, then ϵ-consensus is achieved.*

In Section III, we give an interpretation of our results in terms of Markov chains. Theorem 1 can be interpreted as providing a sufficient condition for a random walk on a time-varying graph to forget its initial distribution in polynomial time.

In Section IV, we capitalize on the Markov chain interpretation and provide a simple proof that relaxing the assumptions of Theorem 1 even slightly can lead to a convergence time which is exponential in n. Specifically, if we replace the assumption that each $d_i(t)$ is independent of t with the weaker assumption that the sorted degree sequence (say, in non-increasing order) is independent of t (thus allowing nodes to "swap" degrees), exponential convergence time is possible. This was proved earlier by Cao, Spielman, and Morse (although unpublished) [5] and our contribution is to provide a simple proof.

In summary: for undirected B-connected graphs with self-loops, unchanging degrees is a sufficient condition for polynomial time convergence, but relaxing it even slightly by allowing the nodes to "swap" degrees leads to the possibility of exponential convergence time.

C. Previous work

There is considerable and growing literature on the convergence time of consensus algorithms. The recent paper [14] amplified the interest in consensus algorithms and spawned a vast subsequent literature, which is impossible to survey here. We only mention papers that are closest to our own work, omitting references to the literature on various aspects of consensus convergence times that we do not address here.

Worst-case upper bounds on the convergence times of consensus algorithms have been established in [8], [6], [7], [1], [2], [10]. The papers [8], [6], [7] considered a setting slightly more general than ours, and established exponential upper bounds. The papers [1], [2] addressed the convergence times of consensus algorithms in terms of spanning trees that capture the information flow between the nodes. It was observed that in several cases this approach produces tight estimates of the convergence times. We mention also [17] which derives a polynomial-time upper bound on the time and total communication complexity required by a network of robotic agents to implement various deployment and coordination schemes. Reference [10] takes a geometric approach, and considers the convergence time in a somewhat different model, involving interactions between geographically nearest neighbors. It finds that the convergence time is quite high (either singly exponential or iterated exponential, depending on the model). Random walks on undirected graphs such as considered here are special cases of reversible agreement systems considered in the related work [11] (see also [9]). Our proof techniques are heavily influenced by the classic paper [15] and share some similarities with those used in the recent work [21], which used similar ideas to bound the convergence time of some inhomogenuous Markov chains. There are also similarities with the recent work [3] on the cover time of time-varying graphs.

Our work differs from these papers in that it studies time-varying, B-connected graphs and establishes convergence time bounds that are polynomial in n and B. To the best of our knowledge, polynomial bounds on the particular consensus algorithm considered in this paper had previously been derived earlier only in [15] (under the assumption that the graph is fixed, undirected, with self-loops at every node), [18] (in the case when the matrix is doubly stochastic, which in our setting corresponds to a sequence of regular graphs $G(t)$). For the special case of graphs that are connected at every time step ($B = 1$), the result has been apparently discovered independently by Chazelle [12] and the authors [20]. Our added generality allows for both disconnected graphs in which the degrees are kept constant, as well as the case where nodes temporarily disconnect from the network, setting their degree to one.

II. PROOF OF THEOREM 1

As in the statement of Theorem 1, we assume that we are given a sequence of undirected B-connected graphs $G(0), G(1), \ldots$, with self-loops at each node, such that $d_i(t)$ equals either d_i or 1. Observe that $d_i > 1$ for all $i = 1, \ldots, n$, since else the sequence of graphs $G(t)$ could not be B-connected. We will use the notation $\mathcal{G}$ to refer to the class

of undirected graphs with self-loops at every node such that the degree of node i either 1 or d_i. Note that the definition of $\mathcal{G}$ depends on the values $d_1, \ldots, d_n$.

Given an undirected graph G, we define the update matrix $A(G)$ by

$$[A(G)]_{ij} = \begin{cases} 1/d_i(G), & \text{if } j \in N_i(G), \\ 0, & \text{otherwise.} \end{cases}$$

We use $A(t)$ as a shorthand for $A(G(t))$, so that Eq. (1) can be written as

$$x(t+1) = A(t)x(t). \tag{3}$$

Conversely, given an update matrix A of the above form, we will use $G(A)$ to denote the graph G whose update matrix is A. We use $\mathcal{A}$ to denote the set of update matrices $A(G)$ associated with graphs $G \in \mathcal{G}$. We define $\mathbf{d}$ to be the vector $\mathbf{d} = [d_1, d_2, \ldots, d_n]^T$; a simple calculation shows that $\mathbf{d}^T A = \mathbf{d}^T$ for all $A \in \mathcal{A}$. Finally, we use D to denote the matrix whose ith diagonal element is d_i.

We begin by identifying a weighted average that is preserved by the iteration $x(t+1) = A(t)x(t)$. For any vector y, we let

$$\bar{y} = \frac{\mathbf{d}^T y}{\mathbf{d}^T \mathbf{1}} = \frac{\sum_{i=1}^n d_i y_i}{\sum_{i=1}^n d_i},$$

where $\mathbf{1}$ is the vector with entries equal to 1. Observe that for any $A \in \mathcal{A}$,

$$\overline{Ay} = \frac{\mathbf{d}^T A y}{\mathbf{d}^T \mathbf{1}} = \frac{\mathbf{d}^T y}{\mathbf{d}^T \mathbf{1}} = \bar{y}.$$

Consequently, if $x(t)$ evolves according to Eq. (3), then $\overline{x(t)} = \overline{x(0)}$, which we will from now on denote simply by $\bar{x}$.

With these preliminaries in place, we now proceed to the main part of our analysis, which is based on the pair of Lyapunov functions

$$V(x) = x^T D x = \sum_{i=1}^n d_i x_i^2, \quad \text{and} \quad V'(x) = \sum_{i=1}^n d_i (x_i - \bar{x})^2.$$

We will adopt the more convenient notation $V(t)$ for $V(x(t))$ and similarly $V'(t)$ for $V'(x(t))$.

Our first lemma provides a convenient identity for matrices in $\mathcal{A}$.

Lemma 2. *For any $A \in \mathcal{A}$ such that $G(A)$ is connected (and in particular, every node i has degree d_i),*

$$A^T D A = D - \sum_{i<j} w_{ij} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T,$$

where w_{ij} is the (i,j)-th entry of $A^T D A$.

Remark 3. *This was proven in [23] and is a generalized version of a decomposition from [24], [18]. It may be quickly verified by checking that both sides of the equation are symmetric, have identical row sums, and whenever $i < j$, the (i,j)-th element of both sides is w_{ij}. The equality of the two sides then immediately follows.*

Our next lemma quantifies the decrease of $V(\cdot)$ when a vector x is multiplied by some matrix $A \in \mathcal{A}$ associated with a connected graph $G(A)$.

Lemma 4. *Fix $x \in \mathbf{R}^n$ and let $i : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation such that $x_{i[1]} \leq x_{i[2]} \leq \cdots \leq x_{i[n]}$. For any $A \in \mathcal{A}$ such that $G(A)$ is connected,*

$$V(Ax) \leq V(x) - \frac{1}{2} \sum_{l=1}^{n-1} (x_{i[l+1]} - x_{i[l]})^2.$$

Proof. We may suppose without loss of generality that $x_1 \leq x_2 \leq \cdots \leq x_n$. Using Lemma 2,

$$V(Ax) = (Ax)^T D (Ax) = x^T A^T D A x = V(x) - \sum_{i<j} w_{ij}(x_i - x_j)^2.$$

From the definitions of w_{ij}, A, and D, we have that

$$w_{ij} = \sum_{k \in N(i) \cap N(j)} \frac{1}{d(k)},$$

and so

$$V(Ax) = V(x) - \sum_{i<j} (x_i - x_j)^2 \sum_{k \in N(i) \cap N(j)} \frac{1}{d(k)}. \tag{4}$$

Observe that if $l < k$, then

$$(x_k - x_l)^2 \geq (x_{l+1} - x_l)^2 + (x_{l+2} - x_{l+1})^2 + \cdots + (x_k - x_{k-1})^2$$

Applying this to each term of Eq. (4), we have that

$$V(Ax) \leq V(x) - \sum_{i=1}^{n-1} W_i (x_i - x_{i+1})^2,$$

where

$$W_i = \sum_{k \leq i,\ l \geq i+1} \sum_{m \in N(k) \cap N(l)} \frac{1}{d(m)} \tag{5}$$

We finish the proof by arguing that $W_i \geq 1/2$ for all $i \leq n-1$. Indeed, by the connectivity of $G(A)$, there is some node j in $\{1, \ldots, i\}$ such that j is connected to a node in $\{i+1, \ldots, n\}$. Let d^+ be the number of neighbors of node j in $\{i+1, \ldots, n\}$ and d^- be the number of neighbors of node j in $\{1, \ldots, i\}$; naturally, $d_j = d^+ + d^-$ and both d^+, d^- are at least 1: the former by the definition of j, and the latter because node j has a self-loop. Observe that the contribution to W_i in Eq. (5), by running k over all the d^- neighbors of j in $\{1, \ldots, i\}$ and running l over all d^+ neighbors of j in $\{i+1, \ldots, n\}$, is at least

$$d^+ d^- \frac{1}{d_j} \geq \frac{d_j - 1}{d_j} \geq \frac{1}{2},$$

where the final inequality is justified because the connectivity of $G(A)$ implies that $d_j \geq 2$. This concludes the proof. $\square$

Remark 5. *We note that $V(Ax) \leq V(x)$, even if $G(A)$ is not connected; this follows by applying Eq. (4) to each connected component of $G(A)$.*

Lemma 6. *Suppose that $x(t)$ evolves according to Eq. (3), where $G(A(t))$ is a sequence of B-connected graphs from $\mathcal{G}$. Let $i : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation such that $x_{i[1]}(kB) \leq x_{i[2]}(kB) \leq \cdots \leq x_{i[n]}(kB)$. Then,*

$$V(x((k+1)B)) \leq V(x(kB)) - \frac{1}{2} \sum_{l=1}^n (x_{i[l+1]}(kB) - x_{i[l]}(kB))^2.$$

Proof. It suffices to prove this under the assumption that $x_1(kB) < x_2(kB) < \cdots < x_n(kB)$; the general case then follows by a continuity argument. We apply the bound of Lemma 4 at each time $t = kB, \ldots, (k+1)B - 1$ to each connected component of $G(t)$. This yields that

$$V((k+1)B) \leq V(kB) - \frac{1}{2} \sum_{t=kB}^{(k+1)B-1} \sum_{(q,l) \in C(t)} (x_q(t) - x_l(t))^2 \tag{6}$$

Here, $C(t)$ contains all the pairs (q, l) such that there is some component of $G(t)$ containing both q and l, and $x_q(t)$ immediately follows $x_l(t)$ when the nodes in that component are ordered according to increasing values of x.

We then observe that for every $i = 1, \ldots, n-1$ there is a first time t between kB and $(k+1)B-1$ when there is a link between a node in $\{1, \ldots, i\}$ and a node in $\{i+1, \ldots, n\}$. Note that because there have been no links between $\{1, \ldots, i\}$ and $\{i+1, \ldots, n\}$ from time kB to time $t-1$, we have that

$$\max_{j=1,\ldots,i} x_j(t) \leq x_i(kB) < x_{i+1}(kB) \leq \min_{j=i+1,\ldots,n} x_j(t).$$

Moreover, at time t, the sum on the right-hand side of Eq. (6) will contain the term $(x_{i'}(t) - x_{i''}(t))^2$ where $i' \in \arg \max_{j=1,\ldots,i} x_j(t)$ and $i'' \in \arg \min_{j=i+1,\ldots,n} x_j(t)$. We conclude that it is possible to associate with every i some triplet i', i'', t such that $t \in [kB, (k+1)B-1]$, $(i', i'') \in C(t)$ and $(x_i(kB) - x_{i+1}(kB))^2 \leq (x_{i'}(t) - x_{i''}(t))^2$.

To complete the proof, we argue that distinct i are associated with distinct triplets i', i'', t. Indeed, we associate i with i', i'', t only if $x_{i'}(t) = \max_{j=1,\ldots,i} x_j(t)$ and there have been no links between $\{1, \ldots, i\}$ and $\{i+1, \ldots, n\}$ from time kB to time $t-1$. Consequently if two indices $i_1 < i_2$ are associated with the same triplet, it follows that $\arg \max_{j=1,\ldots,i_1} x_j(t) \cap \arg \max_{j=1,\ldots,i_2} x_j(t) \neq \emptyset$ which cannot be: at time kB, $x_{i_2}(kB) \geq x_{i_1+1}(kB) > \max_{j=1,\ldots,i_1} x_j(kB)$ and no link between a node in $\{1, \ldots, i_1\}$ and a node $\{i_1 + 1, \ldots, n\}$ occured from time kB to time $t-1$. $\square$

The following lemma may be verified through a direct calculation.

Lemma 7. *Suppose $u_1, \ldots, u_n$ and $w_1, \ldots, w_n$ are numbers satisfying*

$$\sum_{i=1}^{n} d_i u_i = \sum_{i=1}^{n} d_i w_i.$$

Then

$$\sum_{i=1}^{n} d_i(u_i - z)^2 - \sum_{i=1}^{n} d_i(w_i - z)^2$$

is a constant independent of the number z.

Corollary 8. *Suppose $x(t)$ evolves according to Eq. (3) where $G(A(t))$ is a sequence of B-connected graphs from $\mathcal{G}$. Let $i : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation such that $x_{i[1]}(kB) \leq x_{i[2]}(kB) \leq \cdots \leq x_{i[n]}(kB)$. Then,*

$$V'(x((k+1)B)) \leq V'(x(kB)) - \frac{1}{2} \sum_{k=1}^{n} (x_{i[l+1]}(kB) - x_{i[l]}(kB))^2.$$

Proof. Lemma 6 may be restated as

$$\sum_{i=1}^{n} d_i(x_i(kB) - 0)^2$$

$$-\sum_{i=1}^{n} d_i(x_i((k+1)B) - 0)^2 \leq \frac{1}{2} \sum_{k=1}^{n} (x_{i[l+1]}(kB) - x_{i[l]}(kB))^2$$

But since $\mathbf{d}^T x((k+1)B) = \mathbf{d}^T x(kB)$, we can apply Lemma 7 to obtain

$$\sum_{i=1}^{n} d_i(x_i(kB) - \bar{x})^2$$

$$-\sum_{i=1}^{n} d_i(x_i((k+1)B) - \bar{x})^2 \leq \frac{1}{2} \sum_{k=1}^{n} (x_{i[l+1]}(kB) - x_{i[l]}(kB))^2,$$

which is a restatement of the current corollary. $\square$

Remark 9. *An additional consequence of Lemma 7 is that $V'(Ax) \leq V'(x)$ for all $A \in \mathcal{A}$. Remark 5 had established this property for $V(\cdot)$ and Lemma 7 implies now the same property holds for $V'(\cdot)$.*

Lemma 10. *For any x,*

$$\frac{\sum_{l=1}^{n-1} (x_{i[l+1]} - x_{i[l]})^2}{V'(x)} \geq \frac{1}{n^2 d_{\max}},$$

where $d_{\max}$ is the largest of the degrees d_i.

Proof. We employ a variation of an argument first used in [15]. We first argue that we can make three assumptions without loss of generality: 1) that the components of x are sorted in nondecreasing order, i.e., $x_1 \leq x_2 \leq \cdots \leq x_n$; 2) $\sum_i d_i x_i = 0$, since both the numerator and denominator on the left-hand side are invariant under the addition of a constant to each component of x, and in particular, $V(x) = V'(x)$; 3) $V'(x) = \sum_i d_i x_i^2 = 1$, since the expression on the left-hand side remains invariant under multiplication of each component of x by a nonzero constant.

Let l be such that $d_l x_l^2 = \max_i d_i x_i^2$. Without loss of generality, we can assume that $x_l > 0$; else, we replace x by $-x$. The condition that $\sum_i d_i x_i^2 = 1$ implies that $x_l \geq 1/\sqrt{n d_{\max}}$ while the condition that $\sum_i d_i x_i = 0$ implies $x_1 < 0$. Consequently, $x_l - x_1 \geq 1/\sqrt{n d_{\max}}$. We can write this as

$$(x_2 - x_1) + (x_3 - x_2) + \cdots + (x_l - x_{l-1}) \geq \frac{1}{\sqrt{n d_{\max}}}.$$

Applying the Cauchy-Schwarz inequality, we get

$$(l-1) \sum_{i=1}^{l-1} (x_{i+1} - x_i)^2 \geq \frac{1}{n d_{\max}}.$$

We then use the fact that $l - 1 \leq n$ to complete the proof. $\square$

We can now complete the proof of Theorem 1.

Proof of Theorem 1. From Corollary 8 and Lemma 10, we have that for all integer $k \geq 0$,

$$V'((k+1)B) \leq (1 - \frac{1}{2n^3}) V'(kB).$$

Because the definition of ϵ-consensus is in terms of $S(x)$ rather than $V'(x)$, we need to relate these two quantities. On the one hand, for every x, we have

$$V'(x) = \sum_{i=1}^n d_i(x_i - \bar{x})^2 \leq n \sum_{i=1}^n (x_i - \bar{x})^2 \leq n^2 S^2(x).$$

On the other hand, for every x, we have

$$V'(x) \geq \max_i (x_i - \bar{x})^2 \geq \frac{1}{4}(\max_i x_i - \min_i x_i)^2 = \frac{1}{4}S^2(x).$$

Suppose that $t \geq B + 4Bn^3 \ln(2n/\epsilon)$. Then at least $\lceil 4n^3 \ln 2n/\epsilon \rceil$ time periods[1] of length B have passed, and therefore

$$
\begin{aligned}
S(x(t)) &\leq \sqrt{4V'(x(t))} \\
&\leq 2\Big(1 - \frac{1}{2n^3}\Big)^{4n^3 \ln(2n/\epsilon)(1/2)} \sqrt{V'(x(0))} \\
&\leq 2ne^{-\ln(2n/\epsilon)} S(x(0)) \\
&= \epsilon S(x(0)).
\end{aligned}
$$

(We have used here the inequality $(1-1/x)^x \leq e^{-1}$, for $x \geq 1$ as well as the fact that $V'(\cdot)$ is nonincreasing.) $\square$

III. MARKOV CHAIN INTERPRETATION

In this section, we give an alternative interpretation of the convergence time of a consensus algorithm in terms of inhomogeneous Markov chains; this interpretation will be used in the next section. We refer the reader to the recent monograph [16] for the requisite background on Markov chains and random walks.

We consider an inhomogeneous Markov chain whose transition probability matrix at time k is $A(k)$. We fix t and define

$$P = A(0)A(1)\cdots A(t-1).$$

This is the associated t-step transition probability matrix: the (i,j)-th entry of P, denoted by p_{ij}, is the probability that the state at time t is j, given that the initial state is i. Let $\mathbf{p}_i$ be the vector whose kth component is p_{ik}; thus $\mathbf{p}_i^T$ is the ith row of P.

We address a question which is generic in the study of Markov chains, namely, whether the chain eventually "forgets" its initial state, i.e., whether for all i, j, $\mathbf{p}_i - \mathbf{p}_j$ converges to zero as t increases, and if so, at what rate. We will say that the sequence of matrices $A(0), A(1), \ldots, A(t-1)$ is ϵ-*forgetful* if for all i, j, we have

$$\frac{1}{2}\sum_k |p_{ik} - p_{jk}| \leq \epsilon.$$

The above quantity, $\frac{1}{2}\max_{i,j}\|\mathbf{p}_i - \mathbf{p}_j\|_1$ is known as the *coefficient of ergodicity* of the matrix P, and appears often in the study of consensus algorithms (see, for example, [8]). The result that follows relates the times to achieve ϵ-consensus or ϵ-forgetfulness, and is essentially the same as Proposition 4.5 of [16].

Proposition 11. *The sequence of matrices* $A(0), A(1), \ldots,$ $A(t-1)$ *is ϵ-forgetful if and only if the sequence of matrices*

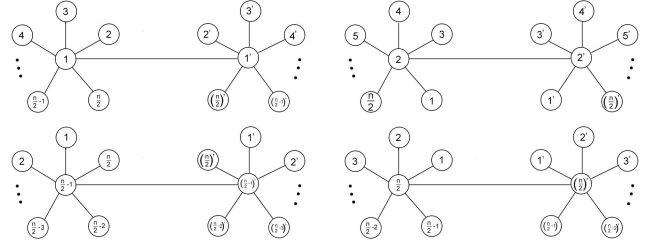[1]The notation $\lceil x \rceil$ means the smallest integer which is at least x.

Fig. 1. The top-left figure shows graph $G(0)$; top-right shows $G(1)$; bottom-left shows $G((n/2) - 2)$; bottom-right shows $G((n/2) - 1)$. As these figures illustrate, $G(t + 1)$ is obtained by applying a circular shift to each half of $G(t)$. Every node has a self-loop which is not shown. For aesthetic reasons, instead of labeling the nodes as $1, \ldots, n$, we label them with $1, \ldots, n/2$ and $1', \ldots, (n/2)'$.

$A(t-1), A(t-2), \ldots, A(0)$ *results in ϵ-consensus (i.e.,* $S(Px) \leq \epsilon S(x)$, *for every vector fx.)*

Proof. Suppose that the matrix sequence $A(0), A(1), \ldots,$ $A(t-1)$ is ϵ-forgetful, i.e., that $\frac{1}{2}\sum_k |p_{ik} - p_{jk}| \leq \epsilon$, for all i and j. Given a vector x, let $c = (\max_k x_k + \min_k x_k)/2$. Note that $\|x - c\mathbf{1}\|_\infty = (\max_k x_k - \min_k x_k)/2 = S(x)/2$. We then have

$$
\begin{aligned}
|[Px]_i - [Px]_j| &= \Big|\sum_k (p_{ik} - p_{jk})(x_k - c)\Big| \\
&\leq \|\mathbf{p}_i - \mathbf{p}_j\|_1 \cdot \|x - c\mathbf{1}\|_\infty \\
&\leq \epsilon S(x).
\end{aligned}
$$

Since this is true for every i and j, we obtain $S(Px) \leq \epsilon S(x)$, and the sequence $A(t-1), A(t-2), \ldots, A(0)$ results in ϵ-consensus.

Conversely, suppose that the sequence of matrices $A(t-1), A(t-2), \ldots, A(0)$ results in ϵ-consensus. Fix some i and j. Let x be a vector whose kth component is $1/2$ if $p_{ik} \geq p_{jk}$ and $-1/2$ otherwise. Note that $S(x) = 1$. We have

$$\frac{1}{2}\|\mathbf{p}_i - \mathbf{p}_j\|_1 = (\mathbf{p}_i^T - \mathbf{p}_j^T)x = [Px]_i - [Px]_j \leq \epsilon S(x) = \epsilon,$$

where the last inequality made use of the ϵ-consensus assumption. Thus, the sequence of matrices $A(0), A(1), \ldots, A(t-1)$ is ϵ-forgetful. $\square$

We will use Proposition 11 for the special case of Markov chains that are random walks. Given an undirected graph sequence sequence $G(0), G(1), \ldots$, we consider the random walk on the state-space $\{1, \ldots, n\}$ which, at time t, jumps to a uniformly chosen random neighbor of its current state in $G(t)$. Proposition 11 allows us to reinterpret Theorem 1 as follows: random walks on time-varying undirected B-connected graphs with self-loops and degree constancy forget their initial distribution in a polynomial number of steps.

IV. A COUNTEREXAMPLE

In this subsection, we show that it is impossible to omit the condition of unchanging degrees in Theorem 1. In particular, if we only impose the slightly weaker condition that the sorted degree sequence (the non-increasing list of node degrees) does not change with time, the time to achieve ϵ-consensus can grow exponentially with n. This is an unpublished result of

Cao, Spielman, and Morse [5]; we provide here a simple proof. We note that the graph sequence used in the proof (see Figure 1) is similar to the sequence used in [3] to prove an exponential lower bound on the cover time of time-varying graphs.

Proposition 12. *Let n be even and let t be an integer multiple of $n/2$. Consider the graph sequence of length $t = kn/2$, consisting of periodic repetitions of the reversal[2] of the length-$n/2$ sequence described in Figure 1. For this graph sequence to result in $(1/4)$-consensus, we must have $t \geq 2^{(n/2)}/8$.*

Proof. Suppose that this graph sequence of length t results in $(1/4)$-consensus. Then Proposition 11 implies that the sequence G' of length t consisting of periodic repetitions[3] of the length $n/2$ sequence described in Figure 1 is $(1/4)$-forgetful. Let p_{ij} be the associated t-step transition probabilities.

Let T be the time that it takes for a random walk that starts at state $n/2$ at time 0 to cross into the right-hand side of the graph, let δ be the probability that T is less than or equal to t, and define R to be the set of nodes on the right side of the graph, i.e., $R = \{1' \ldots (n/2)'\}$. Clearly,

$$\sum_{j' \in R} p_{(n/2),j'} \leq P(T \leq t) = \delta,$$

since a walk located in R at time t has obviously transitioned to the right-hand side of the graph by t. Next, symmetry yields $\sum_{j' \in R} p_{(n/2)',j'} \geq 1 - \delta$. Using the fact that the graph sequence is $(1/4)$-forgetful in the first inequality below, we have

$$
\begin{aligned}
\frac{1}{2} &\geq \sum_{j' \in R} \left| p_{(n/2)',j'} - p_{((n/2),j'} \right| \\
&\geq \sum_{j' \in R} p_{(n/2)',j'} - \sum_{j' \in R} p_{(n/2),j'} \\
&\geq (1 - \delta) - \delta \\
&= 1 - 2\delta,
\end{aligned}
$$

which yields that $\delta \geq 1/4$. By viewing periods of length t as a single attempt to get to the right half of the graph, with each attempt having probability at least $1/4$ to succeed, we obtain $E[T] \leq 4t$.

Next, let us say that node i has *emerged* at time t if node i was the center of the left-star in $G'(t-1)$; for example, node 1 has emerged at time 1, node 2 has emerged at time 2, and so on. By symmetry, T is the expected time until a random walk starting at an emerged node crosses to the right-hand side of the graph. Observe that, starting from an emerged node, the random walk will transition to the right-hand side of the graph if it takes the self-loop $n/2 - 1$ consecutive times and then, once it is at the center, takes the link across; however, if it fails to take the self-loop during the first $n/2 - 1$ times, it then transitions to a newly emerged node. This implies that the expected time to transition to the right hand side from an

emerged node is at least the expected time until the walk takes $n/2 - 1$ self-loops consecutively: $2^{(n/2)-1} \leq E[T]$.

Putting this together with the previous inequality $E[T] \leq 4t$, we immediately have the desired result. $\qquad\square$

REFERENCES

[1] D. Angeli, P.-A. Bliman, "Tight estimates for convergence of some non-stationary consensus algorithms," *Systems and Control Letters,* vol. 57, no. 12, pp. 996-1004, 2008.

[2] D. Angeli, P.-A. Bliman, "Convergence speed of unsteady distributed consensus: decay estimate along the settling spanning-trees," *SIAM Journal on Control and Optimization,* vol. 48, no. 1, pp. 1-32, 2009.

[3] C. Avin, M. Koucky, Z. Lotker, "How to explore a fast-changing world," *Proceedings 35th International Colloquium on Automata, Languages and Programming*, 2008.

[4] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods,* Prentice Hall, 1989.

[5] Ming Cao, personal communication, 2006.

[6] M. Cao, A. S. Morse, B. D. O. Anderson, "Reaching a consensus in a dynamically changing environment: a graphical approach," *SIAM Journal on Control and Optimization,* vol. 47, no. 2, pp. 575-600, 2008.

[7] M. Cao, A. S. Morse, B. D. O. Anderson, "Reaching a consensus in a dynamically changing environment: convergence rates, measurement delays, and asynchronous events," *SIAM Journal on Control and Optimization,* vol. 47, no. 2, pp. 575-600, 2008.

[8] M. Cao, D. Spielman, A. S. Morse, "A lower bound on convergence of a distributed network consensus slgorithm," *Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference,* Madrid, Spain, Dec. 2005.

[9] B. Chazelle, "The convergence of bird flocking," *Proceedings of the 26th Annual Symposium on Computational Geometry,* Snowbird, USA, Jun. 2010.

[10] B. Chazelle, "Natural algorithms," *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms,* New York, USA, Jan. 2009.

[11] B. Chazelle, "'The total s-energy of a multiagent system," *SIAM Journal on Control and Optimization*, vol. 49, no. 4, pp. 1680-1706, 2011.

[12] B. Chazelle, "The total s-energy of a multiagent system," http://arxiv.org/abs/1004.1447, April 2010.

[13] C. Gao, J. Cortes, F. Bullo, "Notes on averaging over acyclic graphs and discrete coverage control," *Automatica,* vol. 44, no. 8, pp. 2120-2127, 2008.

[14] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 3, pp. 988-1001, 2003.

[15] H. J. Landau, A. M. Odlyzko, "Bounds for eigenvalues of certain stochastic matrices," *Linear Algebra and Its Applications,* vol. 38, pp. 5-15, 1981.

[16] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times,* American Mathematical Society, 2008.

[17] S. Martinez, F. Bullo, J. Cortes, E. Frazzoli, "On synchronous robotic networkspart II: time complexity of rendezvous and deployment algorithms," *IEEE Transactions on Automatic Control*, vol. 52, no. 12,pp. 2214-2226, 2007.

[18] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Transactions on Automatic Control,* vol. 54, no. 11, pp. 2506-2517, 2009.

[19] A. Nedic, A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control,* vol. 54, no. 1, pp. 48-61, 2009.

[20] A. Olshevsky, "Degree Flucutations and Convergence Times of Consensus Algorithms," talk delivered at Princeton University, Dec 2010, http://web.engr.illinois.edu/~aolshev2/princetonslides.ps.

[21] L. Saloff-Coste, J. Zuniga, "Convergence of some time inhomogeneous Markov chains via spectral techniques," *Stochastic Processes and their Applications,* vol. 117, pp. 961-979, 2007.

[22] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control,* vol. 31, no. 9, 1986, pp. 803-812.

[23] B. Touri and A. Nedic, "On existence of a quadratic comparison function for random weighted averaging dynamics and its implications," *Proceedings of the 50th IEEE Conference on Decision and Control,* 2011.

[24] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, 53:65-78, 2004

---

[2]That is, we are considering the sequence $G(n/2 - 1), \ldots, G(1), G(0), G(n/2-1), \ldots, G(1), G(0), G(n/2) - 1, \ldots$.

[3]That is, $G'(t)$ is the sequence $G(0), G(1) \ldots, G(n/2 - 1), G(0), G(1), \ldots, G(n/2 - 1), G(0), G(1), \ldots$.