2013 American Control Conference (ACC)
Washington, DC, USA, June 17-19, 2013

Convergence of Distributed Averaging and Maximizing Algorithms Part II: State-dependent Graphs

Guodong Shi and Karl Henrik Johansson

Abstract— **In this paper, we formulate and investigate a generalized consensus algorithm which makes an attempt to unify distributed averaging and maximizing algorithms considered in the literature. Each node iteratively updates its state as a time-varying weighted average of its own state, the minimal state, and the maximal state of its neighbors. In Part I of the paper, time-dependent graphs are studied. This part of the paper focuses on state-dependent graphs. We use a μ-nearest-neighbor rule, where each node interacts with its μ nearest smaller neighbors and the μ nearest larger neighbors. It is shown that $\mu+1$ is a critical threshold on the total number of nodes for the transit from finite-time to asymptotic convergence for averaging, in the absence of node self-confidence. The threshold is 2μ if each node chooses to connect only to neighbors with unique values. The results characterize some similarities and differences between distributed averaging and maximizing algorithms.**

Index Terms— **Averaging algorithms, Max-consensus, Finite-time convergence**

I. Introduction

Distributed averaging algorithms and max-consensus algorithms are two basic models for distributed information processing over networks. In general they tell a same story that nodes exchange information with its neighbors under certain communication graph, update their states based on the information received, and a collective state convergence to a common state will eventually be achieved. Applications for averaging algorithms can be found in engineering [11], [12], [30], computer science [8], [9], and social science [5], [6], [7]. Max-consensus algorithms have been widely used for leader election, network size estimation, and various applications in wireless networks [34], [30], [29].

Central to the study of averaging and maximizing algorithms is the convergence to a consensus. It can be hard to analyze due to the switching underlying communication graph, and various convergence conditions have been established for time-dependent graphs [11], [23], [12], [13], [15], [16], [14], [17], [16]. Asymptotic convergence is common in the study of averaging consensus algorithms [14], [15], [12], while it has been shown that maximizing algorithms converge in general in finite time [34], [35], [36]. Finite-time convergence of averaging algorithms was investigated in [30], [32], [33] for continuous-time models, and recently finite-time consensus in discrete time was discussed in [40] for a special case of gossiping [39].

This work has been supported in part by the NSFC of China under Grant 61120106011, Knut and Alice Wallenberg Foundation, the Swedish Research Council, and KTH SRA TNG.

G. Shi and K. Johansson are with ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm 10044, Sweden.Email: guodongs@kth.se, kallej@kth.se

The switching topology can be dependent on the node states. For instance, in Krause's model, each node is connected only to nodes within a certain distance [21]. Vicsec's model has a similar setting but with higher-order node dynamics [20]. Because the node dynamics is coupled with the graph dynamics for state-dependent graphs, the convergence analysis is quite challenging. Deterministic consensus algorithms with state-dependent graph were studied in [22], [26], and convergence results for state-dependent interactions under probabilistic models were established in [24], [25].

In this paper, we make the simple observation that averaging and maximizing algorithms can be viewed as instances of a more general distributed processing model. Using this model the transition of the consensus convergence can be studied for the two classes of distributed algorithms in a unified way. Each node iteratively updates its state as a weighted average of its own state together with the minimum and maximum states of its neighbors. By special cases for the weight parameters, averaging and maximizing algorithms can be analyzed.

This part of the paper considers time-dependent graphs. In both Krause's [21] and Vicsek's [20] models, nodes interact with neighbors whose distance is within a certain communication range. Recently, it was discovered through empirical data that in a bird flock each bird seems to interact with a fixed number of nearest neighbors, rather than with all neighbors within a fixed metric distance [27]. Nearest-neighbor model has been studied under a probabilistic setting on the graph connectivity for wireless communication networks [28]. From a social network point of view, the evolution of opinions may result from similar models since members tend to exchange information with a fixed number of other members who hold a similar opinion as themselves [5], [26].

We use a μ-nearest-neighbor rule to generate state-dependent graphs, in which each node interacts with its μ nearest smaller neighbors (μ neighbors with smaller state values), and the nearest μ larger neighbors. This model is motivated from recent studies of collective bird behavior [27]. For averaging algorithms without node self-confidence under such state-dependent graphs, we show that $\mu+1$ is a critical value for the total number of nodes: finite-time consensus is achieved globally if the number of nodes is no larger than $\mu+1$, and finite-time consensus fails for almost all initial conditions if the number of nodes is larger than $\mu+1$. Moreover, it is shown that this critical number of nodes is instead 2μ if each node chooses to connect only to neighbors with distinct values in the neighbor rule. Time-

dependent graph models are studied in Part I of the paper [41], and a complete version of the paper can be found in [42].

The rest of the paper is organized as follows. In Section II we introduce the considered network model, the state dependent node interaction, the uniform processing algorithm, and the consensus problem. The main results are presented in Section III. Finally some concluding remarks are given in Section IV.

II. Problem Definition

In this section, we introduce the network model, the considered algorithm, and define the problem of interest.

A. Network

We first recall some concepts and notations in graph theory [1]. A directed graph (digraph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a finite set $\mathcal{V}$ of nodes and an arc set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. An element $e = (i, j) \in \mathcal{E}$ is called an *arc* from node $i \in \mathcal{V}$ to $j \in \mathcal{V}$. If the arcs are pairwise distinct in an alternating sequence $v_0 e_1 v_1 e_2 v_2 \dots e_k v_k$ of nodes $v_i \in \mathcal{V}$ and arcs $e_i = (v_{i-1}, v_i) \in \mathcal{E}$ for $i = 1, 2, \dots, k$, the sequence is called a (directed) *path* with *length* k. If there exists a path from node i to node j, then node j is said to be reachable from node i. Each node is thought to be reachable by itself. A node v from which any other node is reachable is called a *center* (or a *root*) of $\mathcal{G}$. A digraph $\mathcal{G}$ is said to be *strongly connected* if node i is reachable from j for any two nodes $i, j \in \mathcal{V}$; *quasi-strongly connected* if $\mathcal{G}$ has a center [2]. The *distance* from i to j in a digraph $\mathcal{G}$, $d(i, j)$, is the length of a shortest simple path $i \to j$ if j is reachable from i, and the *diameter* of $\mathcal{G}$ is $\mathrm{diam}(\mathcal{G}) = \max\{d(i, j) | i, j \in \mathcal{V},\ j \text{ is reachable from } i\}$. The union of two digraphs with the same node set $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$ is defined as $\mathcal{G}_1 \cup \mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$. A digraph $\mathcal{G}$ is said to be bidirectional if for every two nodes i and j, $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$. A bidirectional graph $\mathcal{G}$ is said to be *connected* if there is a path between any two nodes.

Consider a network with node set $\mathcal{V} = \{1, 2, \dots, n\}$, $n \geq 3$. Time is slotted. Denote the state of node i at time $k \geq 0$ as $x_i(k) \in \mathbb{R}$. Then $x(k) = \big(x_1(k) \dots x_n(k)\big)^T$ represents the network state.

Throughout this paper, we call node j a *neighbor* of node i if there is an arc from j to i in the graph. Each node is supposed to always be a neighbor of itself. Let $\mathcal{N}_i(k)$ represent the neighbor set of node i at time k.

B. State-dependent Communication

In this section, we consider a network model in which nodes interact only with other nodes having a close state value. Consider the following nearest-neighbor rule.

Definition 2.1 (Nearest-neighbor Graph): For a positive integer μ and any node $i \in \mathcal{V}$, there is a link entering i from each node in the set $\mathcal{N}_i^-(k) \cup \mathcal{N}_i^+(k)$, where $\mathcal{N}_i^-(k) = \big\{\text{nearest } \mu \text{ neighbors from } \{j \in \mathcal{V} : x_j(k) < x_i(k)\}\big\}$ denotes the nearest smaller neighbor set, and $\mathcal{N}_i^+(k) = \big\{\text{nearest } \mu \text{ neighbors from } \{j \in \mathcal{V} : x_j(k) > x_i(k)\}\big\}$ denotes the nearest larger neighbor set. The graph defined by this nearest neighbor rule is denoted as $\mathcal{G}^{\mu n}_{x(k)}$, $k = 0, 1, \dots$.

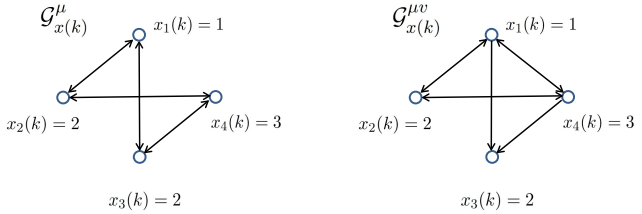

Fig. 1. Examples of nearest-neighbor graph $\mathcal{G}^{\mu n}_{x(k)}$ and nearest-value graph $\mathcal{G}^{\mu v}_{x(k)}$ for $\mu = 2$. Note that for a given set of states, these graphs are in general not unique.

Naturally, if there are less than μ nodes with states smaller than $x_i(k)$, $\mathcal{N}_i^-(k)$ has less that μ elements. Similar condition holds for $\mathcal{N}_i^+(k)$. Hence, the number of neighbor nodes is not necessarily fixed in the nearest-neighbor graph.

Remark 2.1: Note that, at each time k, the nearest-neighbor graph is uniquely determined by the node states. The node interactions are indeed determined by the distance between the node states. In this sense, the nearest-neighbor graph shares similar structure with Krause's model [21], [22], where each node communicates with the nodes within certain radius. This nearest-neighbor graph also fulfills the interaction structure in the bird flock model discussed in [27] since each node communicates with an almost fixed number of neighbors, nearest from above and below.

Note that in the definition of the nearest-neighbor graph, nodes may have neighbors with the same state values. We consider the following nearest-value graph, where each node considers only neighbors with different state values.

Definition 2.2: (Nearest-value Graph) For a positive integer μ and any node $i \in \mathcal{V}$, there is a link entering i from each node in the set $\mathcal{N}_i^-(k) \cup \mathcal{N}_i^+(k)$, where $\mathcal{N}_i^-(k) = \big\{\text{nearest } \mu \text{ neighbors with different values from } \{j \in \mathcal{V} : x_j(k) < x_i(k)\}\big\}$ denotes the nearest smaller neighbor set, and $\mathcal{N}_i^+(k) = \big\{\text{nearest } \mu \text{ neighbors with different values from } \{j \in \mathcal{V} : x_j(k) > x_i(k)\}\big\}$ denotes the nearest larger neighbor set. The graph defined by this nearest neighbor rule is denoted as $\mathcal{G}^{\mu v}_{x(k)}$, $k = 0, 1, \dots$.

An illustration of nearest-neighbor and nearest-value graphs at a specific time instance k is shown in Figure 1 for $n = 4$ nodes and $\mu = 2$.

C. Algorithm

The classical average consensus algorithm in the literature is given by

$$x_i(k+1) = \sum_{j \in \mathcal{N}_i(k)} a_{ij}(k) x_j(k), \quad i = 1 \dots, n. \qquad (1)$$

Two standing assumptions are fundamental in determining the nature of its dynamics:

A1 *(Local Cohesion)* $\sum_{j \in \mathcal{N}_i(k)} a_{ij}(k) = 1$ for all i and k;

A2 *(Self-confidence)* There exists a constant $\eta > 0$ such that $a_{ii}(k) \geq \eta$ for all i and k.

These assumptions are widely imposed in the existing works, e.g., [12], [11], [19], [14], [15], [23]. With A1 and A2, we can always write the average consensus algorithm (1) into the following equivalent form [42]:

$$x_i(k+1) = \eta x_i(k) + \alpha_k^{\langle i\rangle} \min_{j\in\mathcal{N}_i(k)} x_j(k) + \big(1-\eta-\alpha_k^{\langle i\rangle}\big) \max_{j\in\mathcal{N}_i(k)} x_j(k), \qquad (2)$$

where $\alpha_k^{\langle i\rangle} \in [0, 1-\eta]$ for all i and k. Thus, the information processing principle behind distributed averaging is that each node iteratively takes a weighted average of its current state and the minimum and maximum states of its neighbor set.

The standard maximizing algorithm [34], [35], [36] is defined by

$$x_i(k+1) = \max_{j\in\mathcal{N}_i(k)} x_j(k), \qquad (3)$$

so distributed maximizing is each node interacting with its neighbors and simply taking the maximal state within its neighbor set.

In this paper, we aim to present a model under which we can discuss fundamental differences of some distributed information processing mechanisms. We consider the following algorithm for the node updates:

$$x_i(k+1) = \eta_k x_i(k) + \alpha_k \min_{j\in\mathcal{N}_i(k)} x_j(k) + \big(1-\eta_k-\alpha_k\big) \max_{j\in\mathcal{N}_i(k)} x_j(k), \qquad (4)$$

where $\alpha_k, \eta_k \geq 0$ and $\alpha_k + \eta_k \leq 1$. We denote the set of all algorithms of the form (4) by $\mathcal{A}$, when the parameter (α_k, η_k) takes value as $\eta_k \in [0,1], \alpha_k \in [0, 1-\eta_k]$. This model is a special case of (2) as the parameter α_k is not depending on the node index i in (4).

Note that $\mathcal{A}$ represents a uniform model for distributed averaging and maximizing algorithms. Obeying the cohesion and self-confidence assumptions, the set of (weighted) averaging algorithms, $\mathcal{A}_{\rm ave}$, consists of algorithms in the form of (4) with parameters $\eta_k \in (0,1], \alpha_k \in [0, 1-\eta_k]$. The set of maximizing algorithms, $\mathcal{A}_{\max}$, is given by the parameter set $\eta_k \equiv 0$ and $\alpha_k \equiv 0$.

D. Problem

Let $\big\{x(k;x^0) = \big(x_1(k;x^0)\dots x_n(k;x^0)\big)^T\big\}_0^\infty$ be the sequence generated by (4) for initial time k_0 and initial value $x^0 = x(k_0) = \big(x_1(k_0)\dots x_n(k_0)\big)^T \in \mathbb{R}^n$. We will identify $x(k;x^0)$ as $x(k)$ in the following discussions. We introduce the following definition on the convergence of the considered algorithm.

Definition 2.3: (i) Asymptotic consensus is achieved for Algorithm (4) for initial condition $x(k_0) = x^0 \in \mathbb{R}^n$ if there exists $z_*(x^0) \in \mathbb{R}$ such that

$$\lim_{k\to\infty} x_i(k) = z_*, \quad i = 1,\dots,n.$$

Global asymptotic consensus is achieved if asymptotic consensus is achieved for all $k_0 \geq 0$ and $x^0 \in \mathbb{R}^n$.

(ii) Finite-time consensus is achieved for Algorithm (4) for initial condition $x(k_0) = x^0 \in \mathbb{R}^n$ if there exist $z_*(x^0) \in \mathbb{R}$ and an integer $T_*(x^0) > 0$ such that

$$x_i(T_*) = z_*, \quad i = 1,\dots,n.$$

Global finite-time consensus is achieved if finite-time consensus is achieved for all $k_0 \geq 0$ and $x^0 \in \mathbb{R}^n$.

The algorithm reaching consensus is equivalent with that $x(k)$ converges to the manifold

$$\mathrm{C} = \Big\{x = (x_1\dots x_n)^T : \ x_1 = \dots = x_n\Big\}.$$

We call C the consensus manifold. Its dimension is one.

In the following, we focus on the impossibilities and possibilities of asymptotic or finite-time consensus. We will show that the convergence properties drastically change when Algorithm (4) transits from averaging to maximizing.

III. Main Results

In this section, we investigate the convergence of Algorithm (4) for state-dependent graphs. We are interested in a particular set of averaging algorithms, $\mathcal{A}^*_{\rm ave}$, where (α_k, η_k) takes value $\eta_k \equiv 0$, $\alpha_k \in (0,1)$. Algorithms in $\mathcal{A}^*_{\rm ave}$ correspond to the case when the self-confidence assumption A2 does not hold, and are of the form

$$x_i(k+1) = \alpha_k \min_{j\in\mathcal{N}_i(k)} x_j(k) + \big(1-\alpha_k\big) \max_{j\in\mathcal{N}_i(k)} x_j(k). \qquad (5)$$

Algorithms in $\mathcal{A}^*_{\rm ave}$ still have local cohesion. Hence, they fulfill Assumption A1 but not A2. In fact, averaging algorithms without self-confidence have been investigated in classical works on the convergence of product of stochastic matrices, e.g., [3], [4], [5].

A. Basic Lemmas

We first establish two useful lemmas for the analysis of nearest-neighbor and nearest-value graphs. The following lemma indicates that the order of node states is preserved.

Lemma 3.1: For any two nodes $u, v \in \mathcal{V}$ and every algorithm in $\mathcal{A}$, under either the nearest-neighbor graph $\mathcal{G}^{\mu n}_{x(k)}$ or the nearest-value graph $\mathcal{G}^{\mu v}_{x(k)}$, we have

(i) $x_u(k+1) = x_v(k+1)$ if $x_u(k) = x_v(k)$;

(ii) $x_u(k+1) \leq x_v(k+1)$ if $x_u(k) < x_v(k)$.

Proof. When $x_u(k) = x_v(k)$, we have $\{j : x_j(k) < x_u(k)\}$=$\{j : x_j(k) < x_v(k)\}$ and $\{j : x_j(k) > x_u(k)\}$=$\{j : x_j(k) > x_v(k)\}$. Thus, for either $\mathcal{G}^{\mu n}_{x(k)}$ or $\mathcal{G}^{\mu v}_{x(k)}$, both

$$\min_{j\in\mathcal{N}_u(k)} x_j(k) = \min_{j\in\mathcal{N}_v(k)} x_j(k)$$

and

$$\max_{j\in\mathcal{N}_u(k)} x_j(k) = \max_{j\in\mathcal{N}_v(k)} x_j(k)$$

hold. Then (i) follows straightforwardly.

If $x_u(k) < x_v(k)$, it is easy to see that

$$\min_{j\in\mathcal{N}_u(k)} x_j(k) \leq \min_{j\in\mathcal{N}_v(k)} x_j(k)$$

and

$$\max_{j\in\mathcal{N}_u(k)} x_j(k) \le \max_{j\in\mathcal{N}_v(k)} x_j(k)$$

according to the definition of neighbor sets, which implies (ii) immediately. □

Define

$$\Upsilon_k = \Big|\big\{x_1(k),\dots,x_n(k)\big\}\Big|$$

as the number of distinct node states at time k, where $\big|S\big|$ for a set S represents its cardinality. Then Lemma 3.1 implies that $\Upsilon_{k+1} \le \Upsilon_k$ for all $k \ge 0$. This point plays an important role in the convergence analysis.

Moreover, for both the nearest-neighbor graph $\mathcal{G}^{\mu n}_{x(k)}$ and the nearest-value graph $\mathcal{G}^{\mu v}_{x(k)}$, in order to distinguish the node states under different values of neighbors, we denote $x_i^\mu(k)$ as the state of node i when the number of larger or smaller neighbors is μ. Correspondingly, we denote

$$h^\mu(k) = \min_{i\in\mathcal{V}} x_i^\mu(k), \quad H^\mu(k) = \max_{i\in\mathcal{V}} x_i^\mu(k).$$

and $\Phi^\mu(k) = H^\mu(k) - h^\mu(k)$. We give another lemma indicating that the convergence speed increases as the number of neighbors increases, which is quite intuitive because apparently graph connectivity increases as the number of neighbors increases.

Lemma 3.2: Consider either the nearest-neighbor graph $\mathcal{G}^{\mu n}_{x(k)}$ or the nearest-value graph $\mathcal{G}^{\mu v}_{x(k)}$. Given two integers $1 \le \mu_1 \le \mu_2$. For every algorithm in $\mathcal{A}$ and every initial value, we have $\Phi^{\mu_1}(k) \ge \Phi^{\mu_2}(k)$ for all k.

Proof. Fix the initial condition at time k_0. Let $m \in \mathcal{V}$ be a node satisfying $x_m^{\mu_1}(k_0) = h^{\mu_1}(k_0)$ and $x_m^{\mu_2}(k_0) = h^{\mu_2}(k_0)$. The order preservation property given by Lemma 3.1 guarantees that $x_m^{\mu_1}(k) = h^{\mu_1}(k)$ and $x_m^{\mu_2}(k) = h^{\mu_2}(k)$ for all $k \ge k_0$. It is straightforward to see that $x_m^{\mu_1}(k_0+1) \le x_m^{\mu_2}(k_0+1)$ if $\mu_1 \le \mu_2$, and continuing we know that $x_m^{\mu_1}(k_0+s) \le x_m^{\mu_2}(k_0+s)$ for all $s \ge 2$. Thus, we have $h^{\mu_1}(k) \le h^{\mu_2}(k)$ for all $k \ge k_0$. A symmetric analysis leads to $H^{\mu_1}(k) \ge H^{\mu_2}(k)$ for all k and the desired conclusion thus follows. □

B. Convergence for Nearest-neighbor Graph

For algorithms in the set $\mathcal{A}^*_{\text{ave}}$, we present the following result under nearest-neighbor graph.

Theorem 3.1: Consider the nearest-neighbor graph $\mathcal{G}^{\mu n}_{x(k)}$.

(i) When $n \le \mu + 1$, each algorithm in $\mathcal{A}^*_{\text{ave}}$ achieves global finite-time consensus;

(ii) When $n > \mu + 1$, each algorithm in $\mathcal{A}^*_{\text{ave}}$ fails to achieve finite-time consensus for almost all initial values;

(iii) When $n > \mu + 1$, each algorithm in $\mathcal{A}^*_{\text{ave}}$ achieves global asymptotic consensus if $\{\alpha_k\}$ is monotone.

Proof. (i) When $n \le \mu + 1$, the communication graph is the complete graph. Thus, consensus will be achieved in one step following (4) for every algorithm in $\mathcal{A}^*_{\text{ave}}$.

(ii) Let $n > \mu + 1$. We define two index set

$$\mathcal{I}_k^- = \big\{i :\ x_i(k) = h(k) = \min_{i\in\mathcal{V}} x_i(k)\big\},$$

and

$$\mathcal{I}_k^+ = \big\{i :\ x_i(k) = H(k) = \max_{i\in\mathcal{V}} x_i(k)\big\}.$$

Claim. Suppose both $\mathcal{I}_k^-$ and $\mathcal{I}_k^+$ contain one node only. Then so do $\mathcal{I}_{k+1}^-$ and $\mathcal{I}_{k+1}^+$.

Let u and v be the unique element in $\mathcal{I}_k^-$ and $\mathcal{I}_k^+$, respectively. Take $m \in \mathcal{V}\backslash\{u\}$. Noting the fact that $x_m(k) > x_u(k)$ and $\mu \le n-2$, we have

$$\min_{j\in\mathcal{N}_u(k)} x_j(k) \le \min_{j\in\mathcal{N}_m(k)} x_j(k)$$

and

$$\max_{j\in\mathcal{N}_u(k)} x_j(k) < \max_{j\in\mathcal{N}_m(k)} x_j(k).$$

This leads to $x_m(k+1) > x_u(k+1)$. Therefore, u is still the unique element in $\mathcal{I}_{k+1}^-$. Similarly we can prove that v is still the unique element in $\mathcal{I}_{k+1}^+$. The claim holds.

Now observe that

$$\Delta \doteq \bigcup_{u\ne v} \big\{x = (x_1\dots x_n)^T :\ x_u < \min_{m\in\mathcal{V}\backslash\{u\}} x_m \text{ and } x_v > \max_{m\in\mathcal{V}\backslash\{v\}} x_m,\big\}$$

has measure zero with respect to the standard Lebesgue measure on $\mathbb{R}^n$. For any initial value not in Δ, we have both $\mathcal{I}_k^-$ and $\mathcal{I}_k^+$ contain one unique element, and thus finite-time consensus is impossible. The desired conclusion follows.

(iii) Recall that

$$\Upsilon_k = \Big|\big\{x_1(k),\dots,x_n(k)\big\}\Big|.$$

Since $\Upsilon_{k+1} \le \Upsilon_k$ holds for all k according to Lemma 3.1, there exists two integers $0 \le m \le n$ and $T \ge 0$ such that

$$\Upsilon_k = m, \tag{6}$$

for all $k \ge T$. Thus, we can sort the possible node states for all $k \ge T$ as

$$y_1(k) < y_2(k) < \cdots < y_m(k).$$

Apparently $m \ne 1, 2$ since otherwise the graph is complete for time ℓ with $\Upsilon_\ell = 1, 2$ and consensus is reached after one step. We assume $m \ge 3$ in the following discussions.

Algorithm (5) can be equivalently transformed to the dynamics on $\{y_1(k),\dots,y_m(k)\}$. Moreover, based on Lemma 3.2, we only need to prove asymptotic consensus for the case $\mu = 1$.

Let $\mu = 1$ and $k \ge T$. For algorithms in $\mathcal{A}^*_{\text{ave}}$, the dynamics of $\{y_1(k),\dots,y_m(k)\}$ can be written:

$$\begin{cases} y_1(k+1) = \alpha_k y_1(k) + (1-\alpha_k) y_2(k); \\ y_2(k+1) = \alpha_k y_1(k) + (1-\alpha_k) y_3(k); \\ \qquad\vdots \\ y_{m-1}(k+1) = \alpha_k y_{m-2}(k) + (1-\alpha_k) y_m(k); \\ y_m(k+1) = \alpha_k y_{m-1}(k) + (1-\alpha_k) y_m(k). \end{cases} \tag{7}$$

Now let $\{\alpha_k\}$ be monotone, say, non-decreasing. Then we have $\alpha_k \geq \alpha_T > 0$. Therefore, for all $k \geq T$, we have

$$\begin{aligned} y_1(k+1) &= \alpha_k y_1(k) + (1-\alpha_k) y_2(k) \\ &\leq \alpha_T y_1(k) + (1-\alpha_T) y_m(k), \end{aligned} \tag{8}$$

and continuing we know that

$$y_1(k+s) \leq \alpha_T^s y_1(k) + (1-\alpha_T^s) y_m(k), s \geq 1. \tag{9}$$

Similarly for $y_2(k)$, we have

$$\begin{aligned} y_2(k+2) &= \alpha_{k+1} y_1(k+1) + (1-\alpha_{k+1}) y_3(k+1) \\ &\leq \alpha_T^2 y_1(k) + (1-\alpha_T^2) y_m(k) \end{aligned} \tag{10}$$

and

$$y_2(k+s) \leq \alpha_T^s y_1(k) + (1-\alpha_T^s) y_m(k),\ s \geq 2. \tag{11}$$

Proceeding the analysis, eventually we arrive at

$$y_i(k+n-1) \leq \alpha_T^{n-1} y_1(k) + (1-\alpha_T^{n-1}) y_m(k), \tag{12}$$

for all $i = 1, \ldots, n$, which yields

$$\Phi(k+n-1) leq (1-\alpha_T^{n-1}) \Phi(k). \tag{13}$$

Thus, global asymptotic consensus is achieved. The other case with $\{\alpha_k\}$ being non-increasing can be proved using a symmetric argument. The desired conclusion follows.

This completes the proof of the theorem. □

Remark 3.1: In Theorem 3.1, the asymptotic consensus statement relies on the condition that $\{\alpha_k\}$ is monotone. From the proof of Theorem 3.1 we see that this condition can be replaced by that there exists a constant $\varepsilon \in (0,1)$ such that either $\alpha_k \geq \varepsilon$ or $\alpha_k \leq 1-\varepsilon$ for all k. In fact, we conjecture that the asymptotic consensus statement of Theorem 3.1 holds true for all $\{\alpha_k\}$, i.e., we believe that asymptotic consensus is achieved for all algorithms in $\mathcal{A}_{\text{ave}}^*$ under nearest-neighbor graphs.

Remark 3.2: Theorem 3.1 indicates that $\mu+1$ is a critical number of nodes for nearest-neighbor graphs: for algorithms in $\mathcal{A}_{\text{ave}}^*$, finite-time consensus holds globally if $n \leq \mu+1$, and fails almost globally if $n > \mu+1$. Note that $n \leq \mu+1$ implies that the communication graph is the complete graph, which is rare in general. Recalling that in Part I of the paper, it was showed that finite-time consensus fails almost globally for algorithms in $\mathcal{A}_{\text{ave}}$ [41], we conclude that finite-time consensus is generally rare for averaging algorithms in $\mathcal{A}$, no matter with ($\mathcal{A}_{\text{ave}}$) or without ($\mathcal{A}_{\text{ave}}^*$) the self-confidence assumption.

For algorithms in $\mathcal{A}_{\max}$, we present the following result.

Theorem 3.2: Consider the nearest-neighbor graph $\mathcal{G}_{x(k)}^{\mu n}$. Algorithms in $\mathcal{A}_{\max}$ achieve global finite-time consensus in no more than $\lceil \frac{n}{\mu} \rceil$ steps, where $\lceil z \rceil$ represents the smallest integer no smaller than z.

Proof. Without loss of generality, we assume that $x_1(0), \ldots, x_n(0)$ are mutually different. We sort the initial values of the nodes as $x_{i_1}(0) < x_{i_2}(0) < \cdots < x_{i_n}(0)$. Here i_m denotes node with the m'th largest value initially.

Observing that i_n is a right-hand side neighbor of nodes $i_{n-\mu}, i_{n-\mu+1}, \ldots, i_{n-1}$, we have

$$x_{i_\tau}(1) = x_{i_n}(0),\ \tau = n-\mu, \ldots, n.$$

This leads to $\Upsilon_1 = \Upsilon_0 - \mu$. Proceeding the same analysis we know that consensus is achieved in no more than $\lceil \frac{n}{\mu} \rceil$ steps. The desired conclusion follows. □

C. Convergence for Nearest-value Graph

In this subsection, we study the convergence for nearest-value graphs. Since nearest-value graph $\mathcal{G}_{x(k)}^{\mu v}$ indeed increases the connectivity of $\mathcal{G}_{x(k)}^{\mu n}$, the asymptotic consensus statement of Theorem 3.1 also holds for $\mathcal{G}_{x(k)}^{\mu v}$. The main result for finite-time consensus of nearest-value graphs is presented as follows. It turns out that the critical number of nodes for nearest-value graphs is 2μ.

Theorem 3.3: Consider the nearest-value graph $\mathcal{G}_{x(k)}^{\mu v}$.

(i) When $n \leq 2\mu$, algorithms in $\mathcal{A}_{\text{ave}}^*$ achieve global finite-time consensus in no more than $\lceil \log_2(2\mu+1) \rceil$ steps;

(ii) When $n > 2\mu$, algorithms in $\mathcal{A}_{\text{ave}}^*$ fail to achieve finite-time consensus for almost all initial conditions.

Proof. (i) Suppose $n \leq 2\mu$. Based on Lemma 3.2, without loss of generality, we assume $n = 2\mu$ and the initial values of the nodes are mutually different. Now we have $\Upsilon_0 = \big|\{x_1(0), \ldots, x_n(0)\}\big| = 2\mu$. We fist show the following claim.

Claim. If $\Upsilon_k = 2\mu - A$ with $A \geq 0$ an integer, then $\Upsilon_{k+1} \leq \Upsilon_k - A - 1$.

We order the node states at time k and denote them as

$$Y_1 < Y_2 < \cdots < Y_{\Upsilon_k}.$$

When $\Upsilon_k = 2\mu - A$, it is not hard to find that the for all $m = \mu - A, \ldots, \mu+1$, each node with value Y_{Υ_m} will connect to some node with value Y_1, and some other node with value Y_{Υ_k}. Therefore, the nodes with value $Y_{\Upsilon_m}, m = \mu - A, \ldots, \mu+1$ will reach the same state after the k'th update. The claim holds.

Therefore, by induction we have $\Upsilon_k = \max\{0, \Upsilon_0 - \sum_{m=0}^{k-1} 2^m\} = \max\{0, 2\mu - (2^k - 1)\}$. The conclusion (i) follows straightforwardly.

(ii) Suppose $n > 2\mu$. Let $x_1(0), \ldots, x_n(0)$ be mutually different. Then it is not hard to see that for any two nodes u and v with $x_u(0) < x_v(0)$, at least one of $\min_{j \in \mathcal{N}_u(0)} x_j(0) < \min_{j \in \mathcal{N}_v(0)} x_j(0)$ or $\max_{j \in \mathcal{N}_u(0)} x_j(0) < \max_{j \in \mathcal{N}_v(0)} x_j(0)$ holds. This immediately leads to $x_u(1) < x_v(1)$. Because u and v are arbitrarily chosen, we can conclude that $\Upsilon_1 = \Upsilon_0$. By an induction argument we see that $\Upsilon_k = \Upsilon_0 = n$ for all $k \geq 0$, or equivalently, consensus cannot be achieved in finite time. Now observe that $\bigcup_{i \geq j} \big\{x = (x_1 \ldots x_n)^T : x_i = x_j\big\}$ has measure zero with respect to the standard Lebesgue measure on $\mathbb{R}^n$. The desired conclusion thus follows. □

IV. Conclusions

This paper focused on a uniform model for distributed averaging and maximizing. Each node iteratively updated its state as a weighted average of its own state, the minimal

state, and maximal state among its neighbors. This part of the paper studied state-dependent graphs defined by a μ-nearest-neighbor rule, where each node interacts with its μ nearest smaller neighbors and the μ nearest larger neighbors, we showed that $\mu + 1$ is a critical number of nodes when consensus transits from finite time to asymptotic convergence in the absence of node self-confidence: finite-time consensus disappears suddenly when the number of nodes is larger than $\mu + 1$. This critical number of nodes turned out to be 2μ if each node chooses to connect to nodes with different values. The results revealed the fundamental connection and difference between distributed averaging and maximizing, but more challenges still lie in the principles underlying the two types of algorithms, such as their convergence rates.

REFERENCES

[1] C. Godsil and G. Royle. *Algebraic Graph Theory.* New York: Springer-Verlag, 2001.

[2] C. Berge and A. Ghouila-Houri. *Programming, Games, and Transportation Networks*. John Wiley and Sons, New York, 1965.

[3] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proc. Amer. Math. Soc.*, vol. 15, pp. 733-736, 1963.

[4] J. Hajnal, "Weak ergodicity in non-homogeneous Markov chains," *Proc. Cambridge Philos. Soc.*, no. 54, pp. 233-246, 1958.

[5] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118-121, 1974.

[6] P. M. DeMarzo, D. Vayanos, J. Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, vol. 118, no. 3, pp. 909-968, 2003.

[7] B. Golub and M. O. Jackson, "Naïve learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, vol. 2, no. 1, pp. 112-149, 2007.

[8] S. Muthukrishnan, B. Ghosh, and M. Schultz, "First and second order diffusive methods for rapid, coarse, distributed load balancing," *Theory of Computing Systems*, vol. 31, pp. 331-354, 1998.

[9] R. Diekmann, A. Frommer, and B. Monien, "Efficient schemes for nearest neighbor load balancing," *Parallel Computing*, vol. 25, pp. 789-812, 1999.

[10] N. Malpani, J. L. Welch and N. Vaidya, "Leader election algorithms for mobile ad hoc networks," *Proc. 4th international workshop on Discrete algorithms and methods for mobile computing and communications*, pp. 96-104, 2000.

[11] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, pp. 803-812, 1986.

[12] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules, *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988-1001, 2003.

[13] R. Olfati-Saber and R. Murray, "Consensus problems in the networks of agents with switching topology and time dealys," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520-1533, 2004.

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

[15] W. Ren and R. Beard, "Consensus seeking in multi-agent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655-661, 2005.

[16] L. Moreau, "Stability of multi-agent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, pp. 169-182, 2005.

[17] G. Shi and Y. Hong, "Global target aggregation and state agreement of nonlinear multi-agent systems with switching topologies," *Automatica*, vol. 45, no. 5, pp. 1165-1175, 2009.

[18] G. Shi, Y. Hong and K. H. Johansson, "Connectivity and set tracking of multi-agent systems guided by multiple moving leaders," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 663-676, 2012.

[19] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and qantization effects," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2506-2517, 2009.

[20] T. Vicsek, A. Czirok, E. B. Jacob, I. Cohen, and O. Schochet. Novel type of phase transitions in a system of self-driven particles. *Physical Review Letters*, vol. 75, 1226-1229, 1995.

[21] U. Krause, "Soziale dynamiken mit vielen interakteuren eine problemskizze," In *Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren*, pp. 37-51, 1997.

[22] V. Blondel, J. M. Hendrickx and J. Tsitsiklis, "On Krause's consensus formation model with state-dependent connectivity," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2506-2517, 2009

[23] V. Blondel, J. M. Hendrickx, A. Olshevsky and J. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," *IEEE Conf. Decision and Control*, pp. 2996-3000, 2005.

[24] G. Tang and L. Guo, "Convergence of a class of multi-agent systems in probabilistic framework," *Journal of Systems Science and Complexity*, 20(2), pp. 173-197, 2007.

[25] Z. Liu and L. Guo, "Synchronization of multi-agent systems without connectivity assumptions," *Automatica*, vol. 45, no. 12, pp. 2744-2753, 2009.

[26] V. D. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "Continuous-time average-preserving opinion dynamics with opinion-dependent communications," *SIAM Journal on Control and Optimization*, vol. 48, no. 8, pp. 5214-5240, 2010.

[27] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale and V. Zdravkovic, "Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study," *PNAS*, vol. 105, no. 4, pp. 1232-1237, 2008.

[28] F. Xue and P. R. Kumar, "On the θ-coverage and connectivity of large random networks," *IEEE Trans. Infom. Theory*, vol. 52, no. 6, pp. 2289-2299, 2006.

[29] J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Distributed anonymous discrete function computation," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2276-2289, 2011.

[30] J. Cortés, "Finite-time convergent gradient flows with applications to network consensus," *Automatica*, vol. 42, no.11, pp. 1993-2000, 2006.

[31] J. Cortés, "Analysis and design of distributed algorithms for χ-consensus," *IEEE Conference on Decision and Control*, pp. 3363-3368, 2006.

[32] Q. Hui, W. M. Haddad, and S. P. Bhat, "Finite-Time semistability and consensus for nonlinear dynamical networks," *IEEE Trans. Autom. Contr.*, vol.53, pp. 1887-1900, 2008.

[33] L. Wang and F. Xiao, "Finite-time consensus problems for networks of dynamic agents," *IEEE Trans. Autom. Contr*, vol. 55, no.4, pp. 950-955, 2010.

[34] B. M. Nejad, S. A. Attia and J. Raisch, "Max-consensus in a max-plus algebraic setting: The case of fixed communication topologies," in *International Symposium on Information, Communication and Automation Technologies*, pp.1-7, 2009.

[35] B. M. Nejad, S. A. Attia and J. Raisch, "Max-consensus in a max-plus algebraic setting: The case of fixed communication topologies," in *International Symposium on Information, Communication and Automation Technologies*, pp.1-7, 2009.

[36] F. Iutzeler, P. Ciblat, and J. Jakubowicz, "Analysis of max-consensus algorithms in wireless channels," *IEEE Transactions on Signal Processing*, vol. 60, no.11, pp. 6103-6107, 2012.

[37] D. Varagnolo, G. Pillonetto, and L. Schenato, "Distributed statistical estimation of the number of nodes in sensor networks, in *IEEE Conference on Decision and Control*, 2011

[38] R. Carli, F. Fagnani, A. Speranzon and S. Zampieri, "Communication constraints in the average consensus problem," *Automatica*, vol.44, no. 3, pp. 671-684, 2008.

[39] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2508-2530, 2006.

[40] G. Shi, M. Johansson and K. H. Johansson, "When do gossip algorithms converge in finite time?" avaiable at http://arxiv.org/pdf/1206.0992v1.pdf.

[41] G. Shi and K. H. Johansson, "Convergence evolution of distributed averaging and maximizing algorithms: Part I: time-dependent graphs," *America Control Conference*, 2013.

[42] G. Shi and K. H. Johansson, "Finite-time and asymptotic convergence of distributed averaging and maximizing algorithms," submitted for journal publication, available at http://arxiv.org/pdf/1205.1733v2.pdf.