Dynamic Median Consensus Over Random Networks

Shuhua Yu, Yuan Chen, Soummya Kar

Abstract— This paper studies the problem of finding the median of N distinct numbers distributed across networked agents. Each agent updates its estimate for the median from noisy local observations of one of the N numbers and information from neighbors. We consider a undirected random network that is connected on average, and a noisy observation sequence that has finite variance and almost surely decaying bias. We present a *consensus+innovations* algorithm with *clipped innovations*. Under some regularity assumptions on the network and observation model, we show that each agent's local estimate converges to the set of median(s) almost surely at an asymptotic sublinear rate. Numerical experiments demonstrate the effectiveness of the presented algorithm.

I. INTRODUCTION

The past decades have seen increasing interests in decentralized control and coordination of large scale networked systems. A canonical problem in decentralized control is consensus. The objective of the consensus problem is to ensure that networked agents reach agreement on a common decision. This paper focuses on dynamic *median* consensus, where the local observations are dynamic and noise corrupted, and the considered network is random and connected on average.

The problem of consensus over multi-agent networks has a rich literature. In particular, the problem of average (or mean) consensus, i.e., finding the mean of the initial states of network agents has been investigated in [1]–[3]. The convergence of average consensus algorithms have been studied in switching topology and networks with time-delays [2], networks with random link failures [4]; and [3], [5] study topology and weight matrix designs for fast convergence respectively. However, average consensus protocols can be vulnerable to attacks in large scale networks like IoT [6], e.g., a single attack on the initial state of one agent can arbitrarily manipulate the network average.

As a result, consensus on a more robust statistical measure, like median, has been of research interest [7]–[10]. Median consensus also finds applications in multi-robot systems [11]. Specifically, [8], [12] propose a continuous time protocol that finds the median value of networked agents, whereas [9] studies median consensus problem in the presence of matching perturbations to the agents' dynamics. The paper [10] studies dynamic median consensus where agents have locally time-varying signals and the network is open in the sense that agents may come and leave during the protocol execution.

S. Yu and S. Kar are with Department of Electrical and Computer Engineering, Carnegie Mellon University, 15213 Pittsburgh, The PA {shuhuay, soummyak}@andrew.cmu.edu

In this paper, we consider a local observation model in which the network agents obtain noisy measurements of N real numbers (local values or parameters) with decaying bias (expected observation error) and white noise. This observation model arises in scenarios where local measurement perturbations are believed to be decaying sublinearly in expectation while the white noises from measuring devices or environment perturbations are unavoidable. This model also subsumes static or white noise corrupted observations as special cases. The inter-agent communication network we consider is undirected, random, independently and identically distributed (i.i.d.) and only required to be *connected on average*.

Our contributions are as follows. We present an algorithm, under the above observation and network model, that converges to the set of median(s) almost surely and characterize the asymptotic sublinear convergence rate. Moreover, we tackle the observation noise by a carefully designed recursive averaging scheme, and develop a technical lemma of independent interest to show diminishing local averaging errors by carefully balancing the effects of decaying bias and white noise.

To the best of our knowledge, existing works that are most similar to this paper are references [9] and [10]. The authors of [9] study median consensus in static networks. They consider a setup where each agent perfectly observes its local value, and the agents collaborate to compute the median of all local values. The agents' interactions are subject to disturbances that are *deterministically* bounded in magnitude. Reference [10] proposes an algorithm to track the *instantaneous* median of a set of local reference signals over time-varying graphs. Agents perfectly observe their local reference signals, the reference signals have determinstically bounded time-derivatives, and, the graph must be connected at (almost) all times.

In contrast to [9] and [10], in this paper, we consider the scenario where agents are unable to perfectly observe their local values. Each agent instead makes a measurement of its local value, subject to measurement noise, which, unlike [9] and [10], is *not* determinstically bounded in magnitude. Further, unlike [10], we present a median consensus algorithm that does *not* require the network's graph topology to be connected at all times.

The rest of the paper is organized as follows. In section II, we formulate the median consensus problem with assumptions. In section III we present our algorithm and the main result. Section IV is concerned with intermediate lemmas and is concluded with the proof of the main result. In section IV we provide some numerical experiments on networks with

varying connectivity.

Notation: We specify the communication network of agents. Agents may exchange messages over a time-varying network denoted as $G(t) = (V, E(t))$ at discrete time t. The set of agents V is fixed and $|V| = N$, and we use $[N]$ to denote all indices in the sequel. The set of communication links $E(t)$ is time-varying. For each agent n, $\Omega_n(t)$ denotes the set of neighbors at time t. We denote the Laplacian of network $G(t)$ as $L(t) = D(t) - A(t)$ where $D(t)$ is a diagonal matrix whose i-th diagonal entry is the number of neighbors of agent i and $A_{i,j}(t) = 1$ if there exists a link between node i and node j at time t, otherwise $A_{i,j}(t) = 0$. A network is connected if and only if the second largest eigenvalue of its Laplacian matrix is positive [13]. We use $\|\cdot\|_2$ to denote Euclidean norm for vectors. The term $\mathbf{1}_N$ represents an N-dimensional column vector whose elements are all 1.

II. PROBLEM FORMULATION

We consider the problem that a set of N agents aim to find the median(s) of a set of N distinct numbers $\{\theta_1, \theta_2, \ldots, \theta_N\}$ through local time-sequential observations and and neighboring information exchange over a random network. Each agent n can only access noisy observations of θ_n. We consider a dynamic observation model for each agent n,

$$\theta_n(t) = \theta_n + w_n(t) + v_n(t) \tag{1}$$

where $\{w_n(t)\}_{n \in [N], t \geq 0}$ is an i.i.d. noise sequence with zero mean and finite variance, and the $v_n(t)$ is decaying almost surely in that $\mathbb{P}\left(|v_n(t)| \leq v_0(t+1)^{-\delta}\right) = 1$ for some positive constants δ and v_0.

This formulation subsumes several models of local observations. Generally, this model considers observation errors with decaying bias and white noise. The decaying bias may be a result of local computation, or when the local observations are a result of another underlying iterative and convergent computational process, whereas, white noise is due to unavoidable measurement errors from physical devices, environment, etc. Moreover, by taking δ to be arbitrarily large we recover the special model corresponding to unbiased local observations, whereas, by setting $w_n(t)$ to be 0 almost surely the problem is reduced to the simplest static case.

For convenience of notation, let us assume $\theta_1 < \theta_2 < \ldots < \theta_N$, and define $d_{\min}$ as the minimal gap between distinct elements in $\{\theta_n\}_{n=1,\ldots,N}$,

$$d_{\min} = \min\{|\theta_i - \theta_j| : i \neq j, i, j \in [N]\} > 0. \tag{2}$$

We define the set of medians Θ

$$\Theta = \begin{cases} \{\theta_{\frac{N+1}{2}}\}, & N \text{ is odd,} \\ \left[\theta_{\frac{N}{2}}, \theta_{\frac{N}{2}+1}\right], & N \text{ is even,} \end{cases}$$

and the distance function $\text{dist}(x, \Theta) := \min_{\theta \in \Theta} |x - \theta|$.

We now formally state our assumptions on the inter-agent communication network and the observation model.

Assumption 1. *The time-varying inter-agent communication network is given by a random graph sequence $\{G(t)\}$. For all*

t, *$G(t)$ is an undirected graph and the Laplacian sequence $\{L(t)\}$ associated with $\{G(t)\}$ is an i.i.d. sequence whose expectation, denoted as $\bar{L}$, exists and satisfies $\lambda_2(\bar{L}) > 0$.*

Remark. *The assumption clearly subsumes the typical case of a static network $G = (V, E)$ where G is connected. The above assumption is quite general and subsumes phenomena such as random link failures, i.e., in which the links in G have failure probabilities in $[0, 1)$, which models many practical networks such as wireless networks. Again, this assumption contrasts this work with [10] in that this assumption allows models such as gossip based protocols in which no network instance (the stochastic realization) is connected, while [10] requires each network instance to be connected.*

Assumption 2. *The observation noise $w_n(t)$ is i.i.d. distributed over time and independent across agents. For any n and t, $\mathbb{E}(w_n(t)) = 0$ and $\text{Var}(w_n(t)) < \infty$. For any n, the perturbation $v_n(t)$ decays a.s. in that $\mathbb{P}\left(|v_n(t)| \leq v_0(t+1)^{-\delta}\right) = 1$ for some positive constants δ and v_0.*

Let $(\Omega, \mathscr{F})$ be the probability space where random variables $L(t), w_n(t)$ are defined, and let $\{\mathscr{F}(t)\}$ be the corresponding natural filtration, i.e., $\mathscr{F}(t)$ is the sigma algebra $\sigma\left(\{L(t')\}_{t'=0,\ldots,t}, \{w_n(t)\}_{n \in [N]}\right)$. In this paper, unless otherwise stated, all inequalities involving random variables hold almost surely (a.s.).

III. ALGORITHM AND MAIN RESULT

We present the following algorithm to estimate Θ. Let $x_n(t)$ be the estimate of agent n at time t, we update its estimate as follows,

$$x_n(t+1) = x_n(t) - \beta_t \sum_{m \in \Omega_n(t)} (x_n(t) - x_m(t))) \\ - \alpha_t k_n(t)(x_n(t) - \bar{\theta}_n(t)), \tag{3}$$

where $\bar{\theta}_n(t)$ is some recursive weighted average of observations $\{\theta_n(t')\}_{0 \leq t' \leq t}$ given by, for some $0 < c_\mu \leq 1, 0 < \mu < 1$,

$$\bar{\theta}_n(t+1) = (1 - \frac{c_\mu}{(t+1)^\mu})\bar{\theta}_n(t) + \frac{c_\mu}{(t+1)^\mu}\theta_n(t), \tag{4}$$

and the step sizes

$$\alpha_t = \frac{\alpha_0}{(t+1)^{\tau_1}}, \quad \beta_t = \frac{\beta_0}{(t+1)^{\tau_2}}, \tag{5}$$

satisfy $\alpha_0, \beta_0 > 0, 0 < \tau_2 < \tau_1 < 1$. The term $k_n(t)$ in (3) is the clipping operator, defined as,

$$k_n(t) = \begin{cases} 1, & \text{if } |x_n(t) - \bar{\theta}_n(t)| \leq \gamma_t, \\ \gamma_t |x_n(t) - \bar{\theta}_n(t)|^{-1}, & \text{otherwise,} \end{cases} \tag{6}$$

for

$$\gamma_t = \frac{\gamma_0}{(t+1)^{\tau_3}}, \tag{7}$$

with $\gamma_0 > 0$ and

$$\tau_3 < \min\{1 - \tau_1, 0.5\delta_0\}, \tag{8}$$

where $\delta_0 = 1 - \bar{\varepsilon}$ and $\mu = \delta_0$ for any $0 < \bar{\varepsilon} < 1$ if $\delta \geq 1$, otherwise $\delta_0 = \delta$ and $\delta \leq \mu < 1$.

We next refer to the algorithm given by (3)-(8) as DMED for short, a Distributed Median Estimator for Dynamic observations, which is a *consensus+innovations* type estimator [14] equipped with *clipped innovations*.

Remark. *DMED is similar to the SAGE algorithm in [15]. The computation goal in this paper, however, is different from that of [15]. SAGE is an algorithm for resilient distributed estimation, where a network of agents measure the same underlying parameter θ and attempt to recover its value from these measurements. That is, in [15], the value of θ is the same for each agent. In contrast, this paper addresses distributed dynamic median consensus, where the agents have different local values of θ_n. As such, the analysis of the convergence of the DMED algorithm requires new techniques not found in [15], which we will present in Section IV.*

Remark. *The DMED algorithm essentially mimics the behavior of decentralized subgradient descent with errors for an l_1 minimization problem $\min_\theta \sum_{n=1}^N |\theta_n - \theta|$ whose optimal solution set is Θ.*

We present the main result on the convergence of DMED.

Theorem 1. *Under Assumptions 1-2, the local median estimate of every agent $n \in [N]$ in Algorithm (3) converges to Θ a.s. in that $\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\tau_3} dist(x_n(t), \Theta) = 0\right) = 1$ for all n simultaneously.*

Theorem 1 states that, when N is odd, i.e., when the median is unique, all local estimates simultaneously converge to the same unique median almost surely. When N is even, simultaneously, the distances between each local estimate and the set of medians will converge to 0 almost surely. However, when N is even, local estimates are not guaranteed to converge to a particular point in the set of medians in that the estimates may wander within the set of medians asymptotically. Even when N is even, as Lemma 2 suggests, all local estimates are still guaranteed to reach consensus almost surely. The sample wise convergence rate we obtain is $\mathcal{O}((t+1)^{-\tau_3})$; for instance, if $|v_n(t)| \leq (t+1)^{-1}$, δ_0 can be taken to be 1, and by choosing $\tau_1 = 0.5, \tau_2 = 0.3$, we can set $\tau_3 = 0.4$, to ensure a $\mathcal{O}((t+1)^{-0.4})$ convergence rate.

IV. PROOF OF THEOREM 1

To prove Theorem 1, we first bound the local observation errors in Lemma 1, then bound the consensus errors in Lemma 2. Lemma 3 shows that there exists a local contraction for the distance between network average and the the set of medians. Lemma 4 states that the network average will always enter the local contraction region. Combing these arguments we prove the theorem.

We use the following independent lemma to upper bound the local observation errors.

Lemma 1. *Let $\{z_t\}$ be a $\mathbb{R}$ valued discrete time process*

$$z_{t+1} = (1 - r_1(t))z_t + r_1(t)(r_2(t) + w(t)) \qquad (9)$$

where the sequence $\{r_1(t)\}$ is deterministic with $r_1(t) = a_1(t+1)^{-\mu} \leq 1$ for $a_1 > 0$, sequence $\{r_2(t)\}$ is almost surely bounded $|r_2(t)| \leq a_2(t+1)^{-\delta}$ for $a_2 > 0$, and $\{w(t)\}$ is i.i.d. random noise with $\mathbb{E}(w(t)) = 0$ and $\mathrm{Var}(w(t)) = \sigma^2 < \infty$. For $\delta \geq 1$, define $\delta_0 = 1 - \bar{\varepsilon}$ for any $0 < \bar{\varepsilon} < 1$ and take $\mu = \delta_0$; for $0 < \delta < 1$, define $\delta_0 = \delta$ and take μ such that $\delta \leq \mu < 1$. Then, for any $0 < \varepsilon_0 < \delta_0$, we have $\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\delta_0 - \varepsilon_0} z_t^2 = 0\right) = 1$.

Proof. We consider sample paths where $|r_2(t)| \leq a_2(t+1)^{-\delta}$ holds for all t. We first show that $\mathbb{E}(|z_t|) < \infty$. Taking absolute value on the recursion (9) gives

$$|z_{t+1}| \leq (1 - \frac{a_1}{(t+1)^\mu})|z_t| + \frac{a_1}{(t+1)^\mu}\left[\frac{a_2}{(t+1)^\delta} + |w(t)|\right].$$

By Jensen's inequality we have $\mathbb{E}(|w(t)|) \leq \sqrt{\mathbb{E}(w^2(t))} = \sigma$. Then, there exists some constant $c_1 > a_1\delta$ such that for sufficiently large t,

$$\mathbb{E}(|z_{t+1}|) \leq (1 - \frac{a_1}{(t+1)^\mu})\mathbb{E}(|z_t|) + \frac{c_1}{(t+1)^\mu}.$$

By Lemma 4.1 in [16] we have $\mathbb{E}(|z_t|) < \infty$ for all t. Given the independence condition of $\{w(t)\}$, for sufficiently large t, there exist constants $c_2, c_3 > 0$ such that

$$
\begin{aligned}
\mathbb{E}\left(z_{t+1}^2\right) &\leq \mathbb{E}\left(\left((1 - \frac{a_1}{(t+1)^\mu})z_t + \frac{a_1 a_2}{(t+1)^{\mu+\delta}}\right)^2\right) \\
&\quad + \frac{a_1^2}{(t+1)^{2\mu}}\mathbb{E}\left(w^2(t)\right) \\
&\leq (1 - \frac{a_1}{(t+1)^\mu})^2\mathbb{E}\left(z_t^2\right) + \frac{a_1^2 a_2^2}{(t+1)^{2\mu+2\delta}} \\
&\quad + (1 - \frac{a_1}{(t+1)^\mu})\frac{2a_1 a_2}{(t+1)^{\mu+\delta}}\mathbb{E}(|z_t|) \\
&\quad + \frac{a_1^2\sigma^2}{(t+1)^{2\mu}} \\
&\leq (1 - \frac{c_2}{(t+1)^\mu})\mathbb{E}\left(z_t^2\right) + \frac{c_3}{(t+1)^{\mu+\delta_0}}.
\end{aligned}
\qquad (10)
$$

The last inequality is due to the definition of δ_0 which implies $\mu + \delta_0 = \min\{2\mu, \mu + \delta\}$ and $0 < \mu < 1$. By Lemma 4.2 in [16], relation (10) leads to that for any $0 < \varepsilon_0 < \delta_0$, we have

$$\lim_{t\to\infty}(t+1)^{\delta_0 - \varepsilon_0}\mathbb{E}\left(z_t^2\right) = 0. \qquad (11)$$

Now we fix ε_0. The definition of δ_0 implies $0 < \delta_0 < 1$ and thus $(t+1)^{\delta_0 - \varepsilon_0}$ is concave in t. Hence, we have

$$(t+2)^{\delta_0 - \varepsilon_0} \leq (t+1)^{\delta_0 - \varepsilon_0}[1 + (\delta_0 - \varepsilon_0)(t+1)^{-1}].$$

Combining this fact with relation (10) gives that for sufficiently large t,

$$
\begin{aligned}
&(t+2)^{\delta_0 - \varepsilon_0}\mathbb{E}\left(z_{t+1}^2\right) \\
&\leq \left[1 - \frac{c_2}{(t+1)^\mu} + \frac{\delta_0 - \varepsilon_0}{t+1} - \frac{c_2(\delta_0 - \varepsilon_0)}{(t+1)^{\mu+1}}\right] \\
&\quad \cdot (t+1)^{\delta_0 - \varepsilon_0}\mathbb{E}\left(z_t^2\right) + \frac{c_3}{(t+1)^{\mu+\varepsilon_0}}\left(1 + \frac{\delta_0 - \varepsilon_0}{t+1}\right) \\
&\leq (1 - \frac{c_5}{(t+1)^\mu})(t+1)^{\delta_0 - \varepsilon_0}\mathbb{E}\left(z_t^2\right) + \frac{c_6}{(t+1)^{\mu+\varepsilon_0}}
\end{aligned}
\qquad (12)
$$

5697

for some constants $c_5, c_6 > 0$. Define the process

$$V(t) = (t+1)^{\delta_0 - \varepsilon_0} z_t^2$$
$$- \sum_{i=0}^{t-1} \left[\left(\Pi_{j=i+1}^{t-1} (1 - \frac{c_5}{(j+1)^\mu}) \right) \frac{c_6}{(i+1)^{\mu + \varepsilon_0}} \right].$$

An application of Lemma 25 in [14] leads to

$$\lim_{t \to \infty} \sum_{i=0}^{t-1} \left[\left(\Pi_{j=i+1}^{t-1} (1 - \frac{c_5}{(j+1)^\mu}) \right) \frac{c_6}{(i+1)^{\mu + \varepsilon_0}} \right] = 0. \quad (13)$$

Also note that we can split

$$\sum_{i=0}^{t} \left[\left(\Pi_{j=i+1}^{t} (1 - \frac{c_5}{(j+1)^\mu}) \right) \frac{c_6}{(i+1)^{\mu + \varepsilon_0}} \right]$$
$$= \left[1 - \frac{c_5}{(t+1)^\mu} \right] \sum_{i=0}^{t-1} \left[\left(\Pi_{j=i+1}^{t-1} (1 - \frac{c_5}{(j+1)^\mu}) \right) \frac{c_6}{(i+1)^{\mu + \varepsilon_0}} \right]$$
$$+ \frac{c_6}{(t+1)^{\mu + \varepsilon_0}}.$$

Denote the filtration $\mathscr{F}_z(t)$ the natural filtration of the process $\{(t+1)^{\delta_0 - \varepsilon_0} z_t^2\}$, and note that $V(t)$ is adapted to this filtration. Then, by the independence condition,

$$\mathbb{E}(V(t+1) \mid \mathscr{F}_z(t))$$
$$= \mathbb{E} \left((t+2)^{\delta_0 - \varepsilon_0} z_{t+1}^2 \mid \mathscr{F}_z(t) \right)$$
$$- \sum_{i=0}^{t} \left[\left(\Pi_{j=i+1}^{t} (1 - \frac{c_5}{(j+1)^\mu}) \right) \frac{c_6}{(i+1)^{\mu + \varepsilon_0}} \right]$$
$$\leq [1 - \frac{c_5}{(t+1)^\mu}](t+1)^{\delta_0 - \varepsilon_0} \mathbb{E}(z_t^2) + \frac{c_6}{(t+1)^{\mu + \varepsilon_0}}$$
$$- \sum_{i=0}^{t} \left[\left(\Pi_{j=i+1}^{t} (1 - \frac{c_5}{(j+1)^\mu}) \right) \frac{c_6}{(i+1)^{\mu + \varepsilon_0}} \right]$$
$$= (1 - \frac{c_5}{(t+1)^\mu}) V(t)$$
$$\leq V(t).$$

Therefore, $\{V(t)\}$ is a supermartingale. By (13), $\{V(t)\}$ is bounded below. It follows that there exists a finite random variable V_* such that $\lim_{t \to \infty} V(t) = V_*$ almost surely. Thus, we have $\lim_{t \to \infty} (t+1)^{\delta_0 - \varepsilon_0} z_t^2 = V_*$ almost surely. Then, by Fatou's lemma and (11) we have

$$0 \leq \mathbb{E} \left(\lim_{t \to \infty} (t+1)^{\delta_0 - \varepsilon_0} z_t^2 \right) \leq \liminf_{t \to \infty} (t+1)^{\delta_0 - \varepsilon_0} \mathbb{E}(z_t^2) = 0.$$

Thus, we have $\mathbb{P}\left(\lim_{t \to \infty} (t+1)^{\delta_0 - \varepsilon_0} z_t^2 = 0 \right) = 1$. $\square$

We study the behavior of *all* of the agents' estimates together, so, for convenience, we introduce the following notation. Let $K_t = \text{diag}([k_1(t), \dots, k_N(t)]) \in \mathbb{R}^{N \times N}$, $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^\top \in \mathbb{R}^N$, $\bar{\boldsymbol{\theta}}(t) = [\bar{\theta}_1(t), \dots, \bar{\theta}_N(t)]^\top \in \mathbb{R}^N$, then (3) can be rewritten as

$$\mathbf{x}(t+1) = (I - \beta_t L(t)) \mathbf{x}(t) - \alpha_t K_t (\mathbf{x}(t) - \bar{\boldsymbol{\theta}}(t)). \quad (14)$$

The following result (from [15]) establishes that the agents reach consensus. Define $P_N = N^{-1} \mathbf{1}_N \mathbf{1}_N^\top$.

Lemma 2 (Lemma 1 in [15]). *Under Assumption 1, for every $0 \leq \varepsilon_1 < \tau_1 - \tau_2 + \tau_3$, the iterates $\mathbf{x}(t)$ in (14) satisfies $\mathbb{P}(\lim_{t \to \infty} (t+1)^{\tau_1 - \tau_2 + \tau_3 - \varepsilon_1} \|\mathbf{x}(t) - P_N \mathbf{x}(t)\|_2 = 0) = 1$.*

Let $\bar{x}(t) = N^{-1} \mathbf{1}_N^\top \mathbf{x}(t)$ denote the network average at time t. The following lemma analyze the local contraction of $\bar{x}(t)$.

Lemma 3. *Define the auxiliary threshold*

$$\bar{\gamma}_t = \gamma_t - \frac{c_\delta}{(t+1)^{\delta_1}} - \frac{c_\eta}{(t+1)^\eta}$$

where

$$\delta_1 = 0.5(\delta_0 - \varepsilon_0), \eta = \tau_1 - \tau_2 + \tau_3 - \varepsilon_1$$

for arbitrarily small

$$0 < \varepsilon_0 < \delta_0, 0 < \varepsilon_1 < \tau_1 - \tau_2.$$

Then, almost surely, there exists a finite T_0 and positive constants c_δ, c_η such that if $dist(\bar{x}(T_1), \Theta) \leq \bar{\gamma}_{T_1}$ for some $T_1 \geq T_0$, then $dist(\bar{x}(t), \Theta) \leq \bar{\gamma}_t$ for all $t \geq T_1$.

Proof. Substitute (1) into (4), we obtain, for each $n \in [N]$,

$$\bar{\theta}_n(t) - \theta_n = (1 - \frac{c_\mu}{(t+1)^\mu})(\bar{\theta}_n(t) - \theta_n)$$
$$+ \frac{c_\mu}{(t+1)^\mu}(w_n(t) + v_n(t)).$$

Let $z_t = \bar{\theta}_n(t) - \theta_n$ and by applying Lemma 1 for each $n \in [N]$

$$\mathbb{P} \left(\lim_{t \to \infty} (t+1)^{\delta_0 - \varepsilon_0} |\bar{\theta}_n(t) - \theta_n|^2 = 0 \right) = 1. \quad (15)$$

By Lemma 2 we have for all $n \in [N]$,

$$\mathbb{P} \left(\lim_{t \to \infty} (t+1)^\eta |\bar{x}(t) - x_n(t)| = 0 \right) = 1. \quad (16)$$

We perform the derivations in a sample path $\omega \in \Omega$ such that there exist positive constants $c_{\delta,\omega}, c_{\eta,\omega}$ such that

$$|\bar{\theta}_n(t, \omega) - \theta_n| \leq \frac{c_{\delta,\omega}}{(t+1)^{\delta_1}},$$
$$|\bar{x}(t, \omega) - x_n(t, \omega)| \leq \frac{c_{\eta,\omega}}{(t+1)^\eta} \quad (17)$$

hold for all $n \in [N]$. As a consequence of (15)(16), the set of all such sample paths has probability 1. Define

$$e_n(t, \omega) \triangleq x_n(t, \omega) - \bar{x}(t, \omega) + \theta_n - \bar{\theta}_n(t, \omega). \quad (18)$$

Then,

$$|e_n(t, \omega)| \leq c_{\delta,\omega}(t+1)^{-\delta_1} + c_{\eta,\omega}(t+1)^{-\eta}. \quad (19)$$

Now that we have bounded the consensus errors and the local observation errors, we start to analyze the dynamics of the network average $\bar{x}(t, \omega)$. Multiplying $N^{-1} \mathbf{1}^\top$ on both sides of (14) leads to

$$\bar{x}(t+1, \omega) = \bar{x}(t, \omega) - \frac{\alpha_t}{N} \sum_{n=1}^{N} k_n(t)(x_n(t, \omega) - \bar{\theta}_n(t, \omega)). \quad (20)$$

Step 1, we analyze the net effect of agent n where $\theta_n \notin \Theta$, i.e., $\sum_{\theta_n \notin \Theta} k_n(t)(x_n(t, \omega) - \bar{\theta}_n(t, \omega))$. Since $\tau_3 < \delta_1, \tau_3 < \eta$, we can take t_1 as the least t such that $\bar{\gamma}_{t,\omega} > 0$. Also, recall the definition of $d_{\min}$ in (2), we take t_2 as the least $t \geq t_1$ such that

$d_{\min} > 2\gamma_t$ and $\alpha_t < N$. For $t \geq t_2, \theta_n \notin \Theta$, if $\text{dist}(\bar{x}(t,\omega),\Theta) \leq \bar{\gamma}_{t,\omega}$, we have

$$
\begin{aligned}
|x_n(t,\omega) - \bar{\theta}_n(t,\omega)| \\
\geq |\bar{x}(t,\omega) - \theta_n| - |e_n(t,\omega)| \\
\geq \text{dist}(\theta_n, \Theta) - \text{dist}(\bar{x}(t,\omega),\Theta) - |e_n(t,\omega)| \\
\geq d_{\min} - \gamma_t > \gamma_t.
\end{aligned} \tag{21}
$$

Since $x_n(t,\omega) - \bar{\theta}_n(t,\omega) = e_n(t,\omega) + \bar{x}(t,\omega) - \theta_n$, and by the definition of t_1 we have $|e_n(t,\omega)| < \gamma_t$. Combining this with (21) we have

$$
k_n(t)(x_n(t,\omega) - \bar{\theta}_n(t,\omega)) = \gamma_t \text{sign}(\bar{x}(t,\omega) - \theta_n). \tag{22}
$$

By the hypothesis, $\text{dist}(\bar{x}(t,\omega),\Theta) \leq \bar{\gamma}_t < \gamma_t < 0.5 d_{\min}$,

$$
\begin{aligned}
\theta_{\frac{N-1}{2}} < \bar{x}(t,\omega) < \theta_{\frac{N+3}{2}}, \quad N \text{ is odd}, \\
\theta_{\frac{N}{2}-1} < \bar{x}(t,\omega) < \theta_{\frac{N}{2}+2}, \quad N \text{ is even}.
\end{aligned} \tag{23}
$$

Then, by the definition of median we have zero net effect from $\theta_n \notin \Theta$, i.e.,

$$
\sum_{\theta_n \notin \Theta} k_n(t,\omega)(x_n(t,\omega) - \theta_n) = \sum_{\theta_n \notin \Theta} \gamma_t \text{sign}(\bar{x}(t,\omega) - \theta_n) = 0.
$$

Step 2, we analyze the effects of all agents n with $\theta_n \in \Theta$, i.e., $\sum_{\theta_n \in \Theta} k_n(t)(x_n(t,\omega) - \bar{\theta}_n(t,\omega))$. We define m as the index that $\theta_m \in \Theta$ and $\text{dist}(\bar{x}(t,\omega),\Theta) = |\bar{x}(t,\omega) - \theta_m|$ if it exists. If N is odd, $m = (N+1)/2$; if N is even and $\bar{x}(t,\omega) \notin \Theta$, $m = N/2$ or $N/2+1$. By the hypothesis and (19),

$$
\begin{aligned}
|x_m(t,\omega) - \bar{\theta}_m(t,\omega)| \leq |\bar{x}(t,\omega) - \theta_m| + e_m(t,\omega) \\
\leq \bar{\gamma}_{t,\omega} + |e_m(t,\omega)| \leq \gamma_t.
\end{aligned}
$$

Thus, $k_m(t,\omega) = 1$. We next discuss all three cases: N is odd; N is even and $\bar{x}(t,\omega) \notin \Theta$; N is even but $\bar{x}(t,\omega) \in \Theta$.

Step 2a, when N is odd, by step I, (20) reduces to

$$
\bar{x}(t+1,\omega) = \bar{x}(t,\omega) - \frac{\alpha_t}{N}\left(x_{\frac{N+1}{2}}(t,\omega) - \bar{\theta}_{\frac{N+1}{2}}(t,\omega)\right).
$$

It follows that

$$
\begin{aligned}
&\text{dist}(\bar{x}(t+1,\omega),\Theta) \\
&= |\bar{x}(t,\omega) - \theta_{\frac{N+1}{2}} - \frac{\alpha_t}{N}[x_{\frac{N+1}{2}}(t,\omega) - \bar{\theta}_m(t,\omega)]| \\
&\leq (1 - \frac{\alpha_t}{N})|\bar{x}(t,\omega) - \theta_{\frac{N+1}{2}}| + \frac{\alpha_t}{N}|e_{\frac{N+1}{2}}(t,\omega)| \\
&\leq (1 - \frac{\alpha_t}{N})\bar{\gamma}_{t,\omega} + \frac{\alpha_t}{N}\frac{c_{\delta,\omega}}{(t+1)^{\delta_1}} + \frac{\alpha_t}{N}\frac{c_{\eta,\omega}}{(t+1)^{\eta}}.
\end{aligned} \tag{24}
$$

We next show $\text{dist}(\bar{x}(t+1),\omega),\Theta) \leq \bar{\gamma}_{t+1,\omega}$. Define

$$
\Delta_t = (t+1)^{\tau_3}\bar{\gamma}_{t,\omega} = \gamma_0 - \frac{c_{\delta,\omega}}{(t+1)^{\delta_1-\tau_3}} - \frac{c_{\eta,\omega}}{(t+1)^{\eta-\tau_3}}.
$$

Since $\tau_3 < \delta_1, \tau_3 < \eta$, Δ_t is increasing with t and thus

$$
\bar{\gamma}_{t+1,\omega} = \frac{\Delta_{t+1}}{(t+2)^{\tau_3}} \geq \frac{\Delta_t}{(t+2)^{\tau_3}} = \left(\frac{t+1}{t+2}\right)^{\tau_3}\bar{\gamma}_{t,\omega}.
$$

To show that $\text{dist}(\bar{x}(t+1),\omega),\Theta) \leq \bar{\gamma}_{t+1,\omega}$, it suffices to show that

$$
(1 - \frac{\alpha_t}{N})\bar{\gamma}_{t,\omega} + \frac{\alpha_t}{N}\frac{c_{\delta,\omega}}{(t+1)^{\delta_1}} + \frac{\alpha_t}{N}\frac{c_{\eta,\omega}}{(t+1)^{\eta}} \leq \left(\frac{t+1}{t+2}\right)^{\tau_3}\bar{\gamma}_{t,\omega},
$$

rearranging the above relation gives that

$$
1 - \frac{\alpha_t}{N}\left(\underbrace{1 - \frac{c_{\delta,\omega}}{\Delta_t(t+1)^{\delta_1-\tau_3}} - \frac{c_{\eta,\omega}}{\Delta_t(t+1)^{\eta-\tau_3}}}_{\mathscr{T}_1}\right) \leq \left(\frac{t+1}{t+2}\right)^{\tau_3}.
$$

As $\mathscr{T}_1$ is increasing in t, we can take t_3 as the least $t \geq t_2$ such that $\mathscr{T}_1 \geq 0$. Since $1 - x \leq e^{-x}$ for $x \geq 0$, it suffices to show

$$
\frac{\alpha_t}{N\tau_3}\left(\frac{c_{\delta,\omega}}{\Delta_t(t+1)^{\delta_1-\tau_3}} + \frac{c_{\eta,\omega}}{\Delta_t(t+1)^{\eta-\tau_3}} - 1\right) \leq \ln\left(\frac{t+1}{t+2}\right).
$$

Take t_4 as the least $t \geq t_3$ such that $\Delta_t \geq \gamma_0/2$. Since $\ln\left(\frac{t+1}{t+2}\right) \geq 1 - \frac{t+2}{t+1} = -\frac{1}{t+1}$, for $t \geq t_4$, it suffices to show

$$
\frac{\alpha_t}{N\tau_3}\left(\frac{2c_{\delta,\omega}}{\gamma_0(t+1)^{\delta_1-\tau_3}} + \frac{2c_{\eta,\omega}}{\gamma_0(t+1)^{\eta-\tau_3}} - 1\right) \leq -\frac{1}{t+1},
$$

which rearranges to

$$
\frac{2c_{\delta,\omega}}{\gamma_0(t+1)^{\delta_1-\tau_3}} + \frac{2c_2}{\gamma_0(t+1)^{\eta-\tau_3}} + \frac{N\tau_3}{\alpha_0(t+1)^{1-\tau_1}} \leq 1.
$$

Since the left hand side is monotonically decreasing to 0, we can take t_5 as the least $t \geq t_4$ such that the above relation holds. Taking $T_0 = t_5$ completes this case.

Step 2b: when N is even and $\bar{x}(t,\omega) \notin \Theta$, without loss of generality, we consider $\bar{x}(t,\omega) < \theta_{N/2}$, then for $t \geq t_2$

$$
\begin{aligned}
&|x_{\frac{N}{2}+1}(t,\omega) - \bar{\theta}_{\frac{N}{2}+1}(t,\omega)| \\
&\geq |\bar{x}(t,\omega) - \theta_{\frac{N}{2}+1}| - |e_{\frac{N}{2}+1}(t,\omega)| \\
&\geq |\theta_{\frac{N}{2}} - \theta_{\frac{N}{2}+1}| - |\bar{x}(t,\omega) - \theta_{\frac{N}{2}}| - |e_{\frac{N+1}{2}}(t,\omega)| \\
&\geq |\theta_{\frac{N}{2}} - \theta_{\frac{N}{2}+1}| - \gamma_t \geq \gamma_t.
\end{aligned}
$$

Hence, $k_{N/2+1}(t,\omega)(x_{N/2+1}(t,\omega) - \bar{\theta}_{N/2+1}(t,\omega)) = -\gamma_t$ by the same argument in (21)(22). Then, (20) reduces to

$$
\bar{x}(t+1,\omega) = \bar{x}(t,\omega) - \frac{\alpha_t}{N}\left(x_{\frac{N}{2}}(t,\omega) - \bar{\theta}_{\frac{N}{2}}(t,\omega) - \gamma_t\right). \tag{25}
$$

It follows that $|\bar{x}(t+1,\omega) - \bar{x}(t,\omega)| \leq 2N^{-1}\alpha_t\gamma_t$. Take t_6 as the least $t \geq t_2$ such that $2N^{-1}\alpha_t\gamma_t < |\theta_{N/2} - \theta_{N/2+1}|$. Then, for $t \geq t_6$, we either have $\bar{x}(t+1,\omega) \in \Theta$ or $\text{dist}(\bar{x}(t+1,\omega),\Theta) = |\bar{x}(t+1,\omega) - \theta_{N/2}| > 0$. In the first case, $\text{dist}(\bar{x}(t+1,\omega),\Theta) = 0$. In the second case, $\bar{x}(t+1,\omega) < \theta_{N/2}$ and

$$
\begin{aligned}
&\text{dist}(\bar{x}(t+1,\omega),\Theta) \\
&\leq |\bar{x}(t,\omega) - \frac{\alpha_t}{N}\left(-\gamma_t + x_{\frac{N}{2}}(t,\omega) - \bar{\theta}_{\frac{N}{2}}(t,\omega)\right) - \theta_{\frac{N}{2}}| \\
&\leq |(1 - \frac{\alpha_t}{N})(\bar{x}(t,\omega) - \theta_{\frac{N}{2}}) + \frac{\alpha_t\gamma_t}{N}| + \frac{\alpha_t}{N}|e_{\frac{N}{2}}(t,\omega)| \\
&\leq \max\{(1 - \frac{\alpha_t}{N})|\bar{x}(t,\omega) - \theta_{\frac{N}{2}}|, \frac{\alpha_t\gamma_t}{N}\} + \frac{\alpha_t}{N}|e_{\frac{N}{2}}(t,\omega)|.
\end{aligned} \tag{26}
$$

If $(1 - N^{-1}\alpha_t)|\bar{x}(t,\omega) - \theta_{N/2}| \geq N^{-1}\alpha_t\gamma_t$, the above relation falls into the same pursuit of (24) and thus there exists $t_7 \geq t_6$ such that $\text{dist}(\bar{x}(t+1,\omega),\Theta) \leq \bar{\gamma}_{t+1,\omega}$ for $t \geq t_7$. Otherwise, to show $\text{dist}(\bar{x}(t+1,\omega),\Theta) \leq \bar{\gamma}_{t+1,\omega}$ it suffices to show

$$
\frac{\alpha_t\gamma_t}{N} + \frac{\alpha_t}{N}|e_{\frac{N}{2}}(t,\omega)| \leq \bar{\gamma}_{t+1,\omega}.
$$

Substitute (19) into the display above, it suffices to show

$$\frac{\alpha_0 \gamma_0}{(t+1)^{\tau_1}} + \frac{\alpha_0 c_{\eta,\omega}}{(t+1)^{\eta+\tau_1-\tau_3}} + \frac{\alpha_0 c_{\delta,\omega}}{(t+1)^{\delta_1+\tau_1-\tau_3}} + \frac{N c_{\eta,\omega}}{(t+2)^{\eta-\tau_3}}$$
$$+ \frac{N c_{\delta,\omega}}{(t+2)^{\delta_1-\tau_3}} \leq N\gamma_0,$$

The left hand side is decreasing to 0 as $\tau_3 < \delta_1, \tau_3 < \eta$, so we can take t_8 as the least $t \geq t_7$ such that the above relation holds. Taking T_0 as t_8 addresses this case.

Step 2c, when N is even and $\bar{x}(t,\omega) \in \Theta$, we show that $\bar{x}(t+1,\omega) \in \Theta$. In this case, (20) reduces to

$$\begin{aligned}
&\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \\
&= -\frac{\alpha_t}{N} k_{\frac{N}{2}}(t) \underbrace{(x_{\frac{N}{2}}(t,\omega) - \bar{\theta}_{\frac{N}{2}}(t,\omega))}_{\mathscr{T}_2} \\
&\quad - \frac{\alpha_t}{N} k_{\frac{N}{2}+1}(t) \underbrace{(x_{\frac{N}{2}+1}(t,\omega) - \bar{\theta}_{\frac{N}{2}+1}(t,\omega))}_{\mathscr{T}_3}.
\end{aligned} \tag{27}$$

We discuss two possibilities, one is that $|\mathscr{T}_2| \leq \gamma_t$ or $|\mathscr{T}_3| \leq \gamma_t$, the other is both $|\mathscr{T}_2| > \gamma_t$ and $|\mathscr{T}_2| > \gamma_t$. First, consider $|\mathscr{T}_2| \leq \gamma_t$ ($|\mathscr{T}_3| \leq \gamma_t$ is a symmetric case). Since

$$\begin{aligned}
|\bar{x}(t,\omega) - \theta_{\frac{N}{2}}| &\leq |\mathscr{T}_2| + |e_{\frac{N}{2}+1}(t,\omega)| \\
&\leq \gamma_t + \frac{c_{\delta,\omega}}{(t+1)^{\delta_1}} + \frac{c_{\eta,\omega}}{(t+1)^{\eta}},
\end{aligned}$$

and $\bar{x}(t,\omega) \in \Theta$, we obtain

$$\begin{aligned}
-\mathscr{T}_3 &= \theta_{\frac{N}{2}+1} - \bar{x}(t,\omega) - e_{\frac{N}{2}+1}(t,\omega) \\
&\geq \theta_{\frac{N}{2}+1} - \theta_{\frac{N}{2}} - |\bar{x}(t,\omega) - \theta_{\frac{N}{2}}| - |e_{\frac{N}{2}+1}(t,\omega)| \\
&\geq d_{\min} - \gamma_t - \frac{2c_{\delta,\omega}}{(t+1)^{\delta_1}} - \frac{2c_{\eta,\omega}}{(t+1)^{\eta}}.
\end{aligned}$$

Take t_9 as the least $t \geq t_8$ so that the right hand side on the last line of the above relation is larger than γ_t. Then, substituting the values of $k_{N/2}(t)$ and $k_{N/2+1}(t)$ into (27) we have $\bar{x}(t+1,\omega) \geq \bar{x}(t,\omega)$. Take t_{10} as the least $t \geq t_9$ such that $2N^{-1}\alpha_t \gamma_t \leq d_{\min} - (\gamma_t + c_{\delta,\omega}(t+1)^{-\delta_1} + c_{\eta,\omega}(t+1)^{-\eta})$, then by (28), for $t \geq t_{10}$ we have

$$\begin{aligned}
&\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \\
&\leq \frac{2\alpha_t \gamma_t}{N} \leq \theta_{\frac{N}{2}+1} - \theta_{\frac{N}{2}} - |\bar{x}(t,\omega) - \theta_{\frac{N}{2}}| \\
&\leq \theta_{\frac{N}{2}+1} - \bar{x}(t,\omega),
\end{aligned} \tag{28}$$

so we have $\bar{x}(t,\omega) \leq \bar{x}(t+1,\omega) \leq \theta_{N/2+1}$, and thus $\bar{x}(t+1,\omega) \in \Theta$. Second, if both $|\mathscr{T}_2| > \gamma_t$ and $|\mathscr{T}_3| > \gamma_t$, where it is easy to check that under $t \geq t_{10}$, $\mathscr{T}_2$ and $\mathscr{T}_3$ are of opposite signs and thus $\bar{x}(t+1,\omega) = \bar{x}(t,\omega) \in \Theta$. Suppose both $\mathscr{T}_2, \mathscr{T}_3 < -\gamma_t$, then

$$\begin{aligned}
\bar{x}(t,\omega) - x_{\frac{N}{2}} &= \mathscr{T}_2 - e_{\frac{N}{2}}(t,\omega) \\
&< -\gamma_t + \frac{c_{\delta,\omega}}{(t+1)^{\delta_1}} + \frac{c_{\eta,\omega}}{(t+1)^{\eta}} \\
&= -\bar{\gamma}_t < 0,
\end{aligned}$$

which contradicts with $\bar{x}(t,\omega) \in \Theta$. Similar argument applies for $\mathscr{T}_2, \mathscr{T}_3 > \gamma_t$. Thus, taking $T_0 = t_{10}$ addresses this case.

Step 3, in all there cases, in the sample path ω such that (19) holds, there exist some finite T_0 such that if for some $T_1 \geq T_0$ we have $\text{dist}(\bar{x}(T_1),\Theta) \leq \bar{\gamma}_{T_1}$ for some $T_1 \geq T_0$, then $\text{dist}(\bar{x}(T_1+1),\Theta) \leq \bar{\gamma}_{T_1+1}$. Since such sample paths ω has probability 1, T_0 exists a.s. $\qquad\square$

The following lemma asserts that hypothesis of Lemma 4 will always be true, and thus establishes the global convergence of Algorithm (3).

Lemma 4. *For $T_0, \bar{\gamma}_t$ established in Lemma 3, there exists a finite $T_1 \geq T_0$ such that $\text{dist}(\bar{x}(T_1),\Theta) \leq \bar{\gamma}_{T_1}$.*

Proof. We prove this lemma by contradiction. We work on sample paths ω as in Lemma 3. Suppose, on the contrary, for all $t \geq T_{0,\omega}$, $\text{dist}(\bar{x}(t,\omega),\Theta) > \bar{\gamma}_{t,\omega}$. We now show that this leads to $\lim_{t\to\infty} \text{dist}(\bar{x}(t,\omega),\Theta) = -\infty$, a contradiction.

Step 1, we first discuss the value of each *clipped innovation* $k_n(t)[x(t,\omega) - \bar{\theta}_n(t,\omega)]$. We show that there exists some finite $T_2 \geq T_{0,\omega}$ such that $\forall t \geq T_2$, there exists at most one $n \in [N]$ such that $|x_n(t,\omega) - \bar{\theta}_n(t,\omega)| \leq \gamma_t$, which implies there exists at most one $n \in [N]$ such that $k_n(t) = 1$. Take T_2 as the smallest $t \geq T_0$ such that $2\gamma_t < d_{\min} - 2c_\delta(t+1)^{-\delta_1} - 2c_{\eta,\omega}(t+1)^{-\eta}$. Suppose there exist two different $m,n \in [N]$ such that for both $i = m,n, |x_i(t,\omega) - \bar{\theta}_i(t,\omega)| \leq \gamma_t$. Then, for $i = m,n$,

$$\gamma_t \geq |x_i(t,\omega) - \bar{\theta}_i(t,\omega)| \geq |\bar{x}(t,\omega) - \theta_i| - |e_i(t,\omega)|.$$

Summing above relations for $i = m,n$, combining with (19) we have $2\gamma_t \geq |\theta_m - \theta_n| - 2c_\delta(t+1)^{-\delta_1} - 2c_{\eta,\omega}(t+1)^{-\eta}$, which contradicts with the choice of T_2. If it exists, for the agent p that satisfies $|x_p(t,\omega) - \bar{\theta}_p(t,\omega)| \leq \gamma_t$ and $\theta_p \in \Theta$, take T_3 as the least $t \geq T_2$ such that $\bar{\gamma}_{t,\omega} \geq c_{\delta,\omega}(t+1)^{-\delta} + c_{\eta,\omega}(t+1)^{-\eta}$. Thus, by the contradiction hypothesis,

$$|e_p(t,\omega)| \leq \bar{\gamma}_{t,\omega} < \text{dist}(\bar{x}(t,\omega),\Theta) = |\bar{x}(t,\omega) - \theta_p|. \tag{29}$$

Decomposing $x_p(t,\omega) - \bar{\theta}_p(t,\omega) = \bar{x}(t,\omega) - \theta_p + e_p(t,\omega)$ and combing with (29) we have

$$\text{sign}(x_p(t,\omega) - \bar{\theta}_p(t,\omega)) = \text{sign}(\bar{x}(t,\omega) - \theta_p). \tag{30}$$

For any agent m such that $|x_m(t,\omega) - \bar{\theta}_m(t,\omega)| > \gamma_t$, with the same reasoning in (21)(22) we have

$$k_m(x_m(t,\omega) - \bar{\theta}_m(t,\omega)) = \gamma_t \text{sign}(\bar{x}(t,\omega) - \theta_m). \tag{31}$$

We next consider $t \geq T_3$ and assume $\bar{x}(t,\omega)$ is smaller than median(s) without loss of generality. We discuss all three possible cases.

Step 2a, if there exists p such that $|x_p(t,\omega) - \bar{\theta}_p(t,\omega)| \leq \gamma_t$ and $\theta_p \in \Theta$. Then,

$$\begin{aligned}
&\bar{x}(t,\omega) - \theta_{p-1} \\
&= x_p(t,\omega) - \bar{\theta}_p(t,\omega) + e_p(t,\omega) + \theta_p - \theta_{p-1} > 0,
\end{aligned}$$

since by the definition of T_2,

$$\begin{aligned}
\theta_p - \theta_{p-1} &\geq d_{\min} > c_{\delta,\omega}(t+1)^{-\delta_1} + c_{\eta,\omega}(t+1)^{-\eta} + \gamma_t \\
&\geq |e_p(t,\omega)| + |x_p(t,\omega) - \bar{\theta}_p(t,\omega)|.
\end{aligned}$$

Thus, we have $\theta_{p-1} < \bar{x}(t,\omega) < \theta_p$. By (31), and the definition of median

$$\sum_{n \neq p} k_n(t,\omega)(x_n(t,\omega) - \bar{\theta}_n(t,\omega)) = \begin{cases} 0, & N \text{ is odd,} \\ -\gamma_t, & N \text{ is even.} \end{cases} \quad (32)$$

Thus, (20) reduces to

$$\begin{aligned} &\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \\ &= \begin{cases} -\frac{\alpha_t}{N}(x_p(t,\omega) - \bar{\theta}_p(t,\omega) - \gamma_t), & N \text{ is even,} \\ -\frac{\alpha_t}{N}(x_p(t,\omega) - \bar{\theta}_p(t,\omega)), & N \text{ is odd.} \end{cases} \\ &\geq 0. \end{aligned} \quad (33)$$

where the last inequality follows from (30) and $\bar{x}(t,\omega) < \theta_p$. By the contradiction hypothesis,

$$\begin{aligned} &|x_p(t,\omega) - \bar{\theta}_p(t,\omega)| \\ &\geq |\bar{x}(t,\omega) - \theta_p| - |e_p(t,\omega| \\ &> \bar{\gamma}_t - c_{\delta,\omega}(t+1)^{-\delta} - c_{\eta,\omega}(t+1)^{-\eta}, \end{aligned}$$

it follows from (33) that for both even and odd N,

$$\begin{aligned} &\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \\ &\geq \frac{\alpha_t}{N}\left(\gamma_t - \frac{2c_{\delta,\omega}}{(t+1)^{\delta_1}} - \frac{2c_{\eta,\omega}}{(t+1)^\eta}\right). \end{aligned} \quad (34)$$

Step 2b, consider the case that there exists one q such that $|x_q(t,\omega) - \bar{\theta}_q(t,\omega)| \leq \gamma_t$, but $\theta_q \notin \Theta$. Since $\bar{x}(t,\omega)$ is smaller than the median(s), θ_q is also smaller than the median(s), otherwise $|\bar{x}(t,\omega) - \theta_q| \geq d_{\min} > \gamma_t + c_{\delta,\omega}(t+1)^{-\delta} + c_{\eta,\omega}(t+1)^{-\eta}$ by the definition of T_2, and implies the contradiction that

$$|x_q(t,\omega) - \bar{\theta}_q(t,\omega)| \geq |\bar{x}(t,\omega) - \theta_q| - |e_q(t,\omega)| > \gamma_t.$$

Then, from (31)

$$\begin{aligned} &\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \\ &= -\frac{\alpha_t \gamma_t}{N} \sum_{n \neq q} \text{sign}\left[x_n(t,\omega) - \bar{\theta}_n(t,\omega)\right] \\ &\quad - \frac{\alpha_t}{N}(x_q(t,\omega) - \bar{\theta}_q(t,\omega)) \\ &= -\frac{\alpha_t \gamma_t}{N} \sum_{n \neq q} \text{sign}\left[\bar{x}(t,\omega) - \theta_n\right] \\ &\quad - \frac{\alpha_t}{N}(x_q(t,\omega) - \bar{\theta}_q(t,\omega)). \end{aligned} \quad (35)$$

Since $\bar{x}(t,\omega), \theta_q$ are smaller than the median(s), by counting θ_n on both sides of $\bar{x}(t,\omega)$ we have

$$\sum_{n \neq q} \text{sign}\left[\bar{x}(t,\omega) - \theta_n\right] \leq -2.$$

Hence, it follows from (35) that

$$\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \geq \frac{\alpha_t \gamma_t}{N}. \quad (36)$$

Step 2c, if for each agent n, $|x_n(t,\omega) - \bar{\theta}_n(t,\omega)| > \gamma_t$, then from (31) we have

$$\begin{aligned} &\bar{x}(t+1,\omega) - \bar{x}(t,\omega) \\ &= -\frac{\alpha_t \gamma_t}{N} \sum_{n \in [N]} \text{sign}\left[x_n(t,\omega) - \bar{\theta}_n(t,\omega)\right] \\ &= -\frac{\alpha_t \gamma_t}{N} \sum_{n \in [N]} \text{sign}\left[\bar{x}(t,\omega) - \theta_n\right]. \end{aligned} \quad (37)$$

Since $\bar{x}(t,\omega)$ is less than the median(s), by counting number of θ_n on both sides of $\bar{x}(t,\omega)$ we still obtain (36).

Apply similar argument as on (25). There exists finite $T_4 \geq T_3$ such that $\bar{x}(t+1,\omega)$ is still less than the medians as $\bar{x}(t,\omega)$ is less than the medians. Take T_5 as the least $t \geq T_4$ such that $\bar{\gamma}_t - c_{\delta,\omega}(t+1)^{-\delta} - c_{\eta,\omega}(t+1)^{-\eta} \geq c_\gamma \gamma_t$ for some $0 < c_\gamma < 1$. Then in either *Step 2a, 2b, 2c*, by (34)(36) we obtain

$$\text{dist}(\bar{x}(t,\omega),\Theta) - \text{dist}(\bar{x}(t+1,\omega),\Theta) \geq \frac{c_\gamma \alpha_t \gamma_t}{N}.$$

Summing over all $t \geq T_5$ leads to a contradiction in that $\text{dist}(\bar{x}(T_5,\omega)) - \lim_{t\to\infty} \text{dist}(\bar{x}(t),\omega) = c_\gamma N^{-1} \sum_{t=T_6}^{\infty} \alpha_t \gamma_t = \infty$ by the choice $\tau_3 < 1 - \tau_1$, and thus establishes the desired assertion. $\qquad \square$

Proof of Theorem 1. By Lemma 3, 4, there exists some finite T_1 such that for all $t \geq T_1, \text{dist}(\bar{x}(t)) \leq \bar{\gamma}(t)$ a.s. Thus, we have

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\tau_3}\text{dist}(\bar{x}(t),\Theta) = 0\right) = 1.$$

By Lemma 2 we have

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\tau_1-\tau_2+\tau_3-\varepsilon_1}\|\mathbf{x}(t) - P_N\mathbf{x}(t)\|_2 = 0\right) = 1$$

for every $0 < \varepsilon_1 < \tau_1 - \tau_2$. For any $n \in [N]$, by the triangle inequality, we have

$$\begin{aligned} \text{dist}(x_n(t),\Theta) &\leq |x_n(t) - \bar{x}(t)| + \text{dist}(\bar{x}(t),\Theta) \\ &\leq \|\mathbf{x}(t) - P_N\mathbf{x}(t)\|_2 + \text{dist}(\bar{x}(t),\Theta). \end{aligned}$$

Then, it follows that for all n, we have

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\tau_3}\text{dist}(x_n(t),\Theta) = 0\right) = 1.$$

$\qquad \square$

V. NUMERICAL EXPERIMENTS

We generate two random geometric graphs Graph 1 and Graph 2. Both graphs consist of 40 nodes but have different connectivities measured by the second largest eigenvalue of the graph Laplacians. The Laplacians of Graph 1 and Graph 2 are $\lambda_2(L_1) \approx 1.8$, $\lambda_2(L_2) \approx 7.2$ respectively, and from Figs. 1 and 2 below it is clear that Graph 2 is more densely connected. Additionally, both graphs are undirected and connected. We simulate random networks by assigning dropout probability for each link. In our experiments, we use dropout probabilities 0.1 and 0.5.
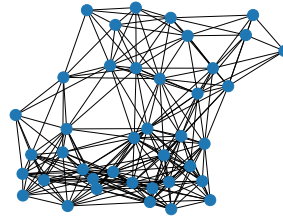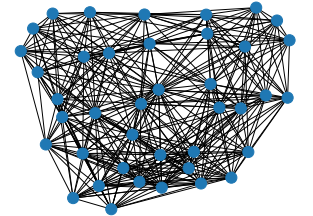

Fig. 1. Graph 1, $\lambda_2(L_1) \approx 1.8$ Fig. 2. Graph 2, $\lambda_2(L_2) \approx 7.2$

We consider the problem setting $\theta_n = n$ for $n = 1,\ldots,40$, $v_n(t) = 10/(t+1)$, and $w_n(t) \sim \mathcal{N}(0,4)$. Note that we

5701

consider perturbations as the sum of a deterministic sequence and i.i.d. white noises. The deterministic sequence is not known to the agents as a prior. Since the largest possible deterministic errors as as tolerable by DMED are used, this problem is the hardest in the problem class we consider.
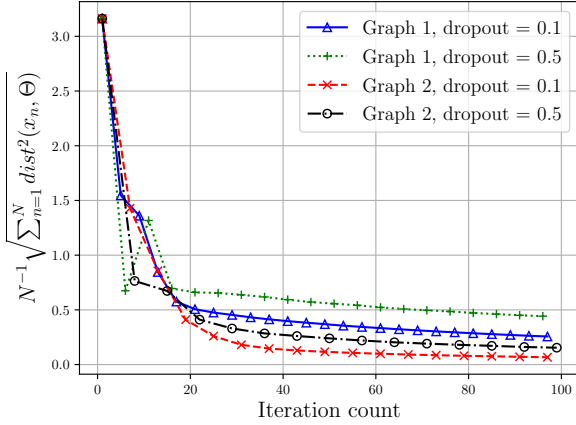

Fig. 3. Iteration count t vs $N^{-1}\sqrt{\sum_{n=1}^{N}\text{dist}^2(x_n,\Theta)}$ with two networks with different dropout probability

By assigning two different dropout probabilities 0.1 and 0.5 for each link in Graph 1 and Graph 2 respectively, we conduct experiments on 4 random networks. In all 4 random networks, we set the same parameters $\alpha_t = (t+1)^{-0.6}, \beta_t = (t+1)^{-0.2}/10, \gamma_t = 20(t+1)^{-0.3}, c_\mu = 10, \mu = 0.9$, and all local estimates start from 0. Given the random nature of the considered networks, we average the network average distance to a set of medians, i.e., $N^{-1}\sqrt{\sum_{n=1}^{N}\text{dist}^2(x_n,\Theta)}$ over 100 network instances for each of 4 experiments, and present the experiments results in Fig. 3.

The simulation results in Fig. 3 demonstrate our theoretical findings that each local estimate converges to the set of medians sublinearly. The results validates the advantage of DMED over previous works that rely on connected networks, in that if a large proportion of links drop out at each time instance the DMED still converges to the set of medians sublinearly. From Fig. 3 we also observe that better connectivity and lower dropout probabilities could benefit the convergence speed of DMED. This empirical finding about convergence rate characteristics is not formally investigated in our analysis, while the intuition behind it is that better connectivity (measured by the second eigenvalue of the Laplacian) and lower dropout probabilities tend to speed up consensus type processes.

VI. CONCLUSION

In this paper, we have studied the problem of dynamic median consensus over random networks that are required to connected only on average. We have considered a multi-agent networked setup in which each agent makes local observations, corrupted by decaying bias and white noise, of a *distinct* value. The agents' objective is to estimate the median of the local values. We presented DMED, a *consensus+innovations* type algorithm with *clipped innovations* to address this problem. Under the DMED algorithm, the agents' local iterates converge almost surely to the set of medians at a sublinear rate. Finally, we validate the performance of our algorithm through numerical simulations.

REFERENCES

[1] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pp. 482–491, IEEE, 2003.

[2] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on automatic control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[3] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[4] S. Kar and J. M. F. Moura, "Distributed average consensus in sensor networks with random link failures," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 2, pp. II–1013, IEEE, 2007.

[5] S. Kar, S. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, 2008.

[6] H. J. LeBlanc, H. Zhang, S. Sundaram, and X. Koutsoukos, "Consensus of multi-agent networks in the presence of adversaries using only local information," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, pp. 1–10, 2012.

[7] W. Ben-Ameur, P. Bianchi, and J. Jakubowicz, "Robust distributed consensus using total variation," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1550–1564, 2015.

[8] M. Franceschelli, A. Giua, and A. Pisano, "Finite-time consensus on the median value with robustness properties," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1652–1667, 2016.

[9] A. Pilloni, A. Pisano, M. Franceschelli, and E. Usai, "Robust distributed consensus on the median value for networks of heterogeneously perturbed agents," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6952–6957, IEEE, 2016.

[10] Z. A. Z. Sanai Dashti, C. Seatzu, and M. Franceschelli, "Dynamic consensus on the median value in open multi-agent systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3691–3697, IEEE, 2019.

[11] G. Vasiljević, T. Petrović, B. Arbanas, and S. Bogdan, "Dynamic median consensus for marine multi-robot systems using acoustic communication," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5299–5306, 2020.

[12] M. Franceschelli, A. Giua, and A. Pisano, "Finite-time consensus on the median value by discontinuous control," in *2014 American Control Conference*, pp. 946–951, IEEE, 2014.

[13] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The laplacian spectrum of graphs," *Graph theory, combinatorics, and applications*, vol. 2, no. 871-898, p. 12, 1991.

[14] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[15] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed parameter estimation with heterogeneous data," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 4918–4933, 2019.

[16] S. Kar, J. M. F. Moura, and H. V. Poor, "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies," *SIAM Journal on Control and Optimization*, vol. 51, no. 3, pp. 2200–2229, 2013.