Convergence Speed of the Consensus Algorithm with Interference and Sparse Long-Range Connectivity

arXiv:1011.5452v1 [cs.IT] 24 Nov 2010

S. Vanka*
Department of Electrical Engineering
218 Cushing Hall
University of Notre Dame
Notre Dame, IN 46556, USA
E-mail:svanka@nd.edu
Phone:(574)-631-8245

M. Haenggi
Department of Electrical Engineering
274 Fitzpatrick Hall
University of Notre Dame
Notre Dame, IN 46556, USA
E-mail:mhaenggi@nd.edu
(574)-631-6103

V. Gupta
Department of Electrical Engineering
270 Fitzpatrick Hall
University of Notre Dame
Notre Dame, IN 46556, USA
E-mail:vgupta2@nd.edu
(574)-631-2294

Abstract

We analyze the effect of interference on the convergence rate of average consensus algorithms, which iteratively compute the measurement average by message passing among nodes. It is usually assumed that these algorithms converge faster with a greater exchange of information (i.e., by increased network connectivity) in every iteration. However, when interference is taken into account, it is no longer clear if the rate of convergence increases with network connectivity. We study this problem for randomly-placed consensus-seeking nodes connected through an interference-limited network. We investigate the following questions: (a) How does the rate of convergence vary with increasing communication range of each node? and (b) How does this result change when each node is allowed to communicate with a few selected far-off nodes? When nodes schedule their transmissions to avoid interference, we show that the convergence speed scales with r^{2-d}, where r is the communication range and d is the number of dimensions. This scaling is the result of two competing effects when increasing r: Increased schedule length for interference-free transmission vs. the speed gain due to improved connectivity. Hence, although one-dimensional networks can converge faster from a greater communication range despite increased interference, the two effects exactly offset one another in two-dimensions. In higher dimensions, increasing the communication range can actually degrade the rate of convergence. Our results thus underline the importance of factoring in the effect of interference in the design of distributed estimation algorithms.

Keywords–Average Consensus, Wireless Networks, Scaling Laws, MAC Protocols.

I. INTRODUCTION

A. Motivation

The advent of wireless sensor and ad hoc networks has motivated the need for distributed information processing algorithms, which allow each node to operate only on local information. A well-studied algorithm that allows distributed averaging is the average consensus algorithm, wherein the global average of a set of initial sensor observations can be computed based on purely local computations at each sensor. Starting from a set of initial measurements, the average consensus algorithm allows a set of nodes to communicate by a (possibly time-varying) topology to iteratively compute the global average of the initial measurements, see e.g., [1]–[9] and references therein. The connectivity properties of the topologies that ensure convergence have been well-studied (e.g., [10], [11]). Of late, the focus has shifted to studying convergence in the face of communication constraints, like quantization [12]–[14], packet drops [15] and noise [16]. A closely associated algorithm is the gossip algorithm [2], [17], [18]. In particular, the recent work [18] proposes and studies a probabilistic version of the broadcast gossip algorithm [17]. The idea is to exploit channel fluctuations to enable opportunistic longer-range message-passing. Since only one node is allowed to transmit at any given time, the question of interference does not arise.

In this paper, unlike prior work, we study the effect of interference, which becomes important in the formation of more general message-passing topologies. We explicitly model the effect of interference on the rate of topology formation—and hence convergence—of the average consensus algorithm. This important effect—which crucially depends on network geometry—has been largely ignored. In wireless networks, depending on the physical proximity of a to d and c to b, the transmission from a to b and c to d may interfere with one another; hence two time slots may be needed to establish edges $\overrightarrow{(a,b)}$ and $\overrightarrow{(c,d)}$. The network thus has two time-scales of interest: that of establishing individual communications among the desired set of nodes and that of the iterations of the distributed algorithms, which occur only when all the desired nodes have successfully communicated. One may thus, view the underlying communication network as constructing the desired message passing graphs from several feasible sub-graphs, each of which satisfies half-duplex, fading and interference constraints. The union of all these sub-graphs is the desired message passing graph.

To illustrate this, consider the formation of a simple linear 6-node network shown in Fig. 1. Suppose the estimation algorithm requires nearest-neighbor communication (shown as bidirectional

edges). However, due to interference constraints, only every third node can transmit. In this case, we see that forming the the desired topology requires at least three time-slots, as shown. In other words, for these interference constraints, this topology's *fastest rate of formation* is three time slots. Clearly, a topology's intrinsic benefit *and* the fastest rate of its formation determine its true utility.
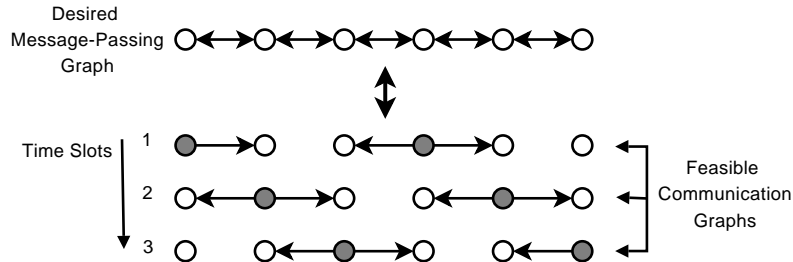

Figure 1. An example illustrating the constraints introduced by interference. If the nodes are physically placed as shown, interference limits the number of nodes that can communicate concurrently. Assuming a spatial re-use factor of two, the message-passing graph can be formed as a union of three feasible sub-graphs, each of them satisfying interfefence constraints. We consider this TDMA schedule *feasible*.

The performance of the underlying (real-time) estimation algorithm is therefore *coupled* with algorithms for channel access and routing. In our previous work [19], we studied the coupling with channel access for the average consensus algorithm for a certain class of deterministic network topologies. Using a simple protocol model [20] for reception, we were able to show that the effect of increasing network connectivity depends crucially on its dimension. In our recent work [21] we exploited the well-known parallels between the convergence of the average consensus algorithm and Markov chain mixing (e.g., [2] and the references therein) to study consensus on disk graphs [22] using the more refined physical model. We examined the scaling behavior of the fastest rate of topology formation with interference, captured by the shortest feasible TDMA schedules that construct the graph.

We note here that implementing inter-node communication in a network will require some additional overhead. For example, one possible protocol that establishes point-to-point communication can have nodes tag their packets with their uniquely assigned address. A receiver reads this address and decodes a packet only if the address is that of one of its intended transmitters. In this work, we neglect this additional overhead. However, we show that even when this overhead is neglected, increased interference alone is enough to significantly lower the rate of topology formation.

B. Main Contributions

In this paper, we study networks with short-range and networks with both short-range and limited long-range communication. Although remarkable improvements in convergence rate have been reported [23]–[25] for consensus on graphs with a few long-range edges (as in small-world graphs [26]), it is not clear if these benefits will carry over to a wireless setting, where long-range links come at a cost of increased interference. Motivated by this fact, we study the average consensus problem in graphs formed by overlaying long-range edges onto an existing "short-range" disk graph. We derive the scaling law for the spectral gap as well as that of the fastest rate of topology formation in the presence of interference. To the best of our knowledge, this is the first such attempt.

We find that the spectral gap scales quadratically in the communication range r, independently of the network dimension d, but the length of the shortest TDMA schedule that constructs such graphs scales as r^d. Thus when interference is factored in, the benefit of a greater communication range depends crucially on the network dimension:

- For one-dimensional networks ($d = 1$), topologies with increased communication range can converge faster despite greater interference.
- For two-dimensional networks, the rate of convergence scales *independently* of the communication range.
- For three- (and higher-) dimensional networks, increasing the communication range can actually slow down convergence.

Furthermore, these results hold whether each node only communicates with all other nodes within its communication range, or, additionally, with a small number of far-away nodes. Thus our results significantly change many optimistic results obtained by analyzing the consensus problem in an abstract graph-theoretic setting.

The remainder of this paper is organized as follows. In Section II, we provide some standard definitions and results used in this paper. In Section III, we specify our system model and formulate the problem using the terminology developed in Section II. In Section IV, we discuss convergence results for the disk graph model. In Section V. we study the effect of selective long-range communication and provide the relevant scaling results. Section VI concludes the paper.

II. DEFINITIONS AND NOTATION

To make this paper self-contained, we formally state the following standard definitions and facts about Markov chains and introduce some notation and other relevant terminology.

1) Basic Definitions from Markov Chain Theory: Consider a connected undirected graph G, with n vertices $V = \{1, 2, \ldots, n\}$ and a set of edges E. We assume G also contains all self-loops, i.e., $i \in V \implies (i, i) \in E$. Let d_i denote the degree of vertex i. For more information, see [27].

Definition 1. (Random walk on a graph) A random walk $\mathcal{X}(G) = (X_k)_{k \in \mathbb{Z}}$ on V is characterized by the $n \times n$ transition probability matrix $\mathbf{P}(G) = [p_{ij}]$, with $p_{ij} \triangleq \mathbb{P}(X_{k+1} = i \mid X_k = j)$, and $p_{ij} > 0$ only if $(i, j) \in E$, with $\sum_j p_{ij} = 1 \ \forall i \in V$.

Observe that P is stochastic.

Definition 2. (Symmetric random walk) A random walk is symmetric if $p_{ij} = p_{ji}$.

For a symmetric random walk $\mathbf{P}$ is doubly stochastic.

Fact 3. *A random walk on G is a Markov chain with state space V. Given an initial distribution $\boldsymbol{\pi}(0)$ over V, the distribution $\boldsymbol{\pi}(k+1)$ after $k+1$ steps satisfies $\boldsymbol{\pi}(k+1) = \mathbf{P}\boldsymbol{\pi}(k)$ for $k = 0, 1, \ldots$*

Definition 4. (Stationary distribution of a Markov chain) A stationary distribution $\boldsymbol{\pi}^*$ satisfies $\boldsymbol{\pi}^* = \mathbf{P}\boldsymbol{\pi}^*$, i.e., remains invariant with time.

Definition 5. (Reversible Markov chain) A Markov chain $\mathcal{X} = (X_k)_{k \in \mathbb{Z}}$ is said to be *reversible* if for all states i, $\pi_j^* p_{ij} = \pi_i^* p_{ji}$.

Fact 6. *An irreducible and aperiodic Markov chain has a unique stationary distribution.*

Definition 7. (Natural random walk) A natural random walk on G is a random walk with

$$p_{ij} = \begin{cases} 1/2d_i, & (i, j) \in E, i \neq j \\ 1/2 & i = j. \end{cases}$$

Fact 8. *The natural random walk is reversible, irreducible and aperiodic with a unique stationary distribution $\pi_i^* = \frac{d_i}{\sum_i d_i}$. When G is regular, a natural random walk is also symmetric and has a uniform stationary distribution.*

Definition 9. (Mixing time of a random walk) For a random walk $\mathcal{X}$ with a unique stationary distribution π^*, consider the Total Variational (TV) distance[1] (cf. [27, Chap. 4]) $d_{\mathrm{TV},i}(t; \pi_0) \triangleq \frac{1}{2} \sum_i |\mathbb{P}(X_t = i, \pi_0) - \pi_i^*|$ for an initial distribution π_0. Then the mixing time of $\mathcal{X}$ is defined as

$$T_{\mathrm{mix}}(\epsilon; P) \triangleq \sup_{\pi(0)} \inf\{t : d_{\mathrm{TV}}(t; P, \pi(0)) \leq \epsilon\}.$$

2) Asymptotic Notation: We use the following asymptotic notation. For two functions f and g of a variable n, as $n \to \infty$, we write

- $g = \mathcal{O}(f)$ if the ratio g/f is asymptotically finite. Further, $g = o(f)$ if this limit is zero.
- $g = \Omega(f)$ if $f = \mathcal{O}(g)$. Further, $g = \omega(f) \iff f = o(g)$.
- $g = \Theta(f)$ if $g = \mathcal{O}(f)$ and $g = \Omega(f)$.

When f and g are random, these relations are defined to hold with probability one.

3) Graph Sequences and the Asymptotic Regime: Consider a sequence of (possibly random) undirected graphs (G_n), whose n^{th} member G_n has n vertices $V_n = \{1, 2, \ldots, n\}$ and a set of edges E_n. We assume each graph contains all self-loops. Denote the maximum and minimum node degrees of G_n by $d_{\max}(G_n)$ (shortened to $d_{\max}$) and $d_{\min}$ (shortened to $d_{\min}$) respectively. We provide some standard definitions below.

Definition 10. (Asymptotically regular graph) G_n is asymptotically regular if $d_{\max}(G_n) - d_{\min}(G_n) = o(1)$.

Definition 11. (Asymptotically almost sure validity) A property $\mathcal{P}$ is true asymptotically almost surely (a.a.s.) for a sequence of random objects (X_n), if $\lim_{n \to \infty} \mathbb{P}(X_n \text{ has property } \mathcal{P}) = 1$.

We obtain scaling results for the convergence of the average consensus algorithm in large networks by mapping the problem to the scaling of mixing times of natural random walks on a sequence of graphs that are connected and regular asymptotically almost surely.

III. PROBLEM FORMULATION

A. Average Consensus and Random Walks

Consider a set of sensor nodes $V_n = \{1, 2, \ldots, n\}$. Associate with the ith sensor an initial observation $z_i(0)$. Given a realization of a random *message-passing* graph G_n with vertices V_n

---

[1]The TV distance between two distributions μ and ν over a countable set $\mathcal{S}$ is defined as $\|\mu - \nu\|_{TV} \triangleq \frac{1}{2} \sum_{i \in \mathcal{S}} |\mu_i - \nu_i|$ (essentially the ℓ_1 norm).

and edges E_n, suppose that all the vertices $i \in V_n$ synchronously update their observations as

$$z_i(k+1) = \frac{1}{2}z_i(k) + \frac{1}{2d_i} \sum_{j \in N_i(G_n)} (z_j(k) - z_i(k)), \qquad k = 0, 1, \ldots \tag{1}$$

Here $N_i(G_n)$ denotes the neighborhood of vertex i in G_n. By stacking the individual observations z_i to form the observation vector $\mathbf{z}$, the $(k+1)^{\text{th}}$ update starting from an initial observation vector $\mathbf{z}(0)$ can be written as

$$\mathbf{z}(k+1) = \mathbf{W}_n \mathbf{z}(k). \tag{2}$$

where we have defined the *update matrix* $\mathbf{W}_n \triangleq (\mathbf{I}_n - \boldsymbol{\Delta}_n \mathbf{L}_n)/2$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, $\boldsymbol{\Delta}_n \triangleq \mathrm{diag}[d_i^{-1}]$ and $\mathbf{L}_n$ is the graph Laplacian. Notice that $\mathbf{W}_n$ depends on the realization of the random graph G_n, which remains the same for all iterations. We will analyze the speed of convergence for specific families of random graphs in the scaling limit $n \to \infty$, by deriving properties of interest that hold a.a.s. for all realizations of G_n.

Without loss of generality, let $z_i(0) > 0$, and define $z_i'(0) \triangleq z_i(0)/\sum_i z_i(0)$ as the normalized initial observation vector. In the light of Fact 3 and Definition 7, the iteration $\mathbf{z}'(k+1) = \mathbf{W}_n \mathbf{z}'(k)$ can now be interpreted as time-evolution of the node occupancy distribution of a natural random walk over G_n with a transition probability matrix $\mathbf{W}_n$ [2], [23].

If G_n is also connected, this equivalence with a natural random walk ensures (from Fact 6) that the value of each vertex asymptotically reaches $\frac{1}{n}\sum_i z_i(0) = \mathbf{1}^{\mathrm{T}}\frac{\mathbf{z}(0)}{n}$ (a more general result for a time-varying case was studied in [10]). Interpreting each vertex as a sensor and the initial values $(z_i(0))_{i \in V_n}$ as sensor measurements, this algorithm allows each sensor to iteratively compute the average $\frac{1}{n}\sum_i z_i(0)$ of the initial measurement set by exchanging messages as described in (1). We will sometimes also refer to G_n as the *message-passing network*.

The rate of convergence of (2) to its steady state value can be understood in terms of the mixing time of the natural random walk described by $\mathbf{W}_n$. Indeed, by expressing z_i' in terms of z_i, we can write from Definition 9:

$$T_{\mathrm{mix}}(\epsilon; \mathbf{W}_n) = \sup_{\mathbf{z}(\mathbf{0})} \inf\{k : \|\mathbf{z}(k) - n^{-1}\mathbf{1}z_0\|_{\mathrm{TV}} \le \epsilon z_0\} \tag{3}$$

where $z_0 \triangleq \sum_i z_i(0)$.

When G_n is a.a.s. connected and regular, we know from Fact 8 that the stationary distribution of the random walk is uniform a.a.s., thereby implying convergence to average consensus a.a.s.

In this paper, we analyze random graphs based on the disk graph [22], which are parameterized by the disk radius (see Section III-B). For this family of graphs, it is well-known that the graphs are a.a.s. connected if and only if the radius remains large enough with n (i.e., in the "supercritical" regime [28], see, e.g., [29] for a proof). In this regime, the asymptotic regularity property was formally shown to hold a.a.s. in [2, Lemma 10]. In fact, in [2] these two properties were used to establish scaling laws for the mixing time of both the natural and the fastest mixing reversible random walks on these graphs to the uniform distribution.

It is well-known that the mixing time of a random walk can be characterized by the second-largest eigenvalue of $\mathbf{W}_n$. Denoting the eigenvalues of $\mathbf{W}_n$ by $\mu_1 = 1 > \mu_2 > \cdots > \mu_n > 0$, the asymptotic convergence of the iteration (2) is determined by μ_2. The result below formally establishes this dependence:

Theorem 12. *[30]. The $\epsilon-$mixing time of a random walk with a doubly stochastic positive definite transition matrix $\mathbf{W}_n$ on a connected graph G_n is bounded as*

$$\frac{\mu_2 \log(2\epsilon)^{-1}}{2(1-\mu_2)} \le T_{\mathrm{mix}}(\epsilon; \mathbf{W}_n) \le \frac{\log n - \log \epsilon}{1 - \mu_2},$$

where $1 - \mu_2$ is called the spectral gap *of G_n.*

Remark: Observe that the spectral gap controls the mixing time. In the scaling limit $n \to \infty$, the scaling of ϵ also becomes important. The logarithmic dependence on ϵ^{-1} suggests three meaningful possibilities:

1) Polynomial scaling: $\epsilon = 1/n^\delta$ for some fixed $\delta > 0$.
2) Exponential scaling: $\epsilon = \exp(-\delta'n)$ for some fixed $\delta' > 0$.
3) Constant error: $\epsilon \ll 1$ is constant.

For polynomial and exponential error scaling, it is clear that the bounds in Theorem 12 are of the same order, and are $\Theta((1 - \mu_2)^{-1} \log n)$ and $\Theta((1 - \mu_2)^{-1}n)$ respectively. For constant error, the upper bound scales $\log n$ times faster than the lower bound, i.e., $T_{\mathrm{mix}} = \Omega((1 - \mu_2)^{-1})$ and $T = \mathcal{O}((1 - \mu_2)^{-1} \log n)$.

In the sequel we assume polynomial scaling, as was done in [2]. It will become clear in the later sections that the scaling laws for exponential scaling follow from a substitution $\log n \mapsto n$.

Spectral Gap and Cheeger's Inequality:

Intuition suggests that the mixing time of a Markov chain depends on how "easy" it is to move

out of any specified region in the state space. This property can be formalized with the notion of *conductance*. The conductance of a reversible Markov chain on a state space $\Omega = V$ on a graph G_n with an equilibrium distribution π^* is defined as follows [31]:

$$h = \min_{S \subset \Omega, \pi^*(S) \leq 1/2} \frac{Q(S, \bar{S})}{\pi^*(S)}, \tag{4}$$

where $\pi^*(S) \triangleq \sum_{i \in S} \pi^*(i)$ and $\bar{S} = \Omega \backslash S$, and $Q(S, \bar{S}) \triangleq \sum_{i \in s, j \in \bar{S}} \pi^*(i) \mathbb{P}(X_{n+1} = j | X_n = i)$. Viewed in graph-theoretic terms, the numerator (4) measures the effective weighted flow across the cut $(S, \bar{S})$, while the denominator measures the weighted capacity of S. Intuitively, we would expect a larger conductance to correspond to a smaller mixing time, or equivalently from Theorem 12, a larger $1 - \mu_2$ of the underlying graph G_n. This is indeed the case, as Cheeger's Inequality shows:

Theorem 13. *[30]. The spectral gap of a reversible Markov chain satisfies*

$$\frac{h^2}{2} \leq 1 - \mu_2 \leq 2h,$$

where h is the conductance of the Markov chain.

Once we know how h scales with n for a (random) sequence of graphs (G_n), we can use Theorem 13 to find the scaling law for their spectral gap. This, in turn, permits the use Theorem 12 in deriving scaling laws for the mixing time for iterations of the form (2) on these sequences of graphs. In the following, motivated by the need to capture the distance-dependence and randomness in the connectivity of the nodes, we present random geometric graph models for G_n.

B. Network Models

Each point $i \in \{1, 2, \ldots, n\}$ is placed uniformly randomly in a $d-$dimensional torus $\mathcal{T}_d$ on $[0, 1]^d$, i.e., the vertices form a binomial point process [32] $\Phi = \{x_i\}$, $i = 1, 2, \ldots n$, on $\mathcal{T}_d$. Each element of (G_n) is based on the well-known disk graph model [22], [28]. In the following let $b_d(x, r) \equiv b(x, r)$ denote a Euclidean ball centered at $x \in \mathbb{R}^d$ and radius r, and $|b(x, r)|$ denote its volume.

1) Networks with Short-Range Communication: In this case, G_n is the $d-$dimensional disk graph parameterized by the common *communication range* r of each node. The neighborhood of

node $x_i \in \Phi$ that will be used for implementing (1) is

$$N_{x_i}(r) \triangleq \{x_j \in \Phi : \|x_j - x_i\| \leq r\},$$

where $\|\cdot\|$ denotes the Euclidean norm. In this paper, we will always operate in the super-critical regime, i.e., $r = \omega(r_c)$, where $r_c \triangleq (\frac{\log n}{n})^{1/d}$ to ensure asymptotic connectivity and regularity of (G_n) [20]. We label this family of graphs as $G_n^{\mathrm{sh}}(r, d) \equiv G_n^{\mathrm{sh}}(r)$, and the update matrix by $\mathbf{W}_n^{\mathrm{sh}}$. We refer to the points of Φ either by their location $x_i \in \mathbb{R}^d$ or by their index $i \in \mathbb{N}$.

2) Networks with both Short- and Selective Long-Range Communication: We start with a disk graph $G_n^{\mathrm{sh}}(r)$ and add long-range edges of length $s = \Theta(r^\gamma)$. The parameter γ controls the distance over which long-range communication occurs: for a given r a node can communicate with nodes farther away as $\gamma \to 0$. We add the long edges as follows.

For some $r, \eta > 0$ and $0 < \gamma < 1$, tile the torus with hypercubes of side length ηr. Let c denote one of these hypercubes. Along each dimension $m = 1, 2, \ldots d$, let c_m^+ and c_m^- denote the farthest hypercubes from c that are less than distance $s/2$ away from c along the mth coordinate axis, the distance being measured in terms of the separation between their farthest edges. We call these hypercubes as the *partner hypercubes* of c. Figure 2 illustrates the case of $d = 2$. It is easy to see that from any vertex in c, any vertex in c_m^+ and c_m^- is at a distance of at most $\sqrt{(d-1)\eta^2 r^2 + s^2/4} \leq \frac{s}{\sqrt{2}}$ for a small enough η.

Since $r = \omega(r_c)$, every tile c contains $n\eta^2 r^2$ nodes a.a.s. Without loss of generality, let x_1 be one of these nodes. Now add an edge between x_1 and every vertex in c_m^+, c_m^- for $m = 1, 2, \ldots, d$. Thus each of these nodes becomes a *long-range partner* of x_1. Repeat this procedure for every node in Φ, and count duplicate edges only once. Thus for $r = \omega(r_c)$, every node in every tile is additionally connected to $nr^2|b(0,1)| + 2dn\eta^2 r^2 + o(1)$ nodes a.a.s., i.e., G_n is regular asymptotically almost surely. Hence an iteration of the form (2) on this graph will converge to a uniform distribution a.a.s. We define the resultant graph as $G_n^{\mathrm{l}}(r, s, d) \equiv G_n^{\mathrm{l}}(r, s)$ and the corresponding update matrix by $\mathbf{W}_n^{\mathrm{l}}$.

Notice that this model adds long edges selectively to each node; it is motivated by the observation that a small number of long edges added to a graph can greatly increase its spectral gap, as is the case in small-world graphs (cf. [33, Chap. 14]). We have adapted this idea to a wireless network. Instead of adding a single additional edge to a node as is normally the case in abstract graph-theoretic models, the inherent broadcast nature of the wireless channel allows a transmitter to
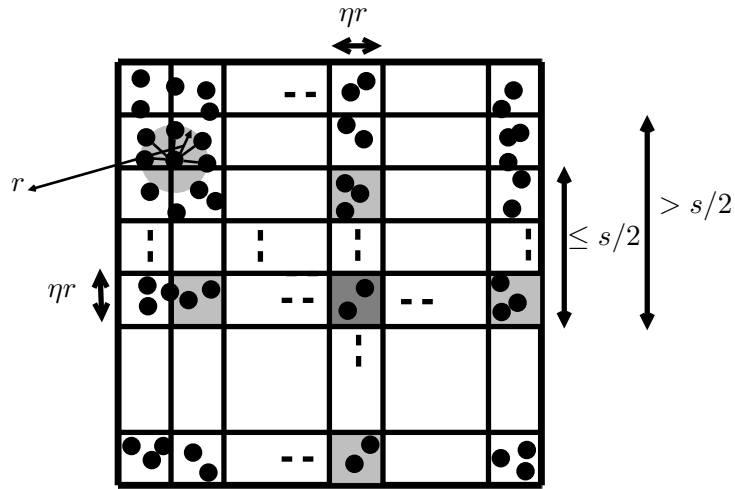
Figure 2. An illustration of the geometric random graph models for $d = 2$. The vertices are shown as black circles. In $G_n^{\mathrm{s}}(r)$, an edge exists between any two nodes iff they are at most at a distance r (the communication range) away from each other. This is shown, for example, for the node at the center of the circle. G_n^{l} contains all edges in G_n^{s}. Additionally each node communicates with its long-range partners. For example, for each node in the dark gray square, all nodes in the lightly shaded squares are long-range partners. These partner squares are chosen such that the distance between their farthest edges is less than $s/2$. Note that there are 4 such partner squares, two along each coordinate axis.

broadcast its information to several receivers that are in close proximity to one another with very little overhead. This allows multiple communication paths to form simultaneously.

We now describe the communication model, which is a well-accepted model in the study of wireless networks.

C. Communication Model

We make the following assumptions on the communication model:

- All edges in G_n^{sh} and G_n^{l} are established by wireless links that operate in the same frequency band (normalized to unit bandwidth).

- Each node encodes its message in $K \gg 1$ nats, such that there is negligible quantization error. These messages are sent using a point-to-point capacity-achieving AWGN channel code with SNR threshold β (i.e., $R = \log(1 + \beta)$). Transmissions are slotted with K/R channel uses allowed per slot.

- There is no fading. The path-loss exponent α is greater than the dimension d of the network, so that the interference remains finite a.s. as the network size grows.

- A packet from node i can be received at j iff the Signal-to-Interference-Ratio (SIR) at node

j, SIR_{ij}, is greater than a known constant $\beta > 0$. Therefore for any sender i and receiver j, the link $i \to j$ will be in outage iff

$$\frac{\|x_j - x_i\|^{-\alpha}}{\sum_{k \in \mathcal{S} \setminus \{i\}} \|x_j - x_k\|^{-\alpha}} < \beta. \tag{5}$$

where $\mathcal{S}$ is the set of all senders that transmit in the same slot as i. This is the well-known interference-limited physical model [20][2].

- The medium-access scheme is TDMA with spatial re-use.

Thus the successful formation of each edge in a graph G_n is mapped to a successful link formation in each direction. Notice that (5) models that fact that there is a limit to the number of edges that can be formed simultaneously, and consequently on the maximum rate at which a given message-passing graph can be established. For a given TDMA protocol, the rate of topology formation is thus determined by its *schedule length* in time-slots. Since we investigate networks in the scaling limit, we will investigate the scaling properties of the fastest TDMA protocols that can establish a given sequence of random graphs (G_n) (i.e., have the smallest schedule length a.a.s.)

D. Quantifying the Effective Speed of Convergence

Note that the mixing time, which is a function of the update matrix $\mathbf{W}_n$, the smallest *number of iterations* to converge to an ϵ−ball around the average consensus point. This is different from the *time taken* to taken to converge to this ball with a finite rate of topology formation in each iteration. For example, in Fig. 1, due to interference constraints, the shortest schedule to construct this topology has at least three time slots. Hence message-passing iterations using this topology can occur no faster than once in every three time slots.

Thus for a topology G_n and an update matrix $\mathbf{W}_n$, the *smallest* effective time to converge is the *product* of the mixing time $T_{\text{mix}}(\epsilon; \mathbf{W}_n)$ of a topology and the length $T^*(G_n, \beta)$ of the *shortest* TDMA schedule that constructs the topology in each iteration. We call this the *Slot Mixing Time*. We formally state it below for future reference:

Definition 14. (Slot Mixing Time) The Slot Mixing Time $T_{\text{slots}}(G_n) \equiv T_{\text{slots}}(G_n, \mathbf{W}_n, \beta, \epsilon)$ is

---

[2]It is possible to derive our scaling results by including both noise and interference in the SINR model at the cost of making equations and derivations more cumbersome while distracting from the main message of the paper, which is the analysis of the performance with interference. Hence we focus on the interference-limited case.

defined as the product

$$T_{\mathrm{slots}}(G_n) \triangleq T_{\mathrm{mix}}(\epsilon; \mathbf{W}_n) \cdot T^*(G_n, \beta),$$

where $T_{\mathrm{mix}}(\epsilon; \mathbf{W}_n)$ is the $\epsilon-$mixing time of iterations using a message-passing graph G_n and an update matrix $\mathbf{W}_n$ and $T^*(G_n, \beta)$ is the length of the shortest TDMA schedule that constructs G_n in time slots.

Notice that in general $T_{\mathrm{slots}}(G_n)$ depends on the realization of the random graph G_n. We will analyze the scaling of $T_{\mathrm{slots}}(G_n)$ for the families of random geometric graphs described in Section III-B.

E. Asymptotic Behavior

From Sections III-A and III-C we notice that the problem involves:

- The network size n.
- The short link distance r.
- The parameter γ that controls the length of long links.

We will study the mixing time in an interference-limited network in the regime $n \to \infty$.

IV. CONVERGENCE IN NETWORKS WITH SMALL COMMUNICATION RANGE

A. Characterizing the Spectral Gap

The spectral gap for the disk graph is known to be $\Theta(r^2)$, independent of network dimension [2]. Using Cheeger's Inequality (Theorem 12), it was shown that the mixing time of the *fastest mixing reversible random walk* with a uniform distribution on $G_n^{\mathrm{sh}}(r)$, for polynomial scaling $\epsilon = 1/n^\delta$, $\delta > 0$ scales as

$$T_{\mathrm{mix}}(\mathbf{W}_n^{\mathrm{sh}}) = \Theta(r^{-2} \log n). \tag{6}$$

It was also shown therein that the mixing time for the natural random walk on G_n^{sh} is also $\Theta(r^{-2} \log n)$. We will now use combine the scaling law for the mixing time with the fastest rate of topology formation implied by the communication model in Section III-C.

B. Interference-Limited Topology Formation

We now prove two results that follow from the assumptions made in Section III-C.

Proposition 15. *Consider a system of n nodes on a $d-$dimensional torus with a short-range communication range r, that communicate using point-to-point codes with SINR threshold β, with $\alpha > d$ being the path-loss exponent. Assuming the short-range network model in Section III-B1 and the communication model described in III-C, the length of the shortest TDMA schedule that constructs G_n^{sh} has no fewer than $C_1 n r^d \beta^{d/\alpha}$ slots a.a.s., for some positive constant C_1.*

Proof: Let $\mathcal{S}$ be the set of concurrent transmitters at any given time. Suppose node j is an intended receiver of a transmitter $i \in \mathcal{S}$. Then i's message is decoded correctly iff (5) is satisfied. Thus for all $k \in \mathcal{S}\backslash\{i\}$,

$$\|x_j - x_k\| \quad \geq \quad \beta^{1/\alpha}\|x_j - x_i\|. \tag{7}$$

Clearly this is true even for the *farthest* intended receiver. It is easy to show that such a receiver lies a.a.s. in a ring of inner radius $s(1-\delta'')$ for some fixed $\delta'' > 0$. We thus conclude $\|x_k - x_j\| \geq r(1-\delta'')\beta^{1/\alpha} \triangleq r_{\min}$ a.a.s.

This suggests that any TDMA protocol allowing i to pass a message to its farthest node j needs to set up a guard zone of radius no smaller than $r_{\min}$ around j. Since every node inside this guard zone must transmit at least once to form the required message passing graph, any TDMA protocol that constructs the message passing graph G_n^{sh} requires least $\sum_{x \in \Phi} \mathbf{1}_{x \in \Phi \cap b(0, r_{\min})}$ slots. Here the indicator $\mathbf{1}_{x \in \Phi \cap b(0, r_{\min})}$ is used to indicate the existence of the point $x \in \Phi$ inside the ball $b(0, r_{\min})$. The summation is over all points $x \in \Phi$.

For $r = \omega(r_c)$, each such ball has $n|b(0, r_{\min})| = nr^d \beta^{d/\alpha}(1-\delta)^d|b(0,1)| + o(1) \geq C_1 nr^d \beta^{d/\alpha}$ a.a.s., where $C_1 = 0.5(1-\delta)^d|b(0,1)|$. ∎

Proposition 16. *Consider the network model in Section III-B1 and the communication model described in III-C. The length of the shortest TDMA schedule that constructs G_n^{sh} has at most $C_2 nr^d \beta^{d/\alpha}$ slots a.a.s., for some positive constant C_2.*

Proof: The proof involves construction of a feasible TDMA schedule whose length is $C_2 nr^d \beta^{d/\alpha}$ Let $x \triangleq \theta r$ for some fixed $\theta > 1$. Consider the lattice $\mathbb{L}$ that consists of points on the scaled integer lattice $x\mathbb{Z}^2$ that also lie on the torus. In other words, $\mathbb{L} = x\mathbb{Z}^2 \cap \mathcal{T}_2(n)$. Partition $\mathbb{L}$ into sublattices as follows:

- $\mathbb{L}_{00} \triangleq \{(ix, jx) \in \mathbb{L} : i \text{ and } j \text{ are even}\}$

- $\mathbb{L}_{01} \triangleq \{(ix, jx) \in \mathbb{L} : i \text{ even}, j \text{ odd}\}$

- $\mathbb{L}_{10} \triangleq \{(ix, jx) \in \mathbb{L} : i \text{ odd}, j \text{ even}\}$

- $\mathbb{L}_{11} \triangleq \{(ix, jx) \in \mathbb{L} : i \text{ and } j \text{ are odd}\}$

With each lattice site $p \in \mathbb{L}$ one can associate the tile $\tau_p = p + [0, x]^2$ that lies within the torus $\mathcal{T}_2(n)$. Denote by $\mathbb{T}_{ij}$ the set of such tiles associated with each of the points in $\mathbb{L}_{ij}$, $i, j = 0, 1$. For example, $\mathbb{T}_{00} \triangleq \{\tau_p : p \in \mathbb{L}_{00}\}$. Thus $\{\mathbb{T}_{ij}\}$ partition the torus $\mathcal{T}_2(n)$.

The idea behind such a partition is to enable spatial re-use. Consider the following four-phase MAC protocol consisting of phases 00, 01, 10, 11. In phase ij at most one node from each tile in $\mathbb{T}_{ij}$ is allowed to transmit. The protocol ensures that each node transmits exactly once.

The next step is to show that this protocol provides the desired connectivity to each node every $C_2 n r^2 \beta^{2/\alpha}$ time slots for some positive C_2. To this end, we first show that the interference at each intended receiver is bounded from above and can be made smaller than any $\beta > 0$ by a suitable choice of θ.

Consider one such transmission in phase 00. Let $\mathcal{S} \subset \mathbb{T}_{00} \cap V_n$ be the set of all transmitters. Consider a transmitting node i in tile τ_p where $p = (0, 0)$, i.e., a tile at the origin. To remain feasible, the protocol must satisfy (5) for each successful link. For any i, j, k, it is clear that

$$
\begin{aligned}
\|x_k - x_j\| &= \|x_k - x_i - (x_j - x_i)\| \\
&\geq \|x_k - x_i\| - \|x_j - x_i\| \\
&\geq \|x_k - x_i\| - r,
\end{aligned}
$$

since $\|x_j - x_i\| \leq r$. Therefore for a transmitter at x_i, the interference power at any intended receiver at x_j can be upper bounded as

$$
\sum_{k \in \mathcal{S} \setminus \{i\}} \|x_k - x_j\|^{-\alpha} \leq \sum_{k \in \mathcal{S} \setminus \{i\}} (\|x_k - x_i\| - r)^{-\alpha}, \tag{8}
$$

where the right hand side is independent of j. By the design of the protocol, an interferer k for any intended receiver of the message from i must lie in a tile distinct from $\tau_{(0,0)}$. Moreover, such a tile should lie within $\mathbb{T}_{00}$; thus the protocol imposes a lower bound on the minimum distance between any two concurrent transmitters. Using geometrical arguments (see Figure 3), the right
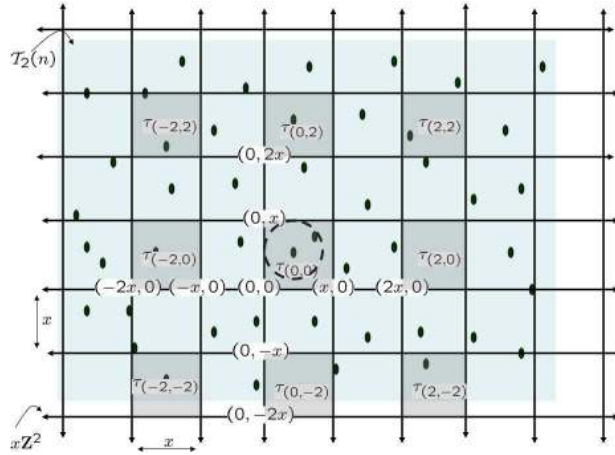
Figure 3. Geometric reasoning underlying the proof of Proposition 16. The location of a typical transmitter in $\tau_{0,0}$ and one of its intended receivers is shown. The nearest interferers and their respective tiles are $\tau_{2,0}, \tau_{2,2}, \tau_{0,2}, \tau_{-2,2}, \tau_{-2,0}, \tau_{-2,-2}, \tau_{0,-2}, \tau_{2,-2}$. The signal power from any of one these interferers at the intended receiver is no larger than that received from the closest interferer allowed by the protocol. The protocol ensures that this nearest distance is no smaller than $x = \theta r$.

hand side of (8) is upper bounded as

$$
\sum_{k \in \mathcal{S} \setminus \{i\}} (\|x_k - x_i\| - r)^{-\alpha}
$$

$$
\leq \sum_{l=1}^{\infty} 8l \left((2l-1)\theta r - r \right)^{-\alpha}
$$

$$
= 8r^{-\alpha} \sum_{l=1}^{\infty} l \left((2l-1)\theta - 1 \right)^{-\alpha}
$$

$$
\leq 8r^{-\alpha} \left(((\theta-1)^{-\alpha} + \sum_{l=2}^{\infty} l((2l-1)\theta - \theta)^{-\alpha} \right)
$$

$$
= 8r^{-\alpha} \left((\theta-1)^{-\alpha} + 2^{-\alpha}\theta^{-\alpha} \sum_{l=2}^{\infty} l(l-1)^{-\alpha} \right)
$$

$$
\leq \xi r^{-\alpha}(\theta-1)^{-\alpha}, \tag{9}
$$

for some fixed $\xi > 0$, since the sum converges for $\alpha > 2$ (in general, for $\alpha > d$, as assumed in the communication model). The SIR condition (5) is guaranteed to be satisfied at every intended

j, if θ is chosen such that

$$
\frac{r^{-\alpha}}{\xi r^{-\alpha}(\theta - 1)^{-\alpha}} \;\geq\; \beta
$$
$$
\implies \theta \;\geq\; 1 + (\xi\beta)^{\frac{1}{\alpha}}.
$$

For a suitable choice of ξ, we can set $\theta = 10(\xi\beta)^{\frac{1}{\alpha}}$.

For $s = \omega(s_c)$, the number of nodes in each tile is $nx^2 + o(1)$ a.a.s. Hence as $n \to \infty$, the protocol constructed requires $4nx^2 + o(1) \leq C_2 ns^2 \beta^{2/\alpha}$ transmissions almost surely to establish the necessary connectivity to each node in the network, where $C_2 \geq 400\xi^{2/\alpha}$. By optimality, the number of slots T^* in the shortest TDMA schedule cannot exceed this number. ∎

The results from Propositions 15 and 16 lead to the following corollary.

Corollary 17. *If $T^*(G_n^{\mathrm{sh}}(r), \beta)$ denotes the length of the shortest TDMA schedule, then as $n \to \infty$, a.a.s.:*

1) *For fixed β, $T^*(G_n^{\mathrm{sh}}, \beta) = \Theta(nr^d)$.*
2) *When $\beta \equiv \beta(n) = \Omega(1)$, $T^*(G_n^{\mathrm{sh}}, \beta) = \Omega(nr^d\{\beta(n)\}^{\frac{d}{\alpha}})$.*

Proof: Claim 1 is evident from the results of Propositions 15 and 16.

For some constants C_1 and C_2, we have from Propositions 15 and 16, a.a.s. for large n and a fixed β,

$$
C_1 nr^d \beta^{d/\alpha} \leq T^*(G_n^{\mathrm{sh}}, \beta) \leq C_2 nr^d \beta^{d/\alpha}.
$$

Since C_1 (but not C_2) is independent of β, we can write for $n \to \infty$, when $\beta \equiv \beta(n) = \Omega(1)$
$T^*(G_n^{\mathrm{sh}}) = \Omega(nr^d\beta^{\frac{d}{\alpha}}(n))$. ∎

If all nodes had independent point-to-point channels between one another, the rate of topology formation would be $\Theta(1)$. For a wireless channel, however, Corollary 17 suggests that it requires $\Theta(1/nr^d)$ even with optimum spatial re-use. Thus better-connected disk graphs are penalized by a smaller rate of topology formation. We combine the mixing time result (6) to examine the scaling law for the effective *time* necessary for convergence in the next section.

C. Rate of Convergence

1) Slot Mixing Time: We now analyze the asymptotic convergence behavior of the distributed averaging algorithm (2) in a dense network as $n \to \infty$. From the earlier sections, we know the

scaling laws for this regime for:

1) The *number of iterations* necessary to a.s. reach an $\epsilon-$ball (from (3)).

2) The *shortest TDMA schedule length* to a.s. realize G_n^{sh} in each iteration (from Corollary 17).

Thus from Definition 14, for fixed β, the Slot Mixing Time scales as

$$T_{\text{slots}}(G_n^{\text{sh}}) \triangleq T_{\text{mix}}(\mathbf{W}_n^{\text{sh}}) \cdot T^*(G_n^{\text{sh}}, \beta) = \Theta(nr^{d-2} \log n) \tag{10}$$

slots a.a.s., for $\epsilon = 1/n^{\delta}$.

From Proposition 15 and the Gaussian signaling assumption, when we also allow β to depend on n such that $\beta(n) = \Omega(1)$, the *time* to reach this ball scales as

$$\Omega\left(nr^{d-2} \frac{e^{R(n)d/\alpha}}{R(n)} \log n\right) \qquad \text{a.a.s.}$$

where $R(n) \equiv \log(1 + \beta(n))$.

2) Choice of Communication Range: For a fixed β the mixing time in (10) scales polynomially in r for $d > 1$. Interestingly, for $d = 1$, the time slots to mix scales as the inverse of r. This suggests that increasing r can improve the rate of convergence. For $d = 2$, however, this quantity scales *independently* of r, suggesting that these two effects exactly cancel each other, a rather non-intuitive result. For higher dimensions, the scaling law has a positive exponent in r—implying that the increasing r can actually slow down mixing.

This dependence on network dimension can be understood as follows. If the network is one-dimensional, although a transmitter is an isotropic radiator, its effect on the network is seen only along the line $[0, 1]$. Although the throughput provided by the optimal TDMA protocol only scales as $\Theta(n^{-1}r^{-1})$ for a given β from Corollary 17, the spectral gap scales as $\Theta(r^{-2})$, offsetting this loss. In $d-$dimensions, however, while the the fastest rate of topology formation scales as $\Theta(n^{-1}r^{-d})$, the spectral gap only scales as $\Theta(r^{-2})$. As a result, improving spatial re-use can become more important than increasing connectivity.

3) Effect of Increasing Transmission Rate: On the one hand, higher transmission rate reduces the packet transmission time; on the other, it also restricts spatial re-use. Clearly the benefit of smaller packet transmission times can be outweighed by reduced spatial re-use for large rates R.

V. CONVERGENCE IN NETWORKS WITH SELECTIVE LONG-RANGE CONNECTIVITY

A. Scaling of the Spectral Gap

To derive the scaling law for the mixing time, we need to find the scaling of the spectral gap of G_n^{l}. As we will see, deriving the scaling law for the conductance of G_n^{l} is sufficient to establish the scaling of the spectral gap.

Proposition 18. *The conductance of G_n^{l} with edge weights determined by $\mathbf{W}_n^{\mathrm{l}}$ is $\Theta(r^\gamma)$ a.a.s., for $d = 1, 2, \ldots$*

Proof: We adopt a modified version of the proof in [34]. From (4) we know that

$$h = \min_{S \subset \Omega, \pi^*(S) \leq 1/2} \frac{Q(S, \bar{S})}{\pi^*(S)}.$$

By the symmetry in G_n^{l} induced by the construction in Section III-B2, it can be shown using arguments similar to [34, Appendix G] that the minimum occurs for $\pi^*(S) = 1/2$, and that the minimizing cut $(S, \bar{S})$ is a hyperplane dividing the torus into two halves. Without loss of generality, define $S \triangleq \Phi \cap \{[0, 1/2) \times [0, 1]\}$.

Also for the natural random walk, each edge weight is $\frac{1}{d_i} = \Theta\left(\frac{1}{nr^2}\right)$ (for d dimensions, $\Theta(n^{-1}r^{-d})$), and the equilibrium distribution is $\Theta(\frac{1}{n})$. It is thus sufficient to count the number of edges traversing this cut. The number of short edges was shown in [34] to be $\Theta(n^2 r^3)$ (for d dimensions $\Theta(n^2 r^{d+1})$). Observe that every node in a square of side ηr has $4n\eta^2 r^2$ long-range partners. One quarter of these edges traverse the cut $(S, \bar{S})$; hence the potential number of long edges that can traverse the cut from a given square is $n\eta^2 r^2 \times n\eta^2 r^2 = \Theta(n^2 r^4)$. Since each edge has length at least $s/2 - 2\eta r = \Theta(r^\gamma)$ (since $s = \Theta(r^\gamma)$ and $0 < \gamma < 1$), which is at most s, it is clear that $\Theta(r^{\gamma-1})$ squares from the cut will contribute to the edges that traverse the cut (see Fig. 4). Multiplying this result by the number of rows $\Theta(r^{-1})$ of such squares, the total number of long edges traversing the cut will be $\Theta(n^2 r^4 \times r^{\gamma-1} \times r^{-1}) = \Theta(n^2 r^{2+\gamma})$ (for general d, $\Theta(n^2 r^{2d} \times r^{\gamma-1} \times r^{-d+1}) = \Theta(n^2 r^{d+\gamma})$). Counting both the short and long edges, we have in d dimensions,

$$
\begin{aligned}
Q(S, \bar{S}) &= \Theta\left(\frac{n^2 r^{d+1} + n^2 r^{d+\gamma}}{n^2 r^d}\right) \\
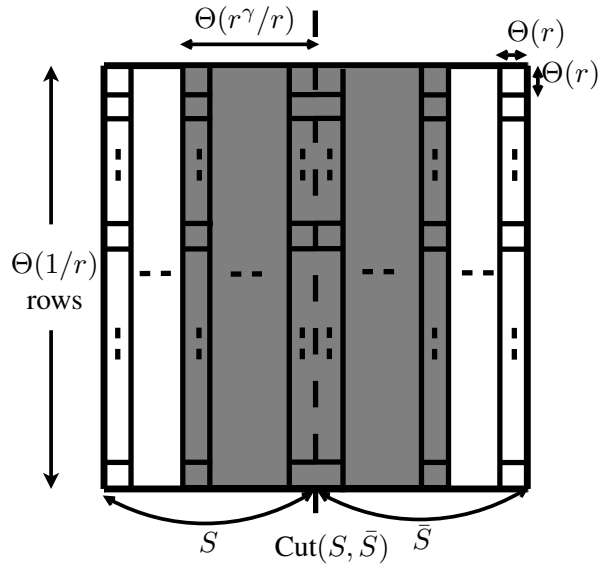&= \Theta(r^\gamma),
\end{aligned}
$$

Figure 4. The geometry behind the proof of Proposition 18 for $d = 2$. The tiling used for the construction of G_n^l is overlaid. By the symmetry induced by the construction, the set $S \subset V_n$ for which $Q(S, \bar{S})/\pi^*(S)$ is minimized corresponds to the left-half of the torus as labeled (it can be argued that this set will have the smallest weighted flow for a given frequency of steady-state occupancy). Since the stationary distribution for this set is $1/2$, finding the scaling law for the number of edges that traverse the cut is sufficient to provide a corresponding scaling result for the conductance. For the short-range communication graph G_n^s (i.e., the disk graph whose edge length is $\mathcal{O}(r)$) only nodes from a finite number of squares from the tiling in either direction from the cut contribute to these edges. For long edges of length $\Theta(r^\gamma)$, a positive fraction of the nodes from $\Theta(r^\gamma/r)$ squares on either side will contribute to these edges. Since there are $\Theta(1/r)$ such rows of squares, the proof lies in finding the scaling law for the number of edges that traverse the cut.

since $\gamma < 1$.

Notice that if a node were allowed to have only a finite number of long-range partners, the contribution of long-edges towards conductance is smaller, without significant interference-reducing benefits. We elaborate on this point in Section V-C3. ∎

We can infer the following from the above result:

Corollary 19. *The spectral gap of G_n^l is $\Omega(r^{2\gamma})$ and $\mathcal{O}(r^\gamma)$.*

Proof: From the lower bound in Theorem 13, we have $1 - \mu_2 = \Omega(r^{2\gamma})$. From the upper bound from the same theorem, we have $1 - \mu_2 = \mathcal{O}(r^\gamma)$. ∎

As noted in Section III-B2, the distance between any two (graph-theoretic) neighbors is no more than $s/\sqrt{2}$. Thus every edge in $G_n^l(r, s, d)$ is also present in the disk graph $G_n^{sh}(s/2)$, i.e., $G_n^l(r, s, d) \subset G_n^{sh}(s/2)$. Hence a reversible random walk on G_n^l with a uniform equilibrium distribution can mix *no faster than the fastest mixing such random walk on $G_n^{sh}(s/2)$*. This key

observation allows us to use a known result that follows from [2, Thm. 8]:

Theorem 20. *The spectral gap corresponding to the transition probability matrix of the fastest mixing reversible random walk on $G_n^{\mathrm{sh}}(r^\gamma)$ with a uniform equilibrium distribution is $\Theta(r^{2\gamma})$ a.a.s.*

Since mixing time decreases with spectral gap, from Theorem 20 we conclude that the spectral gap of G_n^{l} is $\mathcal{O}(r^{2\gamma})$. But we know from Corollary 19 that this gap is also $\Omega(r^{2\gamma})$. Thus we conclude that the spectral gap of G_n^{l} is $\Theta(r^{2\gamma})$, which is formally stated as a theorem:

Theorem 21. *The spectral gap of the natural random walk on G_n^{l} is $\Theta(r^{2\gamma})$.*

This result suggests that the improvement in spectral gap from an increased communication radius from r to r^γ can also be achieved (in the scaling sense) by allowing each node to communicate with a selected number of nodes at a distance $\Theta(r^\gamma)$.

However, as we shall discuss in the next section, such connectivity comes at a price of a lowered rate of topology formation. We find that this loss (as measured by the shortest TDMA schedule length) must be no smaller than the number of nodes in the largest exclusion zone created in the network. Since the longest link distance in both the disk graph $G_n^{\mathrm{sh}}(s/2)$ and G_n^{l} are of the same order, the similarity in the expressions for the spectral gap scaling law suggests that we should expect the same dependence on network dimension as in (10).

B. Convergence with Interference

We will derive bounds for the shortest feasible TDMA schedule for G_n^{l}. In the spirit of the earlier proofs, the lower bound follows from the feasibility constraint (i.e., the schedule constructs the desired message passing graph while satisfying the SINR constraint), while the upper bound is found by bounding the length of the optimum schedule by that of a specific feasible schedule. These results are presented in the following.

Proposition 22. *For a given β, a feasible schedule for G_n^{l} has $C_3 n r^{\gamma d} \beta^{d/\alpha}$ slots a.a.s. for some positive constant C_3. Furthermore, for a given β, this length scales as $\Omega(n r^{\gamma d})$ slots a.a.s.*

Proof: We prove this result for $d = 2$; the proof for $d \neq 2$ is similar. From the system model, it is clear that a TDMA protocol that constructs G_n^{l} must form at least one link of distance at least $s/2 - 2\eta r$. Since $s = \Theta(r^\gamma)$ (i.e., s scales "much slower" than r), at large enough n, the

protocol must create an exclusion zone of radius of at least $s/4$ in the network at least once. All nodes within this exclusion zone must transmit at least once. But $s = \omega(r_c^\gamma)$, which implies we operate the supercritical regime. From a similar argument as in Proposition 15 we can assert that any feasible TDMA protocol must have at least $C_3 n s^d \beta^{d/\alpha}$ slots where C_3 is a positive constant. The scaling law for this length follows from the scaling of s with r. ∎

Proposition 23. *For a given β, the length of the shortest feasible schedule for G_n^l is no more than $C_4(n s^d \beta^{d/\alpha})$ slots a.a.s., for some positive constant C_4. For a given β, this upper bound scales as $\mathcal{O}(n r^d)$ a.a.s.*

Proof: Consider any TDMA protocol that allows each node to communicate with every node within a distance s. Clearly this protocol will also construct G_n^l and is hence feasible. As in Proposition 16, we construct such a four-phase (for $d = 2$, in general a $2d$ phase) TDMA protocol that operates on a tiling of the torus with squares of side $\Theta(s)$. Using an argument similar to Proposition 16, it is clear that the spatial re-use can be adjusted to construct the graph in $C_4 n s^d \beta^{d/\alpha}$ slots a.a.s. for some constant $C_4 > 0$. Using $s = \Theta(r^\gamma)$ we get the scaling law. ∎

Corollary 24. *As $n \to \infty$, the shortest feasible schedule for G_n^l has $T^*(G_n^l, \beta) = \Theta(n r^{\gamma d})$ slots a.a.s., for a fixed β. If we also let $\beta = \beta(n) = \Omega(1)$, $T^*(G_n^l, \beta(n)) = \Omega(n r^{\gamma d} \{\beta(n)\}^{\frac{d}{\alpha}})$.*

Proof: Follows from Propositions 22 and 23. ∎

C. Rate of Convergence with Sparse Long-Range Connectivity

We repeat the analysis in Section IV-C to study the benefit of sparse long-range connectivity for a large number of nodes. From this analysis, we derive a result analogous to (10) for the long-range model. We use this result to discuss the impact of increased communication range.

1) Slot Mixing Time: From Theorem 21, the spectral gap of G_n^l scales as $\Theta(r^{2\gamma})$. Consequently, from the mixing time bounds in Theorem 12, we conclude that the mixing time with $\mathbf{W}_n^l$ scales as

$$T_{\mathrm{mix}}(\mathbf{W}_n^l) = \Theta(r^{2\gamma} \log n) \qquad \text{a.a.s.} \tag{11}$$

iterations for $\epsilon = 1/n^\delta$. On the other hand, from Corollary 24 the shortest TDMA schedule that realizes G_n^l scales as $\Theta(n r^{d\gamma})$ slots.

Multiplying $T_{\text{mix}}(\mathbf{W}_n^{\text{l}})$ and $T^*(G_n^{\text{l}}, \beta)$ we obtain a scaling law analogous to (10) for a network with sparse long links. We state this result as a proposition:

Proposition 25. *As $n \to \infty$, for $\epsilon = 1/n^\delta$ and with the shortest feasible TDMA schedule, the slot mixing time of natural random walks on a sequence of random graphs (G_n^{l}) on a $d-$dimensional torus scales as*

$$T_{\text{slots}}(G_n^{\text{l}}) \triangleq T_{\text{mix}}(\mathbf{W}_n^{\text{l}}) \cdot T^*(G_n^{\text{l}}, \beta) \;\; = \;\; \Theta\left(nr^{(d-2)\gamma} \log n\right) \quad \text{a.a.s.} \tag{12}$$

where r is the short range communication radius, long links are $\Theta(r^\gamma)$ for some $0 < \gamma < 1$, and nodes use point-to-point capacity-achieving AWGN channel codes with SNR threshold β.

From Proposition 15 and the Gaussian signaling assumption in Section III-C, when we also let $\beta \equiv \beta(n) = \Omega(1)$, the *time* to reach this ball scales as

$$\Omega\left(nr^{\gamma(d-2)} \frac{e^{R(n)d/\alpha}}{R(n)} \log n\right), \qquad \text{a.a.s.}$$

where $R(n) \equiv \log(1 + \beta(n))$.

2) Impact of Increasing Communication Range on the Convergence Speed: For a fixed β, from (12) we notice that as with short-range links, the slot mixing time scales polynomially in r for $d > 1$. The parameter s that controls the distance of long-range communication enters the scaling law through r^γ, since $s = \Theta(r^\gamma)$. By comparing (12) and (10) it is clear that its role is identical to that of r in (10). Thus we expect the impact of increased communication range to have the same dependence of the network dimension as in (10). From Proposition 22 and an analysis similar to Section IV-C3, it follows that while one-dimensional networks can converge faster from an increased communication range despite greater interference, the convergence speed of two-dimensional networks scales independently of the communication range. In higher-dimensions the increased interference from a larger communication range can actually lower the rate of convergence.

From the model $s = \Theta(r^\gamma)$, which implies that a larger s can result from either a larger r (communicating with more nearby nodes) or a smaller γ (communicating with nodes farther away). In either case we find that (12) scales faster than (10): when interference is accounted for, selective long-range communications do not improve the rate of convergence.

3) The Importance of Long-Range Clusters: Here we discuss the importance of forming long links from a node to a *cluster* of nodes. Briefly, we argue that adding only a few long edges to a given node does not take full advantage of the broadcast nature of the wireless medium: while these fewer long edges to a node reduce the spectral gap (and can increase mixing time as a result of Theorem 12), forming long links from a node to a *cluster* of nodes causes approximately the same interference as forming a point-to-point link of the same distance. Hence in lowering this cluster size, we do not gain from reduced interference, but can only worsen the spectral gap. So, when interference is factored in, allowing a node to talk to a far-off *cluster* rather than a few far-off nodes allows faster mixing for the same level of interference.

The effect of forming clusters is captured in the long-range model in Section III-B2, which adds all the nodes from a partner hypercube as long-range partners. This maximizes the number of long edges contributed by each hypercube and results in the scaling law in Proposition 18. This is key to deriving Theorem 21.

Suppose we modify the way long edges are added in this model by constructing a new graph $G'_n(s)$ by assigning each node only $\rho_n = \mathcal{O}(nr^d)$ long-range partners in each partner hypercube. Evidently $G'_n(s)$ is regular a.a.s., with node degree $nr^d|b(0,1)| + 2d\rho_n + o(1)$; so iterations as in (1) converge to the average consensus point a.a.s. Denote the corresponding update matrix by $\mathbf{W}'_n$. We will now examine the scaling of the spectral gap of $\mathbf{W}'_n$.

Following the steps in the proof of Proposition 18, the (edge-weighted) conductance of G' is $\Theta(r + r^\gamma(\rho_n/nr^d))$. Since $\rho_n = \mathcal{O}(nr^d)$, the conductance can scale no faster than r^γ.

Therefore, unlike in the case with $G^l_n(s)$, exploiting the inclusion $G'_n(s) \subset G^{\mathrm{sh}}_n(s/2)$ is not enough to conclude the spectral gap of $\mathbf{W}'_n$ to be $\Theta(r^{2\gamma})$. But the inclusion does confirm the spectral gap to be $\mathcal{O}(r^{2\gamma})$. Hence, as one would expect, iterations of the form (1) can converge no faster with $\mathbf{W}'_n$ than with $\mathbf{W}^l_n$.

However, in the scaling limit, the interference resulting from the construction of G^l_n or G'_n are the same: it is obvious from Propositions 22 and 23 that the shortest feasible TDMA schedule for G'_n is also $\Theta(nr^{\gamma d})$ slots. We thus conclude that maximizing the cluster size to include all the nodes inside a partner hypercube speeds up convergence for the same level of interference. However, when this cluster is enlarged to include all nodes within a radius s, we have a disk graph with radius s. From the results in the previous sections, it is clear that the interference penalty to realize this larger disk graph scales similarly but is certainly larger than that of $G^l_n(s)$, which has

only selective long-range links.

VI. Conclusions

We analyzed the convergence rate of average consensus algorithms in the scaling limit of dense wireless networks by combining results from Markov chain theory, random geometric graphs, and wireless networks. When messages in a topology are exchanged over wireless links, the impact of a greater communication range depends crucially on the network dimension. Increased communication range can speed up convergence in one-dimensional networks despite greater interference. In two-dimensional networks, the convergence speed scales independently of the communication range. In three- (and higher-) dimensional networks, forming long links can actually slow down convergence. These results hold whether each node only communicates over short links, or, additionally, with a cluster of far-away nodes.

These results greatly differ from many optimistic results about the benefit of long-range connectivity obtained by analyzing the consensus problem in an abstract graph-theoretic setting. Our results underline the need to accurately account for the cost of interference in designing fast-converging topologies for the average consensus algorithm, or for distributed signal processing problems, in general.

References

[1] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," in *Proc. IEEE Conf. on Decision and Control*, Dec. 2003.

[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Info. Theory*, vol. 52 (6), pp. 2506–2530, 2006.

[3] C. C. Moallemi and B. van Roy, "Consensus propagation," *IEEE Trans. Info. Theory*, vol. 52(11), pp. 4753–4766, Nov. 2006.

[4] V. Blondel, J. Hendrickx, A.Olshevsky, and J. Tsitsiklis., "Convergence in multiagent coordination, consensus and flocking," in *Proceedings of the 44th IEEE Conference on Decision and Control, pp 2996–3000*, 2005.

[5] L. Xiao, S. Boyd, and S. Lall., "A scheme for robust distributed sensor fusion based on average consensus.," in *Proceedings of International Conference on Information Processing in Sensor Networks, pp 63–70*, 2005.

[6] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Sig. Proc.*, vol. 56(7), Part 2, pp. 3315–3326, 2008.

[7] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: efficient aggregation for sensor networks," in *Proceedings of the Fifth International Conference on Information Processing in Sensor Networks*, Apr. 2006.

[8] W. Li, H. Dai, and Y. Zhang, "Location aided fast distributed consensus," *IEEE Trans. Info. Theory (to appear)*. Available at http://www4.ncsu.edu/~hdai/IT08_v10.pdf.

[9] T. C. Aysal and K. E. Barner, "Convergence of consensus models with stochastic disturbances," *IEEE Trans. Info. Theory*, vol. 56(8), pp. 4101–4113, Aug. 2010.

[10] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. on Automatic Control*, vol. vol. 49(9), pp. 1520–1533, Sep. 2004.

[11] W. Ren and R. W. Beard, "Consensus seeking in multi-agent systems using dynamically changing interaction topologies," *IEEE Trans. Auto. Control*, vol. 50(5), pp. 655–61, May 2005.

[12] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. on Auto. Control*, vol. 54(11), pp. 2506–2517, 2009.

[13] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Sig. Proc.*, vol. 56(10), pp. 4905–4918, Oct. 2008.

[14] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43(7), pp. 1192–1203, 2007.

[15] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM J. Control Optim.*, vol. 48(1), pp. 102–33, 2009.

[16] R. Rajagopal and M. J. Wainwright, "Network-based consensus averaging with general noisy channels," tech. rep., UC Berkeley, Dept. of Statistics, Tech. Report 751, May 2008.

[17] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Sig. Proc.*, vol. 57(7), pp. 2748–2761, July 2009.

[18] T. Aysal, A. Sarwate, and A. Dimakis, "Reaching consensus in wireless networks with probabilistic broadcast," in *Proc. 47th Allerton Conf. on Comm., Control and Computation (Monticello, IL)*, Oct. 2009.

[19] S. Vanka, V. Gupta, and M. Haenggi, "Power-Delay Analysis of Consensus Algorithms on Wireless Networks with Interference," *International Journal of Systems, Control, and Communications*, vol. 2, no. 1/2/3, pp. 256–274, 2010.

[20] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Info. Theory*, vol. 46(2), pp. 388–404, Mar. 2000.

[21] S. Vanka, M. Haenggi, and V. Gupta, "Distributed averaging in dense wireless networks," in *2009 IEEE Global Communications Conference (GLOBECOM'09)*, (Honolulu, HI), Dec. 2009.

[22] E. N. Gilbert, "Random plane networks," *J. Soc. for Ind. and Appl. Math.*, vol. 9(4), pp. 533–543, Dec. 1961.

[23] A. T. Salehi and A. Jadbabaie, "Small world phenomenon, rapidly mixing Markov chains, and average consensus algorithms," in *Proc. IEEE Conference on Decision and Control, New Orleans, LA*, Dec. 2007.

[24] S. A. Aldosari and J. M. F. Moura, "Distributed detection in sensor networks: Connectivity graph and small world networks," in *IEEE Asilomar Conf. on Signals, Systems and Compters (Pacific Grove, CA), p. 230-234*, Oct. 30-Nov. 2, 2005.

[25] R. Olfati-Saber, "Ultrafast consensus in small-world networks," in *Proc. American Control Conference 2005 (Portland, OR), p. 2371-2378*, June 8-10, 2005.

[26] M. E. J. Newman and D. J. Watts, "Mean-field solution to the small-world network model," *Phys. Rev. Letters*, vol. 84(14), pp. 3201–3204, 2000.

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

[28] M. Penrose, *Random Geometric Graphs*. Oxford University Press, 2003.

[29] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming, W. M. McEneany, G. Yin, and Q. Zhang, (Eds).*, vol. 46, No. 2, pp. 388–404, Boston, MA: Birkhauser, 1998, pp. 547-566, 1998.

[30] A. Sinclair, "Improved bounds for the mixing rates of markov chains and multicommodity flow," *Combinatorics, Prob. Comput. 1*, pp. 351–370, 1992.

[31] A. Sinclair and M. Jerrum, "Approximate counting, uniform generation and rapidly mixing Markov chains," *Inform. and Comput.*, vol. 82(1), pp. 93–133, 1989.

[32] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications, 2nd Ed.,*. John Wiley and Sons, 1996.

[33] R. Durrett, *Random Graph Dynamics*. Cambridge University Press, 2007.

[34] C. Avin and G. Ercal, "On the cover time and mixing time of random geometric graphs," *Theoretical Computer Science*, vol. 380, pp. 2–22, 2007.