Average Consensus in Dense Wireless Networks with Long-Range Connectivity

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 consider the effect of interference on the convergence rate of a class of distributed averaging algorithms called 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 these topologies are realized on wireless networks, the rate at which they can be formed is in general limited by interference. It is thus not clear if the rate of convergence always increases with network connectivity. We study this problem for randomly-placed consensus-seeking nodes that are connected through a dense (interference-limited) wireless 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 additionally exchange messages with a few selected far-off nodes? When nodes schedule their transmissions to avoid interference, we show that with increasing network dimension, the benefit from increased connectivity remains the same while the time needed to realize these topologies becomes larger. In particular, while increased connectivity improves the convergence rate in one-dimensional networks, it is exactly offset by the longer schedule lengths in two-dimensions. In higher dimensions, greater connectivity actually degrades the rate of convergence. Our results thus provide new insights into the MAC protocol design for this family of distributed averaging algorithms.

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

*Corresponding author. The work of the first two authors was partially supported by NSF (grants CNS 04-47869 and CCF 728763). The work of the third author was supported partially by the NSF award 0846631.
I. INTRODUCTION

The advent of wireless sensor and ad hoc networks has motivated the need for information processing algorithms that are inherently distributed, thereby allowing each node to run the algorithm by utilizing only local information. An archetypal instance of such an algorithm is the distributed averaging algorithm that computes the global average of sensor observations based on purely local computations at each sensor. A well-studied distributed averaging algorithm is the average consensus algorithm, where a set of nodes in a (possibly time-varying) interconnection topology iteratively compute the global average of their initial states, see e.g., [1]–[7] and references therein. The connectivity properties of the graph that ensure convergence have been well-studied (e.g., [8], [9]). Of late, the focus has shifted to studying convergence in the face of communication constraints, like quantization [10], packet drops [11] and noise [12]. A closely associated algorithm is the gossip algorithm, studied for instance in [2], [13].

Most prior work on this problem ignores the fact that in wireless networks, due to half-duplex constraints, it takes more time for two nodes to exchange information than it takes for one-way communication. Moreover, the effect of network geometry has also 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 $(a, b)$ and $(c, d)$. 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.
So the performance of the underlying (real-time) estimation algorithm becomes coupled with algorithms for channel access and routing. In our previous work [14], we study the coupling with channel access for the average consensus algorithm for a certain class of deterministic network topologies. Using a simple protocol model [15] for reception, we were able to show that the effect of increasing network connectivity depends crucially on its dimension. In our recent work [16] we exploit the well-known parallels between the convergence of the average consensus algorithm and Markov chain mixing (see, e.g., [2] and references therein) to study consensus on disk graphs [17] using the more refined physical model. We examine the scaling behavior of the fastest rate of topology formation with interference, captured by the shortest feasible TDMA schedules that construct the graph.

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 [18]–[20] for consensus on graphs with a few long-range edges (as in small-world graphs [21]), 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.

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 $i$.

**Definition 1.** (Random walk on a graph). A random walk $X(G) = (X_k)_{k \in \mathbb{Z}}$ on $V$ is characterized by the $n \times n$ transition probability matrix $P(G) = [p_{ij}]$, with $p_{ij} \triangleq \Pr(X_{k+1} = i \mid X_k = j)$, 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 $P$ is doubly stochastic.

**Fact 3.** A random walk on $G$ is a Markov chain over the state space $V$, with its state distribution $\pi(k+1)$ after $k$ steps satisfying $\pi(k+1) = P\pi(k)$, given an initial state $\pi(0)$.

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

**Definition 5.** (Reversible Markov chain). A Markov chain $X = (X_k)_{k \in \mathbb{Z}}$ 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 $X$ with a unique stationary distribution $\pi^*$, consider the Total Variational (TV) distance$^1$ (cf. [22, Chap. 4]) $d_{\text{TV},i}(t; \pi_0) \triangleq \frac{1}{2} \sum_i |\mathbb{P}(X_t = i, \pi_0) - \pi_i^*|$ for an initial distribution $\pi_0$. Then the mixing time of $X$ is defined as

$$T_{\text{mix}}(\epsilon) \triangleq \sup_{\pi_0} \inf_t \{ t : d_{\text{TV}}(t; \pi_0) \leq \epsilon \}.$$ 

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

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

When $f$ and $g$ are random, these relations 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 $d_{\text{max}}(G_n)$ (shortened to $d_{\text{max}}$) and $d_{\text{min}}$ (shortened to $d_{\text{min}}$) respectively. We provide some standard definitions below.

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

**Definition 11.** (Asymptotically almost sure validity). A property $\mathcal{P}$ is said to be 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 asymptotically almost surely connected and regular graphs.

$^1$The TV distance between two distributions $\mu$ and $\nu$ over a countable $\mathcal{S}$ is defined as $\|\mu - \nu\|_{TV} \triangleq \frac{1}{2} \sum_{i \in \mathcal{S}} |\mu_i - \nu_i|$ (essentially the $\ell_1$ norm).
III. PROBLEM FORMULATION

A. Average Consensus and Random Walks

Consider a graph $G_n$ as in Section II. Associate with each vertex $i$ a state $z_i$ that is updated in a synchronous manner to $z_i^+$ as

$$z_i^+ = \frac{1}{2} z_i + \frac{1}{2d_i} \sum_{j \in N_i(G_n)} (z_j - z_i),$$

(1)

where the superscript $+$ denotes the updated state, $N_i(G_n)$ denotes the neighborhood of vertex $i$. By stacking the individual states $z_i$ to form the column vector $z$, (1) becomes

$$z_k^+ = W z_k,$$

(2)

where $W \triangleq (I - \Delta L)/2$, with $\Delta \triangleq [d_i^{-1}]$ and $L$ being the graph Laplacian. 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 state. In the light of Fact 3 and Definition 7, the iteration $(z')^+ = W z'$ 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 $W$ [2], [18]. If $G_n$ is also connected, this equivalence with a natural random walk ensures (from Fact 6) that the state of each vertex asymptotically reaches $\frac{1}{n} \sum_i z_i(0) = 1^T \frac{z(0)}{n}$ (a more general result for a time-varying case was studied in [8]). 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 $\frac{1}{n} \sum_i z(0)$ of the initial measurement set by exchanging messages as described in (1). We will refer to $G_n$ as the message-passing graph or a 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 $W$. Indeed, by expressing $z_i'$ in terms of the original state $z_i$, we can write from Definition 9:

$$T_{\text{mix}}(\epsilon; W) = \sup_{z(0)} \inf_{z_k \leq \epsilon z(0)} \{k : \|z_k - n^{-1} 1 z_0\|_{TV} \leq \epsilon z_0\}$$

(3)

where $z_0 \triangleq \sum_i z_i(0)$. When $G_n$ is asymptotically almost surely connected and (almost) regular, from Fact 8, we know that stationary distribution of the random walk is (almost) uniform, thus ensuring convergence to (almost) average consensus.

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

**Theorem 12.** [23]. The $\varepsilon$–mixing time of a random walk with a doubly stochastic positive definite transition matrix $W$ on a connected graph $G_n$ is bounded as

$$\frac{\mu_2 \log(2\varepsilon)^{-1}}{2(1 - \mu_2)} \leq T_{\text{mix}}(\varepsilon) \leq \frac{\log n - \log \varepsilon}{1 - \mu_2},$$

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

1) 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 $\pi^*$ is defined as follows [24]:

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

where $\pi^*(S) \triangleq \sum_{i \in S} \pi^*(i)$ and $\bar{S} = \Omega \setminus S$, 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.** [23]. 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 the scaling limit for $h$ 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 labeled point $i \in \{1, 2, \ldots, n\}$ is placed uniformly randomly in a $d$-dimensional torus $T_d$ on $[0,1]^d$, i.e., the vertices form a binomial point process $\Phi = \{x_i\}, i = 1, 2, \ldots, n$, on $T_d$. Each element of $(G_n)$ is based on the well-known disk graph model [17], [26].

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

In this paper, we will always operate in the super-critical regime, i.e., $r = \omega(r_c)$, where $r_c \triangleq (\log n)^{1/d}$ to ensure asymptotic connectivity and regularity of $(G_n)$. We label this family of graphs as $G^s_n(r,d) \equiv G^s_n(r)$. [15]. With a slight abuse of notation, we refer to the points of $\Phi$ 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 $G^s_n(r)$ and add long-range edges as follows. Fix $\eta > 0$ and $0 < \gamma < 1$ and tile the torus with hypercubes of side $\eta r$, and let $c$ denote one of these hypercubes. For some $s = \Theta(r^\gamma)$, 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 $m$th 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 $\eta$.

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 $x_1$ 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 $\Phi$, and count duplicate edges only once. Thus for $r = \omega(r_c)$, every node in every tile is additionally connected to $n\eta^2[b(0,1)] + 2dn\eta^2 r^2 + o(1)$ 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 $G^l_n(r,s,d) \equiv G^l_n(r,s)$. 

July 27, 2010 DRAFT
Figure 2. An illustration of the geometric random graph models for $d = 2$. The vertices are shown as black circles. In $G_n^r(r)$, 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. $G_n^r$ contains all edges in $G_n^r$. 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.

Notice that this model adds long-range edges selectively to each node; it is motivated by the observation that a small number of long-range edges added to a graph can greatly increase its spectral gap, as in the case with small-world graphs (cf. [27, Chap. 14]). We have adapted this idea to a wireless setting. Instead of adding a single additional edge to a node as is normally the case in an abstract graph-theoretic setting, the inherent broadcast nature of the wireless channel allows a transmitter to broadcast its information to several receivers that are in close proximity to one another with very little overhead, allowing the formation of multiple communication paths simultaneously. Moreover, as we will observe in Section V-A, in a random geometric setting, the number of such long edges must be large in order to have a significant effect on the graph conductance. We mention here that since we are interested in the scaling law, all our results hold even if we choose a fraction $\rho$ of the nodes in each partner hypercube as a node’s long-range partners. In the following, we will prove our results for $d = 2$; our proofs can be easily generalized to $d \neq 2$.

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 $G_n^a$ 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 $\beta$ (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 $\alpha$ is greater than the dimension $d$ of the network, such 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 $j$, $\text{SIR}_{ij}$, is greater than a known constant $\beta > 0$. Therefore for any sender $i$ and receiver $j$, the link $i \rightarrow j$ will be in outage iff

$$\frac{\|x_j - x_i\|^{-\alpha}}{\sum_{k \in S \setminus \{i\}} \|x_j - x_k\|^{-\alpha}} < \beta.$$  \hspace{1cm} (5)

where $S$ is the set of all senders that transmit in the same slot as $i$. This is well-known physical model [15].
- 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. Asymptotic Behavior

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

- The network size $n$.
- The short-range link distance $r$.
- The parameter $\gamma$ that controls the length of long-range 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 model was studied in [2], and was found to be $\Theta(r^2)$, independent of network dimension. Using Theorem 12, it was shown that the mixing time of the fastest mixing reversible random walk with a uniform distribution on $G_n^s(r)$, for an error $\epsilon = 1/n^\delta$, $\delta > 0$ scales as

$$T_{\text{mix}}(W) = \Theta(r^{-2} \log n).$$

(6)

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

B. Interference-Limited Link Formation

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

**Proposition 14.** 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 $G_n^s$ has no fewer than $C_1nr^d\beta^{d/\alpha}$ slots a.a.s., for some positive constant $C_1$.

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

$$\|x_j - x_k\| \geq \beta^{1/\alpha}\|x_j - x_i\|.$$ (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 $\|x_k - x_j\| \geq r(1 - \delta)\beta^{1/\alpha} \triangleq r_{\text{min}}$ 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 $r_{\text{min}}$ 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 $G_n^a$ requires least $\sum_{x \in \Phi} 1_{x \in \Phi \cap b(0,r_{\min})}$ slots. 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 $C_1 = 0.5(1 - \delta)^{d}|b(0,1)|$. Hence the result follows.

**Proposition 15.** 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 $G_n^a$ 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, $\mathbb{L} = x\mathbb{Z}^2 \cap T_2(n)$. Partition $\mathbb{L}$ into sublattices as follows:

- $\mathbb{L}_{00} \triangleq \{(i x, j x) \in \mathbb{L} : i \text{ and } j \text{ are even}\}$
- $\mathbb{L}_{01} \triangleq \{(i x, j x) \in \mathbb{L} : i \text{ even, } j \text{ odd}\}$
- $\mathbb{L}_{10} \triangleq \{(i x, j x) \in \mathbb{L} : i \text{ odd, } j \text{ even}\}$
- $\mathbb{L}_{11} \triangleq \{(i x, j x) \in \mathbb{L} : i \text{ and } j \text{ are odd}\}$

With each lattice site $p \in \mathbb{L}$ one can associate the tile $\tau_p = p + [0, x)^{2}$ that lies within the torus $T_2(n)$. Denote by $T_{ij}$ the set of such tiles associated with each of the points in $\mathbb{L}_{ij}$, $i, j = 0, 1$. For example, $T_{00} \triangleq \{\tau_p : p \in \mathbb{L}_{00}\}$. Thus $\{T_{ij}\}$ partition the torus $T_2(n)$.

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 $T_{ij}$ 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 $C_2 nr^{2}\beta^{2/\alpha}$ 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 $\theta$.

Consider one such transmission in phase 00. Let $S \subset T_{00} \cap V_n$ be the set of all transmitters. Consider a transmitting node $i$ in tile $\tau_p$ 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

$$
\|x_k - x_j\| = \|x_k - x_i - (x_j - x_i)\| \\
\geq \|x_k - x_i\| - \|x_j - x_i\| \\
\geq \|x_k - x_i\| - r,
$$

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

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

(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 hand side of (8) is upper bounded as

$$
\sum_{k \in S \setminus \{i\}} (\|x_k - x_i\| - r)^{-\alpha} \\
\leq \sum_{l=1}^{\infty} 8l ((2l - 1)\theta r - r)^{-\alpha} \\
= 8r^{-\alpha} \sum_{l=1}^{\infty} l ((2l - 1)\theta - 1)^{-\alpha} \\
\leq 8r^{-\alpha} \left( (\theta - 1)^{-\alpha} + \sum_{l=2}^{\infty} l((2l - 1)\theta - 1)^{-\alpha} \right) \\
= 8r^{-\alpha} \left( (\theta - 1)^{-\alpha} + 2^{-\alpha} \theta^{-\alpha} \sum_{l=2}^{\infty} l(l - 1)^{-\alpha} \right) \\
\leq \xi r^{-\alpha} (\theta - 1)^{-\alpha},
$$

(9)

for some fixed $\xi > 0$. The SIR condition (5) is guaranteed to be satisfied at every intended $j$. 

July 27, 2010 DRAFT
Figure 3. Geometric reasoning underlying the proof of Proposition 15. 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 $\tau_{2,0}, \tau_{2,2}, \tau_{-2,2}, \tau_{-2,0}, \tau_{0,2}, \tau_{-2,-2}$. The signal power from any of 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 $x = \theta r$.

If $\theta$ is chosen such that

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

$$\Rightarrow \theta \geq 1 + (\xi \beta)^{\frac{1}{\alpha}}.$$

For a suitable choice of $\xi$, 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 14 and 15 lead to the following corollary.

**Corollary 16.** If $T^* = T^*(n, r, \beta)$ denotes the length of the shortest TDMA schedule, then a.a.s.:

1) For fixed $\beta$, $T^* = \Theta(nr^d)$.

2) As $\beta \to \infty$, $T^* = \Omega(\beta^{\frac{d}{\alpha}})$.

**Proof:** Claim 1 is evident from the results of Propositions 14 and 15.

For some constants $C_1$ and $C_2$, we have from Propositions 14 and 15, a.a.s. for large $n$ and a
fixed $\beta$,
\[ C_1 nr^d \beta^{d/\alpha} \leq T^* \leq C_2 nr^d \beta^{d/\alpha}. \]

Since $C_1$ (but not $C_2$) is independent of $\beta$, we can write for $n, \beta \to \infty$, $T^* = \Omega(\beta^{d/\alpha})$. 

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 16 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) **Mixing Time:** We now analyze the asymptotic convergence behavior of the distributed averaging algorithm (2) in a dense network at high transmission rate (large $\beta$). To begin with, (3) bounds the number of iterations necessary to reach an $\epsilon-$ball around the equilibrium distribution. However, from Corollary 16, each iteration using the optimal TDMA protocol requires $\Theta(nr^d)$ time slots to establish the desired connectivity. Neglecting the time taken to compute each state update, the $1/n^\gamma-$mixing time if such an algorithm were to be implemented over a dense wireless network scales as
\[ T_{\text{mix}}(W) = \Theta(r^{d-2} n \log n) \] (10)
slots for some fixed $\gamma > 0$. From Proposition 14 and the Gaussian signalling assumption in Section III-C, we can write for $n, \beta \to \infty$, the time to reach the $\epsilon-$ball scales as
\[ \Omega \left( r^{d-2} n \log n \frac{e^{Rd/\alpha}}{R} \right). \]

2) **Choice of Communication Range:** For a fixed $\beta$, the mixing time in (10) scales polynomially in $r$ for $d > 1$. Interestingly, for $d = 1$, the mixing time scales as the inverse of $r$. This suggests that increasing the interaction radius can improve the rate of convergence. For $d = 2$, however, the mixing time 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 mixing time cannot benefit from increasing $r$.

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 $\beta$ from Corollary 16, the mixing time scales as $\Theta(r^{-2} \log n)$, thereby offsetting this loss. In $d$–dimensions, however, while the throughput scales as $\Theta(n^{-1}r^{-d})$, the mixing time only scales as $\Theta(r^{-2} \log n)$. As a result, improving spatial re-use becomes more important than increasing connectivity.

3) Choice of 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 is always 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 characterize the spectral gap of $G^l_n$. Since finding the exact conductance is a known NP-hard problem, we will instead derive a scaling law for the conductance of $G^l_n$, which is sufficient.

**Proposition 17.** The conductance of $G^l_n$ is $\Theta(r^\gamma)$ a.a.s., and is independent of $d$.

**Proof:** We follow a modified version of the proof in [28] From (4) we know that

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

By the symmetry in $G^l_n$ induced by the construction in Section III-B2, it can be shown using arguments similar to [28, 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, let $S = \{x_i \in \Phi : x_i \in [0, 1/2) \times [0, 1]\}$.

Also for the natural random walk, each edge weight is $\frac{1}{d_i} = \Theta\left(\frac{1}{n^2r^2}\right)$ (for $d$ dimensions, $\Theta(n^{-1}r^{-d})$), and the equilibrium distribution is $\Theta\left(\frac{1}{n}\right)$. It is thus sufficient to count the number of edges traversing this cut. The number of short-range edges was shown in [28] to be $\Theta(n^2r^3)$ (for $d$ dimensions $\Theta(n^2r^{d+1})$). Observe that every node in square of side $\eta r$ has $4n\eta^2r^2$ long-range partners. Thus the total number of edges from a square is $4n\eta^2r^2 \times n\eta^2r^2 = \Theta(n^2r^4)$. Since each edge is at least $s/2 - 2\eta r = \Theta(r^\gamma)$ long, each row of $\Theta(r^{\gamma-1})$ squares (see Fig. 2) will contribute to edges traversing the cut. There are $\Theta(r^{-1})$ such rows. As a result, the total number of long-range edges traversing will be $\Theta(n^2r^{4} \times r^{\gamma-1} \times r^{-1}) = \Theta(n^2r^{2+\gamma})$ edges (for general $d,$
Figure 4. The geometry behind the proof of Proposition 17 for $d = 2$. The tiling used for the construction of $G^1_n$ is overlaid. By the symmetry induced by the construction, the set $S \subseteq V_n$ 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 $G^s_n$ (i.e., the disk graph whose edge length is $O(r)$) only nodes from a finite number of squares from the tiling in either direction from the cut contribute to these edges. For long-range edges of length $O(r^\gamma)$, a positive fraction of the nodes from $O(r^\gamma/r)$ 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.

$$
\Theta(n^2r^{2d} \times r^{\gamma-1} \times r^{-d+1}) = \Theta(n^2r^{d+\gamma}).
$$

Thus for $d$ dimensions we have

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

since $\gamma < 1$.

Notice that if $\rho_n = O(nr^2)$ nodes are each allowed to choose $\rho_n$ long-range partners, the number of long-range edges would have contributed $O(r^\gamma)$ towards conductance, thus incurring a possible spectral gap penalty without any significant interference benefit due to the broadcast nature of the wireless medium. In particular, if a node is allowed to have at most one long-range partner, this contribution falls to $O(\frac{1}{nr^{d-\gamma}})$, again without significant interference-reducing benefits.

We can infer the following from the above result:

**Corollary 18.** The spectral gap of $G^1_n$ is $\Omega(r^{2\gamma})$ and $O(r^\gamma)$.
Proof: Using the lower bound in Theorem 13, we have $1 - \mu_2 = \Omega(r^{2\gamma})$. The upper bound can be similarly proved.

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 $G^l_n(r, s, d)$ is also present in the disk graph $G^n_{\pi}(s/2)$, i.e., $G^l_n(r, s, d) \subseteq G^n_{\pi}(s/2)$. Hence a reversible random walk on $G^l_n$ with a uniform equilibrium distribution can mix no faster than the fastest mixing such random walk on $G^n_{\pi}(s/2)$. This key observation allows us to use a known result from [2]:

**Theorem 19.** The spectral gap of the fastest mixing reversible random walk on $G^n_{\pi}(r^\gamma)$ with a uniform equilibrium distribution is $\Theta(r^{2\gamma})$ a.a.s., and independent of $d$.

Since mixing time decreases with spectral gap, from Theorem 19 we conclude that the spectral gap of $G^l_n$ is $\mathcal{O}(r^{2\gamma})$. But we know from Corollary 18 that this gap is also $\Omega(r^{2\gamma})$. Thus we conclude that the spectral gap of $G^l_n$ is $\Theta(r^{2\gamma})$, which is formally stated as a theorem:

**Theorem 20.** The spectral gap of the natural random walk on $G^l_n$ is $\Theta(r^{2\gamma})$.

This result suggests that the improvement in spectral gap from an increased communication radius from $r$ to $r^\gamma$ can also be achieved (in the scaling sense) by allowing each node to communicate with a selected number of nodes at a distance $\Theta(r^\gamma)$. Indeed, the ratio of the number of neighbors compares as $\Theta(r^{d1 - \gamma})$. Thus each node need not directly communicate with all its neighbors within radius $\Theta(r^\gamma)$ for faster mixing.

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) is proportional to the volume of the largest exclusion zone created in the network. Since the longest link distance in both the disk graph $G^n_{\pi}(s/2)$ and $G^l_n$ 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 TDMA schedule that constructs $G^l_n$. 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 21.** A feasible schedule for $G^l_n$ has $\Omega(n r^\gamma d)$ time slots a.a.s.

*Proof:* Any protocol that constructs $G^l_n$ must form at least one link of distance at least $s/2 - 22r = \Theta(r^\gamma)$. Thus for a given $\beta$, the protocol must thus create an exclusion zone of radius $\Theta(r^\gamma)$ in the network at least once. Since all nodes within this exclusion zone must transmit at least once and $r = \omega(r_c)$, we can conclude that any such TDMA protocol will have at least $\Theta(n r^\gamma d)$ slots. Thus the length of a feasible schedule scales as $\Omega(n r^\gamma d)$. ■

**Proposition 22.** The length of the shortest feasible schedule for $G^l_n$ is $O(n r^\gamma d)$.

*Proof:* Consider any TDMA protocol that allows each node to communicate with every node within a distance $s$. Clearly this protocol will also construct $G^l_n$ and is hence feasible. As in Proposition 15, 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 15, it is clear that the spatial re-use can be adjusted to construct the graph in $O(ns^d) = O(n r^\gamma d)$ slots since $s = \Theta(r^\gamma)$. ■

**Corollary 23.** The shortest feasible schedule for $G^l_n$ has $T^* = \Theta(n r^\gamma d)$ slots a.a.s.

*Proof:* Follows from Propositions 21 and 22. ■

Allowing $s = \Theta(r^\gamma)$ and using these results along with the mixing time from Theorem 20 we obtain the scaling laws for the mixing time in the presence of interference:

**Proposition 24.** As $n \to \infty$, for $\epsilon = 1/n^\delta$ and with the shortest feasible TDMA schedule, the mixing time of (2) on a sequence of random graphs $(G^l_n)$ on a $d$–dimensional torus scales as

$$T_{\text{mix}} = \Theta(n r^{(d-2)\gamma} \log n) \quad \text{a.a.s.}$$

We thus reach similar conclusions as in Section IV-C.

**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 greater network connectivity depends crucially on the network dimension. One of our key results is that while the increase in spectral gap from greater connectivity is dimension-independent, the possible spatial re-use from greater connectivity is.

As network connectivity is increased in one-dimensional networks, the larger spectral gap more than compensates for decreased spatial re-use. These two effects exactly cancel each other in two-dimensional networks, while for three- (and higher-) dimensional networks, forming long-range links actually slows down convergence. 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.

These results greatly differ from many optimistic results obtained by analyzing the consensus problem in an abstract graph-theoretic setting, as opposed to a more real-world wireless setting considered here. Hence these results underline the need to accurately account for the cost of interference in designing fast-converging topologies for the average consensus algorithm, (e.g., small-world graphs) or in general, for distributed signal processing problems.

REFERENCES


