A Proof of Corollary 1

In order to prove the convergence rates, we relate our choice of P to a well-known quantity in spectral graph theory, specifically, the graph Laplacian [3]. Recall the definition of A as the symmetric adjacency matrix of the undirected graph G, the degree of node i as $\delta_i = |N(i)| = \sum_{j=1}^n A_{ij}$, and the diagonal matrix $D = \mathrm{diag}\{\delta_1, \ldots, \delta_n\}$. We assume that the graph is connected, so that $\delta_i \geq 1$ for all i and D is invertible. The *(normalized) graph Laplacian* is given by $\mathcal{L}(G) = I - D^{-1/2} A D^{-1/2}$, a matrix which is always symmetric, positive semidefinite, and satisfies $\mathcal{L}\mathbb{1} = 0$. Therefore, when the graph is degree-regular ($\delta_i = \delta$ for all $i \in V$), the standard random walk with self loops on G given by the matrix $P := I - \frac{\delta}{\delta+1}\mathcal{L}$ is doubly stochastic and is valid for our theory. For non-degree regular graphs, we make the minor modification as in (5) to obtain

$$P_n(G) := I - \frac{1}{\delta_{\max}+1}(D - A) = I - \frac{1}{\delta_{\max}+1} D^{1/2}\mathcal{L}D^{1/2}. \tag{16}$$

As noted earlier, $P_n(G)$ is doubly stochastic, and if the graph is δ-regular, then $P_n(G)$ is simply $(I + A)/(\delta + 1)$. The remainder of the corollary is based on bounding the spectral gap of $P_n(G)$.

Lemma 4. *The matrix P satisfies $\sigma_2(P_n(G)) \leq \max\left\{1 - \frac{\min_i \delta_i}{\delta_{\max}+1}\lambda_{n-1}(\mathcal{L}), \frac{\delta_{\max}}{\delta_{\max}+1}\lambda_1(\mathcal{L}) - 1\right\}$.*

Proof By a theorem of Ostrowski on congruent matrices (cf. Theorem 4.5.9, [10]), we have

$$\lambda_k(D^{1/2}\mathcal{L}D^{1/2}) \in \left[\min_i \delta_i \lambda_k(\mathcal{L}), \max_i \delta_i \lambda_k(\mathcal{L})\right]. \tag{17}$$

Since $\mathcal{L}\mathbb{1} = 0$, we have $\lambda_n(\mathcal{L}) = 0$, and so it suffices to focus on $\lambda_1(D^{1/2}\mathcal{L}D^{1/2})$ and $\lambda_{n-1}(D^{1/2}\mathcal{L}D^{1/2})$. From the definition (16), the eigenvalues of P are of the form $1 - (\delta_{\max} + 1)^{-1}\lambda_k(D^{1/2}\mathcal{L}D^{1/2})$. The bound (17) coupled with the fact that all the eigenvalues of $\mathcal{L}$ are non-negative implies that $\sigma_2(P) = \max_{k<n}\left\{\left|1 - (\delta_{\max}+1)^{-1}\lambda_k(D^{1/2}\mathcal{L}D^{1/2})\right|\right\}$ is upper bounded by the larger of $1 - \frac{\delta_{\min}}{\delta_{\max}+1}\lambda_{n-1}(\mathcal{L})$ and $\frac{\delta_{\max}}{\delta_{\max}+1}\lambda_1(\mathcal{L}) - 1$. □

Computing the upper bound in Lemma 4 requires controlling $\lambda_{n-1}(\mathcal{L})$ and $\lambda_1(\mathcal{L})$. To circumvent this complication, we use the well-known idea of a lazy random walk [3, 14], in which we replace by P by $\frac{1}{2}(I + P)$. The resulting symmetric matrix has the same eigenstructure as P, moreover,

$$\sigma_2\Big(\frac{1}{2}(I+P)\Big) = \lambda_2\Big(I - \frac{1}{2(\delta_{\max}+1)}D^{1/2}\mathcal{L}D^{1/2}\Big) \leq 1 - \frac{\delta_{\min}}{2(\delta_{\max}+1)}\lambda_{n-1}(\mathcal{L}). \tag{18}$$

Consequently, it is sufficient to bound only $\lambda_{n-1}(\mathcal{L})$. The convergence rate implied by the lazy random walk through Theorem 2 is no worse than twice that of the original walk, which is insignificant for the analysis in this paper. The remainder of the proof involves exploiting results from spectral graph theory [3] in order to control the eigenvalues of the Laplacian.

Cycles and paths: Recall the regular k-connected cycle from Figure 1(a), constructed by placing the n nodes on a circle and connecting every node to k neighbors on the right and left. For this graph, the Laplacian $\mathcal{L}$ is a circulant matrix with diagonal entries 1 and off-diagonal non-zero entries $-1/2k$. Using known results on circulant matrices [2, 6], we find that $\lambda_{n-1}(\mathcal{L}) = \Theta\big(\frac{k^2}{n^2}\big)$ for $k = o(n)$. For the regular k-connected path, by computing Cheeger constants [3, Chapter 2] we find that if $k \leq \sqrt{n}$, then $\lambda_{n-1}(\mathcal{L}) = \Theta(k^2/n^2)$. Note also that for the k-connected path on n nodes, $\min_i \delta_i = k$ and $\delta_{\max} = 2k$. Thus, we can combine the above with Lemma 4 to see that for regular k-connected paths or cycles with $k \leq \sqrt{n}$, $\sigma_2(P) = 1 - \Theta\big(\frac{k^2}{n^2}\big)$. Substituting the bound into Theorem 2 yields the claim of Corollary 1(a).

Regular grids: Now consider the case of a $\sqrt{n}$-by-$\sqrt{n}$ grid, focusing in particular on regular k-connected grids, in which any node is joined to every node that is fewer than k horizontal or vertical edges away in an axis-aligned direction. In this case, we use results on Cartesian products of graphs [3, Section 2.6] to analyze the eigen-structure of the Laplacian. In particular, the $\sqrt{n}$-by-$\sqrt{n}$ k-connected grid is simply the Cartesian product of two regular k-connected paths of $\sqrt{n}$ nodes. The second smallest eigenvalue of a Cartesian product of graphs is half the minimum of second-smallest eigenvalues of the original graphs [3, Theorem 2.13]. Thus, based on the preceding discussion of k-connected cycles, we conclude that if $k \leq n^{1/4}$, then we have $\lambda_{n-1}(\mathcal{L}) = \Theta(k^2/n)$, and we use Lemma 4 and (18) to see that the result in Corollary 1(b) immediately follows.

Random geometric graphs: Using the proof of Lemma 10 from Boyd et al. [2], we see that for any $\epsilon > 0$, if $r = \sqrt{\log^{1+\epsilon} n/(n\pi)}$, then with probability at least $1 - 2/n^{c-1}$,

$$\min_i \delta_i \geq \log^{1+\epsilon} n - \sqrt{2}c \log n \quad \text{and} \quad \max_i \delta_i \leq \log^{1+\epsilon} n + \sqrt{2}c \log n. \tag{19}$$

Thus, letting $\mathcal{L}$ be the graph Laplacian of a random geometric graph, if we can bound $\lambda_{n-1}(\mathcal{L})$, (19) coupled with Lemma 4 will control the convergence rate of our algorithm.

Von Luxburg et al. [23] give concentration results on the second-smallest eigenvalue of a geometric graph. In particular, their Theorem 3 says that if $r = \omega(\sqrt{\log n/n})$, then with exceedingly high probability, $\lambda_{n-1}(\mathcal{L}) = \Omega(r)$. Using (19), we see that for $r = (\log^{1+\epsilon} n/n)^{1/2}$, the ratio $\frac{\min_i \delta_i}{\max_i \delta_i} = \Theta(1)$ and $\lambda_{n-1}(\mathcal{L}) = \Omega\big(\frac{\log^{1+\epsilon} n}{n}\big)$ with high probability. Combining this with Lemma 4 and (18), we have $\sigma_2(P) = 1 - \Omega\big(\frac{\log^{1+\epsilon} n}{n}\big)$, the desired result of Corollary 1(c).

Expanders: The constant spectral gap in expanders [3, Chapter 6] removes any penalty due to network communication (up to logarithmic factors) and hence yields Corollary 1(d).

B Proof of Proposition 1

The proof is based on construction of a set of objective functions f_i that force convergence to be slow by using the second eigenvector of the communication matrix P. Recall that $\mathbb{1} \in \mathbb{R}^n$ is the eigenvector of P corresponding to its largest eigenvalue (equal to 1). Let $v \in \mathbb{R}^n$ be the eigenvector of P corresponding to its second singular value, $\sigma_2(P)$. By using the lazy random walk (see Sec. A and (18)), we may assume without loss of generality that $\lambda_2(P) = \sigma_2(P)$. Let $w = \frac{v}{\|v\|_\infty}$ be a normalized version of the second eigenvector of P, and note that $\sum_{i=1}^n w_i = 0$. Without loss of generality, we assume that there is an index i for which $w_i = -1$ (otherwise we can flip signs in what follows); by re-indexing as needed, we can assume that $w_1 = -1$. We set $\mathcal{X} = [-1, 1] \subset \mathbb{R}$ and define the univariate functions $f_i(x) := (c + w_i)x$. The global problem is then to minimize

$$\frac{1}{n} \sum_{i=1}^n f_i(x) = \frac{1}{n} \sum_{i=1}^n (c + w_i)x = cx$$

for some constant $c > 0$ to be chosen. Note that each f_i is $c + 1$-Lipschitz. By construction, we see immediately that $x^* = -1$ is optimal for the global problem.

Consider the evolution of the $\{z(t)\}_{t=0}^\infty \subset \mathbb{R}^n$, as generated by the update (4). By construction, we have $g_i(t) = c + w_i$ for all t. Defining the vector $g = (c\mathbb{1} + w) \in \mathbb{R}^n$, we recall that $P\mathbb{1} = \mathbb{1}$ to see

$$\begin{aligned} z(t+1) = Pz(t) - g = P^2 z(t-1) - Pg - g = \cdots = -\sum_{\tau=0}^{t} P^\tau g \\ = -\sum_{\tau=0}^{t-1} P^\tau (w + c\mathbb{1}) = -\sum_{\tau=0}^{t-1} P^\tau w - ct\mathbb{1} = -\sum_{\tau=0}^{t-1} \sigma_2(P)^\tau w - ct\mathbb{1}. \end{aligned} \tag{20}$$

In order to establish a lower bound, it suffices to show that at least one node is far from the optimum after t steps, and we focus on node 1. Since $w_1 = -1$, the evolution (20) guarantees that

$$z_1(t+1) = \sum_{\tau=0}^{t-1} \sigma_2(P)^\tau - ct = \frac{1 - \sigma_2(P)^{t-1}}{1 - \sigma_2(P)} - ct. \tag{21}$$

Recalling that $\psi(x) = \frac{1}{2}x^2$ for this scalar setting, we have

$$x_i(t+1) = \operatorname*{argmin}_{x \in \mathcal{X}} \Big\{ -z_i(t+1)x + \frac{1}{2\alpha(t)} x^2 \Big\} = \operatorname*{argmin}_{x \in \mathcal{X}} \Big\{ (x - \alpha(t) z_i(t+1))^2 \Big\}$$

Hence $x_1(t)$ is the projection of $\alpha(t) z_1(t+1)$ onto $[-1, 1]$, and unless $z_1(t) < 0$ we have

$$f(x_1(t)) - f(-1) \geq c > 0.$$

If t is overly small, the relation (21) will guarantee that $z_1(t) \geq 0$, so that $x_1(t)$ is far from the optimum. If we choose $c \leq 1/3$, then a little calculation with (21) shows that we require $t = \Omega((1 - \sigma_2(P))^{-1})$ in order to drive $z_1(t)$ below zero.