

# Fast Convergence for Consensus in Dynamic Networks

T-H. Hubert Chan and Li Ning

The University of Hong Kong

**Abstract** We study the convergence time required to achieve consensus in dynamic networks. In each time step, a node's value is updated to some weighted average of its neighbors' and its old values. We study the case when the underlying network is dynamic, and investigate different averaging models. Both our analysis and experiments show that dynamic networks exhibit fast convergence behavior, even under very mild connectivity assumptions.

## 1 Introduction

Natural group behavior is exhibited in many dynamic systems. Typically, each individual or node in the set  $V$  has some number in  $\mathbb{R}$ , which can represent one's opinion. In every time step, an individual observes the opinions of a subset of other individuals and updates one's opinions accordingly. It is observed that in many such systems [17, 18, 6], the values of all nodes converge to the same value (or opinions of individuals reach consensus) after a small number of iterations, even though each node only interacts with a small number of other nodes in each time step.

The *weighted averaging model* [7], used by DeGroot to model consensus of opinions, has been widely studied to explain convergent behavior in such networks. The value  $v_t[i]$  of an individual at time step  $t$  is updated by taking some weighted average of all individuals' values:  $v_{t+1}[i] := \sum_j p_t[i, j] \cdot v_t[j]$ , where each  $p_t[i, j]$  is non-negative and  $\sum_j p_t[i, j] = 1$ . Typically, for each  $i$ , there is only a small number of  $j$ 's such that  $p_t[i, j]$  is non-zero; those correspond to the individuals whose values can be observed by  $i$ . The interactions of individuals in a time step can be represented by a network  $G_t = (V, E_t)$ , where an edge  $\{i, j\} \in E_t$  means the individuals can observe each other's values at time  $t$ . Besides its simplicity, the weighted averaging model has applications in parallel computation [1], control theory [9, 3, 2, 8, 5, 12] and ad hoc networks [11].

In this paper, we study what weighting strategies and what kind of networks can enable fast convergence to achieve consensus. In particular, for different weighting strategies and network properties, we analyze the number of time steps that is sufficient for all nodes' values to be close to one another. The *uniform averaging model* is the case when given a network  $G_t$ , an individual updates its value to the average of its neighbors' and its old values. We consider the case when the underlying network topology is dynamic, i.e. the networks  $G_t$ 's change

over time. To keep our analysis as general as possible, we do not specify how the networks  $G_t$ 's evolve (which may or may not depend on the nodes' values); we only assume general structural properties of the networks such as degree distribution and connectivity.

**Related Work.** The special case of the uniform averaging model with time-invariant network topology is well-understood [7]. Using the theory of stochastic matrices and spectral graph theory, it is known that the convergence time is related to the eigenvalue gap [10] of the transition matrix  $P$  involved. If the underlying network is time-invariant and connected, Olshevsky and Tsitsiklis [16] showed that the convergence time for the uniform averaging model is  $O(n^3)$ .

Relatively little is known about the convergence time when the underlying network is dynamic. Assuming some special structure in the network in each time step, Cao et. al [3] showed that the convergence time is  $n^{O(n)}$ . Olshevsky and Tsitsiklis [16] also considered weak connectivity assumptions: in the given sequence, the union of any  $k$  consecutive networks is connected. In this case, they showed that the convergence time under the uniform averaging model is  $O(kn^{kn})$  and a lower bound of  $\Omega(n)^k$ . Using a "load balancing" algorithm, they can achieve  $O(n^3)$  convergence time. They also showed a convergence time of  $O(n^3)$  for the uniform averaging model with the fixed degree assumption [15]. With the same weak connectivity assumption, Nedić et. al [14] showed convergence time of  $O(\frac{kn^2}{\alpha})$ , for the special case where the transition matrices are doubly stochastic and  $\alpha > 0$  is a lower bound on the non-zero entries.

Vicsek et. al [19] used the weighted averaging model to study interaction between particles, which influence one another's velocities. Two particles can influence each other if their distance is close enough. The system reaches a convergent state when all particles are traveling in nearly the same direction. Jadbabaie et. al [9] gave a theoretical explanation to such convergent behavior. Recently, Chazelle [4] considered a discrete version of the model and showed that the convergence time is  $O(2 \uparrow n)$ .

Other interaction models have also been studied. In [2], directed networks and asynchronous updates were considered. In [8, 5, 12], convergence under non-linear update rules were studied.

**Our Contribution and Results.** In this paper, we give a quantitative analysis between the convergence time and the connectivity of the networks in the given sequence. If convergent behavior is observed at all in real systems, then the number of time steps taken certainly cannot be  $O(n^n)$  or even  $O(2 \uparrow n)$ . Typically, the networks concerned are well-connected and convergence time of  $O(\log n)$  is observed.

For static network, this can be easily explained by the theory of stochastic matrices and spectral graph theory. The update process in each time step corresponds to multiplication by a stochastic matrix  $P$ . Although  $P$  is in general not symmetric (and hence the eigenvectors are not mutually orthogonal), for all positive integers  $t$ , the powers  $P^t$  all have the same eigenvectors, and any eigenvalue gap in  $P$  will be magnified in  $P^t$ . However, if the underlying network is dynamic, then the corresponding transition matrices will not have the same eigenvectors

anymore (apart from the all one's vector), and hence the above argument does not work.

In Section 3, we overcome this technical hurdle by choosing the weights carefully such that in the transformed space the eigenvectors are mutually orthogonal. Assuming that each node has limited degree variation in the given network sequence, and each network is well-connected (as measured by conductance), we can obtain an eigenvalue gap in the transition matrix in each time step. Combining these techniques, we show that convergence time is  $O(\log n)$ . If we just assume that each network is well-connected (without the assumption on limited degree variation), we have  $O(n)$  convergence time. As far as we know, the previous best known convergence time under any weighted averaging model for dynamic connected networks is  $O(n^3)$  [14].

Under the uniform averaging model, we analyze in Section 4 the conditions on the given network sequence such that fast convergence can be obtained. Assuming that each network is degree bounded and for some integer  $k$ , the union of every  $k$  consecutive networks is a vertex expander, we show that the convergence time is polynomial by using the expansion property directly. Furthermore, our techniques can be extended to the probabilistic case where the connectivity condition for each union of  $k$  networks only needs to hold with some positive probability.

On the other hand, our simulations in Section 5 show that for well-connected graphs such as  $G_{n,p}$ , the convergence time under the uniform averaging model grows logarithmically with the network size, suggesting that there is a lot of room for improvement. It would be an interesting open problem to determine the most general conditions on the networks under which the uniform averaging model has fast convergence time.

## 2 Preliminary

Suppose there is a set  $V$  of  $n$  individuals and each one of them holds an opinion which can be represented by a number from  $\mathbb{R}$ . An *opinion configuration* at some time  $t$  is an  $n$ -dimensional vector from  $\mathbb{R}^n$ . We denote the configuration at time  $t$  by  $v_t$ , and the opinion of individual  $i$  by  $v_t[i]$ .

At each time step, the individuals form a network (in this document we use the terms “network” or “undirected graph” interchangeably)  $G_t = (V(G_t), E(G_t))$ , in which the nodes represent the individuals and an edge between two nodes means they can potentially communicate their opinions to each other. We assume all  $G_t$ 's have the same set of nodes, i.e.  $V(G_t) = V$  for all  $t$ . Moreover, we assume that the sequence  $\{G_t\}$  of networks is generated by some process that is, in general, independent of the individuals' opinions  $v_t$ 's.

We use the maximum difference between two individual's numbers to measure how close a configuration reaches consensus.

**Definition 1 ( $\tau$ -Measure).** *Given a configuration vector  $v \in \mathbb{R}^n$ , the  $\tau$ -measure of  $v$  is  $\tau(v) = \max_{i,j} |v[i] - v[j]|$ .*

We say that the vector  $v$  achieves consensus when  $\tau(v) = 0$ ; and for  $\epsilon > 0$ , the vector  $v$  achieves  $\epsilon$ -consensus when  $\tau(v) \leq \epsilon$ .

We next describe the models we use to analyze the convergent behavior for dynamic systems.

## 2.1 Convergence Model for Dynamic Networks

Given a sequence  $\{G_t : t \geq 0\}$  of networks and some initial configuration  $v_0 \in \mathbb{R}^n$ , we describe the update rule for each time step. At time step  $t$ , the nodes are connected by the network  $G_t$ ; we denote the degree of node  $i$  by  $d_t[i]$ . Moreover, each node  $i$  has some positive integral weight  $w_t[i] \geq d_t[i] + 1$ , which indicates how resistant the individual is to others' opinions, with a higher weight indicating higher resistance. The update rule for each node  $i$  at time  $t$  is given by the following equation.

$$v_{t+1}[i] = \left(1 - \frac{d_t[i]}{w_t[i]}\right) \cdot v_t[i] + \frac{1}{w_t[i]} \sum_{j:\{i,j\} \in E(G_t)} v_t[j]. \quad (1)$$

**Matrix Notation.** The update rule can be expressed succinctly using matrix notation. We treat  $w_t$ ,  $v_t$ , and  $d_t$  as  $n \times 1$  column vectors. Given a network  $G_t$ , recall its Laplacian  $L_t$  is defined as the  $n \times n$  matrix such that  $L_t[i, j]$  is  $d_t[i]$  when  $i = j$ ,  $-1$  when  $\{i, j\} \in E(G_t)$ , and  $0$  otherwise.

Given a square matrix  $A$ , we denote its trace by  $\text{tr}(A)$ , i.e., the sum of its diagonal entries. Given a vector  $w \in \mathbb{R}^n$ , we use  $\text{Diag}(w)$  to denote the diagonal matrix such that  $\text{Diag}(w)[i, i] = w[i]$ . Given a weight vector  $w_t \in \mathbb{R}^n$ , let  $W_t = \text{Diag}(w_t)$  and the transition matrix  $P_t = I_n - W_t^{-1}L_t$ , where  $I_n$  is the  $n \times n$  identity matrix. Then, equation (1) can be rewritten as:

$$v_{t+1} = P_t v_t. \quad (2)$$

**Special Cases.** We describe some special cases for the weights  $w_t$ .

- **Static Weight Model.** In this case, there is some fixed weight vector  $w \in \mathbb{R}^n$  such that for all time steps  $t$ ,  $w_t = w$ . Observe that in this case, we need to restrict the networks such that for all  $t$  and all nodes  $i$ , the degree  $d_t[i] \leq w[i] - 1$ . To ensure that each node is still influenced by its neighbors, we normally also assume  $w[i] = O(n)$  for all  $t$ .
- **Uniform Averaging Model.** In this case, for each time step  $t$  and each node  $i$ ,  $w_t[i] = d_t[i] + 1$ . Observe that in this case, the new opinion of a node is simply the average of the sum of its and its neighbors' opinions. Hence, equation (1) reduces to  $v_{t+1}[i] = \frac{1}{d_t[i]+1}(v_t[i] + \sum_{j:\{i,j\} \in E(G_t)} v_t[j])$ .

**Convergence Time.** Given some initial configuration  $v_0 \in \mathbb{R}^n$ , and some convergence process, for  $\epsilon > 0$ , the *convergence time* to achieve  $\epsilon$ -consensus is the minimum  $T$  such that for all  $t \geq T$ ,  $\tau(v_t) \leq \epsilon$ .

**Union Network.** We do not always require each network  $G_t$  to be connected. We can still prove convergence results as long as the union of the networks over a certain period of time is well-connected. Formally, suppose  $I$  is a set of time

indices for the collection of networks  $\{G_t = (V, E_t) : t \in I\}$ . Then, the union network is defined as  $\cup_{t \in I} G_t := (V, \cup_{t \in I} E_t)$ .

## 2.2 Stochastic Matrices

We mention some useful results about *stochastic matrices*. Recall that an  $n \times n$  matrix  $M$  is *row stochastic* (or simply *stochastic*) if all its entries are non-negative and the entries of each row sum to 1. Observe that the transition matrix  $P_t$  in (2) is stochastic. Recall that the product of two stochastic matrices is still stochastic. We define two measures for matrices, which describe how different the rows of a matrix are.

**Definition 2 ( $\tau_1$ - and  $\tau_2$ -Measures).** *Given a matrix  $P$ , the  $\tau_1$ -measure of  $P$  is defined as  $\tau_1(P) = \frac{1}{2} \max_{i,j} \{ \sum_k |P[i,k] - P[j,k]| \}$ ; the  $\tau_2$ -measure of  $P$  is defined as  $\tau_2(P) = \max_{i,j} \{ (\sum_k |P[i,k] - P[j,k]|^2)^{\frac{1}{2}} \}$ .*

Observe that for a column vector  $v$ ,  $\tau(v) = 2\tau_1(v) = \tau_2(v)$ .

Fact 1 states an important relationship between the  $\tau_2$ -measure of the product of two matrices and product of the measures of the corresponding matrices. Its proof is given in [4, 13]. Fact 2 relates the  $\tau_1$ -measure of a stochastic matrix with its smallest entry.

**Fact 1** *For any stochastic matrix  $A$  and any matrix  $B$ , whose dimensions are compatible with  $A$  such that  $AB$  is well-defined, we have  $\tau_2(AB) \leq \tau_1(A)\tau_2(B)$ .*

Observe that any stochastic matrix  $P$  has the property that  $\tau_1(P) \leq 1$ . Hence, it follows that for all  $t$ ,  $\tau(v_{t+1}) = \tau_2(P_t v_t) \leq \tau(v_t)$ .

**Fact 2** *Suppose  $P$  is a stochastic matrix such that all its entries are at least some number  $\alpha > 0$ . Then,  $\tau_1(P) \leq 1 - \alpha$ .*

## 3 Static Weight Model: Limited Degree Variation and Well-connected Dynamic Networks

In this section, we show that fast convergence for the static weight model is achieved if in the given sequence  $\{G_t\}$  of networks, each  $G_t$  is *well-connected*, and for each node  $i$ , the degree  $d_t[i]$  does not vary too much with respect to  $t$ . In particular, we explore a quantitative relationship between the convergence time and the connectivity of the given networks.

The concept *conductance* can be used to measure how connected a graph is.

**Definition 3 (Conductance).** *Given a network  $G = (V, E)$ , and a subset  $S \in V$ , the edge border set of  $S$  is defined as  $\partial(S) = \{\{u, v\} \in E | u \in S, v \in \bar{S} = V \setminus S\}$ . The conductance of  $G$  is defined as  $\Psi(G) = \sup\{\varphi > 0 | \frac{\partial(S)}{\min\{d(S), d(\bar{S})\}} \geq \varphi, \forall S \subset V\}$ , where  $d(S) = \sum_{i \in S} d[i]$ .*

Given any stochastic matrix  $P$ , we sort and label its eigenvalues in the descending order of the eigenvalues' magnitude, i.e.,  $|\lambda_0(P)| \geq |\lambda_1(P)| \geq \dots \geq |\lambda_{n-1}(P)|$ , where  $\lambda_0(P) = 1$ . The following lemma, which is an extension of the Cheeger's Inequality, relates the spectral properties of a transition matrix  $P$  and the conductance of the underlying network. We defer its proof to the full version.

**Lemma 1 (Conductance Implies Eigenvalue Gap).** *Suppose  $G$  is a network with conductance  $\Psi$ , and  $w \in \mathbb{Z}^n$  is a positive weight vector such that for each  $i$ , its degree  $d[i]$  satisfies  $2d[i] \leq w[i] \leq K \cdot d[i]$ . Define the transition matrix  $P := I_n - W^{-1}L$ , where  $W = \text{Diag}(w)$  and  $L$  is the Laplacian of  $G$ . Then,  $|\lambda_1(P)| \leq 1 - \eta$ , where  $\eta = \frac{\Psi^2}{2K}$ .*

**Theorem 1 (Static Weight Model).** *Given a positive weight vector  $w \in \mathbb{Z}^n$ , and a sequence  $\{G_t : t \geq 0\}$  of networks, let the transition matrix  $P_t := I_n - W^{-1}L_t$ , where  $W = \text{Diag}(w)$  and  $L_t$  is the Laplacian of  $G_t$ . Suppose there is some  $0 < \eta < 1$  such that for all  $t$ ,  $|\lambda_1(P_t)| \leq 1 - \eta$ . Then, for any initial configuration vector  $v_0$ , the convergence time to achieve  $\epsilon$ -consensus is  $O(\frac{1}{\eta} \log \frac{\|W^{\frac{1}{2}}v_0\|_2}{\epsilon})$ , which is  $O(\frac{1}{\eta} \log \frac{n\|v_0\|_2}{\epsilon})$ , if for each  $i$ ,  $w[i] = O(n)$ . For the special case when all nodes  $i$  have the same  $w[i]$ , the convergence time can be improved to  $O(\frac{1}{\eta} \log \frac{\|v_0\|_2}{\epsilon})$ .*

*Proof.* Let  $\lambda := 1 - \eta$ . Suppose  $P := I_n - W^{-1}L$  is the transition matrix corresponding to some network with Laplacian  $L$  such that  $|\lambda_1(P)| \leq \lambda$ . Consider  $M := W^{\frac{1}{2}}PW^{-\frac{1}{2}}$  and observe that  $M$  is symmetric and has exactly the same eigenvalues as  $P$ . In particular,  $M$  has eigenvalue 1 with the corresponding eigenvector  $u_0 = \text{tr}(W)^{-\frac{1}{2}}W^{\frac{1}{2}}\mathbf{1}$ , where  $\mathbf{1}$  is the all one's vector. Moreover, since the eigenvectors of a symmetric matrix are mutually orthogonal, we have for each vector  $z$  that is orthogonal to  $u_0$ , the vector  $Mz$  is still orthogonal to  $u_0$  and  $\|Mz\|_2 \leq \lambda\|z\|_2$ . For each  $t$ , we define  $M_t := W^{\frac{1}{2}}P_tW^{-\frac{1}{2}}$ .

Given an initial configuration vector  $v_0$ , we write  $W^{\frac{1}{2}}v_0 = x + y$ , where  $x$  is parallel to  $u_0$  and  $y$  is orthogonal to  $u_0$ . According to our convergence model, we have

$$v_t = P_{t-1} \cdots P_0 v_0 = W^{-\frac{1}{2}} M_{t-1} \cdots M_0 W^{\frac{1}{2}} v_0 = W^{-\frac{1}{2}} x + W^{-\frac{1}{2}} M_{t-1} \cdots M_0 y.$$

We next observe that all entries of  $W^{-\frac{1}{2}}x$  are identical, and hence  $\tau(v_t) = \tau(W^{-\frac{1}{2}}M_{t-1} \cdots M_0 y)$ , which is at most  $2\|W^{-\frac{1}{2}}M_{t-1} \cdots M_0 y\|_2$ , because for any vector  $v$ ,  $\tau(v) \leq 2\|v\|_2$ .

Observing that  $W^{-\frac{1}{2}}$  is a diagonal matrix such that each entry is at most 1, we have  $\|W^{-\frac{1}{2}}M_{t-1} \cdots M_0 y\|_2 \leq \|M_{t-1} \cdots M_0 y\|_2 \leq \lambda^t \|y\|_2 \leq \lambda^t \|W^{\frac{1}{2}}v_0\|_2$ , where the last inequality holds because  $x$  and  $y$  are orthogonal, and the penultimate inequality holds because for each  $0 \leq k \leq t$ , the vector  $M_{k-1} \cdots M_0 y$  remains orthogonal to  $u_0$ , and hence is spanned by eigenvectors (of each  $M_i$ ) whose eigenvalues have absolute values at most  $\lambda$ .

Hence, we have  $\tau(v_t) \leq 2\lambda^t \|W^{\frac{1}{2}}v_0\|_2$ , which is at most  $\epsilon$ , for  $t = \Omega(\frac{1}{\eta} \log \frac{\|W^{\frac{1}{2}}v_0\|_2}{\epsilon})$ , where  $\eta = 1 - \lambda$ .

Finally, observing that for the special case when for all  $i$ ,  $w[i] = \omega$  is the same, we have  $W = \omega I_n$ . Hence, in the above argument the  $W^{\frac{1}{2}}$  and  $W^{-\frac{1}{2}}$  cancel with each other, and we can conclude instead that  $\tau(v_t) \leq 2\lambda^t \|v_0\|_2$ , and so the convergence time becomes  $O(\frac{1}{\eta} \log \frac{\|v_0\|_2}{\epsilon})$ .  $\square$

Hence, from Lemma 1 and Theorem 1, we have the following corollary.

**Corollary 1 (Logarithmic Convergence Time for Limited Degree Variation).** *Given a positive weight vector  $w \in \mathbb{Z}^n$ , and a sequence  $\{G_t : t \geq 0\}$  of networks, let the transition matrix  $P_t := I_n - W^{-1}L_t$ , where  $W = \text{Diag}(w)$  and  $L_t$  is the Laplacian of  $G_t$ . Suppose there is some  $\Psi > 0$  and some  $K > 0$ , such that for all  $t$ ,  $G_t$  has conductance at least  $\Psi$  and for each node  $i$ , its degree  $d_t[i]$  satisfies  $2d_t[i] \leq w[i] \leq Kd_t[i]$ . Then, for any initial configuration vector  $v_0$ , the convergence time to achieve  $\epsilon$ -consensus is  $O(\frac{K}{\Psi^2} \log \frac{n\|v_0\|_2}{\epsilon})$ , if for each  $i$ ,  $w[i] = O(n)$ .*

For the special case when  $w[i] = 2n$  for all nodes  $i$ , we have the following corollary, also using Lemma 1 and the identical weight case in Theorem 1.

**Corollary 2 (Linear Convergence Time).** *Given a sequence  $\{G_t : t \geq 0\}$ , suppose that there is some  $\Psi > 0$  such that for all  $t$ ,  $G_t$  has conductance at least  $\Psi$ . We set the weight vector  $w \in \mathbb{Z}^n$  such that for all  $i$ ,  $w[i] = 2n$ , and consider the transition matrix  $P_t := I_n - W^{-1}L_t$  as before. Then, given  $\epsilon > 0$  and initial configuration vector  $v_0 \in \mathbb{R}^n$ , the convergence time to achieve  $\epsilon$ -consensus is  $O(\frac{n}{\Psi^2} \log \frac{\|v_0\|_2}{\epsilon})$ .*

## 4 Analysis of the Uniform Averaging Model

In this section, we analyze the convergence time for the uniform averaging model. Given a network  $G_t$ , we consider the weight vector  $w_t$  such that  $w_t[i] = d_t[i] + 1$ , the degree of node  $i$  plus 1. The transition matrix is given by  $P_t = I_n - \text{Diag}(w_t)^{-1}L_t$ , where  $L_t$  is the Laplacian of  $G_t$ .

We also assume that each network in the sequence  $\{G_t\}$  has degree bounded by some  $d$ , i.e., for all  $t$  and all  $i$ ,  $d_t[i] \leq d$ . However, we only need weak connectivity assumptions on the given sequence of networks. We do not even require each network to be connected. All we need is that there is some integer  $k$  such that the union of the networks in every  $k$  consecutive time steps is well-connected. Although we only prove convergence time polynomial in  $n$ , experiments in Section 5 suggest that the convergence time for the uniform averaging model is  $O(\log n)$  for well-connected networks.

### 4.1 Weakly Connected Networks

Given a network, the standard notion of vertex expansion can measure its connectivity.

**Definition 4 (Vertex Expansion).** Given a network (undirected graph)  $G = (V, E)$  and a subset  $S \subseteq V$ , the vertex border set of  $S$  is defined as  $\delta(S) = \{v \in V(G) \setminus V(S) \mid \exists u \in S, s.t. \{u, v\} \in E\}$ . The vertex expansion of  $G$  is defined as  $\Phi(G) = \sup\{\phi > 0 \mid \frac{|\delta(S)|}{|S|} \geq \phi, \forall S \subset G, |S| \leq \frac{|V|}{2}\}$ .

**Definition 5 (Union Vertex Expansion).** Given a sequence  $\{G_t : t \geq 0\}$  of networks, and an integer  $k \geq 1$ , we say the sequence has  $k$ -union vertex expansion at least  $\phi$  if for any  $t \geq 0$ ,  $\Phi(\cup_{j=t}^{t+k-1} G_j) \geq \phi$ .

The main result of this section is given in the following theorem, which is a direct consequence of Lemmas 2 and 3

**Theorem 2 (Convergence Time for Union Vertex Expanders).** Suppose the network sequence  $\{G_t\}$  with bounded degree  $d$  has  $k$ -union vertex expansion at least  $\phi > 0$ . Then, given an initial vector  $v_0 \in \mathbb{R}^n$  and  $\epsilon > 0$ , the convergence time to achieve  $\epsilon$ -consensus under the uniform averaging model is  $n^{O(\frac{k}{\phi} \log d)} \log \frac{\tau(v_0)}{\epsilon}$ .

We introduce the idea of hitting diameter of a network sequence, which intuitively measures the number of time steps required for any person's opinion to have some influence over everyone else's.

**Definition 6 ( $\mu$ -Hitting Diameter).** Given a network sequence  $\mathcal{G} = \{G_t : t \geq 0\}$ , let  $P_t$  be the transition matrix associated with  $G_t$  under the uniform averaging model. Let  $0 < \mu < \frac{1}{n}$ . The  $\mu$ -hitting diameter of the sequence, denoted by  $\text{HDiam}_\mu(\mathcal{G})$ , is at most  $T$ , if for every  $t \geq 0$ , every entry of the product  $P_{t+T-1}P_{t+T-2} \cdots P_t$  is at least  $\mu$ .

**Lemma 2 (Hitting Diameter and Convergence Time).** Given a sequence  $\mathcal{G}$  of networks with  $\text{HDiam}_\mu(\mathcal{G}) \leq T$ , an initial configuration  $v_0 \in \mathbb{R}^n$  and  $\epsilon > 0$ , the convergence time to achieve  $\epsilon$ -consensus is  $O(\frac{T}{n\mu} \log \frac{\tau(v_0)}{\epsilon})$ .

*Proof.* By Definition 6 and Fact 2, we have for all  $t \geq 0$ ,  $\tau_1(\prod_{j=t}^{t+T-1} P_j) \leq 1 - n\mu \leq \exp(-n\mu)$ , where we have used the inequality  $1 + x \leq e^x$  for all real  $x$ .

Therefore, by Fact 1 described in the preliminary  $\tau_2(v_t) \leq \exp(-n\mu \lfloor \frac{t}{T} \rfloor) \cdot \tau(v_0)$ , which is at most  $\epsilon$ , when  $t \geq \frac{T}{n\mu} \log \frac{\tau(v_0)}{\epsilon}$ .  $\square$

Next, we show how to use this lemma to derive the convergence time for a specified class of networks.

**Lemma 3 (Hitting Diameter for Union Vertex Expanders).** Suppose a network sequence  $\mathcal{G} = \{G_t : t \geq 0\}$  has bounded degree  $d$  and  $k$ -union vertex expansion at least  $\phi$ . Then, for  $\mu = (\frac{1}{d+1})^{O(\frac{k \log n}{\phi})}$ ,  $\text{HDiam}_\mu(\mathcal{G}) = O(\frac{k \log n}{\phi})$ .

*Proof.* We show that for  $T = O(\frac{k \log n}{\phi})$  and  $\mu = (\frac{1}{d+1})^T$ ,  $\text{HDiam}_\mu(\mathcal{G}) \leq T$ . Hence, it suffices to show that for any  $t \geq T - 1$ , every entry of the product  $\mathcal{P}_t^T := P_t P_{t-1} \cdots P_{t-T+1}$  is at least  $\mu$ .

For  $1 \leq i \leq j$ , we use the notation  $P_{t,k}[i..j]$  to denote the product



$P_{t-(i-1)k}P_{t-(i-1)k-1}\cdots P_{t-jk+1}$ . Observe that if  $T$  is a multiple of  $k$ , then  $\mathcal{P}_t^T = P_{t,k}[1..\frac{T}{k}]$ .

Observe that for each  $t$ , the transition matrix  $P_t$  obtained from  $G_t$  through the uniform averaging model has the following properties.

- For each  $i$ ,  $P_t[i, i] = \frac{1}{d_t[i]+1}$ .
- For  $i \neq j$ ,  $P_t[i, j] > 0$  iff  $\{i, j\} \in G_t$ .
- Since  $G_t$  has bounded degree  $d$ , every non-zero entry of  $P_t$  is at least  $\frac{1}{d+1}$ .

Observe that we can view a matrix  $P$  as a directed graph  $G(P)$ , where  $(u, v) \in G(P)$  iff  $P[u, v] > 0$ . Hence,  $G(P_t)$  is a directed version of  $G_t$  with self-loop at every node added.

Given square matrices  $A_1, A_2, \dots, A_l$ , observe that the  $(u, v)$ -th entry of the product  $A_1A_2 \dots A_l$  is non-zero iff node  $v$  can be reached from node  $u$  in exactly  $l$  steps such that for every  $1 \leq i \leq l$ , only an edge from  $G(A_i)$  can be used in step  $i$ .

Hence, it follows that for every  $i$ ,  $G(P_{t,k}[i])$  is a directed version of  $\cup_{r=t-(i-1)k}^{t-ik+1} G_r$  with self-loops added, which from the hypothesis has vertex expansion at least  $\phi$ .

Fix some nodes  $u$  and  $v$ . It follows from the vertex expansion property that there are at least  $(1 + \phi)$  nodes  $w$  such that the  $(u, w)$ -th entry of  $P_{t,k}[1]$  is non-zero. Repeating this argument  $T_0 := \lceil \log_{1+\phi} \frac{n}{2} \rceil$  times, it follows there are more than  $\frac{n}{2}$  nodes  $w$  such that the  $(u, w)$ -th entry of  $P_{t,k}[1..T_0]$  is non-zero.

By a reverse argument, it follows that there are more than  $\frac{n}{2}$  nodes  $w$  such that the  $(w, v)$ -th entry of  $P_{t,k}[T_0 + 1, 2T_0]$  is non-zero. Hence, by taking  $T = 2T_0k = O(\frac{k \log n}{\phi})$ , we have shown that the  $(u, v)$ -th entry of the product  $\mathcal{P}_t^T = P_{t,k}[1..\frac{T}{k}]$  is non-zero. Observe that  $\mathcal{P}_t^T$  is a product of  $T$  matrices, each of whose non-zero entry is at least  $\frac{1}{d+1}$ . Hence, we conclude that every entry of  $\mathcal{P}_t^T$  is at least  $\mu := (\frac{1}{d+1})^T$ , as required.  $\square$

The connectivity of a network can also be measured by its eigenvalue gap, whose relationship with vertex expansion is given by the following lemma, which is a variation of the Cheeger's Inequality. We defer its proof to the full version.

**Lemma 4 (Eigenvalue Gap Implies Vertex Expansion).** *For a network  $G$  with bounded degree  $d$ , define the weight vector  $w \in \mathbb{Z}^n$  by  $w[i] = d[i] + 1$  for each  $i$ . Let  $P := I_n - W^{-1}L$ , where  $W = \text{Diag}(w)$  and  $L$  is the Laplacian of  $G$ . Suppose there exists  $0 < \eta < 1$ , such that  $|\lambda_1(P)| \leq 1 - \eta$ . Then,  $\Phi(G) = \Omega(\frac{\eta}{d})$ .*

Combining Lemma 4 and Theorem 2 gives the following corollary.

**Corollary 3 (Convergence Time for Networks with Eigenvalue Gap).**

*Given a network sequence  $\{G_t\}$  with bounded degree  $d$ , define the weight vector  $w_t \in \mathbb{Z}^n$  by  $w_t[i] = d_t[i] + 1$  for each  $i$ . Let  $P := I_n - W_t^{-1}L$ , where  $W_t = \text{Diag}(w_t)$  and  $L_t$  is the Laplacian of  $G_t$ . Suppose there exists  $0 < \eta < 1$ , such that  $|\lambda_1(P)| \leq 1 - \eta$ . Then, given an initial vector  $v_0 \in \mathbb{R}^n$  and  $\epsilon > 0$ , the convergence time to achieve  $\epsilon$ -consensus under the uniform averaging model is  $n^{O(\frac{d}{\eta} \log d)} \log \frac{\tau(v_0)}{\epsilon}$ .*

## 4.2 Random Networks

Our analysis for union networks can be easily extended for random networks. Specifically, we only require that the union vertex expansion property holds with some positive probability. Hence, our results also hold for random graphs with expansion property, such as  $G_{n,p}$ .

**Definition 7 (Union Vertex Expansion with Probability  $\xi$ ).** *Given a sequence  $\{G_t : t \geq 0\}$  of networks, and an integer  $k \geq 1$ , we say the sequence has  $k$ -union vertex expansion at least  $\phi$  with probability  $\xi > 0$ , if for any  $t \geq 0$ ,  $\Phi(\cup_{r=t}^{t+k-1} G_r) \geq \phi$  holds with probability at least  $\xi > 0$ ; moreover, the events involving different  $t$ 's are independent, as long as the underlying  $G_r$ 's involved are different.*

**Theorem 3.** *Suppose a network sequence  $\{G_t\}$  has bounded degree  $d$  and there exist  $\phi > 0$  and  $\xi > 0$ , such that the sequence has  $k$ -union vertex expansion at least  $\phi$  with probability  $\xi$ . Then, given an initial vector  $v_0 \in \mathbb{R}^n$  and  $\epsilon > 0$ , with all but negligible probability  $\exp(-n^{\Theta(\frac{k \log d}{\phi \xi})})$ , the convergence time to achieve  $\epsilon$ -consensus is  $n^{O(\frac{k \log d}{\phi \xi})} \log \frac{\tau(v_0)}{\epsilon}$ .*

*Proof.* The theorem is obtained by proving probabilistic versions of Lemmas 3 and 2. We first argue that for  $T = O(\frac{k \log n}{\xi \phi})$ , for every  $t \geq T-1$ , with probability at least  $\frac{1}{2}$ , every entry of the product  $\mathcal{P}_t^T := P_t P_{t-1} \cdots P_{t-T+1}$  is at least  $\mu := (\frac{1}{d+1})^T$ .

Consider a block of networks from the sequence of size  $k$ . By the hypothesis, the union graph over  $k$  networks has vertex expansion at least  $\phi$  with probability at least  $\xi$ . From the proof of Lemma 3, we need  $l := O(\frac{\log n}{\phi})$  such blocks in order to argue that every entry of the corresponding product of transition matrices is non-zero. Hence, by Chernoff Bound, if we have  $\frac{2l}{\xi}$  such blocks, then with probability at least  $1 - e^{-\Theta(l)} \geq \frac{1}{2}$ , at least  $l$  blocks will have the expansion property. Since each block contains  $k$  networks, it follows that the  $\frac{2l}{\xi}$  blocks contain  $T = \frac{2l}{\xi} \cdot k = O(\frac{k \log n}{\xi \phi})$  networks, as required.

Let  $P^{(T)}$  denote the product of a block of  $T$  transition matrices derived from the given sequence. We have just proved that with probability at least  $\frac{1}{2}$ , each entry in  $P^{(T)}$  is at least  $\mu$ . Hence, from Fact 2, we can conclude that with probability at least  $\frac{1}{2}$ ,  $\tau_1(P^{(T)}) \leq 1 - n\mu \leq e^{-n\mu}$ .

Finally, if we are given an initial configuration vector  $v_0$  and  $\epsilon > 0$ , the  $\tau_2$ -measure of the configuration vector after multiplying by  $M$  such blocks of matrices (each block is a product coming from  $T$  transition matrices and each entry of the product is at least  $\mu$ ) is at most  $\tau(v_0) \cdot e^{-Mn\mu}$ , which is at most  $\epsilon$  for  $M = \Omega(\frac{1}{n\mu} \log \frac{\tau(v_0)}{\epsilon})$ .

Hence, using Chernoff Bound again, with probability at least  $1 - e^{-\Theta(M)} \geq 1 - \exp(-n^{\Theta(\frac{k \log d}{\phi \xi})})$ , given  $4M$  blocks of size  $T$  transition matrices, at least  $M$  such blocks will have a product such that every entry is at least  $\mu$ .

Therefore, we conclude with all but negligible probability  $\exp(-n^{\Theta(\frac{k \log d}{\phi \xi})})$ , the convergence time to achieve  $\epsilon$ -consensus is  $O(MT) = n^{O(\frac{k \log d}{\phi \xi})} \log \frac{\tau(v_0)}{\epsilon}$ .  $\square$

## 5 Experiments

In this section, we design experiments to simulate the behavior of our convergence models. In each simulation, we choose  $\epsilon = 0.001$  and record the average convergence time to achieve  $\epsilon$ -consensus. We observe in each case how the convergence time varies with  $n$ , the size of the network.

**(1)  $G_{n,p}$  under the Uniform Averaging Model.** At time  $t$ , the network  $G_t$  is sampled independently from  $G_{n,p}$ , where  $p = \frac{d}{n}$  ( $d = 5, 10, 20$ ). The result is in Figure 1.

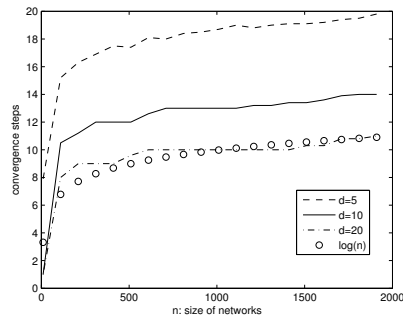


Fig. 1:  $G_{n,p} : p = \frac{d}{n}$  under the Uniform Averaging Model

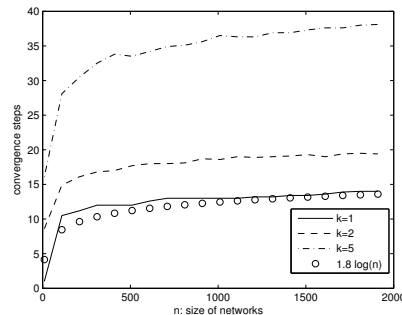


Fig. 2:  $k$ -Union  $G_{n,p}$  (with  $p = \frac{10}{n}$ ) under the Uniform Averaging Model.

**(2)  $k$ -Union  $G_{n,p}$  under the Uniform Averaging Model.** We fix  $p = \frac{10}{n}$  and use the  $G_{n,p}$  as the union network over  $k$  consecutive time steps. In particular, we sample a  $G_{n,p}$  graph as before and divide the edge set (randomly) into  $k$  different sets ( $k = 1, 2, 5$ ), each of which forms the edge set of a network in one time step. The result is in Figure 2.

In all the experiments, we see that the convergence time grows logarithmically with the size of the networks. As a visual aid, the circles in each graph give a reference for logarithmic growth.

**Acknowledgment.** We would like to thank Kurt Mehlhorn for introducing the problem to us, and Konstantinos Panagiotou, Nikolaos Fountoulakis, Hing-Fung Ting and Tak-Wah Lam for useful discussion during the initial stages of this work.

## References

1. D. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.
2. M. Cao, A. Morse, and B. Anderson. Coordination of an asynchronous, multi-agent system via averaging. In *the Proceeding of 16th International Federation of Automatic Control World Congress, IFAC*, 2005.
3. M. Cao, D. Spielman, and A. Morse. A lower bound on convergence of a distributed network consensus algorithm. In *the Proceeding of the 44th IEEE Conference on Decision and Control, and European Control Conference, CDC-ECC*, pages 2356 – 2361, dec. 2005.
4. B. Chazelle. Natural algorithms. In *the Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 422–431, 2009.
5. J. Cortes. Finite-time convergent gradient flows with applications to network consensus. *Automatica*, 42:1993 – 2000, 2006.
6. F. Cucker and S. Smale. Emergent behavior in blocks. *IEEE Transaction on Automatic Control*, 52:852–862, 2007.
7. M. DeGroot. Reaching a consensus. *Journal of American Statistical Association*, 69(345):118–121, 1974.
8. L. Fang and P. Antsaklis. On communication requirements for multi-agent consensus seeking. In *the Workshop on Networked Embedded Sensing and Control*, pages 53–67, 2005.
9. A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.
10. H. J. Landau and A. M. Odlyzko. Bounds for eigenvalues of certain stochastic matrices. *Linear Algebra and its Applications*, 38:5 – 15, 1981.
11. Y. Liu and Y. R. Yang. Reputation propagation and agreement in mobile ad-hoc networks. *Wireless Communications and Networking*, 3:1510 – 1515, 2003.
12. D. A. Lorenz and J. Lorenz. Convergence to consensus by general averaging. In *R. Bru and S. Romero-Vivo(Eds.): Positive Systems, Lecture Notes in Control and Information Sciences 389*, pages 91–99. Springer, 2009.
13. L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50:169–182, 2005.
14. A. Nedić, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Transactions on automatic control*, 54:2506–2517, 2009.
15. A. Olshevsky and J. N. Tsitsiklis. Degree fluctuations and the convergence time of consensus algorithms. *Preprint*.
16. A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009.
17. C. W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25–34, 1987.
18. H. G. Tanner, A. Jadbabaie, and G. J. Pappas. Flocking in fixed and switching networks. *Electronic Communications in Probability*, 52:863–868, 2006.
19. T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Scochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6):1226 – 1229, 1995.