A Detailed Explanation of k-peer Hyperhypercube Graph

In this section, we explain Alg. 1 in more detail. The k-PEER HYPER-HYPERCUBE GRAPH mainly consists of the following five steps.

Step 1. Decompose n as $n = n_1 \times \cdots \times n_L$ with minimum L such that $n_l \in [k+1]$ for all $l \in [L]$.

Step 2. If $L = 1$, we make all nodes obtain the average of parameters in V by using the complete graph. If $L \geq 2$, we split V into disjoint subsets $V_1, \cdots, V_{n_L}$ such that $|V_l| = \frac{n}{n_L}$ for all $l \in [n_L]$ and continue to step 3.

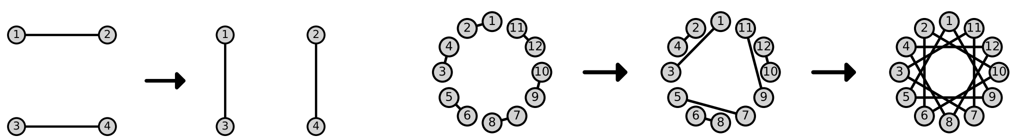
Step 3. For all $l \in [n_L]$, we make all nodes in V_l obtain the average of parameters in V_l by using the k-PEER HYPER-HYPERCUBE GRAPH $\mathcal{H}_k(V_l)$.

Step 4. We take n_L nodes from $V_1, \cdots, V_{n_L}$ respectively and construct a set U_1. Similarly, we construct $U_2, \cdots, U_{n_L}$ such that $U_1, \cdots, U_{n_L}$ are disjoint sets.

Step 5. For all $l \in [n_L]$, we make all nodes in U_l obtain the average of parameters in U_l by using the complete graph. Because the average of parameter U_l is equivalent to the average in V after step 4, all nodes reach the exact consensus.

When $n \leq k+1$, the k-PEER HYPER-HYPERCUBE GRAPH becomes the complete graph because of step 2. When $n > k+1$, we decompose n in step 1 and construct the k-PEER HYPER-HYPERCUBE GRAPH recursively in step 3. Thus, the k-PEER HYPER-HYPERCUBE GRAPH can make all nodes reach the exact consensus by the sequence of L graphs.

Using the example provided in Fig. 10, we explain the k-PEER HYPER-HYPERCUBE GRAPH in a more detailed manner. When $n = 12$, we decompose 12 as $2 \times 2 \times 3$. In step 2, we split $V \coloneqq \{1, \cdots, 12\}$ into $V_1 \coloneqq \{1, \cdots, 4\}$, $V_2 \coloneqq \{5, \cdots, 8\}$, and $V_3 \coloneqq \{9, \cdots, 12\}$. Step 3 corresponds to the first two graphs in Fig. 10b. As shown in Fig. 10a, the subgraphs consisting of V_1, V_2, and V_3 in the first two graphs in Fig. 10b are equivalent to the k-PEER HYPER-HYPERCUBE GRAPH with the number of nodes 4. Thus, all nodes reach the exact consensus by exchanging parameters in Fig. 10b.



(a) $n = 4 (= 2 \times 2)$ (b) $n = 12 (= 2 \times 2 \times 3)$

Figure 10: Illustration of the 2-PEER HYPER-HYPERCUBE GRAPH. In Fig. 10a, all edge weights are $\frac{1}{2}$. In Fig. 10b, edge weights are $\frac{1}{2}$ in the first two graphs and $\frac{1}{3}$ in the last graph.

B Detailed Explanation of Simple Base-$(k+1)$ Graph with $k \geq 2$

In Sec. 4.2, we explain Alg. 2 only in the case where maximum degree k is one. In this section, we explain the details of Alg. 2 in the case with $k \geq 2$.

The SIMPLE BASE-$(k+1)$ GRAPH mainly consists of the following five steps.

Step 1. As in the base-$(k+1)$ number of n, we decompose n as $n = a_1(k+1)^{p_1} + \cdots + a_L(k+1)^{p_L}$ in line 1, and then split V into disjoint subsets $V_1, \cdots, V_L$ such that $|V_l| = a_l(k+1)^{p_l}$ for all $l \in [L]$.

Step 2. For all $l \in [L]$, we split V_l into disjoint subsets $V_{l,1}, \cdots, V_{l,a_l}$ such that $|V_{l,a}| = (k+1)^{p_l}$ for all $a \in [a_l]$ in line 3.

Step 3. For all $l \in [L]$, we make all nodes in V_l obtain the average of parameters in V_l using the k-PEER HYPER-HYPERCUBE GRAPH $\mathcal{H}_k(V_l)$ in line 11. Then, we initialize l' as one.

Step 4. Each node in $V_{l'+1} \cup \cdots \cup V_L$ exchange parameters with $a_{l'}$ nodes in $V_{l'} (= V_{l',1} \cup \cdots \cup V_{l',a_{l'}})$ such that the average in $V_{l',a}$ becomes equivalent to the average in V for all $a \in [a_{l'}]$. We increase l' by one and repeat step 4 until $l' = L$. This procedure corresponds to line 15.

Step 5. For all $l \in [L]$ and $a \in [a_l]$, we make all nodes in $V_{l,a}$ obtain the average in $V_{l,a}$ using the k-PEER HYPER-HYPERCUBE GRAPH $\mathcal{H}_k(V_{l,a})$. Since the average in $V_{l,a}$ is equivalent to the average in V after step 4, all nodes reach the exact consensus. This procedure corresponds to line 25.

The major difference compared with the case where $k=1$ is step 4. In the case where $k=1$, each node in $V_{l'+1} \cup \cdots \cup V_L$ exchange parameters with one node in $V_{l'}$ such that the average in $V_{l'}$ becomes equivalent to that in V, while in the case where $k \geq 2$, each node in $V_{l'+1} \cup \cdots \cup V_L$ exchange parameters such that the average in $V_{l',a}$ becomes equivalent to that in V for all $a \in [a_l]$. Thanks to this step, we can make all nodes reach the exact consensus using k-PEER HYPER-HYPERCUBE GRAPH $\mathcal{H}_k(V_{l,a})$ instead of $\mathcal{H}_k(V_l)$ in step 5, and we can reduce the length of a graph sequence.

Using the example provided in Fig. 11, we explain Alg. 2 in a more detailed manner. Let $G^{(1)}, \cdots, G^{(4)}$ denote the graphs depicted in Fig. 11 from left to right, respectively. First, we split $V := \{1, \cdots, 7\}$ into $V_1 := \{1, \cdots, 6\}$ and $V_2 := \{7\}$, and then split V_1 into $V_{1,1} := \{1,2,3\}$ and $V_{1,2} := \{4,5,6\}$. In step 3, all nodes in V_1 obtain the same parameter by exchanging parameters in $G^{(1)}$ and $G^{(2)}$. In step 4, the average in $V_{1,1}$, that in $V_{1,2}$, and that in V_2 become the same as the average in all nodes V by exchanging parameters in $G^{(3)}$. Thus, in step 5, all nodes reach the exact consensus by exchanging parameters in $G^{(4)}$.

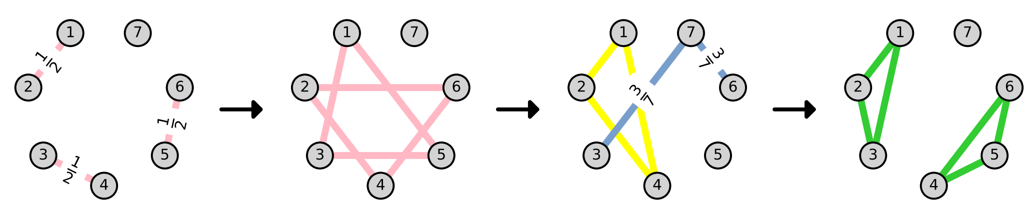


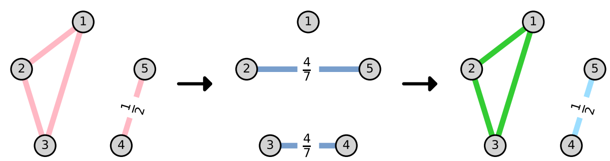
Figure 11: $k=2, n=7 (= 2 \times 3 + 1)$. The value on the edge indicates the edge weight. For simplicity, we omit the edge value when it is $\frac{1}{3}$.

C Illustration of Topologies

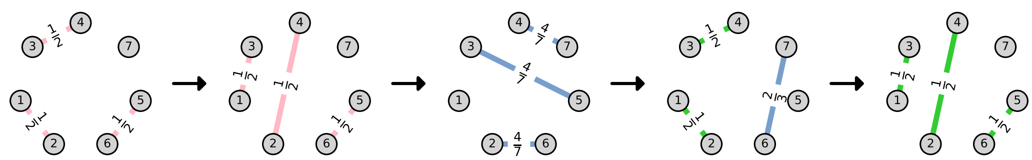
C.1 Examples

Fig. 12 shows the examples of the Simple Base-$(k+1)$ Graph. Using these examples, we explain how all nodes reach the exact consensus.

We explain the case depicted in Fig. 12a. Let $G^{(1)}, G^{(2)}, G^{(3)}$ denote the graphs depicted in Fig. 12a from left to right, respectively. First, we split $V \coloneqq \{1, \cdots, 5\}$ into $V_1 \coloneqq \{1, 2, 3\}$ and $V_2 \coloneqq \{4, 5\}$, and then split V_2 into $V_{2,1} \coloneqq \{4\}$ and $V_{2,2} \coloneqq \{5\}$. After exchanging parameters in $G^{(1)}$, nodes in V_1 and nodes in V_2 have the same parameter respectively. Then, after exchanging parameters in $G^{(2)}$, the average in V_1, that in $V_{2,1}$, and that in $V_{2,2}$ become same as the average in V. Thus, by exchanging parameters in $G^{(3)}$, all nodes reach the exact consensus. Note that edge $(4, 5)$ in $G^{(3)}$, which is added in line 27 in Alg. 2, is not necessary for all nodes to reach the exact consensus because nodes 4 and 5 already have the same parameter after exchanging parameters in $G^{(2)}$; however, it is effective in decentralized learning as we explained in Sec. 4.2.



(a) $k = 2, n = 5 (= 3 + 2)$



(b) $k = 1, n = 7 (= 2^2 + 2 + 1)$

Figure 12: Illustration of the Simple Base-$(k+1)$ Graph. The edge is colored in the same color as the line of Alg. 2 where the edge is added. The value on the edge indicates the edge weight. For simplicity, we omit the edge value when it is $\frac{1}{3}$.

C.2 Illustrative Comparison between Simple Base-$(k+1)$ and Base-$(k+1)$ Graphs

In this section, we provide an example of the Simple Base-$(k+1)$ Graph, explaining the reason why the length of the Base-$(k+1)$ Graph is less than that of the Simple Base-$(k+1)$ Graph.

Let $G^{(1)}, \cdots, G^{(5)}$ denote the graphs depicted in Fig. 13 from left to right, respectively. $(G^{(1)}, G^{(2)}, G^{(3)}, G^{(4)}, G^{(5)})$ is finite-time convergence, but $(G^{(1)}, G^{(2)}, G^{(3)}, G^{(5)})$ is also finite-time convergence because after exchanging parameters in $G^{(3)}$, nodes 3 and 4 already have the same parameters. Then, using the technique proposed in Sec. 4.3, we can remove such unnecessary graphs contained in the Simple Base-$(k+1)$ Graph (see Fig. 4a). Consequently, the Base-$(k+1)$ Graph can make all nodes reach the exact consensus faster than the Simple Base-$(k+1)$ Graph.

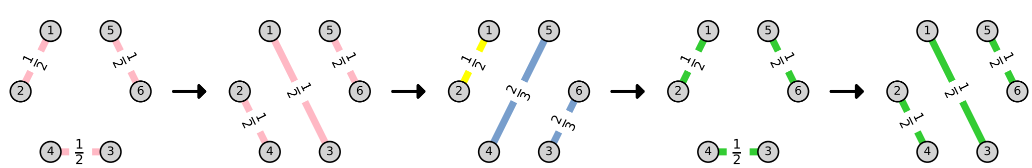
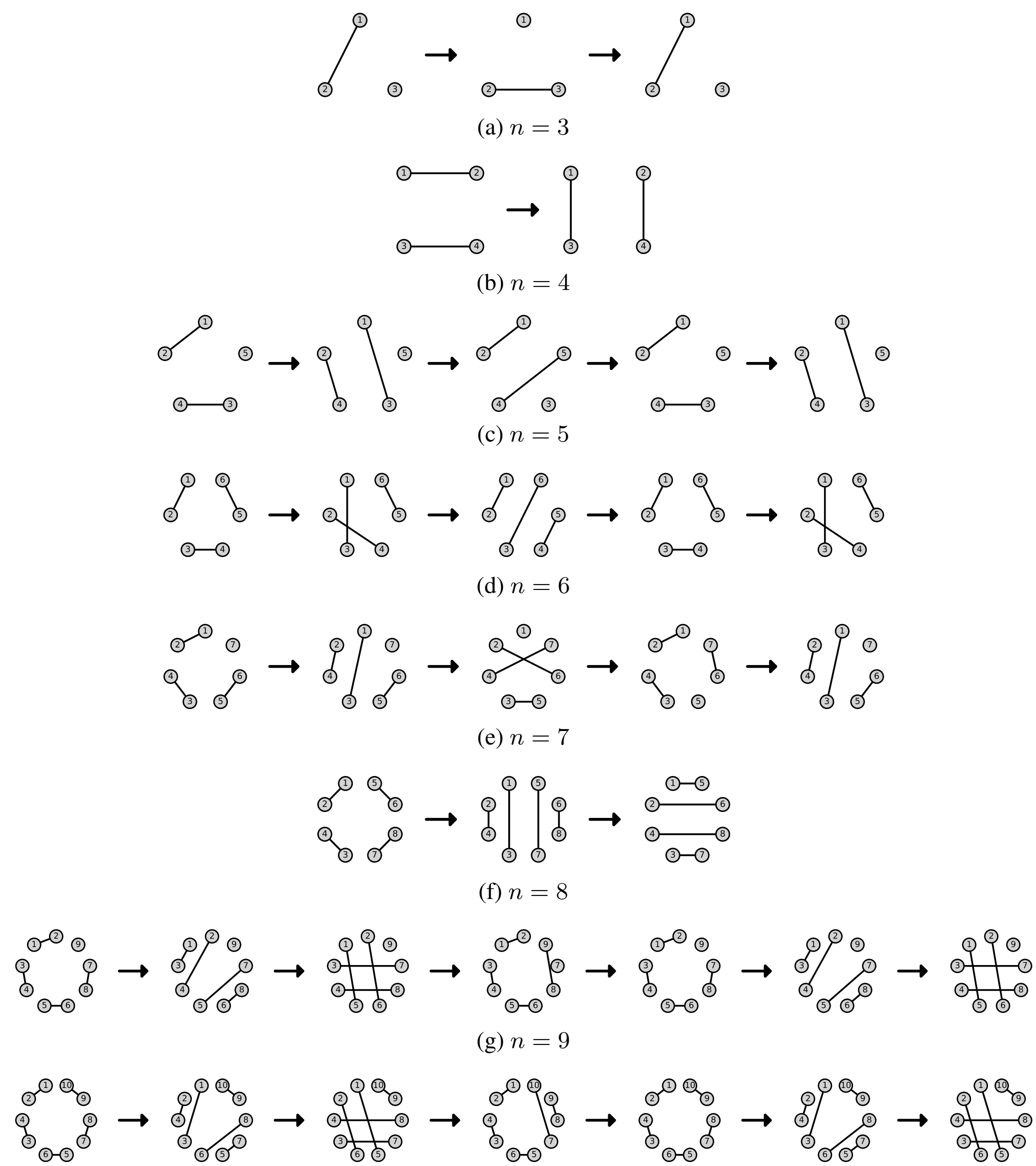


Figure 13: Illustration of the Simple Base-2 Graph with $n = 6 (= 2^2 + 2)$. The edge is colored in the same color as the line of Alg. 2 where the edge is added.

C.3 Additional Examples

C.3.1 Simple Base-$(k+1)$ Graph



(a) $n = 3$

(b) $n = 4$

(c) $n = 5$

(d) $n = 6$

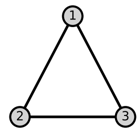
(e) $n = 7$

(f) $n = 8$

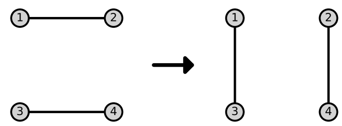
(g) $n = 9$

(h) $n = 10$

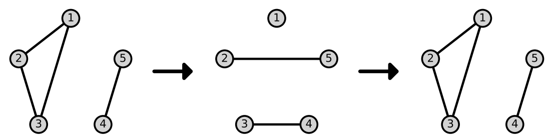
Figure 14: Illustration of the SIMPLE BASE-2 GRAPH with the various numbers of nodes.



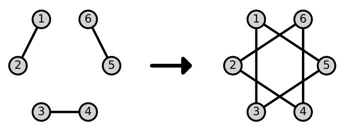
(a) $n = 3$



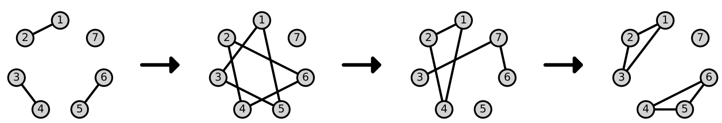
(b) $n = 4$



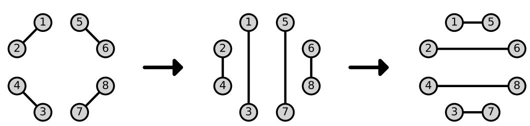
(c) $n = 5$



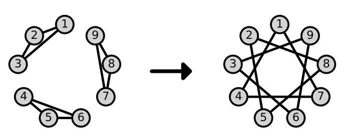
(d) $n = 6$



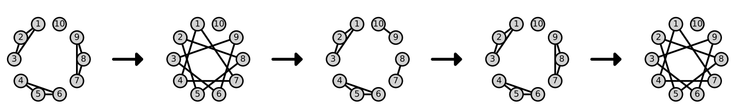
(e) $n = 7$



(f) $n = 8$



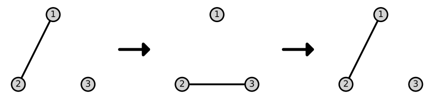
(g) $n = 9$



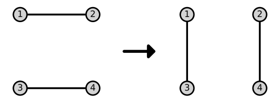
(h) $n = 10$

Figure 15: Illustration of the SIMPLE BASE-3 GRAPH with the various numbers of nodes.

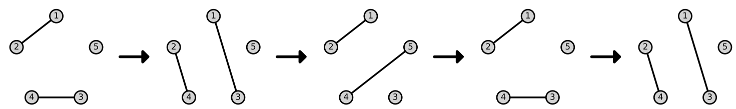
C.3.2 Base-$(k+1)$ Graph



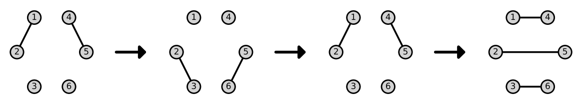
(a) $n = 3$



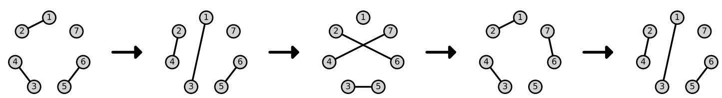
(b) $n = 4$



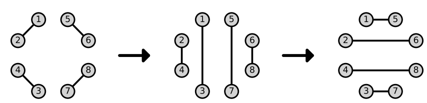
(c) $n = 5$



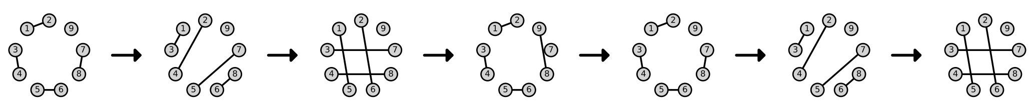
(d) $n = 6$



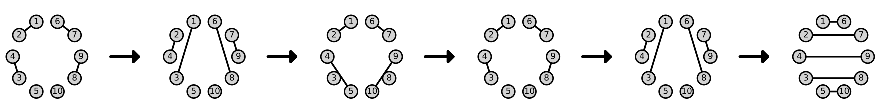
(e) $n = 7$



(f) $n = 8$

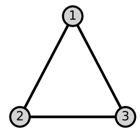


(g) $n = 9$

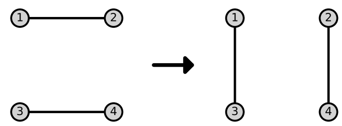


(h) $n = 10$

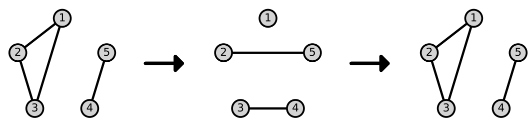
Figure 16: Illustration of the BASE-2 GRAPH with the various numbers of nodes.



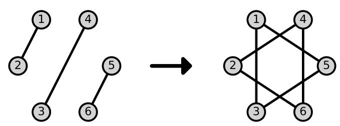
(a) $n = 3$



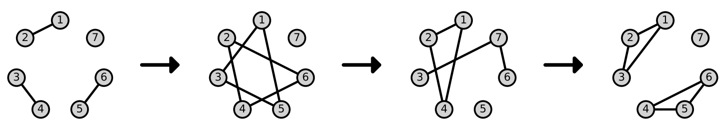
(b) $n = 4$



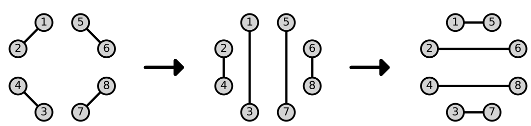
(c) $n = 5$



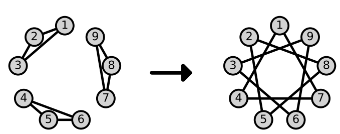
(d) $n = 6$



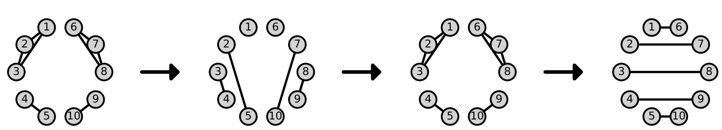
(e) $n = 7$



(f) $n = 8$



(g) $n = 9$



(h) $n = 10$

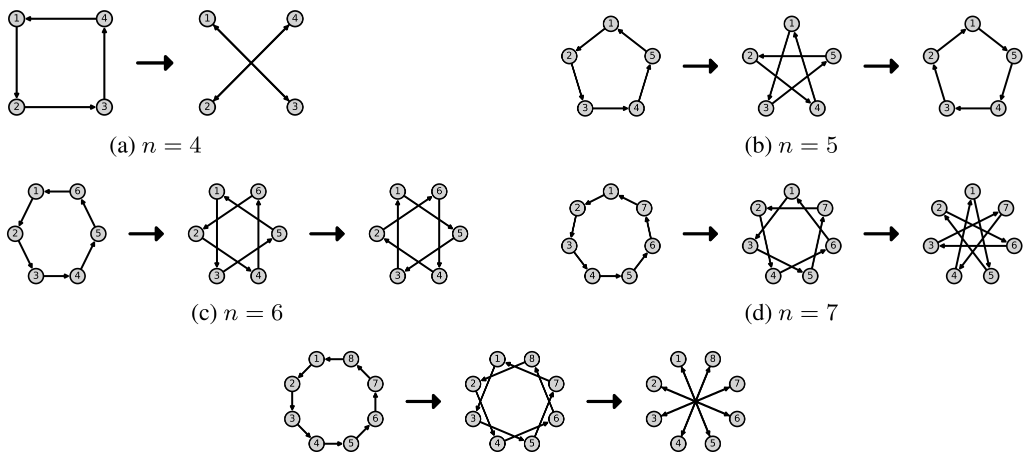
Figure 17: Illustration of the BASE-3 GRAPH with the various numbers of nodes.

C.4 1-peer Hypercube Graph and 1-peer Exponential Graph

For completeness, we provide examples of the 1-peer hypercube [31] and 1-peer exponential graphs [43] in Figs. 19 and 18, respectively.

(a) $n = 4$ (b) $n = 8$

Figure 18: Illustration of the 1-peer hypercube graph. All edge weights are 0.5.



(a) $n = 4$ (b) $n = 5$

(c) $n = 6$ (d) $n = 7$

(e) $n = 8$

Figure 19: Illustration of the 1-peer exponential graph. All edge weights are 0.5.

D Proof of Theorem 1

Lemma 1 (Length of k-PEER HYPER-HYPERCUBE GRAPH)**.** *Suppose that all prime factors of the number of nodes n are less than or equal to $k+1$. Then, for any number of nodes $n \in \mathbb{N}$ and maximum degree $k \in [n-1]$, the length of the k-PEER HYPER-HYPERCUBE GRAPH is less than or equal to* $\max\{1, 2\log_{k+2}(n)\}$.

Proof. We assume that n is decomposed as $n = n_1 \times \cdots \times n_L$ with minimum L where $n_l \in [k+1]$ for all $l \in [L]$. Without loss of generality, we suppose $n_1 \leq n_2 \leq \cdots \leq n_L$. Then, for any $i \neq j$, it holds that $n_i \times n_j \geq k+2$ because if $n_i \times n_j \leq k+1$ for some i and j, this contradicts the assumption that L is minimum.

When L is even, we have

$$n = (n_1 \times n_2) \times \cdots \times (n_{L-1} \times n_L) \geq (k+2)^{\frac{L}{2}}.$$

Then, we get $L \leq 2\log_{k+2}(n)$.

Next, we discuss the case when L is odd. When $L \geq 3$, $n_L \geq \sqrt{k+2}$ holds because $n_{L-2} \times n_{L-1} \geq k+2$. Thus, we get

$$n = (n_1 \times n_2) \times \cdots \times (n_{L-2} \times n_{L-1}) \times n_L \geq (k+2)^{\frac{L-1}{2}} \times n_L \geq (k+2)^{\frac{L}{2}}.$$

Then, we get $L \leq 2\log_{k+2}(n)$ when $L \geq 3$.

Thus, given the case when $L = 1$, the length of the k-PEER HYPER-HYPERCUBE GRAPH is less than or equal to $\max\{1, 2\log_{k+2}(n)\}$. □

Lemma 2 (Length of SIMPLE BASE-$(k+1)$ GRAPH)**.** *For any number of nodes $n \in \mathbb{N}$ and maximum degree $k \in [n-1]$, the length of the* SIMPLE BASE-$(k+1)$ GRAPH *is less than or equal to* $2\log_{k+1}(n) + 2$.

Proof. When all prime factors of n are less than or equal to $k+1$, the SIMPLE BASE-$(k+1)$ GRAPH is equivalent to the k-PEER HYPER-HYPERCUBE GRAPH and the statement holds from Lemma 1. In the following, we consider the case when there exists a prime factor of n that is larger than $k+1$. Note that because when $L = 1$ (i.e., $n = a_1 \times (k+1)^{p_1}$), all prime factors of n are less than or equal to $k+1$, we only need to consider the case when $L \geq 2$. We have the following inequality:

$$\begin{aligned}\log_{k+1}(n) &= \log_{k+1}(a_1(k+1)^{p_1} + \cdots + a_L(k+1)^{p_L}) \\ &\geq p_1 + \log_{k+1}(a_1) \\ &\geq p_1.\end{aligned}$$

Then, because $|V_1| = a_1 \times (k+1)^{p_1}$, it holds that $m_1 = |\mathcal{H}_k(V_1)| \leq 1 + p_1 \leq \log_{k+1}(n) + 1$. Similarly, it holds that $|\mathcal{H}_k(V_{1,1})| = p_1 \leq \log_{k+1}(n)$ because $|V_{1,1}| = (k+1)^{p_1}$. In Alg. 2, the update rule $b_1 \leftarrow b_1 + 1$ in line 22 is executed for the first time when $m = m_1 + 2$ because $L \geq 2$. Thus, the length of the SIMPLE BASE-$(k+1)$ GRAPH is at most $m_1 + |\mathcal{H}_k(V_{1,1})| + 1 \leq 2\log_{k+1}(n) + 2$. This concludes the statement. □

Lemma 3 (Length of BASE-$(k+1)$ GRAPH)**.** *For any number of nodes $n \in \mathbb{N}$ and maximum degree $k \in [n-1]$, the length of the* BASE-$(k+1)$ GRAPH *is less than or equal to* $2\log_{k+1}(n) + 2$.

Proof. The statement follows immediately from Lemma 2 and line 12 in Alg. 3. □

E Convergence Rate of DSGD over Various Topologies

Table 2 lists the convergence rates of DSGD over various topologies. These convergence rates can be immediately obtained from Theorem 2 stated in Koloskova et al. [11] and consensus rate of the topology. As seen from Table 2, the Base-2 Graph enables DSGD to converge faster than the ring and torus and as fast as the exponential graph for any number of nodes, although the maximum degree of the Base-2 Graph is only one. Moreover, for any number of nodes, the Base-$(k+1)$ Graph with $2 \leq k < \lceil \log_2(n) \rceil$ enables DSGD to converge faster than the exponential graph, even though the maximum degree of the Base-$(k+1)$ Graph remains to be less than that of the exponential graph.

Table 2: Convergence rates and maximum degrees of DSGD over various topologies.

Topology	**Convergence Rate**	**Maximum Degree**	**#Nodes** n
Ring [28]	$\mathcal{O}\left(\frac{\sigma^2}{n\epsilon^2} + \frac{\zeta n^2 + \sigma n}{\epsilon^{3/2}} + \frac{n^2}{\epsilon}\right) \cdot LF_0$	2	$\forall n \in \mathbb{N}$
Torus [28]	$\mathcal{O}\left(\frac{\sigma^2}{n\epsilon^2} + \frac{\zeta n + \sigma \sqrt{n}}{\epsilon^{3/2}} + \frac{n}{\epsilon}\right) \cdot LF_0$	4	$\forall n \in \mathbb{N}$
Exp. [43]	$\mathcal{O}\left(\frac{\sigma^2}{n\epsilon^2} + \frac{\zeta \log_2(n) + \sigma \sqrt{\log_2(n)}}{\epsilon^{3/2}} + \frac{\log_2(n)}{\epsilon}\right) \cdot LF_0$	$\lceil \log_2(n) \rceil$	$\forall n \in \mathbb{N}$
1-peer Exp. [43]	$\mathcal{O}\left(\frac{\sigma^2}{n\epsilon^2} + \frac{\zeta \log_2(n) + \sigma \sqrt{\log_2(n)}}{\epsilon^{3/2}} + \frac{\log_2(n)}{\epsilon}\right) \cdot LF_0$	1	A power of 2
1-peer Hypercube [31]	$\mathcal{O}\left(\frac{\sigma^2}{n\epsilon^2} + \frac{\zeta \log_2(n) + \sigma \sqrt{\log_2(n)}}{\epsilon^{3/2}} + \frac{\log_2(n)}{\epsilon}\right) \cdot LF_0$	1	A power of 2
Base-$(k+1)$ **Graph (ours)**	$\mathcal{O}\left(\frac{\sigma^2}{n\epsilon^2} + \frac{\zeta \log_{k+1}(n) + \sigma \sqrt{\log_{k+1}(n)}}{\epsilon^{3/2}} + \frac{\log_{k+1}(n)}{\epsilon}\right) \cdot LF_0$	k	$\forall n \in \mathbb{N}$

F Additional Experiments

F.1 Comparison of Base-$(k+1)$ and Simple Base-$(k+1)$ Graphs

Fig. 20 shows the length of the SIMPLE BASE-$(k+1)$ GRAPH and BASE-$(k+1)$ GRAPH. The results indicate that for all k, the length of the BASE-$(k+1)$ GRAPH is less than the length of the SIMPLE BASE-$(k+1)$ GRAPH in many cases.

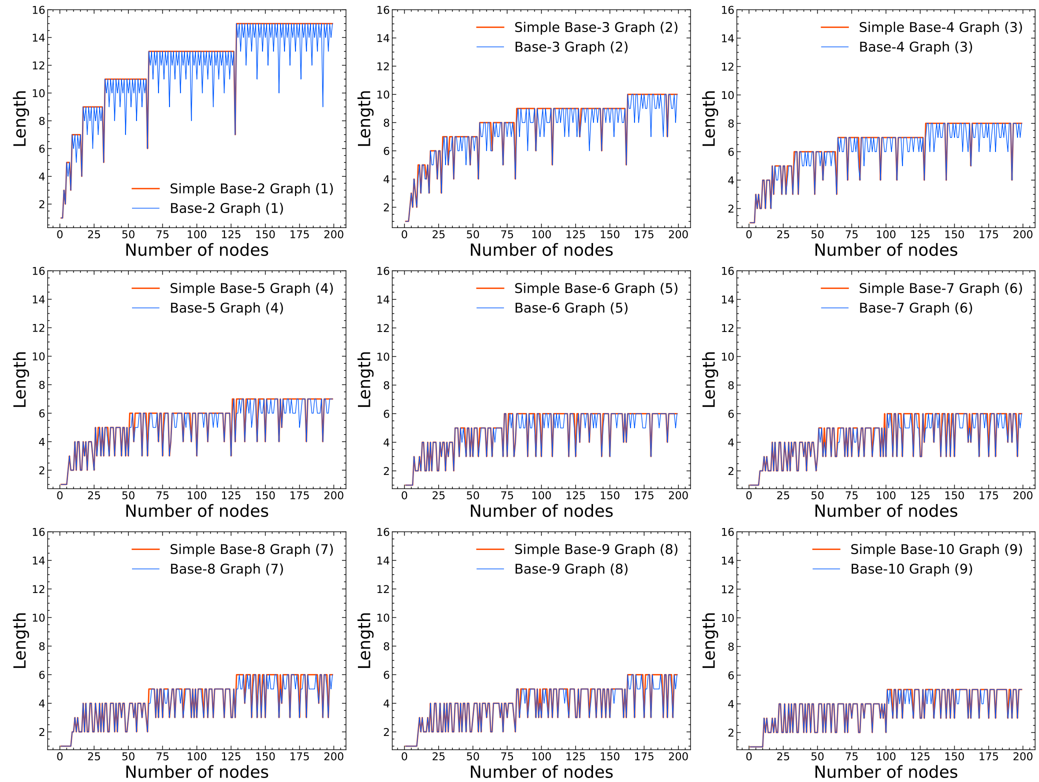


Figure 20: Comparison of the length of the SIMPLE BASE-$(k+1)$ GRAPH and BASE-$(k+1)$ GRAPH.

F.2 Consensus Rate

In Fig. 21, we demonstrate how consensus error decreases on various topologies when the number of nodes n is a power of 2. The results indicate that the BASE-2 GRAPH and 1-peer exponential graph can reach the exact consensus after the same finite number of iterations and reach the consensus faster than other topologies. Note that the BASE-2 GRAPH is equivalent to the 1-peer hypercube graph when n is a power of 2.

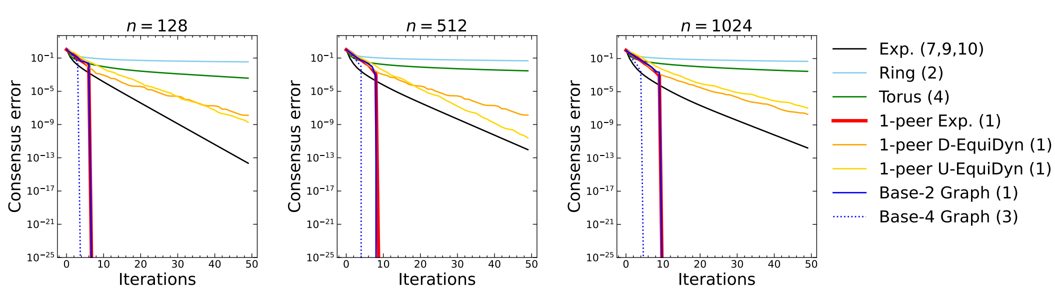


Figure 21: Comparison of consensus rates among different topologies when the number of nodes n is a power of 2. Because the BASE-$\{3, 5\}$ GRAPH are the same as the BASE-$\{2, 4\}$ GRAPH, respectively, when n is a power of 2, we omit the results of the BASE-$\{3, 5\}$ GRAPH.

F.3 Decentralized Learning

F.3.1 Comparison of Base-$(k+1)$ Graph and EquiStatic

In this section, we compared the BASE-$(k+1)$ GRAPH with the {U, D}-EquiStatic [33]. The {U, D}-EquiStatic are dense variants of the 1-peer {U, D}-EquiDyn, and their maximum degree can be set as hyperparameters. We evaluated the {U, D}-EquiStatic varying their maximum degrees; the results are presented in Fig. 22. In both cases with $\alpha = 10$ and $\alpha = 0.1$, the BASE-2 GRAPH can achieve comparable or higher final accuracy than all {U, D}-EquiStatic, and the BASE-$\{3, 4, 5\}$ GRAPH outperforms all {U, D}-EquiStatic. Thus, the BASE-$(k+1)$ GRAPH is superior to the {U, D}-EquiStatic from the perspective of achieving a balance between accuracy and communication efficiency.

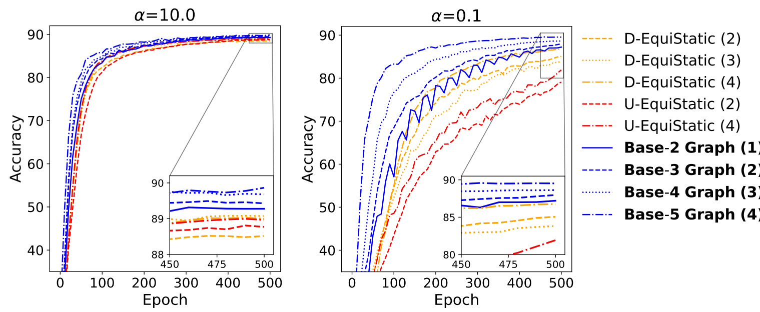


Figure 22: Test accuracy (%) of DSGD with CIFAR-10 and $n = 25$. The number in the bracket is the maximum degree of a topology.

F.3.2 Comparison with Various Number of Nodes

In this section, we evaluated the effectiveness of the BASE-$(k+1)$ GRAPH when varying the number of nodes n. Fig. 24 presents the learning curves, and Fig 23 shows how consensus error decreases when n is 21, 22, 23, 24, and 25. From Fig. 24, the BASE-2 GRAPH consistently outperforms the 1-peer exponential graph and can achieve a final accuracy comparable to that of the exponential graph. Furthermore, the BASE-$\{3, 4, 5\}$ GRAPH can consistently outperform the exponential graph, even though the maximum degree of the BASE-$\{3, 4, 5\}$ GRAPH is less than that of the exponential graph.

In Fig. 25 presents the learning curve for $n = 16$. When the number of nodes is a power of two, the 1-peer exponential graph is also finite-time convergence, and the 1-peer exponential graph and BASE-2 GRAPH achieve competitive accuracy.

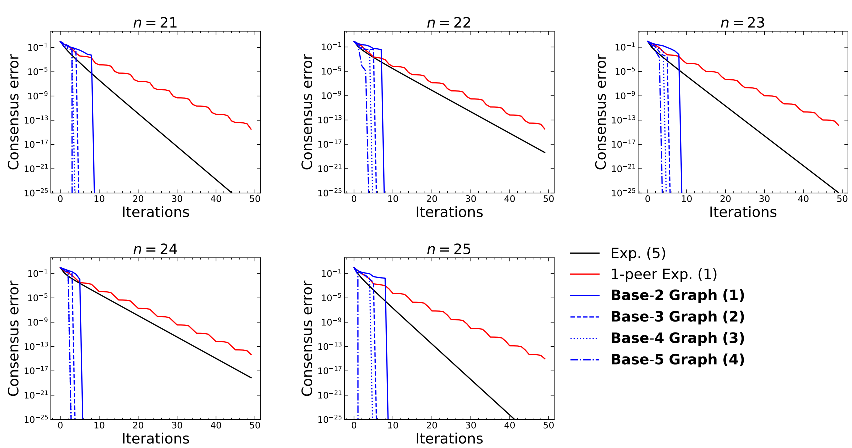


Figure 23: Comparison of consensus rates among different topologies. The number in the bracket denotes the maximum degree of a topology. We omit the results of the BASE-5 GRAPH when $n = 24$ because the BASE-5 GRAPH and BASE-4 GRAPH are equivalent when $n = 24$.

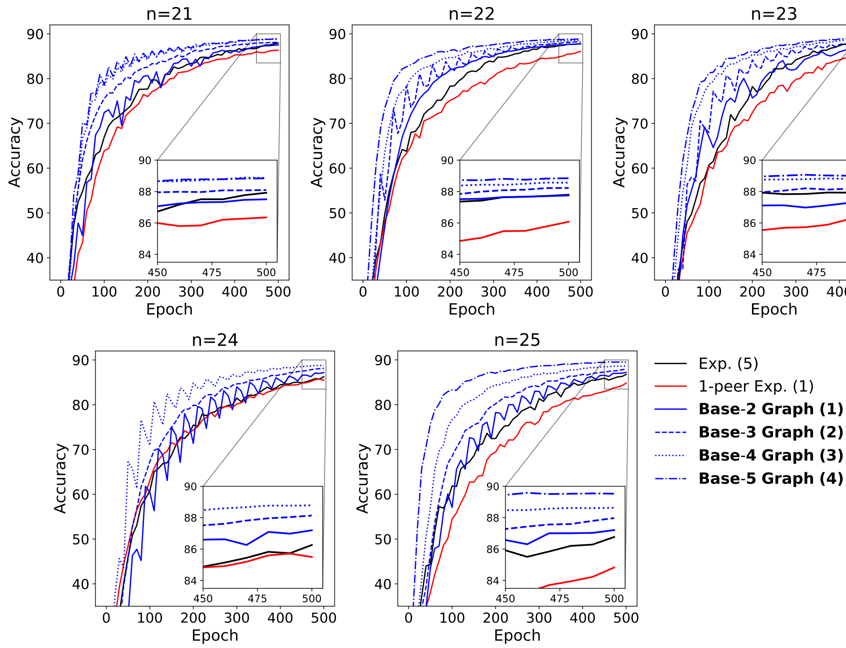


Figure 24: Test accuracy (%) of DSGD with CIFAR-10 and $\alpha = 0.1$. The number in the bracket denotes the maximum degree of a topology. We omit the results of the BASE-5 GRAPH when $n = 24$ because the BASE-5 GRAPH and BASE-4 GRAPH are equivalent when $n = 24$.

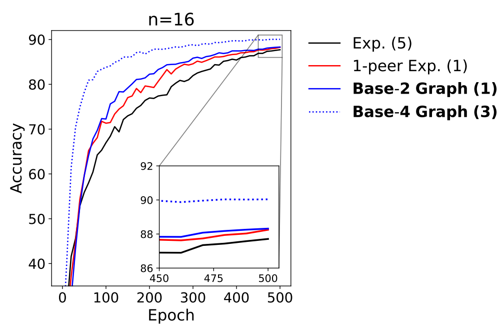


Figure 25: Test accuracy (%) of DSGD with CIFAR-10 and $n = 16$. The number in the bracket is the maximum degree of a topology. We omit the results of the BASE-3 GRAPH and BASE-5 GRAPH because these graphs are equivalent to the BASE-2 GRAPH and BASE-4 GRAPH, respectively.