A COMPLETE PROOF OF THEOREM 2

Theorem 2 (Migration Cost). *Suppose a set of ordered edges is initially split into k partitions via the chunk-based edge partitioning, and the edges are repartitioned into $k+n$ parts by adding n new processes (i.e., scale out). We assume that $|E|$ is much larger than k and n such that $(|E| \mod k+n)/|E| < (k+n)/|E| \approx 0$ and that the ids of new partitions are $k, k+1, ..., k+n-1$.*

Then, the number of migrated edges when applying repartitioning is approximately $\frac{n|E|}{2k(k+n)}\left\lceil\frac{k}{n}\right\rceil\left(\left\lceil\frac{k}{n}\right\rceil+1\right)+\frac{|E|}{k}\left(k-\left\lceil\frac{k}{n}\right\rceil\right)$. The cost for scaling in is the same (i.e., from $k+n$ to k partitions) since it is a reverse operation of scaling out.

Proof. We consider a simple case where $|E| \mod k = 0, |E| \mod (k+1) = 0, |E| \mod (k+2) = 0, ..., |E| \mod (k+n) = 0$. Then, there are two cases in the edge migration for partition $i (i \in [0,k))$: (i) some of the edges in partition i are migrated to other partitions, or (ii) all of the edges in partition i are migrated to other partitions.

Case (i): In this case, for partition i, the edges from $i\frac{|E|}{k}$-th edge to $(i+1)\frac{|E|}{k+n}$-th are kept in partition i, while from $(i+1)\frac{|E|}{k+n}$-th to $(i+1)\frac{|E|}{k}$-th edges are migrated to other partitions.

Thus the number of migrated edges for partition i is represented as follows:

$$(i+1)\frac{|E|}{k}-(i+1)\frac{|E|}{k+n}=(i+1)\frac{|E|n}{(k+n)k}$$

Case (i) happens when $(i+1)\frac{|E|n}{(k+n)k} > \frac{|E|}{k}$.

$$\begin{aligned} & (i+1)\frac{|E|n}{k(k+n)} > \frac{|E|}{k} \\ \Leftrightarrow\quad & (i+1) > \frac{k+n}{n} \\ \Leftrightarrow\quad & i > \frac{k}{n} \end{aligned}$$

Therefore, Case (i) happens when $i > \frac{k}{n}$.

Case (ii): In the other case (i.e., $i \le \frac{k}{n}$), all of the edges in partition i are migrated to other partitions. Thus, the number of migrated edges for partition i is $\frac{|E|}{k}$.

Therefore, to summarize Cases (i) and (ii), the total number of migrated edges from $i=0$ to $i=k-1$ is formalized as follows:

$$\begin{aligned} & \sum_{0\le i<\frac{k}{n}}(i+1)\frac{|E|n}{(k+n)k}+\sum_{\frac{k}{n}\le i<k}\frac{|E|}{k} \\ = \quad & \frac{|E|n}{(k+n)k}\sum_{0\le i<\frac{k}{n}}(i+1)+\frac{|E|}{k}\sum_{\frac{k}{n}\le i<k}1 \\ = \quad & \frac{n|E|}{2k(k+n)}\left\lceil\frac{k}{n}\right\rceil\left(\left\lceil\frac{k}{n}\right\rceil+1\right)+\frac{|E|}{k}\left(k-\left\lceil\frac{k}{n}\right\rceil\right) \end{aligned}$$

The aforementioned simplified proof can be straightforwardly generalized for the case of $|E| \mod k \ne 0, |E| \mod k+1 \ne 0, ..., |E| \mod k+n \ne 0$, based on the assumption $(|E| \mod k+n)/|E| \approx 0$. ■

B COMPLETE PROOF OF THEOREM 3

Theorem 3 (NP-hardness). *The graph edge ordering problem is NP-hard if $|E|$ is much larger than k_{max} so that less than k_{max} edges do not affect the optimized result.*

Proof. We first show that the graph edge ordering problem is NP-hard for single k, i.e., $k_{min}=k_{max}$. We then prove the general case of multiple k, i.e., $k_{min}<k_{max}$.

Case of Single k: Suppose $k_{min}=k_{max}=k$. The objective of the graph edge ordering problem is represented as follows:

$$\min_{\phi\in\Phi}\frac{1}{|V|}\sum_{p=0}^{k-1}\left|V\left(E_{ch}^{\phi}\left(\sum_{x=0}^{p-1}\left\lfloor\frac{|E|+x}{k}\right\rfloor,\left\lfloor\frac{|E|+p}{k}\right\rfloor\right)\right)\right|. \tag{5}$$

Now, we define a function to convert the edge order into the partition, $\texttt{ID2P}_k$: $i \mapsto p$, as Algorithm 2. By using $\texttt{ID2P}_k$, we can generate new edge partitions from the edge orders in linear time.

Algorithm 2: Conversion from Edge ID to Partition

Input : i – Ordered Edge ID
Output: p – Partition ID

```
ID2P_k(i)
    p ← 0; cur ← ⌊(|E|+p)/k⌋
    while i < cur do
        p ← p + 1; cur ← cur + ⌊(|E|+p)/k⌋
    return p
```

Suppose the order ϕ_{opt} is the optimal solution for the graph edge ordering problem. Then, the edge partitions converted from ϕ_{opt} via $\texttt{ID2P}_k$ is also the optimal solution for the edge partitioning problem in a case when $\epsilon \approx 0$ in Definition 1.

The reason is as follows. If the edge partitions converted from ϕ_{opt} via $\texttt{ID2P}_k$ is *not* the optimal solution (more specifically, more than k_{max} edges are in the different partitions from the optimal partitions), then there exist another optimal edge partitions, $\mathcal{E}_k^{opt} := \{\mathcal{E}_k^{opt}[p] \mid 0 \le p < k\}$, which provides a better solution for the edge partitioning problem than ϕ_{opt}. Based on $\mathcal{E}_k^{opt}$, we can generate new edge ordering ϕ' in such a way that for p

$$\mathcal{E}_k^{opt}[p]=\left\{E^{\phi'}[b],E^{\phi'}[b+1],...,E^{\phi'}\left[b+\left\lfloor\frac{|E|+p}{k}\right\rfloor-1\right]\right\},$$

where $b := \sum_{x=0}^{p-1}\left\lfloor\frac{|E|+x}{k}\right\rfloor$. Since $\mathcal{E}_k^{opt}$ provides the optimal solution,

$$\begin{aligned} RF(\mathcal{E}_k^{opt}) &:= \frac{1}{|V|}\sum_{p=0}^{k-1}|V(\mathcal{E}_k^{opt}[p])| \\ &= \frac{1}{|V|}\sum_{p=0}^{k-1}\left|V\left(E_{ch}^{\phi'}\left(\sum_{x=0}^{p-1}\left\lfloor\frac{|E|+x}{k}\right\rfloor,\left\lfloor\frac{|E|+p}{k}\right\rfloor\right)\right)\right| \end{aligned}$$

is the optimal value. On the other hand, ϕ_{opt} provides the optimal value of Eq. (5) as follows:

$$\frac{1}{|V|}\sum_{p=0}^{k-1}\left|V\left(E_{ch}^{\phi_{opt}}\left(\sum_{x=0}^{p-1}\left\lfloor\frac{|E|+x}{k}\right\rfloor,\left\lfloor\frac{|E|+p}{k}\right\rfloor\right)\right)\right|.$$

This is a contradiction to the assumption that $\mathcal{E}_k^{opt}$ provides the better solution than ϕ_{opt}. Thus, ϕ_{opt} can provide the optimal solution for the edge partitioning problem as well.

Therefore, the problem (5) is reducible to the balanced k-way edge partitioning problem, which is an NP-hard problem as proved in [80].

<u>*Case of* $k_{min} < k_{max}$</u>: We explain the case when $k_{min} = 2$ and $k_{max} = 3$. The following discussion can be straightforwardly generalized to any k_{min} and k_{max}.

According to Definition 5, we define a function, $Num(k, p)$, for the normalized number of vertices involved in the chunk of edges as follows:

$$Num(k, p) := \frac{1}{|V|}\left| V\left(E_{ch}^{\phi}\left(\sum_{x=0}^{p-1}\left\lfloor\frac{|E|+x}{k}\right\rfloor, \left\lfloor\frac{|E|+p}{k}\right\rfloor\right)\right)\right|.$$

Suppose $k_{min} = 2$ and $k_{max} = 3$, we will show the NP-hardness of the optimization problem as follows:

$$\min_{\phi\in\Phi}\sum_{k=2}^{3}\sum_{p=0}^{k-1}N(k,p) = \min_{\phi\in\Phi}\{Num(2,0) + Num(2,1) + Num(3,0) + Num(3,1) + Num(3,2)\}. \quad (6)$$

Here, based on the above discussion of the single k, the following optimization problems are already proved to be NP-hard:

$$\min_{\phi\in\Phi}\{Num(2,0) + Num(2,1)\} \quad (7)$$

$$\min_{\phi\in\Phi}\{Num(3,0) + Num(3,1) + Num(3,2)\}. \quad (8)$$

Suppose ϕ_{opt} is the optimal order for (6), then the order can be also the optimal for (7) and (8). Thus, if (6) is not NP-hard, it is a contradiction to the NP-hardness of (7) and (8). Therefore, (6) is also NP-hard. To summarize, the graph edge ordering problem is NP-hard. ∎

C COMPLETE PROOF OF LEMMA 2

Lemma 2. *Suppose $|E|$ is much larger than k_{max} such that $w := \left\lfloor\frac{|E|}{k}\right\rfloor = \left\lfloor\frac{|E|+\mathtt{ID2P}_k(\cdot)}{k}\right\rfloor$ and $D[v] < \frac{|E|}{k_{max}}$ for $\forall v \in V$. Then,*

$$\forall v, u \in V_{rest} \cap V(X^{\phi}), \quad p(v) > p(u) \Rightarrow F_v > F_u,$$

where F_v and F_u are the value of Eq. (3) for $X^{\phi}+N(v)$ and $X^{\phi}+N(u)$ respectively, as shown in Line 10 of Algorithm 3.

Proof. Suppose $Xv^{\phi} := X^{\phi} + N(v)$, $Xu^{\phi} := X^{\phi} + N(u)$.

$$\begin{aligned}
& F_v > F_u \quad \Leftrightarrow \quad F_v - F_u > 0 \\
\Leftrightarrow \quad & \sum_{k=k_{min}}^{k_{max}}\sum_{i=0}^{|E|-1}\left\{f\left(Xv^{\phi}, i, w\right) - f\left(Xu^{\phi}, i, w\right)\right\} > 0 \\
\Leftarrow \quad & \sum_{k=k_{min}}^{k_{max}}\sum_{i=0}^{|E|-1}\left\{\left|V\left(Xv_{ch}^{\phi}(i-w+1,w)\right)\right| - \left|V\left(Xu_{ch}^{\phi}(i-w+1,w)\right)\right|\right\} > 0 \\
\Leftrightarrow \quad & \sum_{k=k_{min}}^{k_{max}}\sum_{i=|X|^{\phi}}^{|E|-1}\{\Delta V(v,i) - \Delta V(u,i)\} > 0, \quad (9)
\end{aligned}$$

where $\Delta V(v,i) := |V(Xv_{ch}^{\phi}(i-w+1,w))| - |V(X_{ch}^{\phi}(i-w+1,w))|$.

Next, we will calculate $\Delta V(v,i)$ for $i \geq |X^{\phi}|$. Intuitively, $\Delta V(v,i)$ means the number of additional replicated vertices in a chunk when we select v to expand the ordered edges. For each chunk determined by i, each additional replicated vertex comes from v or $N(v)$. Thus, $\Delta V(v,i)$ can be represented by the sum of two functions:

$$\Delta V(v,i) = \chi(i) + n(i),$$

where $\chi(i)$ is the number of replicated vertices caused by v; $n(i)$ is caused by $N(v)$.

First, $\chi(i)$ is the indicator function. If $X_{ch}^{\phi}(i-w+1,w)$ already involves v, then the number of replicated vertices does not increase due to the additional v. Therefore, $\chi(i)$ is 0. Specifically, this case appears if $i > M[v]+w$, because $X_{ch}^{\phi}(i-w+1,w)$ involves an edge e whose order is $M[v]$ (i.e., $\phi(e) = M[v]$). Otherwise, v's replication is newly added to the chunk $Xv_{ch}^{\phi}(i-w+1,w)$, and thus $\chi(i)$ is 1. Therefore, $\chi(i)$ is represented as follows:

$$\chi(i) = \begin{cases} 1 & \text{if } i \in [M[v]+w, |X^{\phi}| + D[v] + w) \\ 0 & \text{if } i \notin [M[v]+w, |X^{\phi}| + D[v] + w) \end{cases}$$

where we also consider a case that i is larger so that $Xv_{ch}^{\phi}(i-w+1,w)$ is empty. In this case, $\chi(i)$ is obviously 0.

Second, $n(i)$ is the number of the additional vertices derived from $N(v)$. Its value can be represented as follows:

$$n(i) = \begin{cases} i - |X^{\phi}| & (|X^{\phi}| \leq i < |X^{\phi}| + D[v]) \\ D[v] & (|X^{\phi}| + D[v] \leq i < |X^{\phi}| + w) \\ D[v]-i+|X^{\phi}|+w & (|X^{\phi}| + w \leq i < |X^{\phi}| + D[v] + w) \\ 0 & (|X^{\phi}| + D[v] + w < i) \end{cases}$$

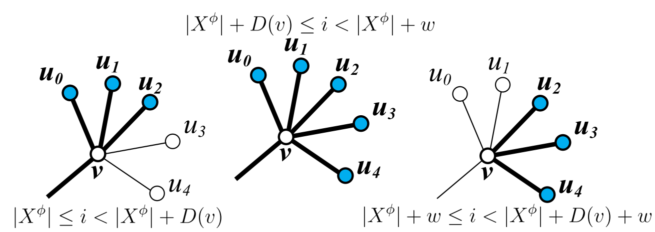

Figure 14: The value of $n(i)$: # of the additional vertices derived from $N(v) = \{u_0, u_1, ..., u_4\}$. Blue vertices are the additional when v is selected for the expansion.

Figure 14 shows an example of these cases. Suppose v is selected in the greedy algorithm and new edge orders are assigned to v's neighbor edges, e_{v,u_0}, e_{v,u_1}, e_{v,u_2}, e_{v,u_3}, and e_{v,u_4}. Then, if $|X^{\phi}| \leq i < |X^{\phi}| + D[v]$, a part of $N(v)$ are added (e.g., u_0, u_1, u_2 in Figure 14). If $|X^{\phi}| + D[v] \leq i < |X^{\phi}| + w$, all vertices in $N(v)$ are added (e.g., u_0, u_1, u_2, u_3, u_4 in Figure 14). If $|X^{\phi}| + w \leq i < |X^{\phi}| + D[v] + w$, also a part of $N(v)$ are added (e.g., u_2, u_3, u_4 in Figure 14). If $|X^{\phi}| + D[v] + w \leq i$, then $Xv_{ch}^{\phi}(i-w+1,w)$ involves no vertices.

Therefore,

$$\sum_{i=|X^{\phi}|}^{|E|-1}\Delta V(v,i) = \sum_{i=|X^{\phi}|}^{|E|-1}\chi(i) + \sum_{i=|X^{\phi}|}^{|X^{\phi}|+D[v]-1}\left\{i - |X^{\phi}|\right\} + \sum_{i=|X^{\phi}|+D[v]}^{|X^{\phi}|+w-1}D[v]$$

$$+\sum_{i=|X^{\phi}|+w}^{|X^{\phi}|+w+D[v]-1}\left\{D[v]-i+|X^{\phi}|+w\right\}$$

$$= \quad wD[v]+|X^{\phi}|+D[v]-M[v]$$

Let $\Delta D := D[v]-D[u]$ and $\Delta M := M[v]-M[u]$.

$$\sum_{i=|X^{\phi}|}^{|E|-1}\{\Delta V(v,i)-\Delta V(u,i)\} = w\Delta D+\Delta D-\Delta M$$

$$\sim w\Delta D-\Delta M \;\; (\because w > \frac{|E|}{k_{max}} \gg 1)$$

Therefore,

$$p(v) > p(u)$$

$$\Leftrightarrow \quad \alpha\cdot D[v]-\beta\cdot M[v] > \alpha\cdot D[u]-\beta\cdot M[u]$$

$$\Leftrightarrow \quad \sum_{k=k_{min}}^{k_{max}}(w\cdot D[v]-M[v]) > \sum_{k=k_{min}}^{k_{max}}(w\cdot D[u]-M[u])$$

$$\Leftrightarrow \quad \sum_{k=k_{min}}^{k_{max}}(w\Delta D-\Delta M) > 0$$

$$\Leftrightarrow \quad \sum_{k=k_{min}}^{k_{max}}\sum_{i=|X^{\phi}|}^{|E|-1}\{\Delta V(v,i)-\Delta V(u,i)\} > 0$$

$$\Rightarrow \quad F_v > F_u. \quad (\because (9))$$

Thus, the lemma is proved. ■