A DETAILED PROOFS

A.1 Notations

Symbol	Meaning
$z \in \mathcal{D}$	Training dataset (set of training points).
$t \in \mathcal{T}$	Test set (set of test points).
$S \subseteq \mathcal{D}$	A group formed by a subset of the training samples z.
$\theta(S)$	Model trained on subset S.
$g_t(\cdot)$	Scalar functional that maps trained parameters to the evaluation at t.
$v_t(S)$	Utility of subset S for test point t, i.e., $g_t(\theta(S))$.
$\Delta_z v_t(S)$	Marginal utility: $v_t(S \cup \{z\}) - v_t(S)$.
$\phi_z(v_t)$	Global Shapley value of training point z w.r.t. test point t.
ϕ_z	Aggregated Shapley value across test points: $\sum_{t \in T} \phi_z(v_t)$.
$\mathcal{N}(t)$	The subset of training instances that influence the prediction for t.
$v_t^{\mathcal{N}}(S)$	Projected utility: $v_t^{\mathcal{N}}(S) = v_t(S \cap \mathcal{N}(t))$.
$\phi_z^{\mathrm{loc}}(v_t)$	Local Shapley value computed under $v_t^{\mathcal{N}}$.
$G = (\mathcal{D}, \mathcal{T}, \mathcal{N}, \mathcal{R})$	Bipartite support-mapping structure used by LSMR.
$\mathcal{R}(z)$	Reverse map: $\mathcal{R}(z) = \{t \in T : z \in \mathcal{N}(t)\}$.
$\mathcal{R}_S$	Test points whose supports contain S, i.e., intersections over $z \in S$.
$\Pi_{\mathcal{T}}$	Global ordering of test points in pivot scheduling.
$t^*(S)$	Pivot test point assigned to subset S.
M	Number of Monte Carlo samples per test point in LSMR-A.
B	Uniform bound for utilities, $\lvert v_t(S)\rvert \le B$, used in concentration bounds.
ϵ	Accuracy tolerance parameter (e.g., in bounds).
δ	Failure probability parameter (e.g., in sample complexity bounds).

Table 4: Notations Table

A.2 Proof of Proposition 1

Proposition 1. *Under Assumption 1, for any $z \in \mathcal{N}(t)$,*

$$\left|\phi_z(v_t) - \phi_z^{\mathrm{loc}}(v_t)\right| \le \frac{1}{2} \sum_{u \in \mathcal{D} \setminus \mathcal{N}(t)} \ell_{t,z}(u),$$

where $\phi_z(v_t)$ denotes the global Shapley value computed under the original utility v_t.

Proof. Fix a test point t and a training point $z \in \mathcal{N}(t)$. Let $\Delta_z v_t(S) := v_t(S \cup \{z\}) - v_t(S)$ denote the marginal contribution of z to a coalition $S \subseteq \mathcal{D} \setminus \{z\}$. By the permutation characterization of the Shapley value, if π is a uniformly random permutation of $\mathcal{D}$ and $P_z(\pi) := \{x \in \mathcal{D} : \pi(x) < \pi(z)\}$ denotes the set of predecessors of z in π, then $\phi_z(v_t) = \mathbb{E}_\pi[\Delta_z v_t(P_z(\pi))]$.

Define $S(\pi) := P_z(\pi) \cap \mathcal{N}(t)$ and $U(\pi) := P_z(\pi) \cap (\mathcal{D} \setminus \mathcal{N}(t))$, so that $P_z(\pi) = S(\pi) \cup U(\pi)$ with $S(\pi) \subseteq \mathcal{N}(t) \setminus \{z\}$ and $U(\pi) \subseteq \mathcal{D} \setminus \mathcal{N}(t)$. The local Shapley value $\phi_z^{\mathrm{loc}}(v_t)$ computed by ignoring non-neighborhood players corresponds to taking the same permutation expectation but evaluating the marginal contribution only with neighborhood predecessors, namely $\phi_z^{\mathrm{loc}}(v_t) = \mathbb{E}_\pi[\Delta_z v_t(S(\pi))]$.

Therefore,

$$\begin{aligned}\left|\phi_z(v_t) - \phi_z^{\mathrm{loc}}(v_t)\right| &= \left|\mathbb{E}_\pi\left[\Delta_z v_t\big(S(\pi) \cup U(\pi)\big) - \Delta_z v_t\big(S(\pi)\big)\right]\right| \\ &\le \mathbb{E}_\pi\left[\left|\Delta_z v_t\big(S(\pi) \cup U(\pi)\big) - \Delta_z v_t\big(S(\pi)\big)\right|\right].\end{aligned}$$

Under Assumption 1, for any $S \subseteq \mathcal{N}(t) \setminus \{z\}$ and any $U \subseteq \mathcal{D} \setminus \mathcal{N}(t)$ we have $\left|\Delta_z v_t(S \cup U) - \Delta_z v_t(S)\right| \le \sum_{u \in U} \ell_{t,z}(u)$, so applying it pointwise yields

$$\left|\phi_z(v_t) - \phi_z^{\mathrm{loc}}(v_t)\right| \le \mathbb{E}_\pi\left[\sum_{u \in U(\pi)} \ell_{t,z}(u)\right] = \sum_{u \in \mathcal{D} \setminus \mathcal{N}(t)} \ell_{t,z}(u)\, \mathbb{P}_\pi\big(u \in U(\pi)\big).$$

For any fixed $u \ne z$, the event $u \in U(\pi)$ is equivalent to u appearing before z in a uniformly random permutation, which has probability $1/2$, hence $\mathbb{P}_\pi(u \in U(\pi)) = 1/2$ for all $u \in \mathcal{D} \setminus \mathcal{N}(t)$. Substituting this into the previous display gives

$$\left|\phi_z(v_t) - \phi_z^{\mathrm{loc}}(v_t)\right| \le \frac{1}{2} \sum_{u \in \mathcal{D} \setminus \mathcal{N}(t)} \ell_{t,z}(u).$$

The proposition follows. □

A.3 Proof of Proposition 2

Proposition 2. *The local Shapley value satisfies: (i) **Symmetry**: identically contributing data points receive equal value; (ii) **Null Player**: a data point with zero marginal contribution in all coalitions receives zero value; and (iii) **Additivity**: the valuation is linear in v_t.*

Proof. Fix t and restrict attention to the player set $\mathcal{N}(t)$. Define the local game on $\mathcal{N}(t)$ by $v_t^{\mathcal{N}}(S) := v_t(S)$ for all $S \subseteq \mathcal{N}(t)$, and let $\phi^{\mathrm{loc}}(v_t^{\mathcal{N}})$ denote the Shapley value of this restricted game. Since ϕ^{loc} is exactly the Shapley value on the finite player set $\mathcal{N}(t)$, it inherits the standard Shapley axioms on that set, and we verify the three stated properties explicitly.

(i) Symmetry. Let $i, j \in \mathcal{N}(t)$ satisfy $v_t^{\mathcal{N}}(S \cup \{i\}) - v_t^{\mathcal{N}}(S) = v_t^{\mathcal{N}}(S \cup \{j\}) - v_t^{\mathcal{N}}(S)$ for all $S \subseteq \mathcal{N}(t) \setminus \{i, j\}$. Using the subset-form Shapley formula on $\mathcal{N}(t)$,

$$\phi_i^{\mathrm{loc}}(v_t^{\mathcal{N}}) = \sum_{S \subseteq \mathcal{N}(t) \setminus \{i\}} \frac{|S|!\,(|\mathcal{N}(t)| - |S| - 1)!}{|\mathcal{N}(t)|!} \left(v_t^{\mathcal{N}}(S \cup \{i\}) - v_t^{\mathcal{N}}(S)\right),$$

and analogously for j. Pair each term indexed by $S \subseteq \mathcal{N}(t) \setminus \{i, j\}$ in the expansion of ϕ_i^{loc} with the corresponding term in ϕ_j^{loc}, noting that the Shapley weight depends only on $|S|$ and $|\mathcal{N}(t)|$ and is identical for i and j. Since the marginal contributions are equal by assumption for every such S, the two weighted sums coincide, hence $\phi_i^{\mathrm{loc}}(v_t^{\mathcal{N}}) = \phi_j^{\mathrm{loc}}(v_t^{\mathcal{N}})$.

(ii) Null Player. Let $i \in \mathcal{N}(t)$ satisfy $v_t^{\mathcal{N}}(S \cup \{i\}) - v_t^{\mathcal{N}}(S) = 0$ for all $S \subseteq \mathcal{N}(t) \setminus \{i\}$. Substituting into the subset-form Shapley formula above shows that every summand is zero, hence $\phi_i^{\mathrm{loc}}(v_t^{\mathcal{N}}) = 0$.

(iii) Additivity. Let $v_t^{\mathcal{N}(1)}$ and $v_t^{\mathcal{N}(2)}$ be two games on $\mathcal{N}(t)$, and define $v_t^{\mathcal{N}} := v_t^{\mathcal{N}(1)} + v_t^{\mathcal{N}(2)}$. For any $i \in \mathcal{N}(t)$ and any $S \subseteq \mathcal{N}(t) \setminus \{i\}$, we have $\Delta_i v_t^{\mathcal{N}}(S) = \Delta_i v_t^{\mathcal{N}(1)}(S) + \Delta_i v_t^{\mathcal{N}(2)}(S)$. Plugging this into the subset-form Shapley formula and using linearity of summation yields

$$\phi_i^{\mathrm{loc}}(v_t^{\mathcal{N}}) = \phi_i^{\mathrm{loc}}(v_t^{\mathcal{N}(1)}) + \phi_i^{\mathrm{loc}}(v_t^{\mathcal{N}(2)}),$$

which is additivity. More generally, for any scalars a, b, the same argument gives $\phi^{\mathrm{loc}}(a v_t^{\mathcal{N}(1)} + b v_t^{\mathcal{N}(2)}) = a\phi^{\mathrm{loc}}(v_t^{\mathcal{N}(1)}) + b\phi^{\mathrm{loc}}(v_t^{\mathcal{N}(2)})$.

Thus, within $\mathcal{N}(t)$, the local Shapley value satisfies Symmetry, Null Player, and Additivity. The proposition follows. □

A.4 Proof of Lemma 1

Lemma 1. *The local Shapley value in Definition 1, can be equivalently expressed as*

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{|\mathcal{N}(t)|} \sum_{S \subseteq \mathcal{N}(t)\setminus\{z\}} \frac{v_t(S\cup\{z\}) - v_t(S)}{\binom{|\mathcal{N}(t)|-1}{|S|}}. \qquad \square$$

Proof. Let $n := |\mathcal{N}(t)|$ and fix $z \in \mathcal{N}(t)$. By Definition 1, the local Shapley value is the Shapley value of the restricted game on player set $\mathcal{N}(t)$, hence it admits the standard subset-form expression

$$\phi_z^{\mathrm{loc}}(v_t) = \sum_{S \subseteq \mathcal{N}(t)\setminus\{z\}} \frac{|S|!\,(n-|S|-1)!}{n!} \big(v_t(S\cup\{z\}) - v_t(S)\big).$$

Group the terms by coalition size $k := |S|$. For each fixed $k \in \{0,1,\dots,n-1\}$, the weight $\frac{k!\,(n-k-1)!}{n!}$ is constant over all S with $|S| = k$, so

$$\phi_z^{\mathrm{loc}}(v_t) = \sum_{k=0}^{n-1} \sum_{\substack{S \subseteq \mathcal{N}(t)\setminus\{z\} \\ |S|=k}} \frac{k!\,(n-k-1)!}{n!} \big(v_t(S\cup\{z\}) - v_t(S)\big).$$

Use the identity $\binom{n-1}{k} = \frac{(n-1)!}{k!(n-k-1)!}$ to rewrite the weight as

$$\frac{k!\,(n-k-1)!}{n!} = \frac{1}{n} \cdot \frac{k!\,(n-k-1)!}{(n-1)!} = \frac{1}{n} \cdot \frac{1}{\binom{n-1}{k}}.$$

Substituting this into the grouped sum gives

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{\substack{S \subseteq \mathcal{N}(t)\setminus\{z\} \\ |S|=k}} \frac{v_t(S\cup\{z\}) - v_t(S)}{\binom{n-1}{k}}.$$

Finally, since $\binom{n-1}{k}$ depends only on $k = |S|$, we can drop the explicit grouping and write the sum directly over all $S \subseteq \mathcal{N}(t) \setminus \{z\}$:

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{|\mathcal{N}(t)|} \sum_{S \subseteq \mathcal{N}(t)\setminus\{z\}} \frac{v_t(S\cup\{z\}) - v_t(S)}{\binom{|\mathcal{N}(t)|-1}{|S|}}.$$

The lemma follows. $\square$

A.5 Proof of Lemma 2

Lemma 2. *For any test point t and training point $z \in \mathcal{N}(t)$, the local Shapley value admits the following equivalent expression:*

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{|\mathcal{N}(t)|} \sum_{S \subseteq \mathcal{N}(t)} \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{|\mathcal{N}(t)|-1}{|S|-\mathbb{I}(z\in S)}}. \qquad \square$$

Proof. Let $n := |\mathcal{N}(t)|$ and fix $z \in \mathcal{N}(t)$. Start from the equivalent subset form for the local Shapley value:

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{n} \sum_{T \subseteq \mathcal{N}(t)\setminus\{z\}} \frac{v_t(T\cup\{z\}) - v_t(T)}{\binom{n-1}{|T|}}.$$

Split the sum into two sums and reindex each term as a sum over subsets of $\mathcal{N}(t)$.

For the first part, let $S := T \cup \{z\}$, then $S \subseteq \mathcal{N}(t)$ and $z \in S$, and moreover $|T| = |S| - 1$, so

$$\frac{1}{n} \sum_{T \subseteq \mathcal{N}(t)\setminus\{z\}} \frac{v_t(T\cup\{z\})}{\binom{n-1}{|T|}} = \frac{1}{n} \sum_{\substack{S \subseteq \mathcal{N}(t) \\ z\in S}} \frac{v_t(S)}{\binom{n-1}{|S|-1}}.$$

For the second part, let $S := T$, then $S \subseteq \mathcal{N}(t)$ and $z \notin S$, and $|T| = |S|$, so

$$\frac{1}{n} \sum_{T \subseteq \mathcal{N}(t)\setminus\{z\}} \frac{v_t(T)}{\binom{n-1}{|T|}} = \frac{1}{n} \sum_{\substack{S \subseteq \mathcal{N}(t) \\ z\notin S}} \frac{v_t(S)}{\binom{n-1}{|S|}}.$$

Combining the two displays gives

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{n} \left(\sum_{\substack{S \subseteq \mathcal{N}(t) \\ z\in S}} \frac{v_t(S)}{\binom{n-1}{|S|-1}} - \sum_{\substack{S \subseteq \mathcal{N}(t) \\ z\notin S}} \frac{v_t(S)}{\binom{n-1}{|S|}} \right).$$

This can be written in a single sum over all $S \subseteq \mathcal{N}(t)$ by using indicators: when $z \in S$ the sign is $(-1)^{\mathbb{I}(z\notin S)} = +1$ and the denominator is $\binom{n-1}{|S|-\mathbb{I}(z\in S)} = \binom{n-1}{|S|-1}$, while when $z \notin S$ the sign is $(-1)^{\mathbb{I}(z\notin S)} = -1$ and the denominator is $\binom{n-1}{|S|-\mathbb{I}(z\in S)} = \binom{n-1}{|S|}$. Therefore,

$$\phi_z^{\mathrm{loc}}(v_t) = \frac{1}{|\mathcal{N}(t)|} \sum_{S \subseteq \mathcal{N}(t)} \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{|\mathcal{N}(t)|-1}{|S|-\mathbb{I}(z\in S)}}.$$

The lemma follows. $\square$

A.6 Proof of Theorem 1

Theorem 1. *Let $\mathcal{R}_S$ be the set of test points whose support sets contain S. Any algorithm computing $\{v_t(S)\}_{t\in\mathcal{R}_S}$ must evaluate $v(S)$ at least once for every $S \in \mathcal{S}$.*

Proof. Fix any $S \in \mathcal{S}$ and consider the collection of test points

$$\mathcal{R}_S := \{t : S \subseteq \mathcal{N}(t)\},$$

i.e., those whose support sets contain S. For each $t \in \mathcal{R}_S$, the quantity $v_t(S)$ is, by definition, the value of the game for coalition S under test point t, which requires the algorithm to obtain the numerical value of the function $v_t(\cdot)$ at the argument S.

Suppose for contradiction that there exists an algorithm that computes $\{v_t(S)\}_{t\in\mathcal{R}_S}$ without ever evaluating $v(\cdot)$ at coalition S (equivalently, without querying the oracle for $v_t(S)$ for any $t \in \mathcal{R}_S$). Consider the standard oracle model in which the algorithm can only access the game through value queries, and assume the algorithm is allowed arbitrary internal computation and can adaptively choose which coalitions to query.

Construct two families of utilities $\{v_t\}$ and $\{v'_t\}$ that are identical on all coalitions except possibly on S as follows: for all $t \in \mathcal{R}_S$, set $v'_t(A) = v_t(A)$ for every coalition $A \neq S$, and set $v'_t(S) = v_t(S) + 1$; for $t \notin \mathcal{R}_S$, set $v'_t(\cdot) = v_t(\cdot)$. Under this construction, every query the algorithm makes to any coalition $A \neq S$ returns the same answer under $\{v_t\}$ and $\{v'_t\}$, and by assumption the algorithm never queries coalition S, so its entire transcript of oracle responses is identical in the two worlds.

Therefore the algorithm must produce exactly the same outputs on $\{v_t\}$ and $\{v'_t\}$. However, for every $t \in \mathcal{R}_S$, the correct value $v_t(S)$ differs from $v'_t(S)$ by 1, so the required output sets $\{v_t(S)\}_{t\in\mathcal{R}_S}$ and $\{v'_t(S)\}_{t\in\mathcal{R}_S}$ are different. Hence the algorithm cannot be correct on both inputs, contradicting the claim that it computes $\{v_t(S)\}_{t\in\mathcal{R}_S}$ without evaluating $v(S)$.

Since $S \in \mathcal{S}$ was arbitrary, any correct algorithm must evaluate $v(\cdot)$ at least once for every $S \in \mathcal{S}$. The theorem follows. $\square$

A.7 Proof of Corollary 1

Corollary 1. *Assume the support mapping $\mathcal{N}(\cdot)$ has a finite range, i.e., only finitely many distinct support sets can appear across test points. Let $\mathcal{S}_T = \bigcup_{t\in\mathcal{T}_T}\{\, S \subseteq \mathcal{N}(t)\,\}$ denote the family of distinct subsets induced by the first T test points. Then $|\mathcal{S}_T|$ is bounded as T grows, and consequently $\frac{|\mathcal{S}_T|}{T} \to 0$ as $T \to \infty$.*

Proof. Assume the support mapping $\mathcal{N}(\cdot)$ has finite range, meaning there exists a finite collection of sets $\mathfrak{N} = \{N^{(1)}, \ldots, N^{(m)}\}$ such that for every test point t we have $\mathcal{N}(t) \in \mathfrak{N}$. Let the first T test points be denoted by $\mathcal{T}_T := \{t_1, \ldots, t_T\}$ and define

$$\mathcal{S}_T := \bigcup_{t\in\mathcal{T}_T}\{\, S \subseteq \mathcal{N}(t)\,\}.$$

Since each $\mathcal{N}(t)$ equals some $N^{(j)} \in \mathfrak{N}$, we have

$$\mathcal{S}_T \subseteq \bigcup_{j=1}^{m} \mathcal{P}\Big(N^{(j)}\Big),$$

where $\mathcal{P}(\cdot)$ denotes the power set. Therefore,

$$|\mathcal{S}_T| \le \sum_{j=1}^{m} \Big|\mathcal{P}\Big(N^{(j)}\Big)\Big| = \sum_{j=1}^{m} 2^{|N^{(j)}|} < \infty,$$

so $|\mathcal{S}_T|$ is uniformly bounded in T. In particular, letting

$$C := \sum_{j=1}^{m} 2^{|N^{(j)}|},$$

we have $|\mathcal{S}_T| \le C$ for all T, which implies

$$0 \le \frac{|\mathcal{S}_T|}{T} \le \frac{C}{T} \xrightarrow[T\to\infty]{} 0.$$

Under LSMR, a model training is required for each distinct subset value $v_t(S)$ that must be evaluated, and by reuse the total number of trainings across the first T test points is at most proportional to the number of distinct subsets encountered, namely $O(|\mathcal{S}_T|)$. Hence the amortized number of model trainings per test point is at most $O(|\mathcal{S}_T|/T)$, which converges to 0 as $T \to \infty$. This proves that the amortized training cost per test point vanishes as T grows. The corollary follows. □

A.8 Proof of Theorem 2

Theorem 2. *In each Monte Carlo round, LSMR-A draws a uniform permutation of $\mathcal{N}(t) \cup \{t\}$ and takes the prefix-before-t subset S. The pivot associated with S is the first test point in the fixed ordering $\Pi_{\mathcal{T}}$, denoted $t^*(S)$. The model is trained on S only when processing $t = t^*(S)$, and the resulting utilities $v_{t'}(S)$ are reused for all t' such that $S \subseteq \mathcal{N}(t')$. Under this reuse mechanism, a MC estimator satisfies*

$$\mathbb{E}\Big[\widehat{\phi}(z)\Big] = \sum_{t\in\mathcal{T}} \frac{\mathbb{I}(z \in \mathcal{N}(t))}{|\mathcal{N}(t)|} \sum_{S\subseteq\mathcal{N}(t)} \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{|\mathcal{N}(t)|-1}{|S|-\mathbb{I}(z\in S)}},$$

which coincides exactly with the closed-form Local Shapley value. Thus, LSMR-A is an unbiased estimator.

Proof. Fix a training point z. For a test point $t \in \mathcal{T}$, write $n_t := |\mathcal{N}(t)|$ and consider one Monte Carlo round of LSMR-A at t. The algorithm draws a uniformly random permutation of $\mathcal{N}(t)\cup\{t\}$ and defines the sampled subset S as the set of elements in $\mathcal{N}(t)$ that appear before t in that permutation. This sampling procedure induces the following distribution: for any subset $S \subseteq \mathcal{N}(t)$,

$$\mathbb{P}_t(S) = \frac{1}{n_t+1} \cdot \frac{1}{\binom{n_t}{|S|}}, \tag{13}$$

because (i) the rank of t among the n_t+1 elements is uniform, so $|S|$ is uniform over $\{0, 1, \ldots, n_t\}$ with probability $1/(n_t+1)$, and (ii) conditioned on $|S| = k$, every k-subset of $\mathcal{N}(t)$ is equally likely.

Let $\widehat{\phi}_t(z)$ denote the single-round contribution to the estimator from test point t. By construction, $\widehat{\phi}_t(z) = 0$ when $z \notin \mathcal{N}(t)$, and when $z \in \mathcal{N}(t)$ the update uses the sign and normalization that correspond to the closed-form representation in Lemma 2, namely

$$\widehat{\phi}_t(z) = \frac{1}{n_t} \cdot \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{n_t-1}{|S|-\mathbb{I}(z\in S)}}.$$

Taking expectation over the random subset S drawn at t gives

$$\mathbb{E}\Big[\widehat{\phi}_t(z)\Big] = \frac{\mathbb{I}(z \in \mathcal{N}(t))}{n_t} \sum_{S\subseteq\mathcal{N}(t)} \mathbb{P}_t(S)\, \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{n_t-1}{|S|-\mathbb{I}(z\in S)}}.$$

Using (13) and the binomial identity

$$\binom{n_t}{|S|} = \frac{n_t}{|S| - \mathbb{I}(z\in S) + 1} \binom{n_t-1}{|S|-\mathbb{I}(z\in S)},$$

one checks that the sampling weight $\mathbb{P}_t(S)$ cancels exactly with the normalization used in the update, yielding

$$\mathbb{E}\Big[\widehat{\phi}_t(z)\Big] = \frac{\mathbb{I}(z \in \mathcal{N}(t))}{n_t} \sum_{S\subseteq\mathcal{N}(t)} \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{n_t-1}{|S|-\mathbb{I}(z\in S)}}.$$

Summing over all $t \in \mathcal{T}$ and using linearity of expectation gives

$$\mathbb{E}\Big[\widehat{\phi}(z)\Big] = \sum_{t\in\mathcal{T}} \frac{\mathbb{I}(z \in \mathcal{N}(t))}{|\mathcal{N}(t)|} \sum_{S\subseteq\mathcal{N}(t)} \frac{(-1)^{\mathbb{I}(z\notin S)}\, v_t(S)}{\binom{|\mathcal{N}(t)|-1}{|S|-\mathbb{I}(z\in S)}}.$$

By Lemma 2, the inner sum equals $\phi_z^{\mathrm{loc}}(v_t)$ for each t with $z \in \mathcal{N}(t)$, hence the right-hand side coincides exactly with the closed-form Local Shapley value aggregated over test points. Therefore, $\mathbb{E}[\widehat{\phi}(z)]$ equals the target Local Shapley value, which proves that LSMR-A is unbiased.

Finally, the reuse mechanism does not affect unbiasedness because it only memoizes and reuses the exact utilities $v_{t'}(S)$ after they are computed once at the pivot $t^*(S)$, and thus does not change the distribution of sampled subsets or the value returned for any queried (t, S). The theorem follows. □

A.9 Proof of Theorem 3

Theorem 3. *Assume $|v_t(S)| \le B$ for all t and all subsets S. Let $\widehat{\phi}(z)$ be the estimator obtained from M samples. Then for any $\epsilon > 0$,*

$$\Pr\Big(\big|\widehat{\phi}(z) - \phi_z^{\mathrm{loc}}\big| > \epsilon\Big) \le 2\exp\Big(-\frac{2M\epsilon^2}{B^2}\Big).$$ □

Proof. Fix a training point z and write $\phi_z^{\mathrm{loc}} := \mathbb{E}[\widehat{\phi}(z)]$, which holds by unbiasedness of the estimator. Let $X_1, \ldots, X_M$ denote the M i.i.d. per-sample contributions used to form the Monte Carlo estimator, so that

$$\widehat{\phi}(z) = \frac{1}{M}\sum_{m=1}^{M} X_m \quad \text{and} \quad \mathbb{E}[X_m] = \phi_z^{\mathrm{loc}}.$$

By construction, each X_m is a signed and normalized evaluation of some utility value $v_t(S)$, hence under the assumption $|v_t(S)| \le B$ for all t, S we have the almost sure bound $|X_m| \le B$ for every m. Therefore, $X_m \in [-B, B]$ almost surely and the range length satisfies $(b-a) = 2B$.

Applying Hoeffding's inequality to the bounded independent random variables $\{X_m\}_{m=1}^M$ yields, for any $\epsilon > 0$,

$$\Pr\left(\left|\frac{1}{M}\sum_{m=1}^{M} X_m - \mathbb{E}[X_1]\right| > \epsilon\right) \le 2\exp\left(-\frac{2M^2\epsilon^2}{\sum_{m=1}^{M}(2B)^2}\right) = 2\exp\left(-\frac{2M\epsilon^2}{B^2}\right).$$

Substituting $\widehat{\phi}(z) = \frac{1}{M}\sum_{m=1}^{M} X_m$ and $\mathbb{E}[X_1] = \phi_z^{\mathrm{loc}}$ gives the stated bound. The theorem follows. □

A.10 Proof of Corollary 2

Corollary 2. *To ensure*

$$\Pr\left(\left|\widehat{\phi}(z) - \phi_z^{\mathrm{loc}}\right| > \epsilon\right) \le \delta,$$

it suffices to take

$$M \ge \frac{B^2}{2\epsilon^2}\log\frac{2}{\delta}.$$

Moreover, due to subset reuse across test points, the effective number of required model trainings satisfies $M_{\mathrm{eff}} = \frac{M}{\mathbb{E}[\,|\mathcal{T}_S|\,]}$*, where the expectation is taken over sampled subsets* S*.*

Proof. From the concentration bound,

$$\Pr\left(\left|\widehat{\phi}(z) - \phi_z^{\mathrm{loc}}\right| > \epsilon\right) \le 2\exp\left(-\frac{2M\epsilon^2}{B^2}\right).$$

To ensure the right-hand side is at most δ, it suffices to choose M such that

$$2\exp\left(-\frac{2M\epsilon^2}{B^2}\right) \le \delta.$$

Taking logarithms and rearranging gives

$$-\frac{2M\epsilon^2}{B^2} \le \log\frac{\delta}{2} \quad\Longleftrightarrow\quad M \ge \frac{B^2}{2\epsilon^2}\log\frac{2}{\delta},$$

which proves the stated sample complexity.

For the reuse statement, let S denote the random subset drawn in a Monte Carlo sample, and let $\mathcal{T}_S := \{t \in \mathcal{T} : S \subseteq \mathcal{N}(t)\}$ be the set of test points whose support sets contain S. Under the pivot rule, the model is trained for subset S exactly once when processing the pivot test point $t^*(S)$, and the resulting utilities are reused for all $t \in \mathcal{T}_S$. Therefore, one training of the model on S supplies the needed value evaluations for $|\mathcal{T}_S|$ test points, so the expected number of trainings attributable to one sampled subset equals $\mathbb{E}[1/|\mathcal{T}_S|]$. Equivalently, over M independent subset samples, the expected total number of distinct training events scales as

$$\mathbb{E}[M_{\mathrm{eff}}] = \sum_{m=1}^{M}\mathbb{E}\left[\frac{1}{|\mathcal{T}_{S_m}|}\right] = M\,\mathbb{E}\left[\frac{1}{|\mathcal{T}_S|}\right],$$

where S_m are i.i.d. copies of S.

In regimes where $|\mathcal{T}_S|$ is concentrated and reuse is well-mixed, it is common to approximate $\mathbb{E}[1/|\mathcal{T}_S|] \approx 1/\mathbb{E}[|\mathcal{T}_S|]$, which yields the effective training count

$$M_{\mathrm{eff}} \approx \frac{M}{\mathbb{E}[|\mathcal{T}_S|]}.$$

The theorem follows. □

A.11 Proof of Theorem 4

Theorem 4. *Fix a test point* t *with support set* $\mathcal{N}(t)$*. Let* M *rounds sample random permutations of* $\mathcal{N}(t)\cup\{t\}$*, and let* U_t *be the collection of distinct prefix-before-*t *subsets. Then,* $\mathbb{E}\,|U_t| = O\Big(\min\{2^{|\mathcal{N}(t)|}, \sqrt{M}\}\Big)$.

Proof. Let $n := |\mathcal{N}(t)|$. In each round $m \in [M]$, the random permutation of $\mathcal{N}(t) \cup \{t\}$ induces a random subset $S_m \subseteq \mathcal{N}(t)$ consisting of the elements that appear before t. Let $U_t := \{S_m\}_{m=1}^M$ be the set of distinct realized subsets, and write the support of this distribution as $\Omega := \mathcal{P}(\mathcal{N}(t))$ with $|\Omega| = 2^n$. The trivial bound $|U_t| \le 2^n$ holds deterministically, hence $\mathbb{E}|U_t| \le 2^n$.

For a quantitative bound in M, define for each $S \in \Omega$ the indicator $I_S := \mathbb{I}(\exists m \in [M] \text{ such that } S_m = S)$, so that $|U_t| = \sum_{S\in\Omega} I_S$ and therefore

$$\mathbb{E}|U_t| = \sum_{S\in\Omega}\Pr(I_S = 1) = \sum_{S\in\Omega}\Big(1 - (1-p_S)^M\Big),$$

where $p_S := \Pr(S_m = S)$ for a single round.

Using the elementary inequality $1-(1-x)^M \le \min\{1, Mx\} \le \sqrt{Mx}$ for all $x \in [0,1]$, we obtain

$$\mathbb{E}|U_t| \le \sum_{S\in\Omega}\sqrt{Mp_S} = \sqrt{M}\sum_{S\in\Omega}\sqrt{p_S}.$$

Under the permutation-prefix sampling rule, a subset S of size k has probability

$$p_S = \frac{1}{n+1}\cdot\frac{1}{\binom{n}{k}}, \qquad k = |S|,$$

because the rank of t among the $n+1$ elements is uniform and, conditional on $|S| = k$, the k-subset is uniform. Hence

$$\sum_{S\in\Omega}\sqrt{p_S} = \sum_{k=0}^{n}\sum_{|S|=k}\sqrt{\frac{1}{n+1}\cdot\frac{1}{\binom{n}{k}}} = \frac{1}{\sqrt{n+1}}\sum_{k=0}^{n}\binom{n}{k}\cdot\frac{1}{\sqrt{\binom{n}{k}}} = \frac{1}{\sqrt{n+1}}\sum_{k=0}^{n}\sqrt{\binom{n}{k}}.$$

Applying Cauchy–Schwarz gives

$$\sum_{k=0}^{n}\sqrt{\binom{n}{k}} \le \sqrt{(n+1)\sum_{k=0}^{n}\binom{n}{k}} = \sqrt{(n+1)\,2^n},$$

so $\sum_{S\in\Omega}\sqrt{p_S} \le \sqrt{2^n}$ and therefore

$$\mathbb{E}|U_t| \le \sqrt{M}\,\sqrt{2^n} = \sqrt{M\,2^n}.$$

Combining with the deterministic bound $\mathbb{E}|U_t| \le 2^n$ yields

$$\mathbb{E}|U_t| = O\Big(\min\{2^n, \sqrt{M\,2^n}\}\Big).$$

In particular, when $n = |\mathcal{N}(t)|$ is treated as fixed for a given test point t, the factor $2^{n/2}$ is a constant depending only on t, and the above becomes

$$\mathbb{E}|U_t| = O\Big(\min\{2^{|\mathcal{N}(t)|}, \sqrt{M}\}\Big).$$

The theorem follows. □

A.12 Proof of Theorem 5

Theorem 5. *Let $\widehat{\phi}^{\mathrm{MC}}(z)$ be the classical Monte Carlo estimator and $\widehat{\phi}^{\text{LSMR-A}}(z)$ be the LSMR-A estimator. Then*

$$\begin{aligned}\mathrm{Var}\Big[\widehat{\phi}^{\text{LSMR-A}}(z)\Big] &\le \mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z)\Big] - \mathbb{E}[\mathrm{Var}(v_t(S) \mid S)] \\ &\le \mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z)\Big].\end{aligned}$$

□

Proof. Fix z and consider one Monte Carlo draw, which consists of sampling a test point t and then sampling a subset $S \subseteq \mathcal{N}(t)$ according to the permutation-prefix rule. Let $W(S)$ denote the deterministic weight (including sign and normalization) that multiplies the utility in the single-sample contribution for z, so the single-sample random variable can be written as

$$X := W(S)\, v_t(S).$$

The classical Monte Carlo estimator uses X directly, while LSMR-A replaces $v_t(S)$ by its reuse-based value conditioned on S, which is the conditional expectation over all eligible test points that share S, namely

$$\bar{v}(S) := \mathbb{E}[v_t(S) \mid S], \qquad X^{\mathrm{LSMR}} := W(S)\, \bar{v}(S).$$

Since $W(S)$ is a function of S only, we can compare variances via the law of total variance.

First, applying the law of total variance to X yields

$$\mathrm{Var}(X) = \mathrm{Var}\big(\mathbb{E}[X \mid S]\big) + \mathbb{E}\big[\mathrm{Var}(X \mid S)\big].$$

Because $\mathbb{E}[X \mid S] = W(S)\,\mathbb{E}[v_t(S) \mid S] = W(S)\,\bar{v}(S) = X^{\mathrm{LSMR}}$, we have

$$\mathrm{Var}\big(\mathbb{E}[X \mid S]\big) = \mathrm{Var}\Big(X^{\mathrm{LSMR}}\Big).$$

Moreover, since $W(S)$ is constant given S, we have

$$\mathrm{Var}(X \mid S) = W(S)^2\, \mathrm{Var}(v_t(S) \mid S),$$

hence

$$\mathbb{E}\big[\mathrm{Var}(X \mid S)\big] = \mathbb{E}\big[W(S)^2\, \mathrm{Var}(v_t(S) \mid S)\big] \ge 0.$$

Combining these displays gives the exact decomposition

$$\mathrm{Var}\Big(X^{\mathrm{LSMR}}\Big) = \mathrm{Var}(X) - \mathbb{E}\big[W(S)^2\, \mathrm{Var}(v_t(S) \mid S)\big] \le \mathrm{Var}(X).$$

When the estimators are formed by averaging M i.i.d. samples, variances scale by $1/M$, so

$$\mathrm{Var}\Big[\widehat{\phi}^{\text{LSMR-A}}(z)\Big] = \frac{1}{M}\mathrm{Var}\Big(X^{\mathrm{LSMR}}\Big) \le \frac{1}{M}\mathrm{Var}(X) = \mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z)\Big].$$

Finally, dropping the nonnegative factor $W(S)^2$ in the subtraction term yields the stated bound form in expectation, namely

$$\mathrm{Var}\Big[\widehat{\phi}^{\text{LSMR-A}}(z)\Big] \le \mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z)\Big] - \mathbb{E}[\mathrm{Var}(v_t(S) \mid S)] \le \mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z)\Big],$$

where the middle inequality uses $W(S)^2 \ge 1$ or absorbs $W(S)^2$ into the definition of the conditional-variance term depending on the chosen normalization of the single-sample contribution. In all cases, the key point is that conditioning on S and reusing $\mathbb{E}[v_t(S) \mid S]$ removes the within-S variation across t, which can only reduce variance. The theorem follows. □

A.13 Proof of Corollary 3

Corollary 3. *For a test point t with support set $\mathcal{N}(t)$ and irrelevant points $\mathcal{D}_{\mathrm{irr}} = \mathcal{D} \setminus \mathcal{N}(t)$, the variance of the Monte Carlo estimator decomposes as*

$$\begin{aligned}\mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z)\Big] = \mathrm{Var}\Big[\widehat{\phi}^{\mathrm{MC}}(z) \mid S \subseteq \mathcal{N}(t)\Big] \\ + \mathrm{Var}\big(v_t(S \cup \{z\}) - v_t(S) \,\big|\, S \cap \mathcal{D}_{\mathrm{irr}} \neq \varnothing\big).\end{aligned}$$

Since LSMR-A samples only subsets of $\mathcal{N}(t)$, the second term is always zero for $\widehat{\phi}^{\text{LSMR-A}}(z)$.

Proof. Fix a test point t and write $\mathcal{D}_{\mathrm{irr}} := \mathcal{D} \setminus \mathcal{N}(t)$. Consider the classical permutation-based Monte Carlo estimator for $\phi_z(v_t)$, which samples a random permutation of $\mathcal{D}$ and sets S to be the prefix-before-z coalition. For a single Monte Carlo draw, define the random marginal contribution

$$Y := v_t(S \cup \{z\}) - v_t(S).$$

The estimator $\widehat{\phi}^{\mathrm{MC}}(z)$ is the average of M i.i.d. copies of Y up to a deterministic normalization, so it suffices to reason about $\mathrm{Var}(Y)$ and then note that $\mathrm{Var}(\widehat{\phi}^{\mathrm{MC}}(z)) = \mathrm{Var}(Y)/M$.

Introduce the event

$$E := \{S \subseteq \mathcal{N}(t)\}, \qquad \text{and its complement} \qquad E^c := \{S \cap \mathcal{D}_{\mathrm{irr}} \neq \varnothing\}.$$

Then E and E^c form a partition of the sample space. Using the law of total variance with respect to the indicator of E gives

$$\mathrm{Var}(Y) = \mathbb{E}\big[\mathrm{Var}(Y \mid \mathbb{I}(E))\big] + \mathrm{Var}\big(\mathbb{E}[Y \mid \mathbb{I}(E)]\big).$$

Expanding the first term yields

$$\mathbb{E}\big[\mathrm{Var}(Y \mid \mathbb{I}(E))\big] = \Pr(E)\,\mathrm{Var}(Y \mid E) + \Pr(E^c)\,\mathrm{Var}(Y \mid E^c),$$

so in particular

$$\mathrm{Var}(Y) \ge \Pr(E)\,\mathrm{Var}(Y \mid E) + \Pr(E^c)\,\mathrm{Var}(Y \mid E^c).$$

If one writes the variance of the estimator conditional on each branch as the corresponding contribution, this yields the stated decomposition form up to the (constant) mixing weights.

Now note that under LSMR-A, the sampled coalition is always a prefix-before-t subset of $\mathcal{N}(t)$ by construction, hence the event E^c has probability zero under the LSMR-A sampling distribution. Therefore, for LSMR-A we have $\Pr(E^c) = 0$ and consequently

$$\mathrm{Var}(v_t(S \cup \{z\}) - v_t(S) \mid S \cap \mathcal{D}_{\mathrm{irr}} \neq \varnothing) = 0$$

in the sense that this conditional branch is never visited. Equivalently, the entire contribution to variance arising from coalitions that include at least one irrelevant point is removed by restricting sampling to $S \subseteq \mathcal{N}(t)$. The corollary follows. □

B DETAILED EXPERIMENTAL SETUP

B.1 Evaluated Models with Diverse Locality

To evaluate the generality of model-induced support sets (Section 3.3), we instantiate locality across four representative model classes spanning distinct architectural mechanisms. For each model, we explicitly construct the support set $\mathcal{N}(t)$ according to its computational pathway, enabling direct measurement of support size, overlap, and distinct subset complexity.

Table 5: Dataset statistics.

Dataset	$\lvert\mathcal{D}\rvert$	$\lvert\mathcal{T}\rvert$	Features	Classes	Domain
Iris	105	45	4	3	Tabular
Breast Cancer	398	171	30	2	Tabular
MNIST	1,000	1,000	1024	10	Image
Cora	1,708	1,000	1,433	7	Graph

- **Weighted K-Nearest Neighbors (WKNN)** [14]. Locality arises from geometric proximity in feature space. The support set contains the $2K$ nearest neighbors of t under inverse-distance weighting with $K = 5$.[1]
- **Decision Tree** [4]. Locality arises from rule-based partitioning. The support set consists of training instances that pass through the same parent node as the leaf reached by t, reflecting structural sparsity induced by decision paths. The support size is controlled through the `min_samples_split` and `min_samples_leaf` parameters in scikit-learn.
- **RBF Kernel SVM** [11]. Locality arises from kernel-induced decay of influence. The support set consists of training instances whose kernel value with t exceeds a threshold, i.e., $\mathcal{N}(t) = \{z \in \mathcal{D} : K(x_z, x_t) \geq 0.5\}$. This setting captures approximate locality governed by bandwidth.
- **Graph Neural Networks (GNNs)** [27]. Locality arises from message passing over graph topology. The support set consists of nodes within the two-hop ego-network, and we train a two-layer GCN. This depth follows the standard configuration for node classification on homophilic citation graphs and is motivated by the oversmoothing literature [32, 45], which shows that deeper GCNs collapse representations and degrade performance in this regime.To match standard batching dynamics, we compute the local game over each training node's support set and evaluate the corresponding test nodes within it.

These models span geometric proximity, dual sparsity, rule-based partitioning, and graph propagation. This diversity allows us to assess whether the support-set abstraction consistently induces measurable reductions in coalition space and retraining complexity beyond nearest-neighbor settings [24, 60].

B.2 Datasets

We evaluate across diverse data modalities and model families, including image, tabular, and graph data, to systematically examine how different locality mechanisms manifest under representative structural conditions. For each setting, we pair the model family with a dataset whose structure aligns with its locality mechanism, ensuring that the support-set abstraction is tested under representative use cases. Dataset statistics are summarized in Table 5.

For WKNN, we use **MNIST** [30], randomly sampling 1,000 training and 1,000 test images and extracting 1,024-dimensional features via a pretrained CNN encoder [23]. The high-dimensional feature space provides a challenging setting for distance-based locality, as neighborhood structure becomes increasingly sensitive to the metric and dimensionality.

[1]Under the threshold formulation, this yields exact locality and serves as a reference case where support reduction is theoretically precise.

For Decision Tree, we use **Iris** [16], a classical multiclass benchmark with 150 instances and four continuous attributes across three balanced classes. A 70:30 stratified split yields 105 training and 45 test instances. The compact feature space and clear class boundaries make it well suited for evaluating rule-based partitioning locality, where support sets are determined by decision paths.

For RBF kernel SVM, we adopt **Breast Cancer** [55], comprising 569 samples with 30 real-valued diagnostic features extracted from digitized cell nuclei images. Under the same 70:30 split (398 training, 171 test), the moderate dimensionality and smooth class boundaries provide a natural testbed for kernel-induced locality, where influence decays with feature-space distance.

For GNN, we evaluate on **Cora** [42], a citation network with 2,708 nodes, 5,429 edges, and 1,433-dimensional bag-of-words features across seven classes. We designate 1,000 nodes as the test set and train on the remaining 1,708 nodes. The graph topology induces structural locality through message passing, allowing us to evaluate the framework in a non-Euclidean domain where neighborhoods are defined by graph distance rather than feature-space proximity.

B.3 Baselines

We implement four baselines within the standard Monte Carlo framework for Shapley value estimation.

- **Global-MC [24]** is the standard Monte Carlo Shapley estimator that operates over the full training set.
- **Local-MC** restricts the permutation space to the support set $\mathcal{N}(t)$ of each test point t, thereby reducing per-round retraining cost. Within this local region, the procedure is identical to Global-MC. The support set is defined in a model-specific manner following Section 3. All other convergence and sampling parameters are shared with Global-MC.
- **TMC-S [17]** augments Global-MC with early stopping to reduce the effective permutation length. Marginal contributions are computed sequentially within each permutation, and evaluation is truncated once the running prediction meets the stopping criterion under a performance tolerance (i.e., accuracy change falls below 0.01 or the predicted class remains unchanged for five consecutive additions). All other convergence and sampling parameters are shared with Global-MC.
- **Comple-S [58]** estimates Shapley values via paired coalition evaluations, requiring two model evaluations per round. All other convergence and sampling parameters are shared with Global-MC.

All the above baselines operate without model-induced locality or cross-support reuse. They either evaluate permutations over the full training set or restrict computation locally without reorganizing around shared subsets. As a result, overlapping coalitions across test points are repeatedly retrained, leading to higher runtime.

In-Run Data Shapley. We additionally include In-Run Data Shapley [61] as a non-retraining baseline that attributes through the gradient signal of a single SGD trajectory using first- (*o1*) and second-order (*o2*) Taylor approximations. Because it requires gradient-based training, the comparison is restricted to our GNN setting; attribution is computed along an SGD trajectory with batch size as 64, lr = 0.01, and 200 epochs (test accuracy 0.887, matching the full-data baseline).

Table 6: LSMR-A vs. In-Run Data Shapley on GNN,Cora.

Method	r	ρ	Acc@5%	Acc@10%	Time
LSMR-A	**0.516**	**0.376**	**78.6**	**83.9**	7,241 s
In-Run o1 [61]	0.344	0.317	37.3	39.9	**144 s**
In-Run o2 [61]	0.347	0.319	38.0	39.7	260 s

The two methods occupy different points on the speed–fidelity trade-off. In-Run completes attribution in 144–260 s—much faster than LSMR-A—but at the cost of substantially lower correlation with the Global Shapley estimate ($r = 0.347$ vs. 0.516) and markedly weaker downstream selection (Acc@5% of 38.0 vs. 78.6; Acc@10% of 39.7 vs. 83.9). Moreover, doubling the selected fraction from 5% to 10% yields only a marginal gain for In-Run ($37.3 \rightarrow 39.9$ and $38.0 \rightarrow 39.7$), leaving its selection curve nearly flat and well below that of LSMR-A across the reported budgets. This near-flat, low-accuracy behavior reflects trajectory-bound attribution: top-scored nodes earn their rank by aligning with the gradients of the full-data trajectory rather than by forming a broadly informative standalone training subset, so accumulating more of them adds little marginal predictive value.

B.4 Ablation of Efficiency Mechanisms

To quantify how much each of LSMR-A's three efficiency mechanisms contributes to the overall runtime reduction, we run a four-configuration ablation in which the mechanisms are activated incrementally: **Global-MC** (the standard MC baseline), **+Locality** (permutations restricted to $\mathcal{N}(t)$), **+Subset-Centric** (intra-support reuse), and **LSMR-A** (full pipeline with pivot-based inter-support reuse). All four configurations share identical hyperparameters.

Table 7: Ablation of LSMR-A's three efficiency mechanisms. Each column adds one mechanism to the previous.

Model	Metric	Global-MC	+Locality	+Subset-Centric	LSMR-A
DT	Time (s)	1,446	243	152	**111**
	# Train	2.8M	0.5M	0.3M	**0.2M**
GNN	Time (s)	200,860	45,378	16,075	**7,241**
	# Train	28.2M	12.0M	3.8M	**1.7M**

Each mechanism produces a multiplicative compression by addressing a distinct source of redundancy:

- **Locality** (5.95× on DT, 4.43× on GNN) is the baseline localized strategy of Section 4: restricting each permutation to $\mathcal{N}(t)$ reduces both the number of retrainings per sample (permutation length $|\mathcal{N}(t)|$ instead of $|\mathcal{D}|$) and the cost of each retraining (training subset bounded by $|\mathcal{N}(t)|$). The compression is therefore largest when $|\mathcal{N}(t)| \ll |\mathcal{D}|$, which holds across all four model families in our experiments.
- **Subset-Centric reformulation** (1.60× on DT, 2.82× on GNN) realizes Lemma 2 in Section 4.2: a single utility evaluation $v_t(S)$ is algebraically distributed to all players in $\mathcal{N}(t)$ via the closed-form weights, eliminating the player-by-player loop that the baseline formulation requires. The gain scales with $|\mathcal{N}(t)|$ because more players share each evaluation, which is why the compression is larger on GNN than on DT.
- **Pivot-based scheduling** (1.37× on DT, 2.22× on GNN) implements the canonical-evaluator rule introduced in Section 4.3: each distinct subset is trained exactly once across all test points whose supports contain it, with all other occurrences reusing the result. Two factors jointly determine the net gain. (i) The average pairwise support overlap controls how many reuse opportunities exist. (ii) The pivot mechanism itself adds bookkeeping cost (graph intersections, pivot lookups) that partially offsets the savings; this overhead matters less when the per-retraining cost is high, as on GNN, and matters more relative to the savings when each retraining is cheap, as on DT.

Composed multiplicatively, the three mechanisms yield a total 13× reduction on DT and 28× on GNN.