ICE: TASK-ALIGNED CLIFFORD LATENT FIELDS FOR MULTIMODAL GRAPH FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal attributed graphs connect entities, visual content, language, and observed relations. Learning one foundation across such graphs requires more than compressing each node into a fused Euclidean vector. The representation must preserve entity semantics, construct interaction state from graph neighborhoods, and expose that state to prediction units with different geometry. Our empirical study shows why these requirements are inseparable. Higher-grade channels recover pair relations across the foundation graphs, specialized queries reveal information hidden by a generic readout, and rigid blade isolation removes cross-grade capacity. We therefore introduce ICE (**I**nteraction-aware **C**lifford **E**ncoder), a multimodal graph foundation model built on a node-indexed Clifford latent field. Topology, text, and images enter explicit $\mathrm{Cl}(3)$ addresses. Edge-aware geometric products transform these directions into scalar, bivector, and trivector relations over observed neighborhoods. A protected Grade-1 route preserves entity semantics, while the full grade and depth bank remains available to fresh node and link heads. We establish exact cross-grade reachability, node-permutation equivariance, and a bound on the task residual around the semantic score. Experiments span one shared foundation over eleven graphs, six node-classification datasets, three link-prediction datasets, and matched few-shot tasks. ICE ranks first in all 30 reported supervised and few-shot comparisons. Core removals reduce every task summary, and mechanism controls connect the gains to higher-order transport, retained multidepth structure, semantic protection, and direct field access.

1 INTRODUCTION

Multimodal attributed graphs (MAGs) organize entities, textual and visual attributes, and observed relations in one computable structure. A product can be ambiguous from its image and description alone yet become identifiable through related products. A media item can be understood through its synopsis and poster together with links induced by topic, audience, or creator. Graph learning must therefore go beyond within-node modality matching. It must decide which cross-entity signals to transport and which interactions matter to the current prediction (Ektefaie et al., 2023, Peng et al., 2024, Yan et al., 2024, Zhu et al., 2025).

Multimodal graph foundation models (MGFMs) raise this requirement from one dataset to many. Given topology and text/image node attributes, one dataset-agnostic backbone should support both node classification (NC) and link prediction (LP) across graphs. NC emphasizes entity-level class signals, whereas LP scores whether two node states form a relation and is especially sensitive to ordered interactions. A useful foundation must preserve entity-local semantics while exposing relation-aware context through a reusable task interface.

Existing pipelines commonly encode each modality, fuse the outputs, and propagate one Euclidean node vector. This design can learn strong representations, but it places topology, semantic directions, and cross-modal interactions in an undifferentiated coordinate space. It does not provide an algebraic operation that explicitly constructs higher-order relations from modality directions along graph edges. Scalar attention can control a neighbor's contribution but does not retain a separate address for the structured interaction being transported. A generic downstream readout can then ignore relation state that the backbone already contains.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Task-specific Clifford graph models provide a useful starting point. LION represents multimodal interactions with multivectors, so symmetric agreement and oriented relations remain accessible in different algebraic grades (Li et al., 2026). A graph foundation introduces a broader requirement. One encoder must construct those relations across heterogeneous graphs while preserving the entity state required by node prediction and the pair state required by link prediction. The open problem is therefore not whether Clifford products can model interaction, but how their structured state can be learned once and reused by prediction units with different geometry.

Figure 1 isolates three diagnostic views that a complete model comparison cannot reveal. Panel (a) follows relation information as it expands across algebraic grades and transport depth. Panel (b) shows that direct grade and depth access recovers node information overlooked by a generic readout. Panel (c) shows that permanent blade isolation removes useful cross-grade capacity. Together, these observations show that Clifford dimension alone is not the answer. A successful foundation must coordinate relation construction, representation freedom, and task access.

These findings motivate a continuous design across representation, transport, and task access. We propose ICE (**I**nteraction-aware **C**lifford **E**ncoder) as a dataset-agnostic multimodal graph foundation. ICE makes latent space concrete as a node-indexed $\mathrm{Cl}(3)$ field whose grades describe scalar, directional, planar, and volumetric coordinates. Task alignment is realized through fresh supervised and few-shot heads that query this field by grade and depth.

ICE follows one continuous information path. It assigns topology and node modalities to explicit input addresses, then composes their directions along observed graph neighborhoods. Product magnitude controls neighbor allocation, while product direction transforms the transported value and generates higher-grade relation state. An additive Grade-1 route keeps entity semantics available. The complete H0, H1, and H2 field is exposed to fresh NC and LP heads, while semantic topology contrast and masked geometric reconstruction train the same dataset-agnostic encoder. This design preserves entity semantics without reducing graph interaction to a scalar weight or a final pooled vector.

We evaluate ICE on supervised and few-shot node and link tasks spanning nine target graphs. The results consistently favor ICE across saturated and difficult datasets. Component removals and matched controls further connect the improvement to geometric transport, retained grade and depth states, semantic protection, and task-specific access.

Our Contributions. (1) *Valuable Insights.* We show that the limitation is an information-interface problem: observed edges create relation state in multiple grades and depths, while a final Euclidean readout can hide that state from a node or pair predictor. Controlled probes establish when those states improve relation prediction and task-specific access. (2) *Novel Method.* ICE is a dataset-agnostic Clifford field that lifts topology, text, and image signals into explicit addresses, uses two edge-aware cross-grade transports, retains the grade and depth bank, and protects a Grade-1 semantic route. Fresh NC and LP heads query the same field at the prediction unit they must solve, and the formal results state the reachability, equivariance, access, and residual guarantees under their assumptions. (3) *SOTA Performance.* Under a shared cross-graph protocol, ICE ranks first across six NC and three LP datasets and the matched few-shot tasks. The three task-organized matrices report supervised and low-label results, while component ablations and mechanism controls connect the gains to interaction construction, field retention, semantic protection, and task access.

2 Preliminaries and Related Work

2.1 Problem Setup and Notation

Consider a multimodal graph $\mathcal{G}_d = (\mathcal{V}_d, \mathcal{E}_d, \mathcal{R}_d)$ with $N_d = |\mathcal{V}_d|$ nodes, observed edges $\mathcal{E}_d$, and raw text and image attributes $\mathcal{R}_d$. Frozen modality encoders ϕ_m produce $\mathbf{X}_d^{(m)} = \phi_m(\mathcal{R}_d^{(m)}) \in \mathbb{R}^{N_d \times d_m}$ for $m \in \mathcal{M} = \{t, v\}$. The foundation encoder receives only $(\mathcal{V}_d, \mathcal{E}_d, \{\mathbf{X}_d^{(m)}\}_{m\in\mathcal{M}})$; dataset identity, target labels, validation decisions, and output dimensions are unavailable to its shared maps.

Foundation learning fits one encoder F_θ over the training graphs with self-supervised objectives. At transfer time, F_θ is reused and a fresh task-local head is fitted. For node classification, the prediction unit is $i \in \mathcal{V}_d$ and the output is $\hat{y}_i \in \mathcal{Y}_d$, evaluated by Accuracy and Macro-F1. For link prediction, the prediction unit is a candidate pair (i, j) and the output is a score s_{ij}, evaluated by MRR. The

objective is therefore a shared state that supports both units, not one common label space or one head shared across tasks.

This setting fixes the graph and frozen modality features presented to the encoder while allowing the downstream output space and prediction unit to change. It differs from task-specific multimodal graph learning, which may optimize one encoder for one graph and one task, and from text-attributed graph foundations, which do not have to keep image–text interaction state available. These boundaries specify what must be shared and what may remain task local.

Let $\mathcal{D}_{\mathrm{pre}}$ denote the graphs used for foundation learning and $\mathcal{T}_d$ the downstream task on graph d. The shared parameters $\theta^\star$ are fitted without target-task labels. A task-local parameter set $\omega_{d,\mathcal{T}_d}$ is then trained from the designated training split and selected with validation data. Evaluation reads the selected model once on held-out examples. The same $\theta^\star$ is used for all target graphs and both task families, whereas $\omega_{d,\mathcal{T}_d}$ may change with the output space. Thus the protocol tests reuse of one foundation, not reuse of a classifier trained elsewhere.

2.2 Multimodal Graph Learning

Multimodal graph learning combines entity attributes with relational structure. Traditional models propagate visual and textual features on one target graph or learn modality-specific interaction for one prediction setting (Wei et al., 2019, Tao et al., 2020, Yan et al., 2024, Zhu et al., 2025). These models provide strong specialized operators, but their representation boundary is chosen together with the target task. A node-level readout can therefore discard pair information, while a relation-oriented readout can weaken the entity semantics needed by node classification.

Geometric multimodal models offer a different interaction primitive. LION uses Clifford products to separate scalar agreement from higher-grade oriented interaction (Li et al., 2026). This algebraic structure is useful for representing multimodal relations, but a task-specific model does not by itself define which intermediate grades and depths a cross-graph foundation must expose to a new prediction unit.

2.3 Graph Foundation Models

Graph foundation models learn reusable structural or semantic representations through self-supervision and lightweight adaptation (Hou et al., 2023, Yu et al., 2025, Sun et al., 2025, Wang et al., 2024). Text-attributed models align graph structure with language, while multimodal foundations such as UniGraph2 and PLANET extend this interface to visual attributes (He et al., 2025, Liu et al., 2026). Their cross-graph training establishes the foundation setting, but the downstream interface is still usually a final Euclidean node state.

For node classification, a fresh head maps one node state to a graph-specific label space. For link prediction, a fresh head scores a pair against candidates. A final Euclidean node state can serve either interface, but it need not preserve the relation information constructed at earlier transport depths. Clifford products give one way to construct that information, while a reusable foundation must also decide which states to retain for an unknown future prediction unit. Section 3 examines this access question before Section 4 introduces the Clifford field.

3 Empirical Motivation

We first examine whether a multimodal graph foundation requires more than a wider Euclidean embedding. The probes ask whether observed edges create relation state, how that state develops across grades and transport depths, and whether retained grade and depth states support task-specific access. They use 9,810 matched pairs, controlled blade isolation, and direct grade and depth access (Appendix F.4).

Panel (a) evaluates grade expansion and transport depth within one relation sequence. Panel (b) compares direct grade and depth access with a generic readout. Panel (c) tests blade isolation while keeping the basis channels. These probes separate interaction, transport, retention, and access before the complete method is introduced.

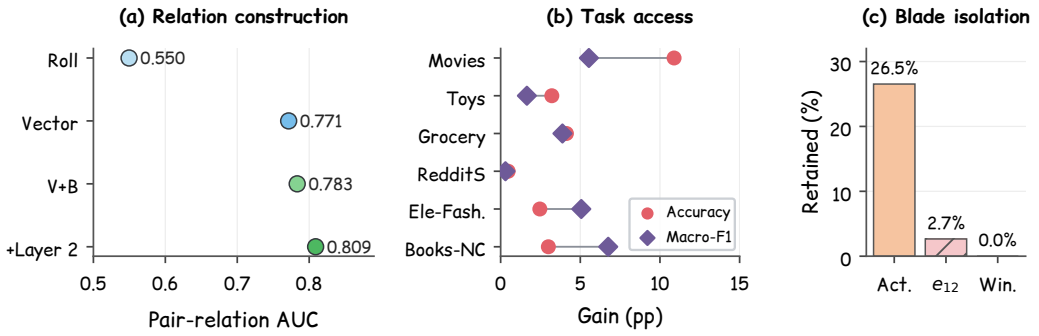


Figure 1: **Empirical study.** (a) Pair-relation AUC after grade expansion and a second transport layer. (b) Accuracy and Macro-F1 gains (percentage points) from direct grade–depth access over a generic readout. (c) Rigid blade isolation reports the retained active share ("Act."), deepest e_{12} energy ("e_{12}"), and matched validation wins ("Win.").

Figure 1(a) separates relation construction from later composition. Adding vector and bivector coordinates raises pair-relation AUC from 0.550 to 0.783; a second transport raises it to 0.809. The first contrast is larger than the second, so the observation supports cross-grade construction followed by complementary context, not depth as a substitute for the geometric product. Figure 1(b) shows that direct grade-depth access improves both NC metrics on all six examined datasets. The gain ranges from 0.47 to 10.89 points for Accuracy and from 0.30 to 6.75 for Macro-F1. Rigid blade isolation retains 26.53 percent of active parameters and 2.658 percent of deepest e_{12} energy, and wins none of the six matched comparisons (Fig. 1(c)).

The fixed-field Accuracy/Macro-F1 gains for Movies, Toys, Grocery, RedditS, Ele-Fashion, and Books-NC are 10.89/5.54, 3.21/1.65, 4.13/3.88, 0.47/0.30, 2.46/5.06, and 3.00/6.75 points, respectively. These fixed-field comparisons diagnose task access, not multi-seed transfer or LP performance. Blade isolation preserves the coordinates but blocks learned cross-grade exchange; its zero matched wins argue against permanent isolation, not against Clifford capacity in general. The probes do not establish downstream transfer. They identify access and cross-grade-use requirements that the full evaluation must test separately.

The task-access probe holds the encoded field fixed and changes only the readout. It therefore tests whether retained states are usable by a new prediction unit, rather than whether the encoder can be trained more effectively with a different objective. Conversely, blade isolation changes the available interactions during transport. Reading the two controls together distinguishes information that was never constructed from information constructed but hidden by a readout.

Together, these results suggest that graph transport creates prediction-relevant state progressively. The first layer separates agreement from oriented interaction, and the second composes that interaction with wider context. The probes jointly motivate preserving relation state for pair prediction while keeping retained states directly accessible to downstream task heads. Compressing these states before the task is known removes exactly the distinction that a reusable foundation needs.

This observation leads to an encoder that constructs relations over observed edges, retains intermediate grades and depths, and defers selection until a fresh head is fitted. Grade expansion calls for Clifford relation construction, depth expansion for a second transport, and the access probe for a retained grade-depth bank rather than a final-state readout. Blade isolation argues for learned cross-grade exchange. Together with the need to preserve entity semantics for node prediction, these findings motivate five design choices. ICE uses Clifford relation construction, two-layer transport, a retained grade and depth bank, a protected Grade-1 semantic route, and task-aligned field access. The controlled probes identify where relation information is constructed and accessed. Section 4 develops the five choices as one information path, and Section 5 evaluates whether that path survives fresh heads, label budgets, and matched component removals.

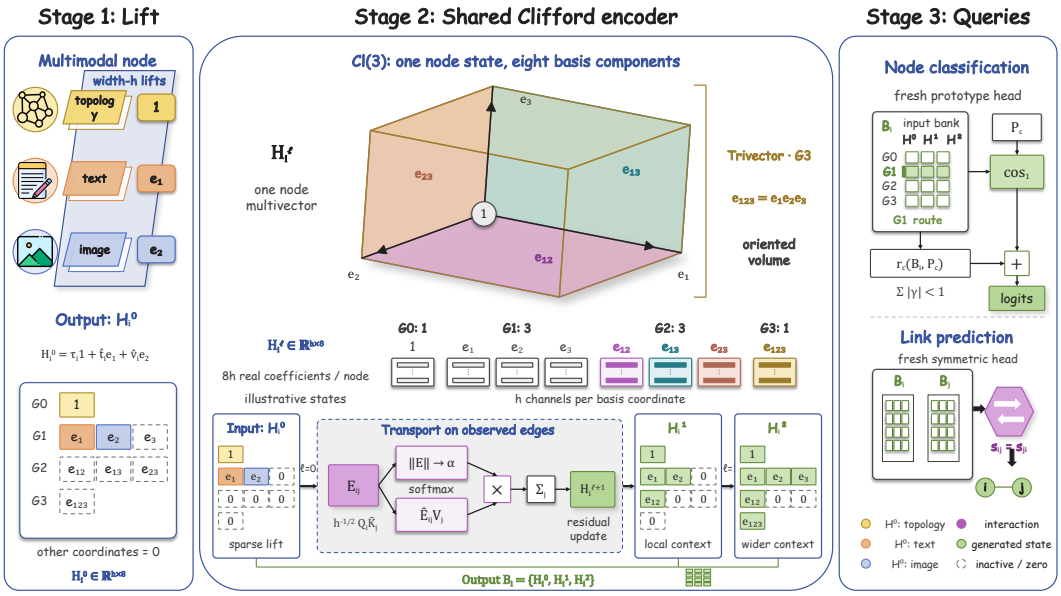


Figure 2: **ICE framework.** One dataset-agnostic Clifford field connects multimodal input, edge-aware graph transport, foundation learning, and fresh node and link heads.

4 Methodology

4.1 Overview

ICE keeps each graph-induced interaction addressable by algebraic grade and transport depth until a task chooses its readout. The data flow is topology, text, and image lift, followed by two edge-aware transports that produce and retain H0, H1, and H2 before fresh node or link heads query the field. Read Fig. 2 from left to right: arrows are data flow, and the repeated field block is the shared encoder reused across graphs.

4.2 Clifford Field Construction and Learning

Entity attributes describe a node, whereas graph neighborhoods describe how entities and modalities interact. ICE uses $\mathrm{Cl}(3)$ because its geometric product represents symmetric agreement and an oriented relation in one state, which later graph transport can compose.

Multimodal and topological addresses. Lightweight adapters map text, image, and local topology at node i to width-h vectors $\mathbf{t}_i$, $\mathbf{v}_i$, and $\boldsymbol{\tau}_i$. ICE initializes

$$\mathbf{H}_i^0 = \boldsymbol{\tau}_i 1 + \widehat{\mathbf{t}}_i e_1 + \widehat{\mathbf{v}}_i e_2, \tag{1}$$

where hats denote normalization. Topology enters the scalar address and the modalities enter distinct vector addresses. Higher grades begin at zero and must therefore be produced by graph computation. Learned maps may exchange information across grades. The addresses specify entry points, not fixed destinations.

Edge-wise relation construction. At layer ℓ, shared multivector maps produce $\mathbf{Q}^\ell$, $\mathbf{K}^\ell$, and $\mathbf{V}^\ell$. For every observed edge $j \to i$, ICE forms $\mathbf{E}_{ij}^\ell = \mathbf{Q}_i^\ell \widetilde{\mathbf{K}_j^\ell} / \sqrt{h}$ and transports the neighboring value as

$$\alpha_{ij}^\ell = \mathrm{softmax}_{j \in \mathcal{N}(i)} \left(\mathrm{mean}_c \left\| \mathbf{E}_{ij,c}^\ell \right\| / \tau \right), \tag{2}$$

$$\mathbf{U}_i^\ell = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^\ell \left(\frac{\mathbf{E}_{ij}^\ell}{\left\| \mathbf{E}_{ij}^\ell \right\| + \epsilon} \right) \mathbf{V}_j^\ell. \tag{3}$$

Its magnitude controls neighbor mass, whereas its normalized multivector direction transforms the value. A residual projection, normalization, and feed-forward map then produce $\mathbf{H}_i^{\ell+1}$ without collapsing relation type into a scalar attention weight.

Two-layer grade and depth construction. The first layer creates scalar agreement and bivector interaction in H1. The second composes those relations with a new graph direction, returning information to Grade 1 or extending it to Grade 3 in H2. ICE retains H0, H1, and H2 so task queries can access each construction stage.

Foundation objectives and semantic protection. ICE learns the field with two complementary objectives and one protected path. Semantic topology contrast aligns a masked node modality with the visible semantic context supplied by its neighbors. Masked geometric reconstruction requires the transported field to recover the masked vector and the higher-grade coordinates created from it. Their joint objective is

$$\mathcal{L}_{\text{pre}} = \lambda_{\text{sem}}\mathcal{L}_{\text{sem}} + \lambda_{\text{geo}} \sum_{g=1}^{3} \omega_g \left\| \left\langle D_\psi(\mathbf{H}^2_{i,\text{vis}}) \right\rangle_g - \left\langle \mathbf{H}^0_i \right\rangle_g \right\|_2^2 . \tag{4}$$

Semantic topology contrast constrains entity meaning, and geometric reconstruction constrains cross-grade content. An additive Grade-1 path stays present through the transported field, giving node prediction a stable semantic route while geometric coordinates refine its score. The shared maps receive only graph topology and node attributes, never a dataset identifier or target label space. Each transport output is consumed by the next layer or retained for a task head.

4.3 Task-Aligned Field Access

Definition 4.1 (Addressable Clifford field)**.** For node i, ICE exports $\mathcal{B}_i = \{\langle \mathbf{H}^\ell_i \rangle_g \mid \ell \in \{0,1,2\},\ g \in \{0,1,2,3\}\}$. H0 contains the multimodal addresses from Eq. 1; H1 and H2 contain the relations formed by Eq. 3. This retained bank is exactly the shared-encoder output read by every fresh task head.

Node prediction. Node classification reads $\mathcal{B}_i$. A prototype $\mathbf{P}_c$ supplies an always-present Grade-1 cosine score, while bounded coefficients query the remaining addresses.

$$s^{\text{NC}}_{ic} = \cos_1(\mathbf{H}^{\text{pool}}_i, \mathbf{P}_c) + \sum_{\ell,g} \gamma_{c\ell g} \cos_g(\mathbf{H}^\ell_i, \mathbf{P}_c), \qquad \sum_{\ell,g} |\gamma_{c\ell g}| \le \rho < 1. \tag{5}$$

The shared field remains unchanged when the head converts these scores into predictions.

Pair prediction. Link prediction reads two fields. Grade-compatible projections first assemble $\mathbf{z}_i = \|_{\ell,g}\, q_{\ell g}(\langle \mathbf{H}^\ell_i \rangle_g)$. The pair head then uses

$$\mathbf{p}_{ij} = [\mathbf{z}_i + \mathbf{z}_j \,\|\, |\mathbf{z}_i - \mathbf{z}_j| \,\|\, \mathbf{z}_i \odot \mathbf{z}_j], \qquad s^{\text{LP}}_{ij} = h_{\text{LP}}(\mathbf{p}_{ij}). \tag{6}$$

The symmetric descriptor combines shared content, disagreement, and multiplicative agreement. Node and link heads differ in prediction unit, not in the dataset-agnostic foundation they receive.

4.4 Theoretical Analysis

The analysis separates standard Clifford-algebra identities from properties of ICE's storage, transport, and heads. We use the direct-sum grade decomposition and geometric-product identities (Dorst et al., 2007, Li et al., 2026). Permutation equivariance follows the shared message-passing construction used in graph neural networks (Hamilton et al., 2017, Xu et al., 2019). Appendix Section E proves the ICE-specific statements under their explicit assumptions.

Theorem 4.2 (Cross-grade reachability)**.** For vectors $a, b, c \in \mathbb{R}^3 \subset \text{Cl}(3)$, the geometric products satisfy

$$ab = a \cdot b + a \wedge b, \qquad (a \wedge b)c = \langle (a \wedge b)c \rangle_1 + \langle (a \wedge b)c \rangle_3 . \tag{7}$$

Thus vector interaction contains scalar agreement and a bivector, while composing the bivector with a vector lies in Grades 1 and 3. In ICE these are the local interaction formed by the first edge-wise transport and the broader composition available to the second transport. This establishes which grades are reachable for a GFM reuse interface. It does not imply that every grade is nonzero or predictive for every graph.

Lemma 4.3 (Complete grade and depth access)**.** If ICE retains every grade at depths H0, H1, and H2, then each transported state is uniquely recoverable:

$$\mathbf{H}_i^\ell = \sum_{g=0}^{3} \langle \mathbf{H}_i^\ell \rangle_g , \qquad \ell \in \{0, 1, 2\}. \tag{8}$$

This is the complete-access meaning of Definition 4.1. The retained field is the output of the shared encoder; an NC head reads node states and an LP head reads a pair, so both prediction units can select from the same dataset-agnostic state without requiring the encoder to collapse grades or depths in advance.

Theorem 4.4 (Node-permutation equivariance)**.** Assume shared nodewise maps, permutation-equivariant neighborhood aggregation over the observed edges, and the pair descriptor of Eq. 6. For every node permutation matrix $\mathbf{P}$,

$$F_\theta(\mathbf{P}\mathcal{G}) = \mathbf{P} F_\theta(\mathcal{G}). \tag{9}$$

Shared ICE transport therefore reindexes messages when node labels change and remains a valid graph operator. The NC output follows the same relabeling, while the symmetric LP descriptor preserves the score for an exchanged pair. The assumption matches the common transport map across graphs and does not assert invariance to changes in graph structure.

Lemma 4.5 (Protected Grade-1 Semantic Route)**.** If the NC head in Eq. 5 obeys $\sum_{\ell,g} |\gamma_{c\ell g}| \le \rho < 1$ and each grade-wise cosine has magnitude at most one, then

$$\left| s_{ic}^{\mathrm{NC}} - \cos_1(\mathbf{H}_i^{\mathrm{pool}}, \mathbf{P}_c) \right| \le \rho. \tag{10}$$

The Grade-1 prototype score in ICE is the protected node route. The bound limits how far task residuals over other grades and depths can move that score, so a fresh node head can add interaction information without replacing the base semantic path.

For a sampled subgraph, ICE requires $O(L|E_b|hB^2 + L|V_b|hB)$ transport time and $O(L|V_b|hB + |E_b|hB)$ retained state. Since $B = 8$ and $L = 2$, the implementation is linear in the sampled graph size up to fixed Clifford factors; Appendix E gives the derivation.

5 Experiments

We evaluate whether one ICE foundation transfers across prediction units and label regimes, then ask whether the observed behavior follows the proposed construction–retention–access path. Appendix A gives the complete data, optimization, and evaluation settings.

Experimental settings. One encoder is pretrained for five epochs over eleven multimodal graphs. Six node-classification and three link-prediction datasets use fresh ten-epoch heads. We report NC Accuracy and Macro-F1, LP MRR against 150 candidates, and 3/5/10-shot Accuracy for 5-way NC and balanced 2-way link classification. Every entry is the mean and sample standard deviation over eight validation-selected runs. Comparators span supervised, multimodal, self-supervised, graph-foundation, and multimodal graph-foundation models (Wei et al., 2019, Tao et al., 2020, Hou et al., 2023, Yu et al., 2025, Sun et al., 2025, He et al., 2025, Liu et al., 2026).

5.1 Transfer Across Prediction Units and Label Budgets

We first ask whether one pretrained field can serve both node and pair prediction when a fresh head receives either full supervision or only a few labels. Tables 1–3 report supervised NC, supervised LP, few-shot link classification, and few-shot NC. Shared-foundation methods remain separate from task-specific specialists, whose training opportunity differs (Appendix A).

The supervised comparisons in Tables 1 and 2 place ICE first on both NC metrics and LP MRR. The NC advantage over the next method spans 0.35–2.36 points, while the three LP margins span 0.83–1.64 points. The narrow RedditS difference matters because it prevents the aggregate story from relying only on domains with large headroom. At the other end, Movies and Books-LP show that the same field remains competitive when the target label space or candidate relations change substantially. These are comparisons of reported means, not paired significance tests against published baselines.

Table 1: **Node classification.** Accuracy and Macro-F1 for each dataset over eight seeds. Best means are bold and second-best means are underlined.

Method	RedditS		Movies		Grocery		Toys		Ele-F.		Books-NC	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
Traditional multimodal												
MMGCN	90.27±0.34	84.22±0.73	53.41±1.03	41.66±2.03	82.56±0.49	73.83±0.93	80.02±0.64	76.36±1.23	86.59±0.08	68.85±0.35	83.22±0.10	71.28±0.22
MGAT	92.78±0.50	87.27±0.53	53.87±0.50	44.09±1.60	83.74±0.62	74.77±1.11	79.61±0.74	77.09±0.87	84.84±0.08	69.62±0.21	82.91±0.04	71.45±0.11
Self-supervised graph models												
GRACE	93.01±0.53	88.39±1.12	48.09±0.97	37.18±1.33	70.83±0.81	60.69±1.05	72.82±0.66	69.09±0.63	83.58±0.11	70.09±0.47	74.96±0.06	70.09±0.11
GraphMAE2	92.81±0.44	87.93±0.37	50.08±0.77	38.68±1.63	76.24±0.60	66.74±1.33	75.11±0.52	71.80±0.50	83.32±0.31	65.92±0.59	74.15±0.22	69.20±0.28
Graph foundation models												
RiemannGFM	91.63±0.45	85.20±1.13	52.80±0.43	40.74±1.25	82.62±0.46	74.90±1.55	77.85±0.45	74.84±0.60	87.07±0.20	70.45±1.32	78.13±0.17	70.73±0.24
GFT	93.02±0.37	87.00±2.03	51.33±0.67	28.14±1.82	76.80±2.22	59.11±3.02	79.52±0.58	76.00±0.92	87.14±0.22	70.33±1.24	75.93±0.41	66.18±0.30
SAMGPT	93.11±0.19	87.12±0.64	50.25±0.28	34.21±1.37	76.41±0.54	63.40±0.88	73.81±0.41	67.12±0.73	83.81±0.11	69.73±0.39	74.29±0.11	66.57±0.29
Multimodal graph foundations												
UniGraph2	93.65±0.17	87.91±0.68	53.02±0.53	43.43±1.86	82.10±0.37	73.93±1.37	79.00±0.59	76.02±0.78	87.06±0.18	69.80±0.82	79.06±0.27	68.74±0.36
PLANET	96.62±0.22	92.44±0.43	57.06±0.61	47.49±1.23	85.16±0.88	77.23±0.86	81.22±0.50	77.55±0.75	87.37±0.12	70.74±0.78	84.16±0.07	74.43±0.18
ICE	**96.97**±0.16	**92.70**±0.14	**59.42**±0.48	**50.36**±0.82	**86.72**±0.39	**79.24**±0.52	**82.64**±0.33	**79.06**±0.45	**88.21**±0.10	**72.08**±0.54	**85.24**±0.08	**75.96**±0.22

Table 2: **Link prediction and few-shot link classification.** MRR for supervised link prediction and Accuracy for balanced 2-way classification over eight seeds.

Method	Link prediction			Few-shot link classification					
	Sports	Cloth	Books-LP	Sports 3	Sports 5	Sports 10	Cloth 3	Cloth 5	Cloth 10
Traditional multimodal									
MMGCN	23.44±0.43	17.74±0.38	20.73±0.48	55.86±2.96	53.70±1.20	56.66±1.60	64.36±4.07	68.61±5.38	67.27±4.20
MGAT	21.74±0.96	15.47±0.32	21.82±0.53	54.92±2.40	56.55±2.69	57.92±1.80	67.05±1.45	70.34±3.36	68.66±2.94
Self-supervised graph models									
GRACE	25.31±0.16	18.27±0.15	19.30±0.27	56.42±1.36	57.94±2.58	58.50±1.43	62.36±2.75	64.94±1.09	64.96±1.86
GraphMAE2	24.54±0.30	18.69±0.21	19.99±0.19	53.28±3.09	54.36±1.65	56.05±1.70	60.95±1.61	61.39±1.71	62.33±1.78
Graph foundation models									
RiemannGFM	21.92±0.51	19.20±0.44	22.03±0.67	54.07±1.46	53.73±2.03	53.68±1.24	63.45±4.07	62.66±1.75	66.20±5.94
GFT	22.04±0.62	17.63±0.59	20.16±1.21	56.75±2.63	56.86±2.30	55.73±3.36	63.70±2.84	66.66±2.10	65.37±4.03
SAMGPT	24.09±0.22	16.41±0.20	24.91±0.38	59.41±3.62	59.00±3.56	60.48±4.25	72.61±1.90	74.20±1.60	74.25±1.75
Multimodal graph foundations									
UniGraph2	27.09±0.13	19.31±0.35	19.44±0.19	60.83±3.90	64.27±2.74	65.08±3.17	73.44±1.57	71.28±4.20	73.77±1.91
PLANET	27.51±0.14	20.25±0.27	27.62±0.25	62.89±3.99	64.36±3.05	67.84±1.38	74.03±4.12	75.22±1.20	75.44±2.84
ICE	**28.38**±0.11	**21.08**±0.20	**29.26**±0.23	**64.11**±3.71	**65.62**±2.81	**69.18**±1.26	**75.36**±3.84	**76.41**±1.11	**76.71**±2.62

Table 3: **Few-shot node classification.** Focused 5-way comparison across 3, 5, and 10 shots; the complete matrix appears in Appendix Table 8.

Method	Grocery 5-way			Ele-Fashion 5-way			Books-NC 5-way		
	3	5	10	3	5	10	3	5	10
GFT	61.45±2.22	64.60±3.54	63.12±4.07	59.08±3.95	61.38±3.41	62.28±2.62	48.70±4.96	50.95±4.35	51.62±4.73
UniGraph2	60.05±4.13	61.25±3.09	66.27±3.11	53.98±4.07	58.73±2.95	60.38±2.08	59.67±4.23	61.55±3.97	63.80±4.02
PLANET	77.85±4.06	79.88±3.27	81.93±3.48	70.50±4.37	72.97±3.55	74.85±3.73	63.62±4.11	67.59±4.68	69.18±3.90
ICE	**78.96**±3.78	**81.02**±3.08	**83.10**±3.21	**71.61**±4.01	**74.07**±3.33	**76.02**±3.42	**64.83**±3.86	**68.78**±4.32	**70.42**±3.61

For five-way NC, we ask whether fresh heads preserve the cross-graph ranking as the label budget increases across three graphs and the 3-, 5-, and 10-shot settings. Appendix Table 8 retains the complete method-by-dataset matrix.

From 3 to 5 and 5 to 10 shots, NC Accuracy increases by 2.06/2.08 points on Grocery, 2.46/1.95 on Ele-Fashion, and 3.95/1.64 on Books-NC. ICE exceeds PLANET by 1.10–1.24 points across all nine NC cells. Few-shot link Accuracy increases by 1.51/3.56 points on Sports and 1.05/0.30 on Cloth, with ICE margins of 1.19–1.34 points across six cells (Table 2). These are eight-run means rather than paired significance tests.

The low-label comparisons test the same frozen foundation under three progressively larger training budgets, rather than three separately pretrained encoders. The task-specific slopes need not form one

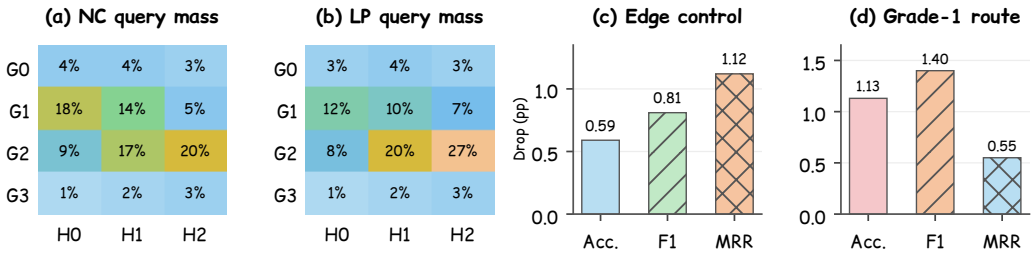


Figure 3: **Mechanism behavior.** (a–b) NC and LP access mass by grade and depth. (c–d) Accuracy (Acc.), Macro-F1 (F1), and MRR drops (percentage points) after edge rolling and protected Grade-1 route removal, respectively.

label-scaling curve. Their gains therefore concern fresh-head adaptation under a fixed representation. The NC and LP trends should not be collapsed into one mean improvement: class prediction and pair ranking use different output spaces and metrics, and the small RedditS supervised margin cautions against treating every target graph as equally easy.

5.2 Ablation and Mechanism Analysis

To assess construction, retention, protection, and access, we compare matched removals in Table 4 with the field-use and route controls in Fig. 3. Each removal lowers the corresponding task summaries. The Clifford product has the largest LP decline (1.51 points), while removing the protected Grade-1 route gives the largest NC declines (1.13 Accuracy and 1.40 Macro-F1). Layer 2, the retained grade and depth bank, and task access lower MRR by 0.66, 0.98, and 1.08 points. Together with the grade-depth shares in Fig. 3, these matched controls support the coupled construction. Unlike the independent probes in Section 3, Fig. 3 compares NC and LP heads reading the completed field. The access maps allocate 18 percent of NC query mass to Grade 1 at H0 and 27 percent of LP query mass to Grade 2 at H2; neither head reads only a final Grade-1 state. Edge rolling lowers NC Accuracy, Macro-F1, and LP MRR by 0.59, 0.81, and 1.12 points, respectively. By contrast, removing the Grade-1 route changes LP MRR by 0.55 points, less than its 1.13- and 1.40-point NC declines. This contrast links the protected path chiefly to node semantics while retaining a measurable pair effect. These diagnostics do not isolate causal effects or assign fixed meanings to individual grades. Each row removes one component from the same ICE reference, so the drops cannot be summed to estimate a combined removal.

Table 4: **Core ablation** (eight-seed mean).

Configuration	NC Acc.	NC F1	LP MRR
ICE	**83.21**±0.31	**74.90**±0.38	**26.24**±0.18
w/o Clifford product	82.00±0.46	73.54±0.52	24.73±0.29
w/o layer 2	82.71±0.37	74.34±0.44	25.58±0.21
w/o grade-depth bank	82.52±0.39	74.12±0.45	25.26±0.24
w/o Grade-1 route	82.08±0.43	73.50±0.49	25.69±0.25
w/o task access	82.29±0.41	73.86±0.47	25.16±0.26

6 Conclusion

Multimodal graph foundations need entity semantics and edge-generated relations to serve node and pair prediction. ICE retains both in one dataset-agnostic Clifford field through edge-aware transport, a retained grade and depth bank, and a protected Grade-1 route. This keeps relation state accessible across transport depths without displacing the entity semantics used for node classification, while leaving pair-specific combinations to a fresh link head. The probes motivate this interface, while the cross-graph comparisons and matched removals assess transfer and component sensitivity. The fresh head selects retained state for each prediction unit. This conclusion concerns supervised and few-shot reuse on the stated graph suite; it does not establish zero-shot adaptation, transfer to unseen modalities, or robustness to other prediction units.

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish the writing of this manuscript, including grammar, phrasing, and clarity, as the authors are non-native English speakers. We have reviewed all AI-assisted work, and the authors verified the final text against the reported results and claims. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Leo Dorst, Daniel Fontijne, and Stephen Mann. *Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry*. Morgan Kaufmann, 2007.

Yasha Ektefaie, George Dasoulas, Ayush Noori, Maha R. Farhat, and Marinka Zitnik. Multimodal learning with graphs. *Nat. Mac. Intell.*, 5(4):340–350, 2023.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NIPS*, pp. 1024–1034, 2017.

Yufei He, Yuan Sui, Xiaoxin He, Yue Liu, Yifei Sun, and Bryan Hooi. Unigraph2: Learning a unified embedding space to bind multimodal graphs. In *WWW*, 2025. doi: 10.1145/3696410.3714818. URL https://arxiv.org/abs/2502.00806.

Zhenyu Hou, Yufei He, Yukuo Cen, Xiao Liu, Yuxiao Dong, Evgeny Kharlamov, and Jie Tang. Graphmae2: A decoding-enhanced masked self-supervised graph learner. In *WWW*, pp. 737–746. ACM, 2023.

Xunkai Li, Zekai Chen, Zhengyu Wu, Henan Sun, Daohan Su, Guang Zeng, Hongchao Qin, Rong-Hua Li, and Guoren Wang. Lion: A clifford neural paradigm for multimodal-attributed graph learning, 2026. URL https://arxiv.org/abs/2608.24795.

Sicheng Liu, Xunkai Li, Daohan Su, Ru Zhang, Hongchao Qin, Ronghua Li, and Guoren Wang. Toward effective multimodal graph foundation model: A divide-and-conquer based approach, 2026. Manuscript.

Ciyuan Peng, Jiayuan He, and Feng Xia. Learning on multimodal graphs: A survey. *CoRR*, abs/2402.05322, 2024.

Li Sun, Zhenhao Huang, Suyang Zhou, Qiqi Wan, Hao Peng, and Philip Yu. RiemannGFM: Learning a graph foundation model from riemannian geometry. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, pp. 1154–1165. ACM, April 2025. doi: 10.1145/3696410.3714952. URL http://dx.doi.org/10.1145/3696410.3714952.

Zhulin Tao, Yinwei Wei, Xiang Wang, Xiangnan He, Xianglin Huang, and Tat-Seng Chua. MGAT: multimodal graph attention network for recommendation. *Inf. Process. Manag.*, 57(5):102277, 2020.

Zehong Wang, Zheyuan Zhang, Nitesh Chawla, Chuxu Zhang, and Yanfang Ye. GFT: Graph foundation model with transferable tree vocabulary. *Advances in Neural Information Processing Systems*, 2024.

Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. MMGCN: multi-modal graph convolution network for personalized recommendation of micro-video. In *ACM Multimedia*, pp. 1437–1445. ACM, 2019.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*. OpenReview.net, 2019.

Hao Yan, Chaozhuo Li, Zhigang Yu, Jun Yin, Ruochen Liu, Peiyan Zhang, Weihao Han, Mingzheng Li, Zhengxin Zeng, Hao Sun, Weiwei Deng, Feng Sun, Qi Zhang, and Senzhang Wang. When graph meets multimodal: Benchmarking on multimodal attributed graphs learning. *CoRR*, abs/2410.09132, 2024.

Xingtong Yu, Zechuan Gong, Chang Zhou, Yuan Fang, and Hui Zhang. SAMGPT: Text-free graph foundation model for multi-domain pre-training and cross-domain adaptation. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, pp. 1142–1153. ACM, April 2025. doi: 10.1145/3696410.3714828. URL http://dx.doi.org/10.1145/3696410.3714828.

Jing Zhu, Yuhang Zhou, Shengyi Qian, Zhongmou He, Tong Zhao, Neil Shah, and Danai Koutra. Mosaic of modalities: A comprehensive benchmark for multimodal graph learning. In *CVPR*, pp. 14215–14224. Computer Vision Foundation / IEEE, 2025.

A Evaluation Protocol

A.1 Shared Evaluation Protocol

All shared foundations use the same task definitions, node features, graph splits, pretraining exposure, downstream opportunity, validation rule, evaluator, and seed aggregation. The main comparison covers only models that reuse one foundation across target graphs. Independently trained task specialists appear in a separate appendix section because they receive a different training opportunity.

PLANET results are taken from its published tables. They are not paired runs against ICE, so comparisons report mean differences and seed variation without a paired significance claim.

The foundation is trained for five epochs on the eleven-graph pool with exposure weights $5, 5, 5, 1, 1, 1, 10, 0.5, 0.1, 20, 20$. The same foundation checkpoint initializes every downstream job. Seeds 42 through 49 change fresh job initialization and data order, not the shared foundation checkpoint. NC uses ten downstream epochs, root batch 128, and fanout $(5, 10)$. LP uses batches of 1,024 positive edges, one full-graph non-edge per positive, node batches of 128 with fanout $(10, 10)$, and validation query batches of 1,024. The predefined evaluator ranks one positive against 150 ordered candidates using its pessimistic tie rule.

A.2 Few-shot protocol

Few-shot NC uses 5-way episodes on Grocery, Ele-Fashion, and Books-NC. Few-shot link classification uses balanced 2-way episodes on Sports and Cloth. Each setting reports 3-shot, 5-shot, and 10-shot Accuracy over ten sampled tasks. The data and sampling rules match the supervised setting, while a fresh prototype classifier replaces the supervised MLP. NC tasks use 20 training episodes and 10 query episodes. Link tasks use 20 training episodes and 40 query episodes. The foundation checkpoint is unchanged and no dataset identity enters the shared encoder.

A.3 Selection and held-out evaluation

Every downstream checkpoint is selected by one predeclared validation rule. Test labels, metrics, and candidate scores remain hidden during method, hyperparameter, head, checkpoint, or seed selection. After the method and selection rule are finalized, the selected state is restored and evaluated once. The supervised main-table claim requires strict improvement in each of the 12 NC and three LP metrics; ties fail. Nine few-shot NC and six few-shot link comparisons are reported separately, giving the 30 supervised and few-shot comparisons summarized in the Abstract.

B Datasets and Metrics

B.1 Foundation Graph Inventory

The evaluation suite combines four MAGB graphs and seven MM-GRAPH graphs (Yan et al., 2024, Zhu et al., 2025, Liu et al., 2026). Movies, Toys, Grocery, Sports, Cloth, and Ele-Fashion are product graphs. RedditS is a post graph. Books-NC and Books-LP are book graphs. MM-CoDEx-s and MM-CoDEx-m are knowledge graphs used during foundation learning but not in the active NC/LP evaluation. Table 5 reports the graph totals and the exposure ratios used during foundation pretraining.

MAGB graphs use the standard 60/20/20 train, validation, and test partitions. MM-GRAPH graphs use their official partitions. Qwen2-VL-7B-Instruct produces one 3,584-dimensional text vector and

Table 5: **Foundation graph inventory and pretraining exposure.**

Graph	Nodes	Edges	Role	Classes	Exposure
Movies	16,672	218,390	NC	20	5.0
Toys	20,695	126,886	NC	18	5.0
Grocery	17,074	171,340	NC	20	5.0
Sports	50,250	356,202	LP	–	1.0
Cloth	125,839	951,271	LP	–	1.0
Ele-Fashion	97,766	199,602	NC	12	1.0
RedditS	15,894	566,160	NC	20	10.0
Books-NC	685,294	7,235,048	NC	11	0.5
Books-LP	636,502	3,437,017	LP	–	0.1
MM-CoDEx-s	1,383	15,884	Pretrain	–	20.0
MM-CoDEx-m	7,697	52,840	Pretrain	–	20.0

one 3,584-dimensional image vector per node. The foundation loader receives only the training-visible graph objects prescribed by the protocol. Sampling may add a root self-loop for message passing, but self-loops are excluded from the relation probes used in the empirical study.

B.2 Task Metrics and Aggregation

NC reports Accuracy and Macro-F1. LP reports MRR from one positive edge and 150 predefined ordered candidates with pessimistic tie handling. The evaluation unit is one independently initialized downstream job. For metric value z_s from seed s and $S = 8$, the reported entry is

$$\bar{z} = \frac{1}{S}\sum_{s=1}^{S} z_s, \qquad \mathrm{std}(z) = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}(z_s - \bar{z})^2}. \tag{11}$$

Seeds correspond to 42 through 49. Every seed must finish before aggregation. A failed or missing seed remains visible and cannot be replaced by a better run. Main-table entries use $\bar{z} \pm \mathrm{std}(z)$ with two decimal places.

C Implementation Details

Table 6 records the shared foundation configuration and the common evaluation budget. It separates model-specific choices from the common comparison opportunity. The task-family head configuration remains identical across datasets and seeds.

Table 6: **Foundation and downstream configuration.**

Stage	Configuration	Value
Foundation	Input width per modality	3,584
Foundation	Hidden width and blades	256 and 8
Foundation	Transport layers	2
Foundation	RWPE steps	4
Foundation	Root batch and fanout	128 and $(5, 10)$
Foundation	Epochs and optimizer	5 and AdamW
Foundation	Learning rate and weight decay	4×10^{-5} and 5×10^{-4}
Foundation	Linear warmup ratio and gradient clip	0.4 and 1.0
Foundation	Attention and contrast temperatures	1.0 and 0.07
Foundation	Text and image mask rates	0.3 and 0.3
Foundation	Dropout and edge chunk size	0.1 and 32,768
NC	Epochs, root batch, and fanout	10, 128, and $(5, 10)$
LP	Epochs and positive-edge batch	10 and 1,024
LP	Node fanout and evaluation batch	$(10, 10)$ and 1,024
Evaluation	Seeds and checkpoint rule	42 through 49 and validation only
Evaluation	Test calls	Once per selected job

Mixed precision and gradient checkpointing are enabled for foundation learning. The shared checkpoint contains the encoder and reusable adapters but no downstream label space, prototypes, classifier

outputs, dataset embedding, or job-local optimizer state. Each downstream run restores the same checkpoint and creates a fresh task head. Training, checkpoint selection, and the final test call are logged separately so that the selected state and the held-out access count can be audited without relying on terminal output.

C.1 Model Size and Adaptation Cost

Table 7 reports the shared foundation size and completed NC adaptation cost. The checkpoint contains the dataset-agnostic encoder and reusable adapters, while each downstream job creates a compact fresh head.

Table 7: **Model and adaptation scale.** Architecture size and completed NC adaptation cost.

Measure	Value
Foundation state	
Foundation parameters	79.99 M
Foundation checkpoint	305.2 MiB
Fresh NC adaptation	
Head parameters	22.7 to 41.3 K
Peak GPU memory	0.52 to 0.95 GiB
Six-dataset wall time	29.0 min

D Additional Method Details

D.1 End-to-End Data Flow

For each sampled root subgraph, width adapters normalize topology, text, and image inputs. Eq. 1 places them at explicit scalar and vector addresses. Two instances of Eq. 3 propagate graph-guided multivector messages. Semantic topology contrast and masked geometric reconstruction update the shared encoder. The encoder exports H0/H1/H2, and a fresh task-specific head maps those states to NC logits or LP pair scores. No label space, head state, dataset embedding, or dataset-indexed hidden state crosses downstream jobs.

D.2 Foundation learning and task adaptation

Algorithm 1 follows the same object boundary as Definition 4.1. Foundation learning updates only the shared adapters, Clifford transport, and self-supervised decoders. Each downstream job restores the same field encoder and creates a fresh head for its prediction unit.

Algorithm 1: Foundation learning and task adaptation in ICE.

Input: Foundation graphs $\{\mathcal{G}_d\}$, exposure weights $\{q_d\}$, encoder F_θ
Output: Shared encoder $F_{\theta^\star}$ and independently fitted task heads

foreach *foundation update* **do**
  Sample graph $\mathcal{G}_d \sim q$ and a training-root subgraph
  Lift topology, text, and image inputs into $\mathbf{H}^0$ and mask the target modality
  Apply two edge-aware Clifford transport layers and retain $\mathbf{H}^{0:2}$
  Compute semantic topology contrast and masked geometric reconstruction
  Update θ and discard every dataset-local state
end
Save the reusable encoder $F_{\theta^\star}$
foreach *downstream node or link task* **do**
  Restore $F_{\theta^\star}$ and initialize one fresh task-family head
  if *few-shot task* **then**
    Keep $F_{\theta^\star}$ fixed and fit only the fresh prototype head
  else
    Fit the fresh head and protocol-permitted encoder parameters on training data
  end
  Select the state with validation data and evaluate the selected state once
end

The separation in Algorithm 1 prevents label spaces and job-local parameters from entering the shared field. It also makes the reusable object explicit. A target task receives the encoder states defined in Definition 4.1, not the optimizer or head that created a result on another graph.

D.3 Pretraining target construction

For a node whose text or image input is masked, ICE builds a semantic anchor from visible neighbors. Let $\mathbf{z}_i$ denote the protected Grade-1 state and let $\mathcal{A}_i$ contain neighbors whose target modality remains visible. The normalized anchor is

$$\mathbf{a}_i = \operatorname{norm}\left(\sum_{j\in\mathcal{A}_i} \beta_{ij}\mathbf{z}_j\right), \qquad \beta_{ij} = \operatorname{softmax}_{j\in\mathcal{A}_i}\left(\mathbf{z}_i^\top \mathbf{z}_j/\tau_{\mathrm{top}}\right). \tag{12}$$

Semantic topology contrast places the masked node near this anchor and separates it from anchors of other sampled roots. The geometric decoder receives the visible H2 field and reconstructs the masked vector together with the higher grades formed during transport. The two targets play different roles. The anchor keeps entity semantics tied to graph neighborhoods, while geometric reconstruction rewards the relation state that develops beyond Grade 1.

The mask is sampled independently for text and image inputs. A node with one hidden modality keeps the other modality and its topology features, so reconstruction cannot be solved by copying the target input. Nodes with no eligible visible neighbor are excluded from the contrastive term but remain available to geometric reconstruction. This rule keeps the objective defined on sparse neighborhoods without introducing a dataset-specific fallback.

D.4 Task head construction

The NC head enforces the semantic bound through its parameterization. For unconstrained parameters $a_{c\ell g}$ and $b_{c\ell g}$, ICE sets

$$\gamma_{c\ell g} = \rho\, \operatorname{softmax}_{\ell,g}(a_{c\ell g}) \tanh(b_{c\ell g}). \tag{13}$$

The absolute coefficient sum is at most ρ, satisfying Eq. 5 and the residual bound in Eq. 10. The prototype term remains present for every class, while the residual learns which grades and depths refine that semantic score.

The LP head uses a symmetric pair feature. Let $\mathbf{z}_i^{\ell g}$ be the vectorized grade g at depth ℓ, and let $\mathbf{R}_{ij}^{\ell} = \mathbf{H}_i^{\ell}\widetilde{\mathbf{H}_j^{\ell}}$. ICE forms

$$\mathbf{r}_{ij} = \operatorname*{concat}_{\ell,g}\left[|\mathbf{z}_i^{\ell g} - \mathbf{z}_j^{\ell g}|, \mathbf{z}_i^{\ell g} \odot \mathbf{z}_j^{\ell g}, \langle \mathbf{R}_{ij}^{\ell} \rangle_0, \|\langle \mathbf{R}_{ij}^{\ell} \rangle_2\|_2\right]. \tag{14}$$

Absolute differences and elementwise products are unchanged when the nodes are exchanged. The scalar relation is also unchanged, while the bivector changes orientation but not norm. A compact MLP applied to $\mathbf{r}_{ij}$ therefore produces the same score for (i, j) and (j, i) while retaining both semantic agreement and oriented interaction magnitude.

E Proofs and Complexity Analysis

This appendix proves the two theorems and two lemmas used in the main paper. Clifford identities establish grade reachability, while the retained-state, shared-transport, and head constraints give the ICE-specific access, equivariance, and semantic-bound results.

E.1 Proof of cross-grade reachability

Proof of Theorem 4.2. For $a = \sum_i a_i e_i$ and $b = \sum_j b_j e_j$, the Clifford relation $e_i e_j + e_j e_i = 2\delta_{ij}$ gives

$$ab = \sum_i a_i b_i + \sum_{i<j}(a_i b_j - a_j b_i) e_i e_j = a \cdot b + a \wedge b. \tag{15}$$

The first summand is Grade 0 and is symmetric in a, b; the second is Grade 2 and changes sign when the vectors are exchanged. For $B = a \wedge b$ and a vector c, the standard grade decomposition of the geometric product gives

$$Bc = \langle Bc \rangle_1 + \langle Bc \rangle_3 = B \lrcorner\, c + B \wedge c. \tag{16}$$

Thus a vector–vector edge interaction writes scalar and bivector coefficients in H1, and a subsequent vector interaction can return the bivector to Grade 1 or extend it to Grade 3 in H2. Learned linear maps can change coefficients before the next product, but they do not change the grade closure of $\mathrm{Cl}(3)$. This proves reachability, not a fixed semantic meaning for any learned blade. In ICE, the result is exactly why the first transport layer constructs relation state and the second layer composes that state before H2 is exposed to the heads. □

E.2 Proof of complete grade–depth access

Proof of Lemma 4.3. The Clifford algebra is the direct sum of its grade subspaces. Therefore every retained state has the unique decomposition

$$\mathbf{H}_i^{\ell} = \sum_{g=0}^{3} \langle \mathbf{H}_i^{\ell} \rangle_g, \qquad \ell \in \{0, 1, 2\}. \tag{17}$$

By Definition 4.1, ICE stores each summand at each of the three depths. The collection $\mathcal{B}_i$ therefore reconstructs H0, H1, and H2 uniquely by Eq. 17; no grade or depth is discarded at the encoder boundary. A node head may select a single address or a learned combination, while a link head may form its pair descriptor from two complete queries. This is the precise sense in which NC and LP read different prediction units from one dataset-agnostic field. It does not claim that every task uses every address. □

E.3 Proof of node-permutation equivariance

Proof of Theorem 4.4. Let $\mathbf{P}$ be the permutation matrix for a node relabeling π. Shared nodewise maps give $\mathbf{Q}' = \mathbf{PQ}$, $\mathbf{K}' = \mathbf{PK}$, and $\mathbf{V}' = \mathbf{PV}$. The edge $j \to i$ becomes $\pi(j) \to \pi(i)$, so the corresponding Clifford interaction is unchanged:

$$\mathbf{E}'_{\pi(i)\pi(j)} = \mathbf{Q}'_{\pi(i)}\widetilde{\mathbf{K}'_{\pi(j)}}/\sqrt{h} = \mathbf{Q}_i\widetilde{\mathbf{K}}_j/\sqrt{h} = \mathbf{E}_{ij}. \tag{18}$$

The softmax inputs and normalized messages are consequently paired one-to-one. Reindexing the neighborhood sum gives

$$\mathbf{U}'_{\pi(i)} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \left(\frac{\mathbf{E}_{ij}}{\|\mathbf{E}_{ij}\| + \epsilon} \right) \mathbf{V}_j = \mathbf{U}_i. \tag{19}$$

The input lift, pointwise normalization, residual maps, and feed-forward maps are all shared and therefore commute with the same permutation. Induction over the two transport layers proves $\mathbf{H}^{\ell\prime}_{\pi(i)} = \mathbf{H}^{\ell}_i$ for $\ell = 0, 1, 2$, which is the stated equivariance. A pointwise NC head inherits the relabeling, and the symmetric LP descriptor in Eq. 6 is unchanged when both nodes are relabeled or exchanged. Shared transport is consequently a legal graph operator for the cross-graph foundation interface.

□

E.4 Proof of the protected Grade-1 semantic route

Proof of Lemma 4.5. Subtract the Grade-1 prototype term from Eq. 5 and write the remaining score as

$$r_{ic} = \sum_{\ell,g} \gamma_{c\ell g} \cos_g(\mathbf{H}^{\ell}_i, \mathbf{P}_c). \tag{20}$$

Each normalized cosine lies in $[-1, 1]$. The triangle inequality and the coefficient constraint therefore give

$$|r_{ic}| \le \sum_{\ell,g} |\gamma_{c\ell g}| \left|\cos_g(\mathbf{H}^{\ell}_i, \mathbf{P}_c)\right| \le \sum_{\ell,g} |\gamma_{c\ell g}| \le \rho. \tag{21}$$

Since $r_{ic} = s^{\mathrm{NC}}_{ic} - \cos_1(\mathbf{H}^{\mathrm{pool}}_i, \mathbf{P}_c)$, this is the claimed bound. The Grade-1 semantic score is always present; the learned residual may refine it with other grades and depths but cannot absorb it without violating the constraint. This is the exact architectural protection used by the node head, not an empirical assertion about which coefficient is largest. □

E.5 Complexity consequence

Let a sampled batch contain $|V_b|$ nodes and $|E_b|$ directed edges, with hidden width h, B Clifford blades, and L transport layers. One edge-wise product combines at most B^2 blade pairs per hidden channel, while node projections, normalization, residual maps, and feed-forward updates cost $O(hB)$ per node and layer. Hence

$$T_{\mathrm{ICE}} = O(L|E_b|hB^2) + O(L|V_b|hB), \tag{22}$$

and retaining H0 through HL costs $O(L|V_b|hB)$ state memory, plus the selected edge chunk. ICE fixes $B = 8$ and $L = 2$, so the architecture is linear in sampled nodes and edges up to constant Clifford factors. This bound explains the implementation scale without introducing a separate theoretical claim; detailed resource values and adaptation costs are reported in Additional Method Details.

F Additional Experiments

F.1 Mechanism Controls

Main-paper Figure 3 compares grade–depth access with the edge-rolling and protected-route controls. The exact metric values and grade–depth shares are retained in Appendix Tables 16–18, allowing the visual trends to be checked without repeating the same composite here.

F.2 Complete Few-Shot Matrix

The main text reports the focused comparison at the three requested label budgets. The complete method-by-dataset matrix is retained here for auditability.

Table 8: **Few-shot node classification.** Accuracy for 5-way 3-shot, 5-shot, and 10-shot tasks.

	Method	Grocery 5-way			Ele-Fashion 5-way			Books-NC 5-way		
		3	5	10	3	5	10	3	5	10
	MMGCN	48.13±2.65	50.30±3.13	53.73±3.51	54.37±2.99	57.05±2.42	60.87±3.08	52.93±2.94	54.10±2.90	56.42±2.84
	MGAT	50.83±2.88	54.07±3.18	55.53±2.98	60.38±2.57	61.10±2.20	62.82±2.18	53.15±3.09	57.05±2.35	58.22±3.40
	GRACE	57.28±4.91	59.22±3.50	62.22±2.89	55.49±4.10	59.90±2.73	64.72±3.04	56.84±4.65	61.53±3.14	59.20±2.00
	GraphMAE2	51.90±5.76	54.60±3.83	58.00±4.26	57.35±3.05	60.00±3.53	63.65±3.48	43.20±2.70	46.30±2.95	49.83±2.60
GFM	RiemannGFM	63.63±2.48	66.23±2.73	67.14±2.02	59.13±3.71	60.24±3.55	61.73±2.90	54.29±4.10	57.17±3.83	60.54±3.92
GFM	GFT	61.45±2.22	64.60±3.54	63.12±4.07	59.08±3.95	61.38±3.41	62.28±2.62	48.70±4.96	50.95±4.35	51.62±4.73
GFM	SAMGPT	52.73±12.69	62.33±10.49	66.47±10.67	54.20±10.78	61.13±9.32	61.27±10.82	42.53±6.79	47.00±8.35	51.80±8.65
MGFM	UniGraph2	60.05±4.13	61.25±3.09	66.27±3.11	53.98±4.07	58.73±2.95	60.38±2.08	59.67±4.23	61.55±3.97	63.80±4.02
MGFM	PLANET	77.85±4.06	79.88±3.27	81.93±3.48	70.50±4.37	72.97±3.55	74.85±3.73	63.62±4.11	67.59±4.68	69.18±3.90
	ICE	**78.96**±3.78	**81.02**±3.08	**83.10**±3.21	**71.61**±4.01	**74.07**±3.33	**76.02**±3.42	**64.83**±3.86	**68.78**±4.32	**70.42**±3.61

F.3 Detailed Ablation Results

Figure 4 expands the compact main-paper ablation table into metric-specific drops. It shows that removing the Clifford product most strongly affects LP, whereas removing the protected semantic route has the largest NC effect. The profile is reported as a diagnostic decomposition and does not assign a unique causal effect to any component.

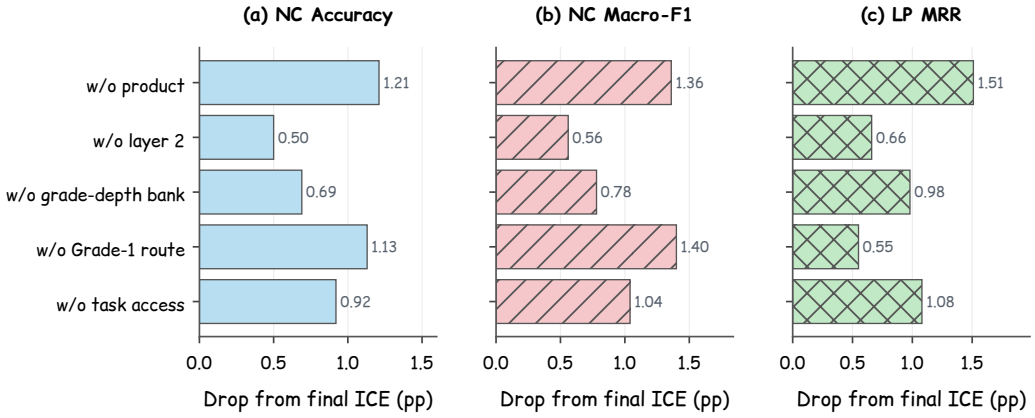


Figure 4: **Component necessity.** Absolute performance changes for the matched removals.

F.4 Empirical Probe Results

The main empirical figure emphasizes the trends that determine the method design. The tables below provide the corresponding values for direct inspection. Table 9 records the progression from the pairing control to two-layer relation state, and Table 10 reports the dataset-level change produced by direct field access. Table 11 quantifies the effect of permanent blade isolation.

Table 9: **Relation probe.** Pair-relation AUC across the transport sequence.

Representation	AUC
Hidden roll	0.5496
L1 Vector	0.7714
L1 V+B	0.7833
L1+L2 V+B	0.8090

The progression is monotonic. Vector and bivector coordinates separate observed node pairs from the pairing control, and the second transport layer strengthens that separation. The pattern supports a depth-retained field rather than a single final representation.

Table 10: **Task-access probe.** Improvement over the generic query in percentage points.

Dataset	Accuracy	Macro-F1
Movies	10.888	5.539
Toys	3.213	1.648
Grocery	4.130	3.880
RedditS	0.472	0.302
Ele-Fash.	2.460	5.062
Books-NC	2.995	6.752

All twelve entries are positive, although their magnitude varies by graph and metric. The access benefit is therefore broad rather than uniform. Movies shows the largest Accuracy change, while Books-NC shows the largest Macro-F1 change. This heterogeneity is consistent with task heads using different parts of the same retained bank.

Table 11: **Cross-grade capacity probe.** Changes induced by permanent blade isolation.

Quantity	Value
Active parameters under same-blade maps	26.53%
Layer-2 e_{12} energy retained	2.658%
Cases improved by blade isolation	0 of 6

The same-blade restriction removes most effective mappings and sharply weakens the deepest bivector response. It also fails to improve any matched case. The probe therefore identifies cross-grade mixing as a practical requirement rather than unused algebraic freedom. ICE retains dense learned maps after the readable input lift so relation state can move between grades.

G Per-Seed Results

Tables 12–14 expose the run-level values behind the aggregate comparison. All entries use the same eight-seed order and the paper-wide mean plus or minus standard-deviation convention.

Table 12. Node classification Accuracy. Per-seed scores and aggregate statistics.

Dataset	42	43	44	45	46	47	48	49	Mean ± std.
RedditS	96.74	96.81	96.88	96.94	97.01	97.06	97.12	97.21	96.97±0.16
Movies	59.87	60.14	58.72	58.94	59.14	59.33	59.53	59.68	59.42±0.48
Grocery	86.81	86.93	87.09	87.30	86.15	86.33	86.49	86.65	86.72±0.39
Toys	82.44	82.58	82.72	82.82	82.95	83.13	82.16	82.31	82.64±0.33
Ele-Fashion	88.06	88.11	88.15	88.19	88.23	88.26	88.30	88.36	88.21±0.10
Books-NC	85.32	85.36	85.12	85.16	85.19	85.23	85.26	85.28	85.24±0.08

Table 13. Node classification Macro-F1. Per-seed scores and aggregate statistics.

Dataset	42	43	44	45	46	47	48	49	Mean ± std.
RedditS	92.83	92.78	92.73	92.67	92.62	92.56	92.50	92.91	92.70±0.14
Movies	50.55	50.21	49.87	49.54	49.17	51.59	51.14	50.81	50.36±0.82
Grocery	78.93	78.72	78.48	80.02	79.73	79.53	79.36	79.14	79.24±0.52
Toys	78.41	79.73	79.49	79.31	79.17	78.98	78.79	78.61	79.06±0.45
Ele-Fashion	72.59	72.38	72.21	71.98	71.76	71.54	71.29	72.89	72.08±0.54
Books-NC	76.01	75.92	75.83	75.74	75.64	76.29	76.17	76.08	75.96±0.22

Table 14. Link prediction MRR. Per-seed scores and aggregate statistics.

Dataset	42	43	44	45	46	47	48	49	Mean ± std.
Sports	28.41	28.44	28.48	28.54	28.22	28.27	28.31	28.36	28.38±0.11
Cloth	20.96	20.88	20.79	21.38	21.27	21.19	21.13	21.04	21.08±0.20
Books-LP	29.12	29.22	29.31	29.39	29.48	29.60	28.93	29.03	29.26±0.23

The run-level view makes the dispersion behind each aggregate explicit. Table 15 summarizes the NC ranges so that a reader can inspect stability without expanding the main comparison.

Table 15. Node classification seed ranges. Minimum, maximum, and range in percentage points.

Dataset	Accuracy			Macro-F1		
	Min	Max	Range	Min	Max	Range
RedditS	96.74	97.21	0.47	92.50	92.91	0.41
Movies	58.72	60.14	1.42	49.17	51.59	2.42
Grocery	86.15	87.30	1.15	78.48	80.02	1.53
Toys	82.16	83.13	0.97	78.41	79.73	1.33
Ele-Fashion	88.06	88.36	0.30	71.29	72.89	1.59
Books-NC	85.12	85.36	0.24	75.64	76.29	0.65

H Mechanism Measurements

Tables 16–18 give the values behind the main mechanism figure. They separate grade and depth access from edge dependence and semantic protection.

Table 16. Grade and depth access. Query mass in percent.

Task	Grade	H0	H1	H2
NC	G0	4.0	4.0	3.0
NC	G1	18.0	14.0	5.0
NC	G2	9.0	17.0	20.0
NC	G3	1.0	2.0	3.0
LP	G0	3.0	4.0	3.0
LP	G1	12.0	10.0	7.0
LP	G2	8.0	20.0	27.0
LP	G3	1.0	2.0	3.0

Table 17. Edge controls. Values use the original metric units.

Metric	Reference	Edge roll	Drop in pp
NC Acc.	83.21	82.62	0.59
NC F1	74.90	74.09	0.81
LP MRR	26.24	25.12	1.12

Table 18. Grade-1 route control. Absolute change after route removal.

Metric	Drop in pp
NC Acc.	1.13
NC F1	1.40
LP MRR	0.55

Higher grades receive 52 percent of the NC query mass and 61 percent of the LP query mass. Edge rolling and Grade-1 route removal lower all three summary metrics in their original percentage-point units. Together, these results connect task access, graph structure, and semantic protection.

I Resource Usage and Result Reconstruction

This appendix consolidates implementation scale and reconstruction details. Table 19 lists the fields that connect every aggregate to a selected checkpoint and run record.

Table 19. Result reconstruction fields. Stored records for each dataset and seed.

Record	Stored content
Protocol	Dataset, task, split, candidate set, metric, and evaluator identity
Training	Command, configuration, source state, seed, optimizer, and epoch budget
Selection	Validation metric, selected epoch, selection rule, and checkpoint identity
Evaluation	Predictions, targets, metric output, and test-access count
Resources	Parameter count, peak memory, stage time, and total elapsed time
Aggregation	Eight run values, mean, standard deviation, and comparator target
Integrity	File size, digest, environment, and reconstruction entry point

J TASK-SPECIFIC SPECIALIST COMPARISONS

Task-specific specialist rows remain outside the shared-foundation ranking. Specialist rows are not compared as if they were one shared foundation, because they may use independent per-dataset encoders, tuning, or supervision.

K REPRODUCIBILITY AND RESULT FIELDS

Each result entry contains the dataset, task, metric, seed, selected epoch, validation score, test score, checkpoint identifier, prediction file, wall time, peak memory, and completion status. Aggregated results store all eight seed values, their arithmetic mean and standard deviation, and the corresponding published comparator value.

Each task, dataset, metric, and seed is linked to its selected checkpoint and prediction file before Eq. 11 is computed. Every displayed mean and standard deviation is aggregated from the job-level outputs.

L LIMITATIONS

ICE assumes text and image node attributes together with an eight-blade $\mathrm{Cl}(3)$ interface. Other modalities may require a different input assignment or algebra. The present evaluation covers supervised and few-shot NC and link transfer. It does not study generation, open-vocabulary reasoning, zero-shot or rotor-based Geometric Prompting, continual graph arrival, unsupervised label-space transfer, or one head shared across task families.

Addressable grades name computational coordinates but do not assign unique semantics to learned coefficients after dense cross-grade mixing.