TRANSFERABLE STRUCTURAL CONCORDANCE FOR GRAPH FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph foundation models (*GFMs*) aim to transfer knowledge across heterogeneous graph domains. Existing approaches largely organize cross-domain transfer at the graph level, treating each graph as a domain-specific structural unit and thereby overlooking substantial node-level structural heterogeneity. We characterize this heterogeneity through *structural concordance*, which measures the local agreement between a node's topological and feature-induced neighborhoods and reveals complementary high- and low-concordance regimes across graph domains. Building on this observation, we propose a *concordance-aware hierarchical coordination* framework for multi-domain GFM pre-training. Within each graph, two complementary coordinators specialize in high- and low-concordance regimes, while node-wise soft routing adapts their influence to each node's concordance level. Graph-specific meta coordinators further integrate concordance-conditioned information and mediate communication across source domains. This hierarchy enables node-adaptive cross-domain coordination while preserving the original graph topologies, following the principle of coordinate without homogenizing. We theoretically establish the informativeness, estimability, and representation stability of concordance-aware coordination. Experiments across diverse graph benchmarks consistently demonstrate improved unseen-domain transfer over existing GFMs.

1 INTRODUCTION

Graph-structured data arise across diverse scientific and real-world domains, including citation networks, biological systems, recommender systems, and social networks. While *graph neural networks* (*GNNs*) have achieved remarkable progress in graph representation learning, they are typically optimized for individual datasets or tasks and generalize poorly beyond their training domains. Inspired by foundation models in language and vision, *graph foundation models* (*GFMs*) aim to learn transferable representations through large-scale self-supervised pre-training (Mao et al., 2024; Wang et al., 2025b;c;d). A central goal of multi-domain GFMs is therefore to exploit heterogeneous source graphs to learn a shared encoder that generalizes to unseen domains.

Early graph pre-training methods learn transferable representations through objectives such as node masking, edge prediction, and contrastive learning, including GCC (Qiu et al., 2020), DGI (Veličković et al., 2019), GPT-GNN (Hu et al., 2020), GraphCL (You et al., 2020), GPPT (Sun et al., 2022), and SimGRACE (Xia et al., 2022). These methods are effective when graphs share feature semantics or structural regularities, but transfer becomes substantially harder when both features and topology vary. The central challenge is therefore not simply how to pre-train across multiple graphs, but what structural information can be meaningfully shared across heterogeneous domains.

Recent multi-domain GFMs primarily address this challenge through domain-specific coordination or alignment. GCOPE (Zhao et al., 2024) introduces dataset-specific virtual coordinators, SAMGPT (Yu et al., 2025), building on MDGPT (Yu et al., 2024), employs shared and domain-specific tokens, and MDGFM (Wang et al., 2025a) explicitly aligns graph topology through structure learning. Related approaches include MDP-GNN (Lin et al., 2025), BRIDGE (Yuan et al., 2025), and OMOG (Liu et al., 2024). Despite their different implementations, these approaches largely organize cross-domain transfer at the graph level, treating each graph as a domain-specific structural unit for coordination, adaptation, or alignment. Such graph-level treatment captures structural variation across domains but

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.

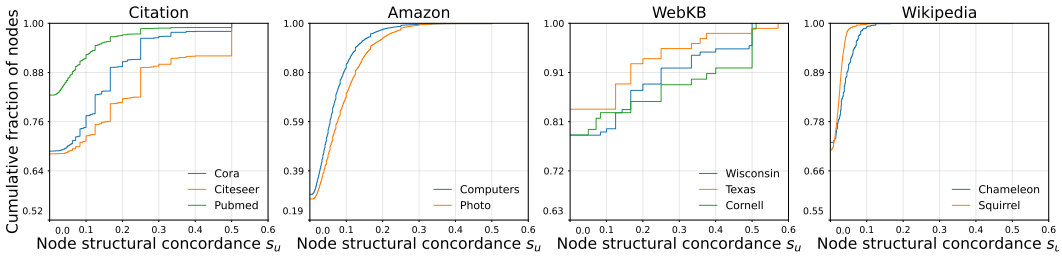

Figure 1: Cumulative distributions of node-level structural concordance across ten graphs from four domains. Concordance distributions vary substantially across domains, while nodes with different concordance levels coexist within individual graphs, revealing intra-domain heterogeneity in local topology–feature relations. This motivates concordance-conditioned node-level coordination rather than treating each graph as a homogeneous structural unit.

largely overlooks heterogeneity within a domain, where individual nodes may exhibit substantially different local structural characteristics.

We take a different perspective along two fundamental dimensions. First, cross-domain coordination should account for node-level structural heterogeneity within each domain rather than treating all nodes in a domain uniformly. Even within the same graph, nodes can exhibit markedly different relationships between graph topology and feature geometry: for some nodes, the topological neighborhood closely agrees with the feature-induced neighborhood, whereas for others, the two substantially diverge. We characterize this heterogeneity through *structural concordance*, which measures the local agreement between a node's topological neighborhood and its feature-induced neighborhood (Zheng et al., 2024). As illustrated in Figure 1, the resulting concordance distributions differ substantially across graph domains, while nodes with different concordance levels also coexist within individual graphs, revealing pronounced node-level structural heterogeneity. More importantly, unlike absolute graph structures whose meanings can vary across domains, structural concordance describes the same local topology–feature relation regardless of domain-specific node, feature, or edge semantics. This provides a common structural coordinate for organizing cross-domain knowledge exchange at the node level, rather than relying solely on domain identity or directly aligning heterogeneous graph structures. Second, node-level structural heterogeneity calls for concordance-conditioned soft coordination. Rather than assigning all nodes in a domain the same coordination signal, we introduce two complementary coordinators corresponding to higher- and lower-concordance patterns and softly route each node between them according to its structural concordance. This allows each node to receive a distinct mixture of coordinator information adapted to its local topology–feature relation. Consequently, cross-domain coordination becomes node-adaptive while retaining a compact domain-level coordination architecture.

Building on these observations, we propose a *concordance-aware hierarchical coordination* framework for multi-domain GFM pre-training. The dual coordinators first organize node-level structural heterogeneity within each graph through concordance-conditioned routing, while a graph-specific meta coordinator integrates the resulting information and mediates communication across source domains. This local-to-global hierarchy enables node-adaptive cross-domain coordination while preserving the original topology of each graph, following a principle we term *coordinate without homogenizing*. We further provide theoretical justification for structural concordance as an effective basis for node-adaptive coordination. Specifically, we show that concordance-conditioned routing achieves a strict approximation advantage over uniform domain-level coordination when concordance is informative of node-specific coordination requirements, that structural concordance can be reliably estimated from finite neighborhoods, and that routing estimation errors induce controlled perturbations in the learned representations. Together, these results characterize the informativeness, estimability, and representation stability of concordance-aware coordination.

Our contributions are summarized as follows: *(i)* We identify an overlooked limitation of existing multi-domain graph pre-training: cross-domain transfer is predominantly organized at the graph level, without explicitly accounting for node-level structural heterogeneity within each domain. We introduce structural concordance, measuring local topology–feature agreement, to characterize this intra-domain heterogeneity and distinguish complementary high- and low-concordance regimes that coexist across graph domains. *(ii)* We develop a concordance-aware hierarchical coordination framework that combines dual concordance-conditioned coordinators with node-wise soft routing,

translating graph-level coordination into node-adaptive information exchange. Graph-specific meta coordinators further enable structured communication across source domains while preserving their original graph topologies. *(iii)* We establish theoretical guarantees for the informativeness, estimability, and representation stability of concordance-aware coordination, and demonstrate consistent improvements in unseen-domain transfer across diverse graph benchmarks.

2 Cross-Domain Transfer in Graph Foundation Models

Multi-Domain Graph Pre-training. We consider M heterogeneous source graphs $\mathcal{D}_{\text{pre}} = \{G_1, \ldots, G_M\}$, where $G_m = (V_m, E_m, X_m)$ consists of node set V_m, edge set E_m, and node features $X_m \in \mathbb{R}^{|V_m| \times d_m}$. The source graphs may differ substantially in feature spaces, semantics, and topology. Without node-label supervision, multi-domain pre-training learns a shared GNN encoder f_θ through $\theta^* = \arg\min_\theta L_{\text{pre}}\big(\{G_m\}_{m=1}^M; \theta\big)$, where L_{pre} denotes a self-supervised objective. The goal is to learn transferable representations that generalize beyond the source domains.

Unseen-Domain Adaptation. Cross-domain transfer is evaluated on an unseen target graph $G_t = (V_t, E_t, X_t, Y_t)$ excluded from pre-training. The pre-trained parameters θ^* initialize the target encoder, which is adapted with a task-specific head g_ϕ using available target labels. Let Y_t^{tr} denote the target training labels. For node classification, $(\theta_t^*, \phi^*) = \arg\min_{\theta,\phi} L_{\text{CE}}\big(g_\phi(f_\theta(G_t)), Y_t^{\text{tr}}\big)$. The adapted model is evaluated on unseen target test nodes.

The central challenge is to identify what structural knowledge remains transferable when graph semantics and topology vary across domains. Absolute structural patterns lack a common coordinate across heterogeneous graphs, while graph-level characterization can further obscure substantial node-level heterogeneity. A transferable structural basis should therefore be both comparable across domains and sensitive to local structural regimes. This motivates *structural concordance*, which characterizes the local agreement between topology and feature-induced neighborhoods and provides a common structural coordinate for organizing node-level structural relations across heterogeneous graphs. Section 3 builds our cross-domain coordination framework around this coordinate.

3 Cross-Domain Coordination via Transferable Structural Concordance

Building on Section 2, we organize cross-domain transfer around two principles: transferable structure should be defined through relations that remain comparable across domains rather than absolute topology, and coordination should operate at the level of local structural regimes rather than graph identity alone. We therefore use structural concordance as a node-level coordinate of topology–feature relations and develop a hierarchy that first organizes nodes by this coordinate within each graph and then exchanges the resulting information across domains. Our framework consists of cross-domain feature alignment, concordance-aware hierarchical coordination, and self-supervised graph pre-training.

3.1 Cross-Domain Feature Compatibility

Structural concordance compares topology with feature-induced neighborhood geometry and thus requires feature relations to be represented in compatible spaces. Since source graphs may differ in feature dimensionality and semantics, we map each graph into a common d-dimensional space using a domain-specific SVD-based projection $\phi_m(\cdot)$: $\widetilde{X}_m = \phi_m(X_m) \in \mathbb{R}^{|V_m| \times d}$, yielding $\widetilde{G}_m = (V_m, E_m, \widetilde{X}_m)$. The domain-specific projections accommodate heterogeneous inputs and enable a shared graph encoder, without requiring direct alignment of graph structures across domains.

3.2 Concordance-Aware Hierarchical Coordination

As illustrated in Figure 2, our mechanism follows a local-to-global hierarchy: structural concordance identifies node-level structural regimes, concordance-conditioned routing organizes them within each graph, and meta coordination enables information exchange across domains.

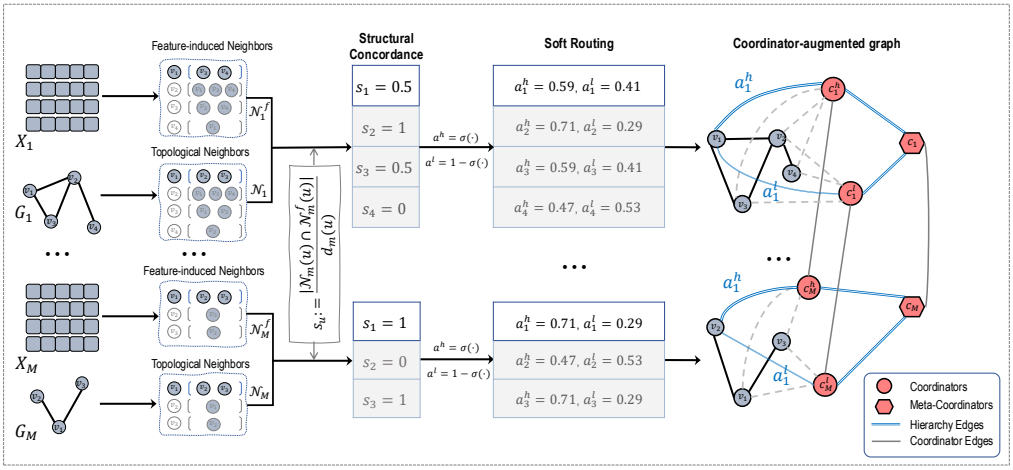

Figure 2: **Overview of concordance-aware hierarchical coordination.** For each source graph, we measure node-level structural concordance s_u as the local agreement between the topological and feature-induced neighborhoods, revealing heterogeneous high- and low-concordance regimes within the same domain. Rather than assigning all nodes in a domain to a shared communication channel, two domain-specific virtual coordinators, c_m^h and c_m^l, specialize in the high- and low-concordance regimes, and each node is softly routed between them according to its own concordance score s_u. Corresponding coordinators facilitate communication across domains for nodes with similar topology–feature agreement patterns, while graph-specific meta coordinators c_m further integrate concordance-conditioned information and enable hierarchical cross-domain exchange. The resulting coordinator-enhanced graph thus supports node-adaptive cross-domain transfer while preserving the original topology of each source graph, realizing the principle of coordinate without homogenizing.

Structural Concordance. For node $u \in V_m$, let $\mathcal{N}_m(u)$ denote its one-hop topological neighborhood and $d_m(u) = |\mathcal{N}_m(u)|$ its degree. Using the aligned features $\widetilde{X}_m$, we construct a degree-matched feature-induced neighborhood containing the $d_m(u)$ most feature-similar nodes:

$$\mathcal{N}_m^f(u) := \mathrm{TopK}_{v \in V_m \setminus \{u\}} \big(\mathrm{sim}(\widetilde{X}_u, \widetilde{X}_v),\, d_m(u)\big), \tag{1}$$

where $\mathrm{sim}(X, X') = X^\top X' / (\|X\|_2 \|X'\|_2)$ for any $X, X' \in \mathbb{R}^d$. Since $|\mathcal{N}_m^f(u)| = |\mathcal{N}_m(u)| = d_m(u)$, their overlap reflects neighborhood composition rather than neighborhood size. We define the *structural concordance* of node u as

$$s_u := \frac{|\mathcal{N}_m(u) \cap \mathcal{N}_m^f(u)|}{d_m(u)} \in [0, 1]. \tag{2}$$

A larger s_u indicates stronger agreement between observed connectivity and feature-induced proximity. Unlike absolute topology, s_u describes a relative topology–feature relation independent of specific edge semantics, providing a node-level coordinate for comparing local structural regimes across heterogeneous graphs.

As shown in Figure 1, structural concordance exhibits substantial variation not only across graph domains but also among nodes within the same graph. This empirical heterogeneity suggests that a single graph-level coordination signal may be insufficient to capture the diverse local topology–feature relations within a domain, motivating concordance-conditioned node-level coordination.

Concordance-Conditioned Dual Coordination. Because different concordance regimes may coexist within the same graph, a single graph-level coordinator cannot adapt its signal to such node-level heterogeneity. We therefore introduce two graph-specific virtual coordinators, c_m^h and c_m^l, referred to as the *high-* and *low-concordance coordinators*, respectively, softly route each node between them:

$$a_v^h = \sigma((s_v - \gamma)/\tau), \quad a_v^l = 1 - a_v^h, \tag{3}$$

where γ specifies the concordance reference point, $\tau > 0$ controls routing smoothness, and $\sigma(\cdot)$ denotes the sigmoid function. Specifically, each node $v \in V_m$ is connected to both c_m^h and c_m^l, with

the routing coefficients a_v^h and a_v^l serving as the corresponding node–coordinator edge weights:

$$E_{\text{coord}} = \bigcup_{m=1}^{M} \left\{ (v, c_m^h, a_v^h), (v, c_m^l, a_v^l) \mid v \in V_m \right\}. \tag{4}$$

Through message passing, these weighted connections provide node v with the effective coordinator signal $a_v^h \boldsymbol{c}_m^h + a_v^l \boldsymbol{c}_m^l$. Rather than imposing a discrete partition, the two coordinators define a continuous coordination space indexed by structural concordance. Each node therefore receives a coordinator signal adapted to its local structural regime, shifting coordination from graph-level identity to node-level structural regimes.

Cross-Domain Meta Coordination. Dual coordination organizes structural heterogeneity within each graph, but does not yet enable knowledge transfer across source domains. Directly connecting nodes from different graphs is undesirable, as their local structures and semantics are domain-specific. We therefore perform cross-domain communication at a higher level of abstraction. Specifically, for each graph G_m, we introduce a graph-specific meta coordinator c_m^{meta} that aggregates information from its high- and low-concordance coordinators: $E_{\text{local}} = \left\{ (c_m^{\text{meta}}, c_m^h), (c_m^{\text{meta}}, c_m^l) \mid m = 1, \ldots, M \right\}$. We then connect the meta coordinators and dual coordinators across source graphs to enable cross-domain information exchange: $E_{\text{cross}}^{\text{meta}} = \{ (c_m^{\text{meta}}, c_{m'}^{\text{meta}}) \mid m \neq m' \}$, $E_{\text{cross}}^{\text{dual}} = \left\{ (c_m^h, c_{m'}^h), (c_m^l, c_{m'}^l) \mid m \neq m' \right\}$. This construction yields a hierarchical coordination mechanism: structural concordance first identifies comparable local regimes at the node level; the dual coordinators then organize these regimes within each graph; and the meta coordinators finally exchange the organized graph-level information across domains. In this way, cross-domain transfer operates on structurally organized representations rather than directly coupling domain-specific node structures.

Combining the original graphs and coordinators yields the augmented graph $G' = (V', E', X')$, where $V' = (\bigcup_{m=1}^{M} V_m) \cup \{c_m^h, c_m^l\}_{m=1}^{M} \cup \{c_m^{\text{meta}}\}_{m=1}^{M}$, $E' = (\bigcup_{m=1}^{M} E_m) \cup E_{\text{coord}} \cup (E_{\text{local}} \cup E_{\text{cross}}^{\text{meta}} \cup E_{\text{cross}}^{\text{dual}})$. The original intra-domain edges remain unchanged, while the coordinator edges form a compact hierarchical communication backbone. Hence, domain-specific topology is preserved while cross-domain transfer is organized through a shared node-level structural coordinate.

3.3 Coordinator Representation Learning and Graph Pre-training

We jointly learn the coordinator representations and shared graph encoder over the augmented communication structure.

Coordinator Initialization. We initialize the coordinators by concordance-weighted aggregation of aligned node features: $\boldsymbol{c}_m^h = \sum_{v \in V_m} a_v^h \widetilde{X}_v / \sum_{v \in V_m} a_v^h$, $\boldsymbol{c}_m^l = \sum_{v \in V_m} a_v^l \widetilde{X}_v / \sum_{v \in V_m} a_v^l$. The meta coordinator is initialized by integrating the two anchors: $\boldsymbol{c}_m^{\text{meta}} = (\boldsymbol{c}_m^h + \boldsymbol{c}_m^l)/2$. All coordinator representations are subsequently learnable and jointly optimized with the shared encoder.

Self-Supervised Graph Pre-training. Following the graph contrastive learning paradigm of GraphCL (You et al., 2020), we perform contrastive pre-training over subgraphs sampled from the unified augmented graph G'. For a batch of sampled subgraphs $\{G_b\}_{b=1}^{B}$, we independently construct two correlated views G_b^1 and G_b^2 of each subgraph and encode them as $H_b^1 = f_\theta(G_b^1)$ and $H_b^2 = f_\theta(G_b^2)$. A permutation-invariant readout followed by a projection head $g(\cdot)$ gives $\boldsymbol{z}_b^1 = g\big(\text{READOUT}(H_b^1)\big)$ and $\boldsymbol{z}_b^2 = g\big(\text{READOUT}(H_b^2)\big)$. Views of the same subgraph form positive pairs, while views from different subgraphs serve as negatives. We optimize the contrastive objective $L_{\text{CONT}} = -\frac{1}{B} \sum_{b=1}^{B} \log\{\exp(\text{sim}(\boldsymbol{z}_b^1, \boldsymbol{z}_b^2)/\tau_c) / [\sum_{j=1}^{B} \exp(\text{sim}(\boldsymbol{z}_b^1, \boldsymbol{z}_j^2)/\tau_c)]\}$, where $\tau_c > 0$ is the contrastive temperature. The encoder and coordinator representations are jointly optimized as $\min_{\theta, \mathcal{C}} L_{\text{CONT}}$, where $\mathcal{C}$ collects all learnable coordinator representations.

Unlike standard graph contrastive pre-training, the contrastive objective here operates on the coordinator-augmented graph G', whose topology determines how information is organized and exchanged during message passing. Structural concordance organizes node-level structural information within each domain, while meta coordination enables its exchange across domains. Thus, the standard contrastive objective learns representations over the proposed cross-domain coordination structure rather than directly aligning domain-specific graph topologies.

4 THEORETICAL FOUNDATIONS OF CONCORDANCE-AWARE COORDINATION

We theoretically characterize the structural-concordance-based coordination mechanism introduced in Section 3. Our analysis addresses three questions: whether concordance-conditioned routing is informative beyond graph-level uniform coordination, whether it is estimable from finite neighborhoods, and whether its estimation error remains stable under GNN message passing. Together, these results provide an end-to-end justification for structural concordance as a node-level coordination coordinate.

4.1 APPROXIMATION ADVANTAGE OF CONCORDANCE-CONDITIONED ROUTING

We first ask when node-level concordance provides useful coordination information beyond a shared graph-level coordinator. Consider N nodes $v_1, \dots, v_N$, where $q_i \in \mathbb{R}^d$ denotes the ideal coordinator message for node v_i. The collection $\{q_i\}_{i=1}^N$ captures heterogeneous node-specific coordination requirements.

Single-Coordinator Approximation. A shared coordinator supplies the same message $c \in \mathbb{R}^d$ to every node, with optimal error $\mathcal{E}_{\text{single}} := \min_c \sum_{i=1}^N \|q_i - c\|_2^2$. Let $\bar{q} = N^{-1} \sum_{i=1}^N q_i$. The optimum is $c^* = \bar{q}$, yielding $\mathcal{E}_{\text{single}} = \sum_{i=1}^N \|q_i - \bar{q}\|_2^2$.

Concordance-Conditioned Soft Coordination. Let $\alpha_i^* \in [0,1]$ denote the ideal population routing coefficient induced by the structural concordance of v_i. Dual coordination supplies the signal of $m_i = \alpha_i^* c_H + (1 - \alpha_i^*) c_L$, where $c_H, c_L \in \mathbb{R}^d$ are complementary coordinator anchors. Its optimal approximation error is $\mathcal{E}_{\text{soft}} := \min_{c_H, c_L} \sum_{i=1}^N \|q_i - \alpha_i^* c_H - (1 - \alpha_i^*) c_L\|_2^2$.

Theorem 4.1 (Approximation advantage of concordance-conditioned soft routing)**.** *Assume that the routing coefficients are non-degenerate, $\sum_{i=1}^N (\alpha_i^* - \bar{\alpha})^2 > 0$, where $\bar{\alpha} = N^{-1} \sum_{i=1}^N \alpha_i^*$. Then,*

$$\mathcal{E}_{\text{soft}} = \mathcal{E}_{\text{single}} - \frac{\| \sum_{i=1}^N (\alpha_i^* - \bar{\alpha})(q_i - \bar{q}) \|_2^2}{\sum_{i=1}^N (\alpha_i^* - \bar{\alpha})^2}. \tag{5}$$

Consequently, $\mathcal{E}_{\text{soft}} \le \mathcal{E}_{\text{single}}$, with strict inequality whenever $\sum_{i=1}^N (\alpha_i^* - \bar{\alpha})(q_i - \bar{q}) \neq 0$.

Theorem 4.1 precisely characterizes when concordance is useful: the gain is determined by the cross-covariation between concordance-induced routing and node-specific coordination requirements. Thus, the improvement does not follow merely from adding another coordinator; it arises when concordance provides an informative direction along which coordination requirements vary.

4.2 CONCENTRATION OF STRUCTURAL-CONCORDANCE ROUTING

The previous result assumes access to population routing coefficients. We next establish whether they can be reliably approximated from finite neighborhoods. For node v_i, let $d_i = |\mathcal{N}(i)|$ and define $Z_{ij} = \mathbf{1}\{v_j \in \mathcal{N}^f(i)\}$, $v_j \in \mathcal{N}(i)$. The empirical and population concordance scores are $s_i = d_i^{-1} \sum_{j \in \mathcal{N}(i)} Z_{ij} = |\mathcal{N}(i) \cap \mathcal{N}^f(i)| / d_i$ and $s_i^* = \mathbb{E}\left[Z_{ij} \mid v_j \in \mathcal{N}(i)\right]$, respectively. Consistent with Section 3.2, their routing coefficients are $\alpha_i^* = \sigma((s_i^* - \gamma)/\tau)$, $\alpha_i = \sigma((s_i - \gamma)/\tau)$.

Theorem 4.2 (Concentration of structural-concordance routing)**.** *Assume that, conditioned on node v_i, the indicators $\{Z_{ij} : v_j \in \mathcal{N}(i)\}$ are independent Bernoulli variables with common mean s_i^*. Then, for any $\delta \in (0,1)$, with probability at least $1 - \delta$,*

$$|s_i - s_i^*| \le \sqrt{\frac{\log(2/\delta)}{2d_i}}, \qquad |\alpha_i - \alpha_i^*| \le \frac{1}{4\tau} \sqrt{\frac{\log(2/\delta)}{2d_i}}. \tag{6}$$

The same bound holds for the complementary routing coefficient.

Theorem 4.2 establishes that concordance-induced routing is statistically recoverable from finite neighborhoods. Its error decreases as $O(d_i^{-1/2})$, with routing temperature τ explicitly controlling the sensitivity of routing to concordance estimation error.

4.3 Representation Stability under Concordance Routing

We finally examine how routing estimation error propagates through the shared GNN encoder. At layer l, the population and empirical coordinator messages are

$$m_i^{*l} = \alpha_i^* W_H^l c_H^l + (1-\alpha_i^*) W_L^l c_L^l, \qquad m_i^l = \alpha_i W_H^l c_H^l + (1-\alpha_i) W_L^l c_L^l.$$

The corresponding message-passing dynamics are

$$H^{l+1} = \phi\big(PH^l W^l + M^{*l}\big), \qquad \widehat{H}^{l+1} = \phi\big(P\widehat{H}^l W^l + M^l\big),$$

where P is the propagation operator, M^{*l} and M^l stack the population and empirical coordinator messages.

Theorem 4.3 (End-to-end representation bound). *Assume that ϕ is ρ-Lipschitz, $\|P\|_2 \le 1$, $\|W^l\|_2 \le B_l$, and $\|W_H^l c_H^l - W_L^l c_L^l\|_2 \le B_c^l$. Suppose that $\widehat{H}^0 = H^0 = X$. Let N be the total number of nodes across all source graphs. Then, after L message-passing layers, with probability at least $1-\delta$,*

$$\|\widehat{H}^L - H^L\|_F \le \frac{1}{4\tau}\sqrt{\sum_{i=1}^{N}\frac{\log(2N/\delta)}{2d_i}}\sum_{l=0}^{L-1}\Big(\rho B_c^l \prod_{r=l+1}^{L-1}\rho B_l\Big). \tag{7}$$

Theorem 4.3 shows that concordance estimation error induces a controlled perturbation in the final representations, which decreases with larger neighborhoods and smoother routing. Its propagation through the encoder is governed by the separation between coordinator signals and the amplification of subsequent message-passing layers.

Together, these results establish a complete theoretical chain: structural concordance is informative when it explains variation in node-specific coordination requirements, estimable from finite neighborhoods, and stable in the sense that its estimation error induces a controlled perturbation of learned representations. This provides an end-to-end justification for structural concordance as a node-level, domain-agnostic coordinate for cross-domain graph coordination.

5 Experiments

We empirically evaluate TSCGFM from five perspectives: cross-domain transfer performance on unseen target graphs, the contributions of concordance-conditioned routing and hierarchical meta coordination, transfer under different levels of target supervision, sensitivity to routing hyperparameters, and computational efficiency.

Experimental Setup We consider ten heterogeneous graph benchmarks from four domains: citation networks, including `Cora`, `Citeseer` (Sen et al., 2008), and `PubMed` (Namata et al., 2012); e-commerce networks, including `Computers` and `Photo` (McAuley et al., 2015; Shchur et al., 2018); web networks, including `Cornell`, `Texas`, and `Wisconsin`; and social networks, including `Chameleon` and `Squirrel` (Pei et al., 2020). In the unseen-domain setting, each graph is held out in turn as the target domain, while the remaining nine graphs are used for self-supervised pre-training. The pre-trained encoder is then adapted to the target graph for node classification. We additionally evaluate transfer to `Penn94` (Traud et al., 2012), a large-scale online social network from a distinct domain not represented by the ten source graphs, using all ten benchmarks above for pre-training. Detailed dataset statistics are provided in Appendix B.1.

We compare against task-specific GNNs, including GCN (Kipf & Welling, 2017), GAT (Veličković et al., 2018), and FAGCN (Bo et al., 2021); graph self-supervised methods, including DGI (Veličković et al., 2019) and GraphCL (You et al., 2020); graph prommpt-tuning methods, including GPF (Fang et al., 2023) and GraphPrompt (Liu et al., 2023); and multi-domain GFMs, including GCOPE (Zhao et al., 2024), SAMGPT (Yu et al., 2025), and MDGFM (Wang et al., 2025a). For a consistent comparison, all methods in the latter three categories use FAGCN (Bo et al., 2021) as the backbone. We report the average accuracy over five random runs.

Cross-Domain Transfer Performance Table 1 reports node-classification accuracy on unseen target domains. TSCGFM achieves the highest mean accuracy on all ten datasets, consistently

outperforming task-specific GNNs, graph self-supervised methods, and recent multi-domain GFMs across substantially different graph semantics and topologies. These results support our central premise that effective cross-domain transfer need not rely on direct structural similarity. Instead, organizing pre-training through local topology–feature relations provides a common coordination basis across heterogeneous graph domains.

Table 1: Node classification Accuracy (ACC, %) on unseen target domains.

Method	Cora	Citeseer	PubMed	Computers	Photo	Wisconsin
GCN	21.05±2.96	22.24±3.62	40.66±3.98	19.80±4.35	24.68±3.60	37.74±8.50
GAT	19.88±4.94	22.07±1.63	41.85±4.08	23.67±7.13	28.12±7.77	40.27±8.50
FAGCN	26.78±3.17	36.18±4.15	42.16±1.58	32.84±7.18	41.04±8.31	31.22±9.00
DGI	34.80±3.76	30.28±4.52	49.44±5.73	44.86±2.48	49.55±9.30	27.45±4.64
GraphCL	32.39±5.23	39.73±1.42	45.22±0.95	40.87±4.77	52.99±1.37	35.85±9.27
GPF	34.94±7.31	42.68±4.35	44.60±9.37	39.83±8.30	58.74±4.58	44.71±9.26
GraphPrompt	37.28±5.35	38.94±4.32	38.70±7.33	45.66±5.16	53.66±1.06	48.91±1.38
GCOPE	40.08±3.25	56.75±1.65	45.23±0.20	43.62±2.16	63.41±2.26	67.60±1.78
SAMGPT	42.59±12.17	47.37±8.89	44.66±9.57	48.79±9.68	57.39±10.16	60.45±8.64
MDGFM	39.77±3.75	50.07±1.50	45.68±1.72	44.48±3.95	59.65±4.77	64.80±3.91
TSCGFM (Ours)	**44.40±2.53**	**59.64±2.89**	**46.75±0.75**	**50.21±1.92**	**64.78±3.32**	**71.58±1.38**

Method	Texas	Cornell	Chameleon	Squirrel	Penn94
GCN	33.87±7.84	30.12±8.13	28.70±1.91	26.33±2.18	50.10±2.02
GAT	34.63±19.34	23.00±5.22	29.50±1.96	27.37±1.67	49.46±1.80
FAGCN	43.37±13.41	29.62±9.48	29.99±2.43	30.22±2.34	49.48±1.51
DGI	31.35±16.09	28.65±9.61	29.78±1.82	23.30±3.39	50.43±1.21
GraphCL	39.89±2.50	37.08±2.32	34.59±2.19	25.65±1.06	51.29±1.97
GPF	50.81±8.49	36.73±9.56	27.24±3.68	21.46±1.81	50.11±2.13
GraphPrompt	61.27±0.84	42.35±2.35	27.65±0.16	25.54±0.03	48.08±0.94
GCOPE	79.50±3.67	55.63±1.90	33.86±3.87	31.42±1.50	52.39±0.33
SAMGPT	66.24±8.61	42.09±9.10	34.64±5.21	30.22±5.73	50.80±1.78
MDGFM	72.50±7.68	51.13±3.50	35.38±2.39	32.82±3.05	52.56±0.36
TSCGFM (Ours)	**80.88±1.65**	**60.70±3.38**	**36.54±1.95**	**33.26±1.59**	**53.57±0.10**

Best and second-best mean performances are shown in **bold** and underline, respectively.

Ablation Studies We next examine whether the gains arise from concordance-conditioned organization rather than simply from additional coordinators or communication edges. We compare the full model with five variants: single coordinator, which removes node-level structural specialization; w/o meta coordinators, which retains concordance-conditioned local organization but removes cross-domain meta coordination; hard routing, which assigns each node exclusively to either the high- or low-concordance coordinator according to its concordance; uniform routing, which assigns equal weights to both coordinators; and random routing, which randomly permutes node-level routing coefficients while preserving the same architecture and routing-weight distribution.

Table 2: Ablation study of concordance-conditioned routing and hierarchical coordination.

Method	Cora	Citeseer	PubMed	Computers	Photo
Single coordinator	40.08±3.25	56.75±1.65	45.23±0.20	43.62±2.16	63.41±2.26
w/o meta coordinators	40.35±3.90	53.33±4.55	45.65±2.99	47.05±3.39	64.17±2.07
Uniform routing	42.81±1.38	55.41±2.69	46.05±1.26	40.61±9.13	60.67±2.68
Random routing	41.57±4.54	56.11±3.18	44.42±1.30	37.81±3.37	61.29±2.21
Hard routing	39.82±2.57	58.25±2.41	45.00±3.99	45.00±3.57	64.36±1.05
Ours	**44.40±2.53**	**59.64±2.89**	**46.75±0.75**	**50.21±1.92**	**64.78±3.32**

Table 2 reveals three consistent patterns. First, the full model outperforms the single-coordinator variant across all datasets, supporting the benefit of node-level structural specialization. Second, removing meta coordination consistently degrades performance, indicating that local structural organization benefits from subsequent cross-domain information exchange. Third, the three routing controls isolate complementary aspects of concordance-conditioned organization. Uniform routing underperforms the full model despite retaining the same architecture, showing that dual coordinators alone are insufficient and that node-specific routing matters. Random routing preserves heterogeneous routing weights but breaks their correspondence with node concordance, further demonstrating that the routing signal must be structurally informative. Finally, hard routing retains concordance-aware assignment but discretizes nodes into two structural regimes; its lower performance supports continuously adapting coordination strength to the degree of structural concordance. Together, these results support the complementary roles of concordance-conditioned soft routing and hierarchical meta coordination in cross-domain transfer.

Performance under Different Numbers of Shots To assess the effectiveness of cross-domain transfer under varying levels of target supervision, we compare TSCGFM with the three strongest baselines, i.e., GCOPE, MDGFM, and SAMGPT, under different m-shot settings. As shown in Figure 3, TSCGFM consistently achieves competitive or superior performance across different numbers of labeled nodes per class m. The advantage is particularly pronounced in low-shot settings, where limited target supervision makes effective knowledge transfer from source domains especially important. As m increases, stronger target-domain supervision reduces the dependence on transferred knowledge, and the relative improvement over competing GFMs gradually narrows. These results demonstrate the effectiveness of concordance-conditioned coordination, particularly when labeled data in the unseen target domain are scarce.

Parameter Sensitivity. We examine the sensitivity of TSCGFM to the concordance reference γ, which determines the center of concordance-conditioned routing, and the routing temperature τ, which controls routing smoothness. Both hyperparameters are selected exclusively based on validation performance on a held-out 10% of the labeled nodes from each source domain, without access to target-domain labels. As shown in Figure 4, accuracy remains relatively stable across a wide range of values for both hyperparameters across different datasets. These results demonstrate that TSCGFM is robust to the choices of both γ and τ and does not require narrowly tuned hyperparameters.

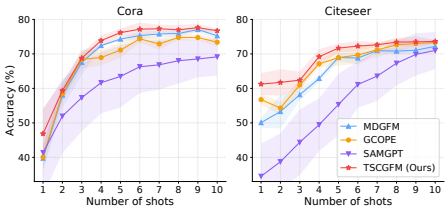

Figure 3: m-shot performance.

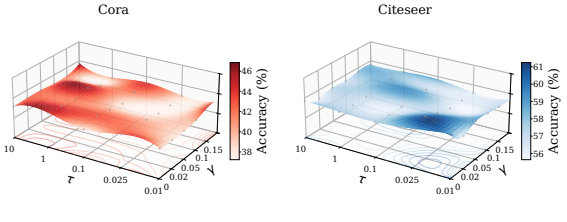

Figure 4: Sensitivity to parameters τ and γ.

Computational Efficiency Finally, we quantify the computational overhead introduced by hierarchical coordination. With M source graphs and N original nodes, TSCGFM introduces two concordance-conditioned coordinators and one global meta-coordinator for each source graph, resulting in $|\mathcal{V}'| = N + 3M$. Node-to-coordinator routing introduces $O(N)$ weighted edges, while connecting the graph-specific coordinators to the meta coordinator requires only $O(M)$ additional edges. Therefore, the augmented graph contains $|E'| = |E| + O(N) + O(M)$ edges. Since $M \ll N$, the coordination structure introduces only a linear number of additional edges with respect to the number of nodes, while retaining efficient message passing over the augmented graph. Figure 5 further compares the empirical running time and accuracy of TSCGFM with GCOPE, MDGFM, and SAMGPT. Although hierarchical coordination introduces additional computation, TSCGFM maintains comparable computational efficiency while achieving consistently higher accuracy, demonstrating a favorable trade-off between efficiency and performance.

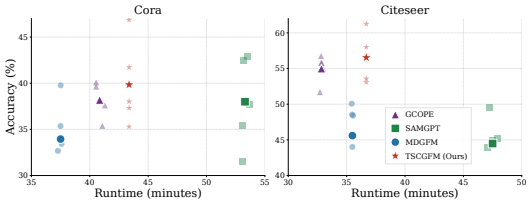

Figure 5: Efficiency analysis.

6 Conclusion

We study cross-domain transfer in graph foundation models by shifting both the basis and the granularity of structural transfer. Rather than directly aligning domain-specific graph structures, we introduce structural concordance—the local agreement between topology and feature-induced neighborhood geometry—as a domain-agnostic coordinate for identifying transferable structural relations. Building on this coordinate, our framework organizes heterogeneous nodes through concordance-conditioned soft routing and subsequently enables cross-domain information exchange through hierarchical meta coordination, thereby preserving local structural diversity without homogenizing graph domains. Our theory establishes the approximation advantage, statistical reliability, and representation stability of concordance-aware routing, while experiments demonstrate improved transfer to unseen domains. Overall, our results suggest that transferable graph knowledge can be organized through shared local structure–feature relations rather than direct structural similarity across domains.

AI USE STATEMENT

In this work, we used generative AI tools to assist with intermediate mathematical derivations and provide ingredients for proving mathematical claims, assist with the writing and verification of proofs, provide feedback on research methodology and experimental design, and assist with the interpretation and presentation of experimental results. We did not use generative AI tools to generate synthetic datasets, develop the theoretical models or conceptual framework of the proposed method, formulate mathematical claims, propose or refine research hypotheses, implement the proposed method, assist with translation, or clean or reformat datasets. Qualitative and thematic data analysis was not applicable to this work.

Additionally, we used generative AI tools for brainstorming, summarizing and discussing relevant literature, suggesting experimental parameters, creating or editing auxiliary software code, drafting and restructuring parts of the manuscript, improving the clarity and readability of the writing, suggesting paper structure and presentation, and assisting with reference formatting.

All AI-assisted content was critically reviewed and revised by the authors. In particular, all AI-assisted mathematical derivations and proofs were independently checked by the authors, experimental choices and interpretations were validated against the actual experimental results, and AI-assisted code was reviewed and tested before use. The research questions, theoretical and conceptual framework, mathematical claims, final methodological choices, theoretical results, and experimental conclusions were determined and verified by the authors. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops a general framework for cross-domain graph representation learning and does not collect new data. Our experiments use publicly available benchmark datasets following their intended research purposes. As with general-purpose graph representation learning methods, models pretrained across heterogeneous graph domains may inherit biases or sensitive structural patterns present in the underlying data. Care should therefore be taken when applying the proposed framework to high-stakes or privacy-sensitive domains, with appropriate evaluation and safeguards for the specific deployment context.

REPRODUCIBILITY STATEMENT

We provide the methodological, theoretical, and experimental details needed to reproduce our results in the appendix and supplementary materials. Section A provides complete proofs of the theoretical results in Section 4, covering the informativeness, estimability, and representation stability of concordance-conditioned routing. Section B provides the experimental details required for reproduction, including the datasets, experimental configurations, and baseline implementations. We additionally provide the implementation code in the supplementary materials.

REFERENCES

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence and Thirty-Third Conference on Innovative Applications of Artificial Intelligence and Eleventh Symposium on Educational Advances in Artificial Intelligence*, volume 35, pp. 3950–3957. AAAI Press, 2021.

Taoran Fang, Yunchao Zhang, Yang Yang, Chunping Wang, and Lei Chen. Universal prompt tuning for graph neural networks. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Proceedings of the 37th International Conference on Neural Information Processing Systems*, volume 36 of *NIPS '23*, pp. 52464–52489, Red Hook, NY, USA, 2023. Curran Associates, Inc.

Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. GPT-GNN: Generative pre-training of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International*

Conference on Knowledge Discovery & Data Mining, KDD '20, pp. 1857–1867, New York, NY, USA, 2020. Association for Computing Machinery.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *The Fifth International Conference on Learning Representations*, 2017.

Mingkai Lin, Xiaobin Hong, Wenzhong Li, and Sanglu Lu. Unified graph neural networks pre-training for multi-domain graphs. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, pp. 12165–12173. AAAI Press, 2025.

Jingzhe Liu, Haitao Mao, Zhikai Chen, Bingheng Li, Wenqi Fan, Mingxuan Ju, Tong Zhao, Neil Shah, and Jiliang Tang. One model for one graph: A new perspective for pretraining with cross-domain graphs. *arXiv preprint arXiv:2412.00315*, 2024.

Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. GraphPrompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the ACM Web Conference 2023*, WWW '23, pp. 417–428, New York, NY, USA, 2023. Association for Computing Machinery.

Haitao Mao, Zhikai Chen, Wenzhuo Tang, Jianan Zhao, Yao Ma, Tong Zhao, Neil Shah, Mikhail Galkin, and Jiliang Tang. Position: Graph foundation models are already here. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 34670–34692. PMLR, 21–27 Jul 2024.

Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pp. 43–52, New York, NY, USA, 2015. Association for Computing Machinery.

Galileo Namata, Ben London, Lise Getoor, and Bert Huang. Query-driven active surveying for collective classification. In *Proceedings of the 10th International Workshop on Mining and Learning with Graphs (MLG)*, pp. 1–8, Edinburgh, Scotland, UK, 2012.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: Geometric graph convolutional networks. In *The Eighth International Conference on Learning Representations*, 2020.

Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. GCC: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pp. 1150–1160, New York, NY, USA, 2020. Association for Computing Machinery.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. In *Relational Representation Learning Workshop (R2L) at NeurIPS*, Montréal, Canada, 2018.

Mingchen Sun, Kaixiong Zhou, Xin He, Ying Wang, and Xin Wang. GPPT: Graph pre-training and prompt tuning to generalize graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, pp. 1717–1727, New York, NY, USA, 2022. Association for Computing Machinery.

Amanda L. Traud, Peter J. Mucha, and Mason A. Porter. Social structure of Facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *The Sixth International Conference on Learning Representations*, 2018.

Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, R Devon Hjelm, et al. Deep graph infomax. In *The Seventh International Conference on Learning Representations*, 2019.

Shuo Wang, Bokui Wang, Zhixiang Shen, Boyan Deng, and Zhao Kang. Multi-domain graph foundation models: Robust knowledge transfer via topology alignment. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64806–64821. PMLR, 13–19 Jul 2025a.

Yuxiang Wang, Wenqi Fan, Suhang Wang, and Yao Ma. Towards graph foundation models: A transferability perspective. *arXiv preprint arXiv:2503.09363*, 2025b.

Zehong Wang, Zheyuan Liu, Tianyi Ma, Jiazheng Li, Zheyuan Zhang, Xingbo Fu, Yiyang Li, Zhengqing Yuan, Wei Song, Yijun Ma, Qingkai Zeng, Xiusi Chen, Jianan Zhao, Jundong Li, Meng Jiang, Pietro Lio, Nitesh Chawla, Chuxu Zhang, and Yanfang Ye. Graph foundation models: A comprehensive survey. *arXiv preprint arXiv:2505.15116*, 2025c.

Zehong Wang, Chuxu Zhang, Jundong Li, Nitesh Chawla, and Yanfang Ye. Graph foundation models: Challenges, methods, and open questions. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 6184–6194, New York, NY, USA, 2025d. Association for Computing Machinery.

Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z. Li. SimGRACE: A simple framework for graph contrastive learning without data augmentation. In *Proceedings of the ACM Web Conference 2022*, WWW '22, pp. 1070–1079, New York, NY, USA, 2022. Association for Computing Machinery.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33 of *NIPS '20*, pp. 5812–5823, Red Hook, NY, USA, 2020. Curran Associates, Inc.

Xingtong Yu, Chang Zhou, Yuan Fang, and Xinming Zhang. Text-free multi-domain graph pre-training: Toward graph foundation models. *arXiv preprint arXiv:2405.13934*, 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, New York, NY, USA, 2025. Association for Computing Machinery.

Haonan Yuan, Qingyun Sun, Junhua Shi, Xingcheng Fu, Bryan Hooi, Jianxin Li, and Philip S. Yu. How much can transfer? BRIDGE: Bounded multi-domain graph foundation model with generalization guarantees. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 73604–73644. PMLR, 13–19 Jul 2025.

Haihong Zhao, Aochuan Chen, Xiangguo Sun, Hong Cheng, and Jia Li. All in one and one for all: A simple yet effective method towards cross-domain graph pretraining. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 4443–4454, New York, NY, USA, 2024. Association for Computing Machinery.

Yilun Zheng, Sitao Luan, and Lihui Chen. What is missing for graph homophily? disentangling graph homophily for graph neural networks. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Proceedings of the 38th International Conference on Neural Information Processing Systems*, volume 37 of *NIPS '24*, pp. 68406–68452, Red Hook, NY, USA, 2024. Curran Associates, Inc.

This appendix provides complete proofs and additional experimental details supporting the main paper. Section A proves the theoretical results in Section 4, following the progression from informativeness to estimability and representation stability. Section B reports the datasets, experimental configurations, routing hyperparameters, and baseline details used in our evaluation.

A PROOFS

This section provides complete proofs of the theoretical results in Section 4. We first establish when structural concordance is informative for node-adaptive coordination, then show that the induced routing is estimable from finite neighborhoods, and finally prove that routing estimation error induces controlled perturbations in GNN representations.

A.1 PROOF OF THE APPROXIMATION ADVANTAGE

Theorem 4.1 characterizes when node-level concordance provides information beyond a shared graph-level coordinator. The key observation is that dual coordination is equivalent to a vector-valued linear regression of ideal coordinator messages on the concordance-induced routing coordinate.

Proof of Theorem 4.1. Recall that $m_i = \alpha_i^* c_H + (1-\alpha_i^*)c_L = c_L + \alpha_i^*(c_H - c_L)$. Let $d = c_H - c_L$. Then,

$$\mathcal{E}_{\text{soft}} = \min_{c_L, d} \sum_{i=1}^{N} \|q_i - c_L - \alpha_i^* d\|_2^2.$$

For fixed d, minimizing over c_L gives $c_L^* = \bar{q} - \bar{\alpha} d$, where $\bar{q} = \frac{1}{N}\sum_{i=1}^{N} q_i$, $\bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i^*$. Substituting c_L^* yields

$$\mathcal{E}_{\text{soft}} = \min_{d} \sum_{i=1}^{N} \|(q_i - \bar{q}) - (\alpha_i^* - \bar{\alpha})d\|_2^2 .$$

This is a least-squares problem with scalar predictor $\alpha_i^* - \bar{\alpha}$ and vector-valued response $q_i - \bar{q}$. Since $\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})^2 > 0$, the optimal direction is

$$d^* = \frac{\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})(q_i - \bar{q})}{\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})^2}.$$

Substituting d^* gives

$$\mathcal{E}_{\text{soft}} = \sum_{i=1}^{N} \|q_i - \bar{q}\|_2^2 - \frac{\|\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})(q_i - \bar{q})\|_2^2}{\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})^2}.$$

Since the optimal single coordinator is $c^* = \bar{q}$, $\mathcal{E}_{\text{single}} = \sum_{i=1}^{N} \|q_i - \bar{q}\|_2^2$, we obtain

$$\mathcal{E}_{\text{soft}} = \mathcal{E}_{\text{single}} - \frac{\|\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})(q_i - \bar{q})\|_2^2}{\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})^2},$$

which proves Eq. (5). The reduction term is nonnegative, so $\mathcal{E}_{\text{soft}} \le \mathcal{E}_{\text{single}}$, with strict inequality whenever $\sum_{i=1}^{N}(\alpha_i^* - \bar{\alpha})(q_i - \bar{q}) \neq 0$. Hence, concordance-conditioned routing yields a strict approximation advantage precisely when the routing coordinate explains nonzero variation in node-specific coordination requirements. □

A.2 PROOF OF CONCORDANCE-ROUTING CONCENTRATION

We next show that concordance-conditioned routing can be reliably estimated from finite neighborhoods by first concentrating the empirical concordance score and then transferring the bound through the sigmoid routing function.

Proof of Theorem 4.2. For node v_i, $s_i = \frac{1}{d_i}\sum_{j\in\mathcal{N}(i)} Z_{ij}$, where $\{Z_{ij} : v_j \in \mathcal{N}(i)\}$ are conditionally independent Bernoulli variables with common mean s_i^*. Since $Z_{ij} \in [0,1]$, Hoeffding's inequality gives, for any $\epsilon > 0$,

$$\mathbb{P}\left(|s_i - s_i^*| \geq \epsilon\right) \leq 2\exp(-2d_i\epsilon^2).$$

Setting $\epsilon = \sqrt{\log(2/\delta)/(2d_i)}$ yields, with probability at least $1-\delta$,

$$|s_i - s_i^*| \leq \sqrt{\frac{\log(2/\delta)}{2d_i}}.$$

Recall that $\alpha_i = \sigma((s_i-\gamma)/\tau)$, $\alpha_i^* = \sigma((s_i^*-\gamma)/\tau)$. Since $\sigma'(x) = \sigma(x)(1-\sigma(x)) \leq 1/4$, the sigmoid is $1/4$-Lipschitz. By the mean value theorem,

$$|\alpha_i - \alpha_i^*| \leq \frac{1}{4\tau}|s_i - s_i^*|.$$

Therefore,

$$|\alpha_i - \alpha_i^*| \leq \frac{1}{4\tau}\sqrt{\frac{\log(2/\delta)}{2d_i}}$$

with probability at least $1-\delta$.

Finally, $|(1-\alpha_i) - (1-\alpha_i^*)| = |\alpha_i - \alpha_i^*|$, so the complementary routing coefficient satisfies the same bound. Hence, both concordance and routing estimation errors decay as $O(d_i^{-1/2})$. □

A.3 Proof of Representation Stability

We now propagate the routing error through the GNN encoder by bounding the perturbation of coordinator messages and unrolling the resulting layer-wise recursion.

Theorem A.1 (Representation stability under concordance-routing error)**.** *Assume that ϕ is ρ-Lipschitz, $\|P\|_2 \leq 1$, $\|W^l\|_2 \leq B_l$, and $\left\|W_H^l c_H^l - W_L^l c_L^l\right\|_2 \leq B_c^l$. Suppose that $\widehat{H}^0 = H^0 = X$. Then, after L message-passing layers,*

$$\|\widehat{H}^L - H^L\|_F \leq \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2 \sum_{l=0}^{L-1}\left(\rho B_c^l \prod_{r=l+1}^{L-1}\rho B_l\right),$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ collect the empirical and population routing coefficients.

Proof of Theorem A.1. Define $\Delta^l = \widehat{H}^l - H^l$. From

$$\widehat{H}^{l+1} = \phi\big(P\widehat{H}^l W^l + M^l\big), \qquad H^{l+1} = \phi\left(PH^l W^l + M^{*l}\right),$$

and the ρ-Lipschitz continuity of ϕ,

$$\|\Delta^{l+1}\|_F \leq \rho\left\|P\Delta^l W^l + M^l - M^{*l}\right\|_F \leq \rho B_l\|\Delta^l\|_F + \rho\|M^l - M^{*l}\|_F,$$

where $\|P\|_2 \leq 1$ and $\|W^l\|_2 \leq B_l$.

For node v_i,

$$m_i^l - m_i^{*l} = (\alpha_i - \alpha_i^*)\left(W_H^l c_H^l - W_L^l c_L^l\right).$$

By assumption, $\left\|W_H^l c_H^l - W_L^l c_L^l\right\|_2 \leq B_c^l$, and therefore

$$\|m_i^l - m_i^{*l}\|_2 \leq B_c^l|\alpha_i - \alpha_i^*|.$$

Aggregating over nodes gives

$$\|M^l - M^{*l}\|_F \leq B_c^l\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2.$$

Hence,

$$\|\Delta^{l+1}\|_F \leq \rho B_l\|\Delta^l\|_F + \rho B_c^l\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2.$$

Since $\Delta^0 = \widehat{H}^0 - H^0 = 0$, unrolling the recursion over L layers yields

$$\|\widehat{H}^L - H^L\|_F \leq \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2 \sum_{l=0}^{L-1}\left(\rho B_c^l \prod_{r=l+1}^{L-1}\rho B_l\right),$$

which proves the result. □

A.4 Proof of the End-to-End Representation Bound

Theorem 4.3 combines finite-neighborhood routing concentration with representation stability.

Proof of Theorem 4.3. By Theorem 4.2, for each node v_i and $\delta_i \in (0, 1)$,

$$|\alpha_i - \alpha_i^*| \leq \frac{1}{4\tau}\sqrt{\frac{\log(2/\delta_i)}{2d_i}}$$

with probability at least $1 - \delta_i$. Set $\delta_i = \delta/N$. By the union bound, with probability at least $1 - \delta$,

$$|\alpha_i - \alpha_i^*| \leq \frac{1}{4\tau}\sqrt{\frac{\log(2N/\delta)}{2d_i}}$$

simultaneously for all nodes. Therefore,

$$\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2 \leq \frac{1}{4\tau}\sqrt{\sum_{i=1}^{N}\frac{\log(2N/\delta)}{2d_i}}.$$

Applying Theorem A.1 yields

$$\|\widehat{H}^L - H^L\|_F \leq \frac{1}{4\tau}\sqrt{\sum_{i=1}^{N}\frac{\log(2N/\delta)}{2d_i}}\sum_{l=0}^{L-1}\left(\rho B_c^l \prod_{r=l+1}^{L-1}\rho B_l\right),$$

which proves the corollary. □

Together, the proofs establish the theoretical chain underlying structural-concordance coordination. Theorem 4.1 establishes informativeness, Theorem 4.2 establishes estimability at rate $O(d_i^{-1/2})$, and Theorem A.1 together with Corollary 4.3 establishes representation stability. These results characterize when structural concordance is useful, how reliably it can be estimated, and how its estimation error propagates through the GNN.

B Experimental Details

This section provides the datasets, experimental configurations, routing hyperparameters, and baseline details supporting Section 5.

B.1 Datasets

We evaluate on ten benchmark graphs spanning four domains: citation networks, Amazon co-purchase networks, WebKB webpage networks, and Wikipedia page networks. These datasets vary substantially in feature semantics, topology, and local homophily, providing a heterogeneous testbed for cross-domain transfer based on relative topology–feature relations.

Citation Networks. `Cora` and `Citeseer` (Sen et al., 2008) contain scientific papers connected by citation links, with bag-of-words node features and topic labels. `PubMed` (Namata et al., 2012) is a biomedical citation network with TF-IDF-based features and diabetes-related labels.

Amazon Networks. `Computers` and `Photo` (McAuley et al., 2015; Shchur et al., 2018) are product co-purchase graphs, where nodes represent products, edges indicate frequent co-purchases, features are derived from product information, and labels denote product categories.

WebKB Networks. `Cornell`, `Texas`, and `Wisconsin` (Pei et al., 2020) are university webpage networks with hyperlink edges, bag-of-words features, and five webpage categories.

Wikipedia Networks. Chameleon and Squirrel (Pei et al., 2020) are Wikipedia page networks with hyperlink edges, word-based node features, and labels derived from average monthly page traffic.

Online Social Networks. Penn94 (Traud et al., 2012) is a large-scale online social network derived from the Facebook100 dataset, where nodes represent university students and the task is to predict their reported gender from profile attributes.

Table 3 summarizes the dataset statistics. Together, these benchmarks span both homophilic and heterophilic structures and exhibit substantial variation in graph semantics and topology, making them suitable for evaluating transfer through relative topology–feature relations.

Table 3: Statistics of the multi-domain graph datasets.

Dataset	Domain	#Nodes	#Edges	#Features	#Classes
Cora	Citation	2,708	10,556	1,433	7
Citeseer	Citation	3,327	9,104	3,703	6
PubMed	Citation	19,717	88,648	500	3
Computers	Co-purchase	13,752	491,722	767	10
Photo	Co-purchase	7,650	238,162	745	8
Wisconsin	WebKB	251	515	1,703	5
Texas	WebKB	183	325	1,703	5
Cornell	WebKB	183	298	1,703	5
Chameleon	Wikipedia	2,277	36,101	2,325	5
Squirrel	Wikipedia	5,201	217,073	2,089	5
Penn94	Online Social	41,554	1,362,229	4814	2

B.2 Experimental Configurations

We follow the leave-one-target-out protocol used in the main experiments. For each target domain, the target graph is excluded from both self-supervised pre-training and feature-routing calibration, while the remaining nine graphs serve as source domains. Table 4 summarizes this split.

Table 4: Leave-one-target-out source visibility.

Target	Cora	Cite.	PubMed	Comp.	Photo	Wisc.	Texas	Corn.	Cham.	Squir.
Cora		✓	✓	✓	✓	✓	✓	✓	✓	✓
Citeseer	✓		✓	✓	✓	✓	✓	✓	✓	✓
PubMed	✓	✓		✓	✓	✓	✓	✓	✓	✓
Computers	✓	✓	✓		✓	✓	✓	✓	✓	✓
Photo	✓	✓	✓	✓		✓	✓	✓	✓	✓
Wisconsin	✓	✓	✓	✓	✓		✓	✓	✓	✓
Texas	✓	✓	✓	✓	✓	✓		✓	✓	✓
Cornell	✓	✓	✓	✓	✓	✓	✓		✓	✓
Chameleon	✓	✓	✓	✓	✓	✓	✓	✓		✓
Squirrel	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Penn94	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

A checkmark (✓) denotes a source dataset, while a blank cell denotes the held-out target.

Experiments use Python 3.10.12, PyTorch 2.1.1, CUDA 11.8, and PyG 2.4.0 on one NVIDIA H200 GPU with 8 GB host memory and 4 CPU cores. Features are aligned to dimension 100 using SVD reduction or zero padding. The backbone is a two-layer FAGCN with hidden dimension 128, dropout 0.2, and $\epsilon = 0.1$. Pre-training uses GraphCL with Adam for 100 epochs, batch size 10, learning rate 10^{-4}, weight decay 10^{-5}, reconstruction weight 0.2 and contrastive temperature $\tau_c = 0.5$. We hold out 10% of the labeled nodes from each source domain for validation and select γ and τ based solely on source-domain validation performance, with $\gamma \in \{0, 0.025, 0.05, 0.1, 0.2\}$ and $\tau \in \{0.01, 0.1, 1, 10\}$. No target-domain data are used for hyperparameter selection.

B.3 Baseline Details

We compare against representative methods from four categories: task-specific GNNs, graph self-supervised learning methods, prompt-based graph fine-tuning methods, and multi-domain graph foundation models.

Task-Specific Graph Neural Networks. These methods are trained from scratch on each downstream dataset.

- **GCN** (Kipf & Welling, 2017): propagates node features over the normalized graph structure using graph convolution.
- **GAT** (Veličković et al., 2018): uses attention weights to adaptively aggregate neighboring representations.
- **FAGCN** (Bo et al., 2021): explicitly models both low- and high-frequency graph signals to improve learning on graphs with diverse homophily levels.

Graph Self-Supervised Learning. These methods pre-train graph representations from unlabeled data.

- **DGI** (Veličković et al., 2019): contrasts node embeddings with a graph-level summary to capture global context.
- **GraphCL** (You et al., 2020): contrasts stochastic augmentations of the same graph.

Prompt-based Graph Fine-tuning. These methods adapt pre-trained graph models to downstream tasks by introducing learnable prompts or reformulating downstream tasks into prompt-compatible objectives.

- **GPF** (Fang et al., 2023): uses learnable graph prompts to adapt pre-trained graph models to downstream tasks.
- **GraphPrompt** (Liu et al., 2023): unifies graph pre-training and downstream adaptation under a prompt-based framework.

Multi-Domain Graph Foundation Models. These methods explicitly target transfer across heterogeneous graph domains.

- **GCOPE** (Zhao et al., 2024): introduces coordinator nodes to facilitate information exchange across graph datasets.
- **SAMGPT** (Yu et al., 2025): learns transferable knowledge across text-free graph domains with prompt-based downstream adaptation.
- **MDGFM** (Wang et al., 2025a): improves multi-domain transfer by aligning topological patterns across heterogeneous graphs.