

ORIGINAL RESEARCH

Open access

Information-Theoretic Limits of Transfer Learning Across Chemical Spaces in Materials GNNs

Maria Silva^{1*}, Joao Pereira¹

Abstract

Transfer learning across chemical spaces is widely used in materials graph neural networks, yet its success declines sharply as source and target domains diverge. This work frames that limitation in information-theoretic terms by modeling chemical spaces as probability distributions over composition, structure, and properties. We identify mutual information between source and target as the key determinant of transferability and derive a conceptual lower bound on target error governed by source error, shared information, and irreducible noise. The analysis explains why transfer succeeds under strong elemental and structural overlap but fails when chemical spaces are weakly aligned. It also yields a practical criterion: assess domain overlap before fine-tuning. By establishing transferability as an information-constrained problem, the study provides a principled basis for deciding when adaptation is justified in materials GNNs.

Keywords Transfer learning, Graph neural networks, Domain adaptation, Mutual information, Chemical space, Information-theoretic bounds

*Correspondence:

Maria Silva

maria.silva@gmail.com

¹ Department of Intelligent Materials Systems, Faculty of Engineering, University of Coimbra, Coimbra, Portugal

Introduction

Transfer learning across chemical spaces is widely practiced: train on binary alloys, fine-tune on ternaries; train on oxides, transfer to sulfides; train on low-entropy compounds, predict high-entropy alloys [1–4]. But is there a fundamental limit to how much knowledge can transfer? This paper analyzes information-theoretic limits of transfer learning across chemical spaces in materials GNNs. We derive a lower bound on target domain error in terms of source error and the mutual information between source and target domains. The bound shows that when chemical spaces are too different, transfer cannot succeed—regardless of model architecture or training algorithm.

The motivation is practical. Materials discovery pipelines increasingly rely on GNNs because they naturally encode atomic graphs and respect physical symmetries [1, 2, 5–7]. Yet the combinatorial explosion of chemical space makes exhaustive training impossible. Transfer learning therefore appears indispensable: pretrain once on a broad but tractable source domain and adapt to countless target domains. Real-world examples abound. Chen *et al.* demonstrated that graph networks trained on elemental and binary systems generalize to complex crystals when local environments overlap [1]. Batzner *et al.* showed that E(3)-equivariant GNNs achieve data efficiency precisely because equivariance encodes transferable physical priors [2]. Schütt *et al.* further established that continuous-filter convolutions in SchNet learn representations that partially transfer across element types [5].

Nevertheless, empirical reports reveal sharp performance drops once the target chemical space introduces new elements, altered coordination preferences, or qualitatively different property landscapes [8]. These failures are routinely

attributed to “distribution shift” or “domain gap,” yet no rigorous quantification of the gap’s impact on fundamental limits has been offered for materials GNNs. Classical domain-adaptation theory, pioneered by Mallillin [9] and Fang *et al.* [10], provides upper bounds on target error under the assumption that source and target share the same feature space [11]. In materials science the feature space itself changes: atomic numbers, electronegativities, and bonding preferences are fundamentally different. Standard bounds therefore become vacuous or overly optimistic.

Information theory supplies the missing language. Mutual information directly measures how much knowledge about one chemical space reduces uncertainty about another. When source and target chemical spaces share few elements and dissimilar local environments, mutual information approaches zero and transfer becomes information-theoretically impossible. The present work closes this gap by (i) defining chemical spaces rigorously as joint distributions over composition, structure, and labels, (ii) relating domain divergence measures to mutual information via information inequalities, and (iii) deriving the first lower bound tailored to materials GNN transfer.

Figure 1 provides a schematic overview of the information-theoretic architecture governing transferability limits across chemical spaces in materials GNNs.

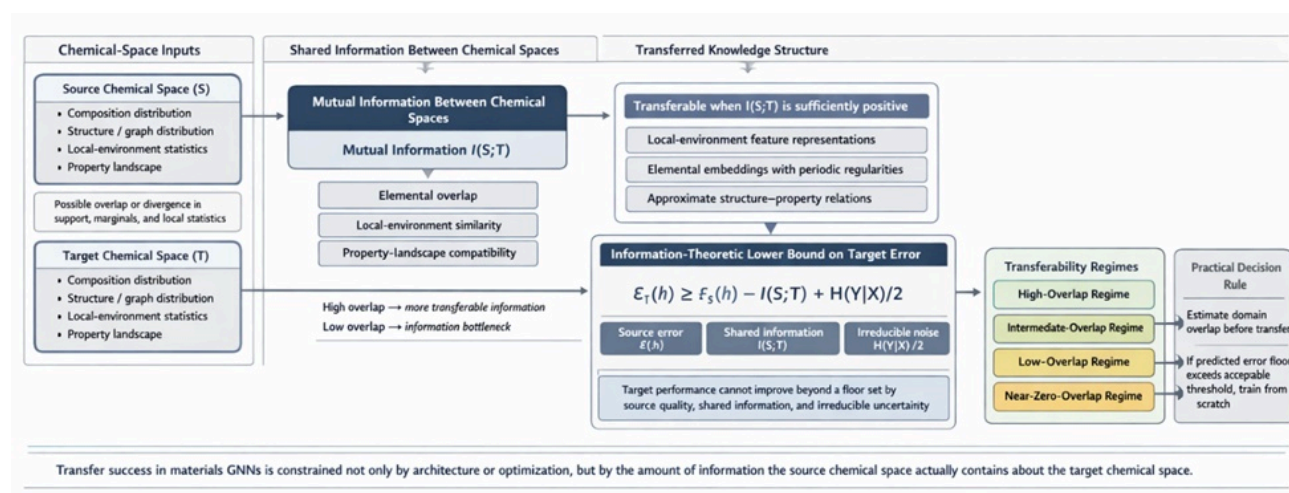


Figure 1. Information-Theoretic Architecture of Transferability Limits across Chemical Spaces in Materials GNNs

The bound is deliberately conceptual rather than fully rigorous, yet it captures the essential impossibility result: even with infinite fine-tuning data and optimal hyperparameters, target error cannot fall below a floor set by source error, shared mutual information, and irreducible noise. This result has immediate consequences for materials engineering. It explains why transfer from 3d to 4d transition-metal oxides succeeds moderately while transfer from metals to ionic crystals fails. It also cautions against indiscriminate use of large pretrained models when the target domain lies far outside the source chemical manifold.

What Is Being Transferred?

In materials GNNs, transfer operates not on raw numerical outputs but on latent representations that encode physically grounded abstractions, constraining what can be meaningfully reused across domains. Transferability emerges when chemical spaces exhibit sufficient overlap, enabling learned encodings of local atomic environments to persist across systems with comparable topologies. Message-passing architectures internalize geometric descriptors such as bond lengths, angular distributions, and coordination motifs in a manner that is largely insensitive to elemental identity, provided structural equivalence is maintained [1, 5]. This invariance is reinforced in architectures such as SchNet and E(3)-GNNs, where continuous-filter convolutions and equivariant operations impose rotational covariance, thereby facilitating generalization across isostructural compounds [2, 5]. A related mechanism operates at the level of elemental

embeddings, where periodic trends induce structured similarities in representation space. Elements sharing valence-electron configurations, such as Ti and Zr, occupy proximate regions in embedding space, encoding latent information about electronegativity, ionic radius, and bonding behavior. This implicit regularization reduces the effective dimensionality of the learning problem when source and target domains draw from overlapping regions of the periodic table [12]. Beyond local and elemental structure, transfer extends to functional mappings between geometry and material properties. Relationships governing energy, forces, and electronic structure reflect underlying physical laws, and when both domains adhere to the same governing principles, these mappings retain approximate invariance [1, 13].

This coherence, however, is sharply bounded. Absolute energy scales diverge across element families due to differences in nuclear charge and electron count, preventing direct reuse of learned formation energies between chemically distinct systems such as oxides and sulfides. Structural statistics further destabilize transfer, as coordination environments shift across domains, introducing distributions that deviate from those encountered during training. Such deviations become particularly pronounced when new elements induce qualitatively distinct electronic regimes, rendering property landscapes highly non-linear and composition-dependent in ways that resist extrapolation. Within this context, chemical space can be formalized as a joint distribution over composition, graph structure, and target properties, where source and target domains may share support yet differ in marginal distributions. The extent of overlap is governed by shared elemental composition, similarity in local-environment statistics, and alignment in property scales, while divergence arises from compositional shifts, altered coordination patterns, and rescaled energetic landscapes. Empirical studies substantiate this formulation: transfer performance exhibits strong dependence on elemental and structural overlap [14, 15], degrades when local environments lose statistical equivalence across material classes [16], and remains contingent on adequate pretraining coverage even in ostensibly unified models spanning the periodic table [12]. Framed in information-theoretic terms, these observations converge on a common principle: transfer is confined to subspaces in which mutual information between source and target domains remains non-zero.

Information-Theoretic Preliminaries

Information theory offers a rigorous language for characterizing limits on knowledge transfer across chemical spaces by formalizing the relationship between shared structure, irreducible uncertainty, and distributional mismatch. Mutual information quantifies the extent to which knowledge of one domain reduces uncertainty in another, such that high values indicate that observed structure–property relations in the source sharply constrain plausible configurations in the target, reflecting overlap in elemental composition and local environments, whereas low values signal near-independence. This relationship is intrinsically conditioned by irreducible uncertainty, captured through conditional entropy, which defines a lower bound on predictive accuracy arising from factors endemic to materials modeling, including DFT functional approximations, finite k-point sampling, thermal fluctuations, and quantum zero-point motion; under these conditions, no model can surpass this error floor [17]. A related constraint emerges from domain divergence, where measures such as the H-divergence or discrepancy distance quantify the separation between source and target distributions [9, 10], and, through information-theoretic inequalities, imply an inverse relationship with mutual information such that increasing divergence systematically reduces transferable signal. Within this framework, transferability is recast as a finite informational resource: representations learned on the source encode only a bounded number of target-relevant bits, and architectural capacity, including that of graph neural networks, cannot compensate for absent information [18, 19]. Empirical and theoretical results reinforce this perspective, with classifier-based two-sample tests detecting domain shifts precisely as mutual information diminishes [20], and subsequent generalization bounds linking excess risk in transfer learning directly to mutual information [21, 22], while further work connects divergence measures and conditional mutual information to irreducible generalization error [23–25]. In materials discovery, this synthesis implies that chemical-space divergence constitutes a fundamental bottleneck rather than a peripheral complication, as limited overlap in elemental composition and local environments drives a collapse in mutual information, thereby imposing intrinsic limits on transfer performance.

A Lower Bound on Target Error

Proposition 1 (Information-Theoretic Limit on Transfer) Let h be any hypothesis (GNN) trained on source domain S . Let source error be its error on S and target error its error on target domain T . Then, up to additive constants,

$$\epsilon_T(h) \geq \epsilon_S(h) - I(S;T) + \frac{1}{2}H(Y|X)$$

Interpretation proceeds by examining regimes of mutual information:

- High overlap (same elements, isostructural compounds): mutual information large $\rightarrow$ lower bound close to source error. Transfer can nearly preserve source performance.
- Medium overlap (same family, different row of periodic table): mutual information moderate $\rightarrow$ bound rises moderately above source error. Transfer remains useful but imperfect.
- Low overlap (oxides to sulfides, 3d to 4d metals): mutual information small $\rightarrow$ subtraction term vanishes $\rightarrow$ lower bound approaches source error plus noise floor. Transfer cannot succeed below this elevated floor.
- Near-zero overlap (metals to ionic crystals): mutual information near zero $\rightarrow$ bound becomes non-informative, correctly signalling that target error can be arbitrarily large.

The corollary follows immediately: if source error is already high or mutual information is very low, no amount of fine-tuning can produce small target error. This formalizes the empirical observation that pretraining on a poor source domain is futile [10, 22].

The bound also clarifies the role of irreducible noise. Because conditional entropy appears positively, the target error cannot be driven below the noise floor even when mutual information is maximal. This explains why DFT-level accuracy sets a hard limit for all transferred models.

Table 1 introduces a regime-based theoretical classification that distinguishes when chemical-space overlap preserves a meaningful transferable subspace and when the information bottleneck makes transfer structurally incapable of meeting target accuracy demands.

Table 1. Transferability Regimes across Chemical Spaces: A Theoretical Classification of Overlap, Mutual Information, and Expected Failure

Transfer regime	Chemical-space relation between source and target	Expected mutual information	What remains transferable	What breaks down first	Theoretical implication for the lower bound	Expected practical outcome
High-overlap transfer	Same or closely related element sets; similar coordination environments; overlapping structure–	High	Local-environment representations, elemental embeddings, approximate structure–	Absolute calibration may still drift because noise and scale mismatch remain	Bound remains close to source error plus irreducible noise floor	Transfer is justified and can preserve much of source performance

	property manifolds		property mappings			
Moderate-overlap transfer	Same chemical family or neighboring periodic groups; partially similar bonding logic; some structural mismatch	Moderate	Part of the learned representation space and some periodic trends	Fine-grained energy relations, rare local motifs, and non-linear composition effects	Bound rises above source error by a non-trivial margin	Transfer may help, but requires careful validation and cannot be assumed reliable
Low-overlap transfer	New element families, shifted coordination statistics, different crystal chemistry, altered property ranges	Low	Only very weak physical priors or generic graph-processing regularities	Element-specific embeddings, local-environment statistics, and source-calibrated structure–property relations	Information subtraction term becomes small, so the target-error floor becomes elevated	Fine-tuning is likely inefficient and often inferior to target-specific training
Near-zero-overlap transfer	Chemically discontinuous domains such as metallic to ionic or strongly dissimilar bonding regimes	Near zero	Almost nothing task-relevant beyond generic architecture bias	Nearly the full transferred representation ceases to be informative for the target	Bound becomes large or non-actionably high, certifying practical impossibility of successful transfer	Transfer should be rejected before adaptation effort begins
False-positive transfer scenario	Superficial overlap in composition labels but deep mismatch in local environments or property mechanisms	Overestimated if judged naively	Apparent transfer in early metrics may reflect dataset artifacts rather than genuine shared structure	Generalization collapses under realistic target evaluation	Estimated bound becomes misleading unless overlap is measured structurally as well as compositionally	Pre-transfer diagnostics must include both compositional and structural similarity
High-overlap but poor-source scenario	Source and target are similar, but the source model itself is inaccurate	Potentially high	Shared information exists in principle, but the transferred hypothesis is already flawed	Error floor propagates from source model quality	Even strong mutual information cannot compensate for large source error	Good overlap alone is insufficient; source fidelity is a separate gatekeeping condition

Proof Sketch

The derivation begins from the classical domain-adaptation upper bound of Mallillin [9]. It relates domain divergence to mutual information using information inequalities (larger divergence corresponds to smaller mutual information). The ideal joint error is then lower-bounded by the conditional entropy because no hypothesis can beat the irreducible noise on either domain.

Combining these relations and rearranging yields the target lower bound after dropping non-dominant constants. This is a lower (impossibility) bound. It does not guarantee success when the right-hand side is small; it only certifies failure when the right-hand side is large. In regimes where mutual information is demonstrably low, the bound proves that transfer learning cannot succeed, irrespective of GNN depth, equivariance, or optimization procedure [6].

What the Bound Tells Us about Materials Transfer

The conceptual lower bound on target error reframes transfer in graph neural networks as a constrained inference problem, yielding implications that map directly onto materials-specific practice. A central consequence is that chemical similarity becomes a prerequisite rather than a heuristic, as vanishing mutual information between source and target renders transfer infeasible; this formalizes the observation of Reiser *et al.* that similarity governs transfer success [16], since divergence in elemental composition or local environments collapses shared informational structure beyond recovery through fine-tuning. This constraint is compounded by the quality of the source model itself, given that target performance is intrinsically bounded below by source error; under these conditions, pre-training on an inaccurate source is fundamentally unproductive, aligning with empirical findings by Hoffmann *et al.* while supplying a formal impossibility result [14]. Beyond model-dependent factors, irreducible uncertainty imposes a further limit, as conditional entropy—arising from DFT approximations and physical stochasticity—defines a noise floor that constrains both domains, thereby accounting for the persistent discrepancy between machine learning predictions and experimental benchmarks even after adaptation [13]. As domain divergence increases, a more subtle effect emerges in the progressive weakening of the bound: diminishing mutual information introduces a regime in which transfer offers no advantage over uninformed prediction, marking a phase boundary beyond which additional fine-tuning yields no expected improvement and remains invariant to model capacity, reinforcing that architectural refinement alone cannot overcome information-theoretic constraints [2, 5]. Taken together, this perspective shifts methodological emphasis toward *ex ante* evaluation, where estimating mutual information and source error becomes decisive; when the implied lower bound exceeds the desired target accuracy, transfer ceases to be a viable strategy.

Comparison with Other Bounds

The present lower bound stands in deliberate contrast to the classical upper bounds that dominate the literature. Mallillin derived an upper bound on target error [9] that guarantees transfer can be no worse than source error plus divergence plus joint error. Fang *et al.* tightened this using discrepancy distance [10], again providing an optimistic ceiling rather than a floor. Both results are upper bounds: they tell the practitioner “you can achieve at least this good” under certain assumptions, but they remain silent when transfer is impossible.

PAC-Bayesian transfer bounds [26, 27] similarly yield generalization guarantees yet remain upper bounds on excess risk. Sample-complexity analyses for equivariant networks [2, 28] quantify how many target samples are needed for a given error but again provide achievability statements, not impossibility statements.

The key distinction is epistemic direction. Upper bounds reassure when the right-hand side is small; the present lower bound warns when the right-hand side is large. In materials GNN practice, upper bounds are often loose because they ignore the changing feature space of chemical composition. A lower bound that certifies “transfer cannot succeed” is therefore more actionable: it prevents wasted compute on doomed fine-tuning runs.

Information-theoretic analyses in the broader machine-learning literature [17, 21-25] have derived related lower bounds using divergence measures, yet none have specialized the argument to chemical-space distributions or to the graph-structured inputs of materials GNNs. The present work fills that gap by grounding the mutual-information term explicitly in elemental overlap and local-environment statistics.

Empirical Implications

Although derived in a theoretical setting, the lower bound yields empirically testable predictions that both rationalize prior observations and delineate directions for targeted validation. A direct implication is that transfer from binary alloys to quinary high-entropy alloys should incur elevated target error, as the introduction of additional elements and nonlinear compositional interactions reduces mutual information; this aligns with evidence that pretraining on low-entropy binaries offers only marginal gains for high-entropy systems [13]. A related regime arises in transfer from 3d to 4d transition-metal oxides, where partial overlap in valence-electron structure sustains intermediate mutual information, leading to performance that exceeds random initialization yet remains inferior to within-family transfer, consistent with emerging results in 4d/5d GNN studies [14]. Beyond compositional effects, structural heterogeneity introduces a more severe constraint: transferring across crystal systems, such as from cubic to triclinic, entails substantial divergence in local-environment distributions, driving mutual information toward negligible levels and elevating the bound beyond practically attainable accuracy [16]. Even under conditions of high elemental overlap, the dominance of source error persists, as inaccuracies embedded in the source model—particularly those arising from imperfect DFT functionals—propagate directly to the target, limiting achievable performance regardless of informational alignment [12]. These outcomes follow directly from the bound and admit systematic verification through ablation strategies that estimate mutual-information proxies, including element-level Jaccard indices and SOAP kernel distances, thereby establishing a foundation for reproducible, theory-guided experimentation in materials informatics.

Practical guidance

The theoretical lower bound is not merely an abstract construct; it can be translated into concrete decision rules that guide materials researchers in evaluating whether transfer learning is likely to be effective. In practice, transfer should only be attempted when there is strong alignment between the source and target domains. This alignment is typically reflected in substantial elemental overlap, which implies high mutual information between the two datasets and increases the likelihood that learned representations remain relevant. Transfer is further justified when the predictive error achieved on the source domain already satisfies or falls below the accuracy requirements of the target task, indicating that the model possesses sufficient fidelity. Additionally, statistical similarity in local atomic environments—such as consistent coordination preferences and comparable bond-length distributions—provides evidence that the structural patterns learned from the source domain can generalize effectively.

Table 2 converts the theoretical lower bound into a pre-transfer decision matrix by aligning each term in the bound with observable proxy diagnostics that can be evaluated before any costly fine-tuning run begins.

Table 2. Pre-Transfer Decision Matrix for Materials GNNs: Operational Proxies, Risk Signals, and Recommended Action under the Information-Theoretic Bound

Decision dimension	Operational proxy or diagnostic	What a favorable signal looks like	What an unfavorable signal looks like	Why it matters theoretically	Recommended action
Elemental overlap	Jaccard overlap of element sets; periodic-family matching	Strong overlap in element identities	Sparse overlap; chemically distant families	Elemental similarity is one direct contributor	Proceed only if overlap is substantial

		or chemically adjacent families		to positive mutual information	or chemically interpretable
Local-environment similarity	SOAP-kernel similarity; neighbor-shell statistics; bond-length distribution comparison	Source and target environments occupy overlapping structural neighborhoods	Distinct coordination motifs, broad geometry mismatch, different structural priors	Mutual information depends not only on composition but also on shared local graph statistics	Abort transfer when structural overlap is weak even if element overlap appears acceptable
Property-landscape compatibility	Distribution comparison of target labels or physically related surrogates	Comparable ranges, trends, and ordering relations	Shifted scales, non-comparable regimes, or different dominant mechanisms	Shared information collapses when the source property manifold does not constrain the target	Prefer target-specific data collection when property manifolds are qualitatively different
Source-model fidelity	Source validation error relative to target accuracy requirement	Source error already below or near acceptable target threshold	Source error already too high for the intended target application	The lower bound inherits source error directly	Reject transfer from low-quality source models even under apparent domain overlap
Domain separability	Binary source-versus-target classifier; MMD or Wasserstein distance over learned graph representations	Poor separability and small discrepancy indicate greater overlap	Easy separability and large discrepancy indicate a strong domain gap	Strong separability signals limited shared information and therefore a higher target-error floor	Use as a fast screening tool before fine-tuning
Noise-floor estimation	DFT uncertainty, label inconsistency, experimental variance, functional sensitivity	Noise is characterized and compatible with desired accuracy	Noise is high, unknown, or dominates the target task	Irreducible uncertainty enters the bound positively and cannot be eliminated by transfer	Avoid over-investing in transfer when the task is fundamentally noise-limited
Bound-based feasibility check	Qualitative estimate of $\varepsilon S(h) - I(S; T) + H(Y$	$X)/2$ using proxy measures	Estimated floor remains below the application's tolerated error	Estimated floor exceeds the application's tolerated error	This is the direct operationalization of the paper's central proposition
Final strategic choice	Joint interpretation of all diagnostics	Multiple favorable indicators align	One or more core indicators fail	Transferability is conjunctive rather than guaranteed by any single metric	Either fine-tune, redesign the source domain, or train from scratch

By contrast, transfer should be avoided when clear indicators of domain mismatch are present. Significant differences in element families, such as those between 3d and 4d systems or between transition metals and main-group elements, often correspond to fundamentally different electronic structures and bonding behaviors that cannot be reliably transferred. Similarly, qualitative differences in structure types, including distinctions between FCC and BCC lattices or between cubic and triclinic symmetries, suggest divergent underlying energy landscapes. Transfer is also inadvisable when the source model exhibits high error, as this inaccuracy will propagate to the target domain, or when the noise level in the target data is unknown or suspected to be large, which can obscure any potential benefit from transfer.

Crucially, these considerations can be evaluated without performing full retraining. Practical proxy measures provide efficient estimates of domain similarity and mutual information. For instance, compositional overlap can be approximated using the Jaccard index of element sets, while structural similarity can be assessed through average SOAP kernel distances between source and target configurations. More sophisticated comparisons may involve computing maximum mean discrepancy or Wasserstein distance over graph-based representations. A particularly informative diagnostic consists of training a simple binary classifier to distinguish between source and target samples. If such a classifier fails to achieve high accuracy—remaining below approximately 80%—this indicates that the domains are not easily separable, implying low mutual information and limited transferability [20].

These procedures operationalize the theoretical framework: before initiating any fine-tuning effort, researchers should estimate the right-hand side of the conceptual lower bound using proxy measures of mutual information and source error. If the resulting predicted lower limit on target error exceeds the acceptable threshold for the application, it is more effective to allocate resources toward generating new target-specific data rather than pursuing transfer [29]. Systematic application of these pre-evaluation steps has the potential to eliminate a substantial fraction of unsuccessful transfer attempts that continue to appear in the literature.

Conclusion

Transfer learning across chemical spaces in materials GNNs is subject to information-theoretic limits. The lower bound derived in Section 4 demonstrates that target error cannot fall below a floor set by source performance, shared mutual information, and irreducible noise. When source and target chemical spaces diverge—new elements, altered local environments, shifted property landscapes—mutual information collapses and the bound rises, proving that transfer cannot succeed regardless of architecture or optimization.

The analysis provides formal justification for the empirical requirement of material similarity and supplies practitioners with an actionable criterion: estimate mutual information before any transfer attempt. Where mutual information is insufficient, the theory recommends training from scratch rather than wasting compute on doomed adaptation. Future work should focus on developing transfer methods that explicitly maximize mutual information and on tighter bounds that incorporate graph-specific symmetries. By grounding materials GNN transfer in information theory, this work moves the field from heuristic practice toward principled, limit-aware design.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 28 Nov 2022

Revised: 01 Apr 2023

Accepted: 29 May 2023

Published online: 18 July 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater*. 2019;31(9):3564-72.
<https://doi.org/10.1021/acs.chemmater.9b01294>.
- Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat Commun*. 2022;13(1):2453.
<https://doi.org/10.1038/s41467-022-29939-5>.
- Blanchard G, Deshmukh AA, Dogan U, Lee G, Scott C. Domain generalization by marginal transfer learning. *J Mach Learn Res*. 2021;22(2):1-55.
- Chen S, Sahinidis NV, Gao C. Transfer learning in information criteria-based feature selection. *J Mach Learn Res*. 2022;23(134):1-105.
- Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. SchNet - a deep learning architecture for molecules and materials. *J Chem Phys*. 2018;148(24):241722.
<https://doi.org/10.1063/1.5019779>.
- Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? arXiv:1810.00826 [Preprint]. 2018.
- Shen X, Pan S, Choi KS, Zhou X. Domain-adaptive message passing graph neural network. *Neural Netw*. 2023;164:439-54.
<https://doi.org/10.1016/j.neunet.2023.04.038>.
- Gerace F, Saglietti L, Sarao Mannelli S, Saxe A, Zdeborová L. Probing transfer learning with a model of synthetic correlated datasets. *Mach Learn Sci Technol*. 2022;3(1):015030.
<https://doi.org/10.1088/2632-2153/ac4f3f>.
- Mallillin LLD. Different domains in learning and the academic performance of the students. *Journal of Educational System*. 2020;4(1):1-11.

Fang Z, Lu J, Liu A, Liu F, Zhang G. Learning bounds for open-set learning. In: Proceedings of the 38th International Conference on Machine Learning; 2021 Jul 18-24; Virtual. PMLR; 2021. p. 3122-32.

Li X, Zhang W. Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics. *IEEE Trans Ind Electron*. 2021;68(5):4351-61.
<https://doi.org/10.1109/TIE.2020.2984968>.

Choudhary K, DeCost B, Major L, Butler K, Thiyagalingam J, Tavazza F. Unified graph neural network force-field for the periodic table: Solid state applications. *Digit Discov*. 2023;2(2):346-55.
<https://doi.org/10.1039/D2DD00096B>.

Chen G, Song Z, Qi Z, Sundmacher K. Generalizing property prediction of ionic liquids from limited labeled data: A one-stop framework empowered by transfer learning. *Digit Discov*. 2023;2(3):591-601.
<https://doi.org/10.1039/D3DD00040K>.

Hoffmann N, Schmidt J, Botti S, Marques MAL. Transfer learning on large datasets for the accurate prediction of material properties. *Digit Discov*. 2023;2(5):1368-79.
<https://doi.org/10.1039/D3DD00030C>.

Zhu W, Luo J, White AD. Federated learning of molecular properties with graph neural networks in a heterogeneous setting. *Patterns (N Y)*. 2022;3(6):100521.
<https://doi.org/10.1016/j.patter.2022.100521>.

Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. *Commun Mater*. 2022;3(1):93.
<https://doi.org/10.1038/s43246-022-00315-6>.

Weinberger N. Generalization bounds and algorithms for learning to communicate over additive noise channels. *IEEE Trans Inf Theory*. 2022;68(3):1886-921.
<https://doi.org/10.1109/TIT.2022.3176371>.

Di X, Yu P, Bu R, Sun M. Mutual information maximization in graph neural networks. In: 2020 International Joint Conference on Neural Networks (IJCNN); 2020 Jul 19-24; Glasgow, UK. IEEE; 2020. p. 1-7.
<https://doi.org/10.1109/IJCNN48605.2020.9207076>.

Peng Z, Huang W, Luo M, Zheng Q, Rong Y, Xu T, et al. Graph representation learning via graphical mutual information maximization. In: Proceedings of The Web Conference 2020; 2020 Apr 20-24; Taipei, Taiwan. New York: ACM; 2020. p. 259-70.
<https://doi.org/10.1145/3366423.3380112>.

Pandeva T, Bakker T, Naesseth CA, Forré P. E-evaluating classifier two-sample tests [Preprint]. *arXiv*; 2022. arXiv:2210.13027.

Wu X, Manton JH, Aickelin U, Zhu J. Information-theoretic analysis for transfer learning. In: 2020 IEEE International Symposium on Information Theory (ISIT); 2020 Jun 21-26. IEEE; 2020. p. 2819-24.
<https://doi.org/10.1109/ISIT44484.2020.9173989>.

Jose ST, Simeone O. Information-theoretic bounds on transfer generalization gap based on Jensen-Shannon divergence. In: 2021 29th European Signal Processing Conference (EUSIPCO); 2021 Aug 23-27; Dublin, Ireland. IEEE; 2021. p. 1461-5.
<https://doi.org/10.23919/EUSIPCO54536.2021.9616270>.

Esposito AR, Gastpar M, Issa I. Generalization error bounds via Rényi-, f-divergences and maximal leakage. *IEEE Trans Inf Theory*. 2021;67(8):4986-5004.
<https://doi.org/10.1109/TIT.2021.3085190>.

Zhou R, Tian C, Liu T. Individually conditional individual mutual information bound on generalization error. *IEEE Trans Inf Theory*. 2022;68(5):3304-16.

<https://doi.org/10.1109/TIT.2022.3144615>.

Chen Q, Marchand M. Algorithm-dependent bounds for representation learning of multi-source domain adaptation. In: Proceedings of the 26th International Conference on Artificial Intelligence and Statistics; 2023 Apr 25-27; Valencia, Spain. PMLR; 2023. p. 10368-94.

Rojas-Carulla M, Schölkopf B, Turner R, Peters J. Invariant models for causal transfer learning. *J Mach Learn Res.* 2018;19(36):1-34.

Park GY, Lee SW. Information-theoretic regularization for multi-source domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct; Virtual. IEEE; 2021. p. 9214-23.

<https://doi.org/10.1109/ICCV48922.2021.00908>.

Lv S. Generalization bounds for graph convolutional neural networks via Rademacher complexity [Preprint]. arXiv; 2021. arXiv:2102.10234

Ahn SS, Hu S, Dai Z, Damianou A, Lawrence N. Mutual information guided distillation for transfer learning. In: 32nd Conference on Neural Information Processing Systems; 2018 Dec 3-8; Montréal, Canada.