When Do Graph Foundation Models Transfer? A Data-Centric Theory

Jiajun Zhu[1], Ying Chen[2], Peihao Wang[1], Yixuan He[2], Pan Li[3], Aditya Akella[1], Zhangyang Wang[1]

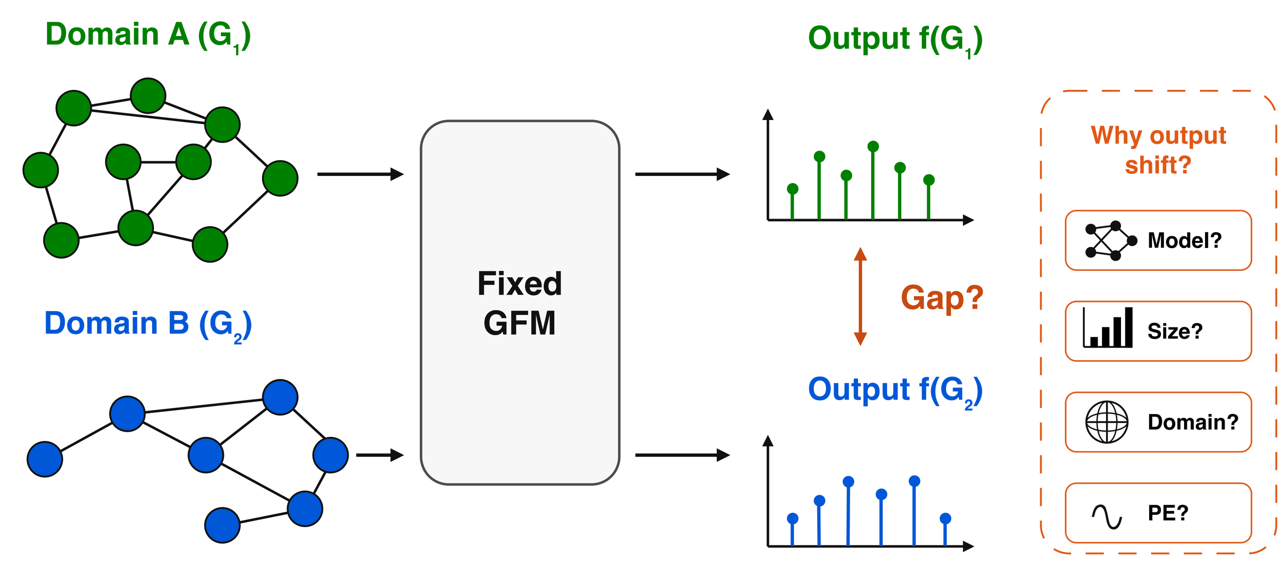
[1]The University of Texas at Austin, [2]Arizona State University, [3]Georgia Tech

ICML International Conference On Machine Learning

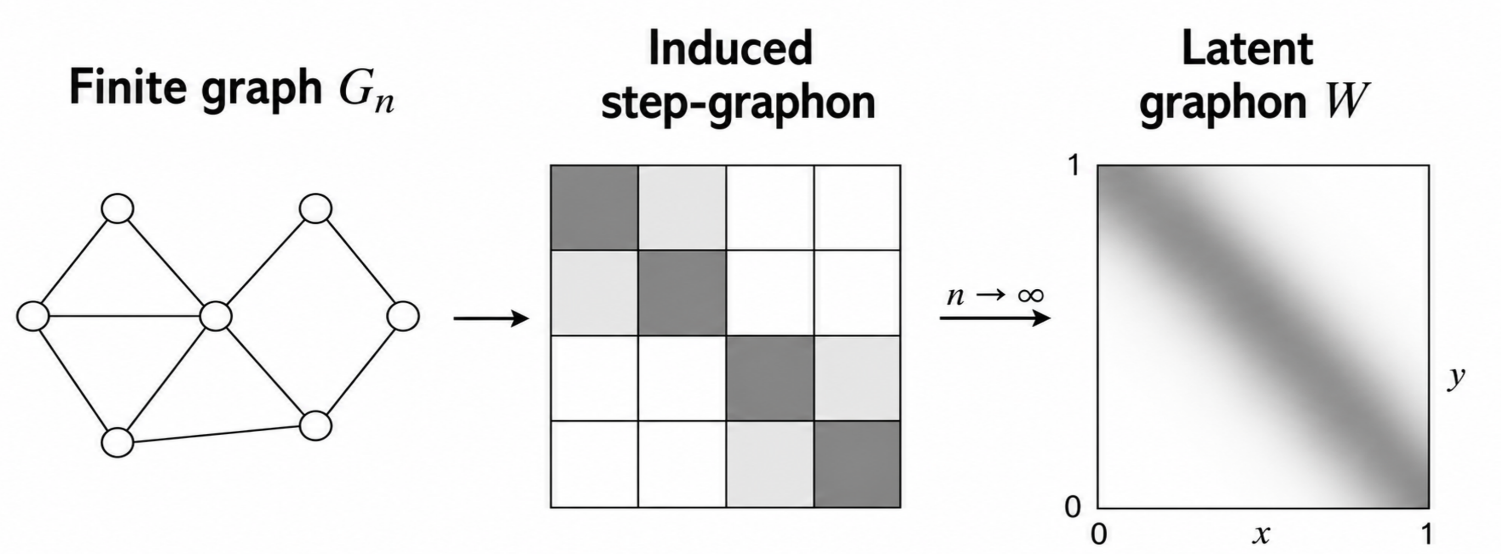
❶ Background

Problem. GFMs promise one backbone across graph domains, but transfer is often uneven and can become negative. Existing work mostly asks how to design or adapt models.

Question. Given a fixed GFM backbone, what data-domain properties determine transfer?



❷ Key Abstraction: Graphons Unify Different Sizes



- Different-sized graphs live in a common graphon space.
- Finite graphs are sample/step-graphon approximations of latent domains, with larger graphs reduce sampling error.
- Measure-preserving alignment handles arbitrary node relabeling.

❸ Main Result: Transfer Gap Decomposition

For a fixed Lipschitz GFM backbone f (set-based or message-passing) with spectral PE tokens, the cross-domain output gap satisfies

Transfer Gap

$$\|f(G_1) - f(G_2)\| \le L_f(1 + C_{\mathrm{PE}})\left(\varepsilon^{(1)}_{\mathrm{sample}} + \varepsilon_{\mathrm{graphon}} + \varepsilon^{(2)}_{\mathrm{sample}}\right).$$

L_f: model Lipschitz constant

C_{PE}: PE stability amplifier (depends on eigengaps & PE dimension)

Sampling / Size Error (Domain 1)
Finite-sample approximation from graph to graphon

$$\varepsilon^{(1)}_{\mathrm{sample}} = O_p\left(\sqrt{\frac{\log n_1}{n_1}}\right)$$

Latent Graphon Mismatch
Intrinsic, relabeling-invariant domain discrepancy

$$\varepsilon_{\mathrm{graphon}} = W_1(W_1, W_2)$$

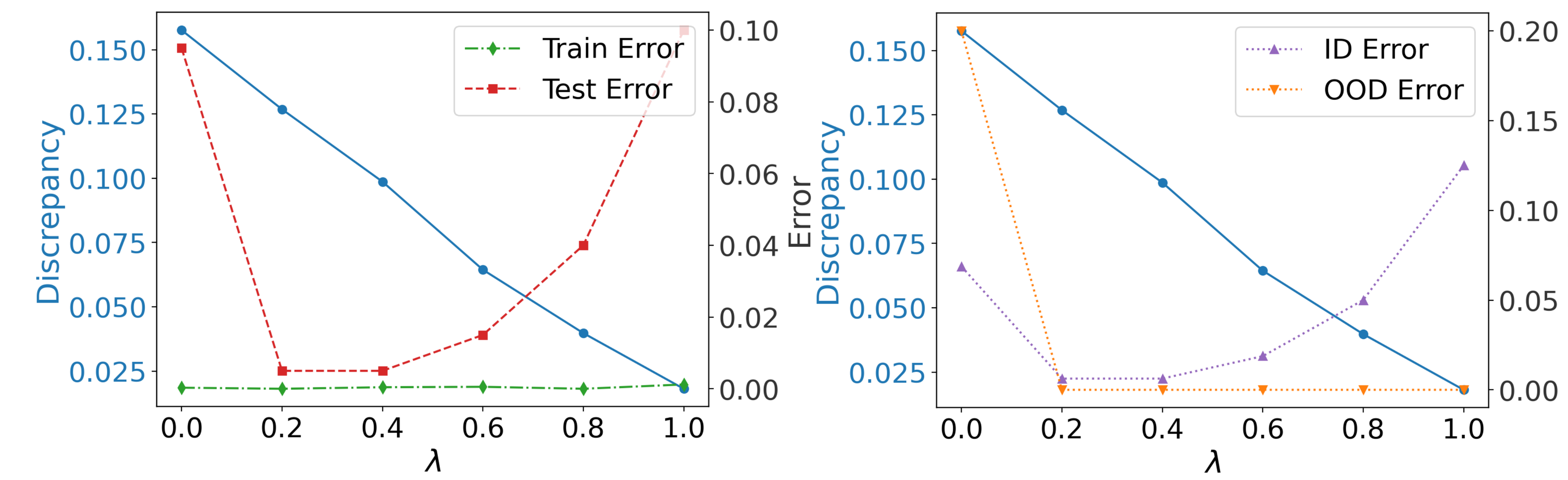
Sampling / Size Error (Domain 2)
Finite-sample approximation from graph to graphon

$$\varepsilon^{(2)}_{\mathrm{sample}} = O_p\left(\sqrt{\frac{\log n_2}{n_2}}\right)$$

Data-Centric Reading
- Larger graphs ↓ sampling error
- Similar graphons ↓ mismatch
- Stable PE prevents small structural shifts from becoming large token shifts

❹ Experiments

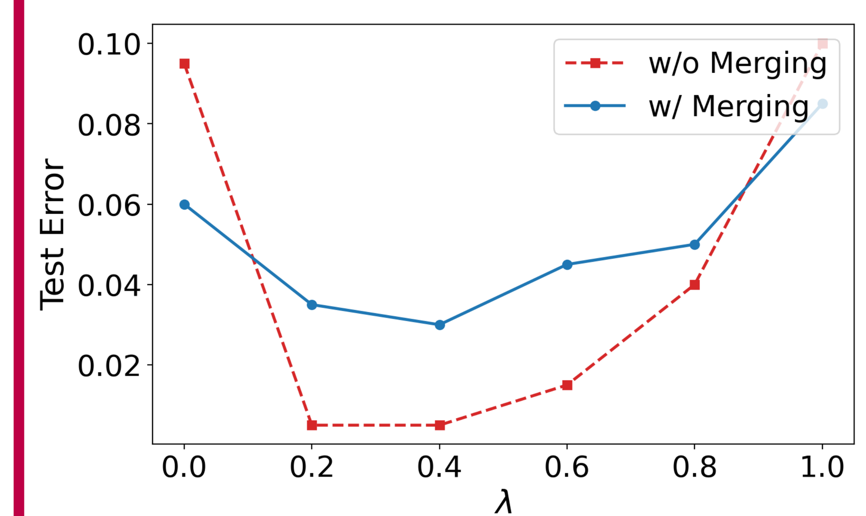
Result 1: Size shift is not monotonic



Note: λ = fraction of training budget allocated to larger graphs.

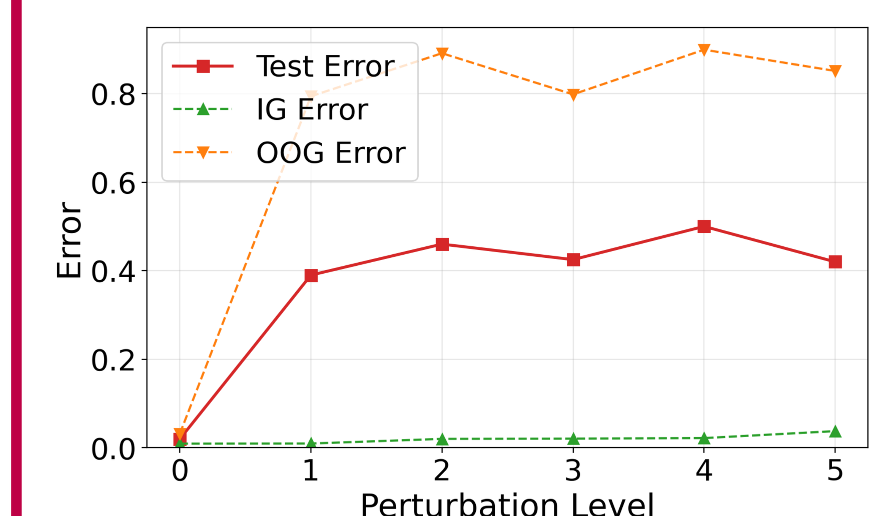
Main Finding: Larger graphs reduce token discrepancy, but test error becomes U-shaped because in-domain size coverage degrades.

Result 2: Graph merging helps when train-test size gap is large



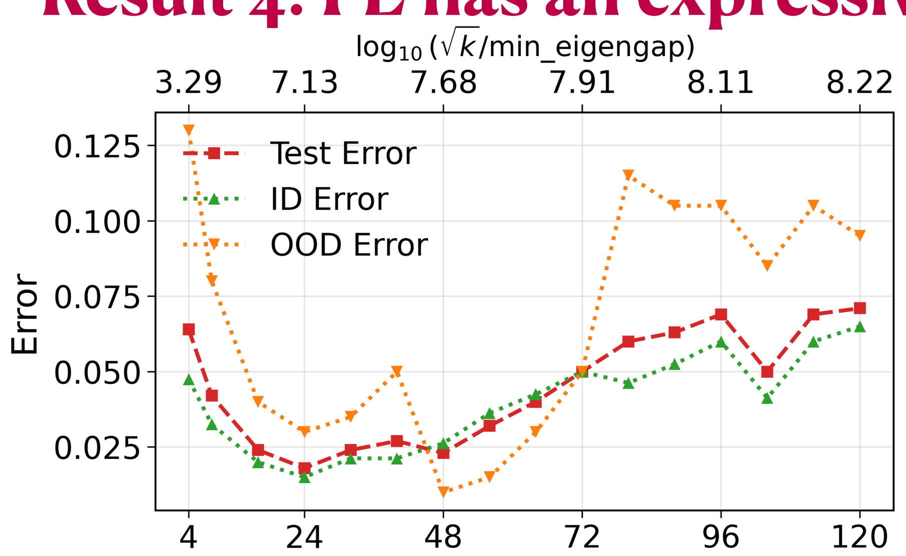
Main Finding: Graph merging helps most when the train–test size gap is large: adding a small fraction of larger synthetic graphs reduces test error at extreme λ. But when the baseline is already aligned, the gains are limited and may hurt the performance.

Result 3: Graphon shift dominates OOD failure



Main Finding: Perturbing the test graphons leaves in-graphon error nearly unchanged, while out-of-graphon error rises sharply and dominates the total error. This supports the latent graphon mismatch term as the main source of OOD failure.

Result 4: PE has an expressivity–stability trade-off



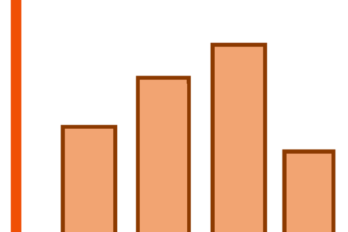
Main Finding: PE dimension (k) controls an expressivity–stability trade-off: small k underfits due to limited token expressivity, while large k becomes less stable as eigengaps shrink.

Takeaways: Data-Centric Guidances For GFM Transfer

Connect

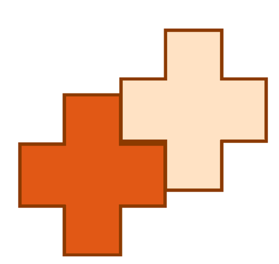
Paper

1. Balance graph sizes.



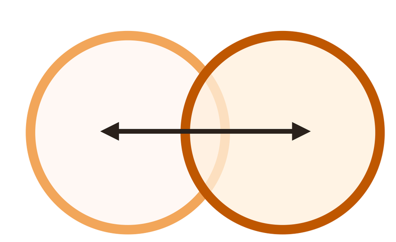
Do not only train on larger graphs; preserve size coverage.

2. Use graph merging selectively.



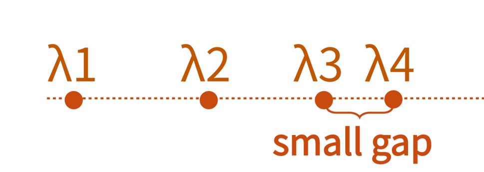
Merging helps at large size gaps, but may hurt when distributions are already aligned.

3. Measure graphon mismatch.



Transfer fails when latent graphons differ, even with enough graph size coverage.

4. Control PE instability.



Choose PE dimension & PE type with eigengap sensitivity in mind.