Distance-Preserving Graph Machine Learning

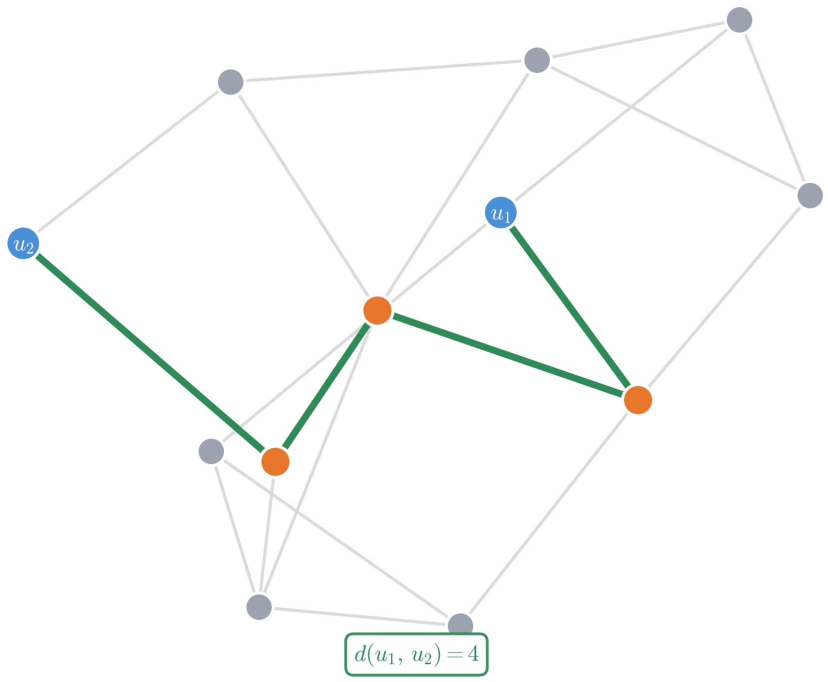
Luana Ruiz Johns Hopkins University

Joint work with My Le (JHU), Souvik Dhara (GATech) and Meher Chaitanya (KTH)

Graph Signal Processing Workshop · 2026

The structural topology of graphs across scales

- At each scale, network topology dictates how nodes sense and encode graph information
- **Local scale:** few-hop neighborhoods encode fundamental invariants and drive decentralized consensus
- **Global scale:** shortest-path distances influence retrieval, routing, and navigation
- **Mesoscopic scale:** multi-hop walks and effective resistance reveal motifs, communities and bottlenecks



Can graph SP/ML embed topology at all scales?

- **Local embeddings** ⇒ handled by graph SP and ML; most graph processing models are local (convolutions, aggregation, message-passing) ✓
- **Global embeddings** ⇒ handled in theory by graph transformers (GTs) but dense attention scales as $O(n^2)$, and storing all shortest distances requires n-dimensional node embeddings; no size generalization ✗
- **Mesoscopic embeddings** ⇒ feasible in the spectral domain given convolutions with large enough receptive field; but challenging to learn in practice, and low generalization over poorly behaved graph families ✗

Can graph SP/ML embed topology at all scales?

local — **mesoscopic** — **global**

- **Local embeddings** ⇒ handled by graph SP and ML; most graph processing models are local (convolutions, aggregation, message-passing) ✓

- **Global embeddings** ⇒ handled in theory by graph transformers (GTs) but dense attention scales as $O(n^2)$, and storing all shortest distances requires n-dimensional node embeddings; no size generalization ✗

- **Mesoscopic embeddings** ⇒ feasible in the spectral domain given convolutions with large enough receptive field; but challenging to learn in practice, and low generalization over poorly behaved graph families ✗

Le, My, Luana **Ruiz**, and Souvik Dhara. "Landmark-Based Node Representations for Shortest Path Distance Approximations in Random Graphs." *arXiv preprint arXiv:2504.08216* (2025).
Ruiz, Luana, Ningyuan Teresa Huang, and Soledad Villar. "A spectral analysis of graph neural networks on dense and sparse graphs." In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 9936-9940. IEEE, 2024.

Idea: preserve appropriate metric for each scale

PART I

Low-distortion shortest-path embeddings using local-global algorithms; local step implemented by a GNN

metric: shortest-path distance

PART II

Maximum adjacency search embeddings counting node co-ocurrences in multi-hop disjoint paths; graph rewiring with bounded effective resistance for GNNs & GTs

metric: effective resistance

PART I

Approximating Global Distances

Landmark embeddings on inhomogeneous random graphs

Approximating shortest-path distances

- Exact all-pairs distances are prohibitive at scale ⇒ $O(n(n+m))$ or worse
- Idea: **embed nodes** so that distances in embedding space **approximate graph distance** ⇒ quality is measured by multiplicative distortion

$$(1-\varepsilon)d(u,v) \le \hat{d}(u,v) \le (1+\varepsilon)d(u,v)$$

$(1 \pm \varepsilon)$*-distortion*

Landmark-based embeddings

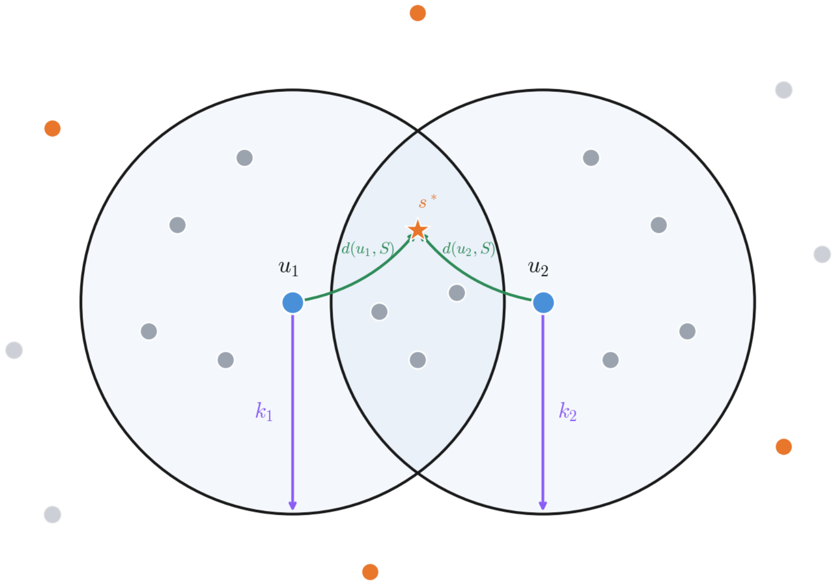
1 **Local step (offline)**

Sample landmark sets $S_0, \ldots S_r$; store each node's distance to its nearest landmark per set:

$$[x_u]_j = \min_{s \in S_j} d(u, s)$$

2 **Global step (online, query-based)**

Bound $d(u, v)$ from these coordinates via the triangle inequality



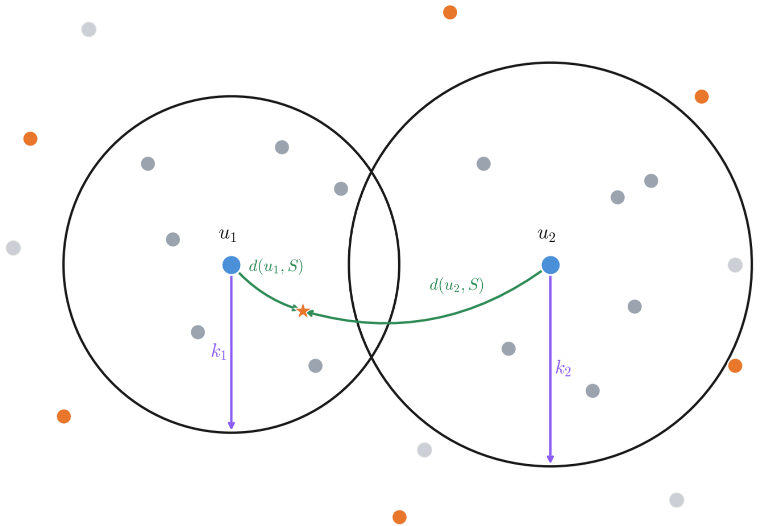
Landmark-based embeddings: lower bound

1

Local step

Sample landmark sets $S_0, \dots S_r$; store each node's distance to its nearest landmark per set:

$$[x_u]_j = \min_{s \in S_j} d(u,s)$$



2

Global step: *Bourgain's algorithm*

$$\underline{d}(u,v) = \| x_u - x_v \|_\infty = \max_j |[x_u]_j - [x_v]_j| \Rightarrow \underline{d}(u,v) \leq d(u,v)$$

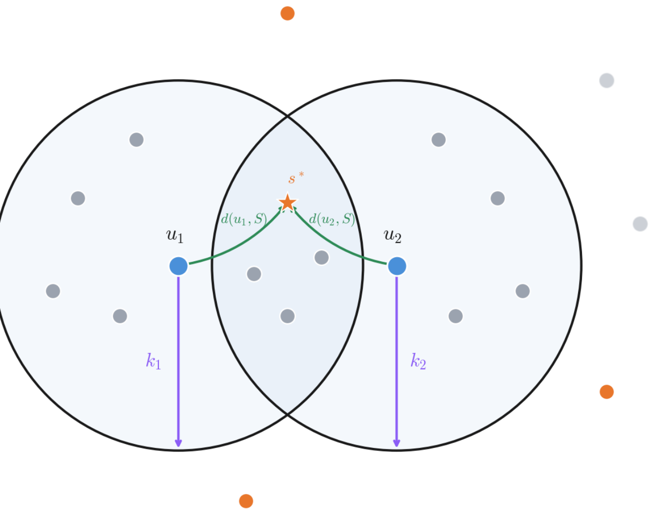
Bourgain, Jean. "On Lipschitz embedding of finite metric spaces in Hilbert space." *Israel Journal of Mathematics* 52, no. 1 (1985): 46-52.

Landmark-based embeddings: upper bound

1 **Local step:** ***Das Sarma et al's algorithm***

Sample landmark sets $S_0, \ldots S_r$; store each node's distance to its nearest landmark per set **and** the node in S_j that achieves it

$$[x_u]_j = \min_{s \in S_j} d(u,s) \quad [s_u]_j = \mathrm{argmin}_{s \in S_j} d(u,s)$$



2 **Global step:** ***Das Sarma et al's algorithm***

$$\overline{d}(u,v) = \min_{[s_u]_j = [s_v]_j} [x_u]_j + [x_v]_j \Rightarrow d(u,v) \leq \overline{d}(u,v)$$

Das Sarma, Atish, Sreenivas Gollapudi, Marc Najork, and Rina Panigrahy. "A sketch-based distance oracle for web-scale graphs." In *Proceedings of the third ACM international conference on Web search and data mining*, pp. 401-410. 2010.

Distortion guarantees: lower bound

- **Lower bound distortion** ⇒ Bourgain's and Matoušek's seminal results on the **distortion realized when embedding metric spaces into Hilbert spaces**

Bourgain's embedding theorem

Let $G = (V, E)$ be a graph with $n \geq 3$nodes and let $u_1, u_2 \in V$. For any $c > 1$, there exist embeddings $x^\star_{u_1}, x^\star_{u_2} \in \mathbb{R}^D$ with $D = \Omega(n^{1/c} \log n)$ such that the lower-bound estimator $\underline{d}(u_1, u_2)$ satisfies

$$\frac{1}{2c-1} d(u_1, u_2) \leq \underline{d}(u_1, u_2) \leq d(u_1, u_2).$$

Bourgain, Jean. "On Lipschitz embedding of finite metric spaces in Hilbert space." *Israel Journal of Mathematics* 52, no. 1 (1985): 46-52.
Matoušek, Jiří. "On the distortion required for embedding finite metric spaces into normed spaces." *Israel Journal of Mathematics* 93, no. 1 (1996): 333-344.

Distortion guarantees: upper bound

- **Upper bound distortion** ⇒ Das Sarma et al. provide guarantees whenever $|S_j| = 2^j$ for $0 \le j \le \lfloor \log n \rfloor$

Theorem (Das Sarma et al.)

Let $G = (V, E)$ be a graph with $n \geq 3$ nodes and let $u_1, u_2 \in V$. For any $c > 1$, there exist embeddings $x^\star_{u_1}, x^\star_{u_2} \in \mathbb{R}^D$ with $D = \Omega(n^{1/c} \log n)$ such that the upper-bound estimator $\overline{d}(u_1, u_2)$ satisfies

$$d(u_1, u_2) \le \overline{d}(u_1, u_2) \le (2c - 1) d(u_1, u_2).$$

\end{equation*}

Das Sarma, Atish, Sreenivas Gollapudi, Marc Najork, and Rina Panigrahy. "A sketch-based distance oracle for web-scale graphs." In *Proceedings of the third ACM international conference on Web search and data mining*, pp. 401-410. 2010.

Existing distortion results are worst case in nature

- Classical guarantees need **embedding dimension polynomial in n ⇒ tight** in the **worst case**

$(1-\varepsilon)$-distortion (lower bound)

$$\Omega\left(n^{\frac{2(1-\varepsilon)}{2-\varepsilon}}\log n\right)$$

$(1+\varepsilon)$-distortion (upper bound)

$$\Omega\left(n^{\frac{2}{2+\varepsilon}}\log n\right)$$

- Prohibitive at scale, and pessimistic for the structured graphs we actually have

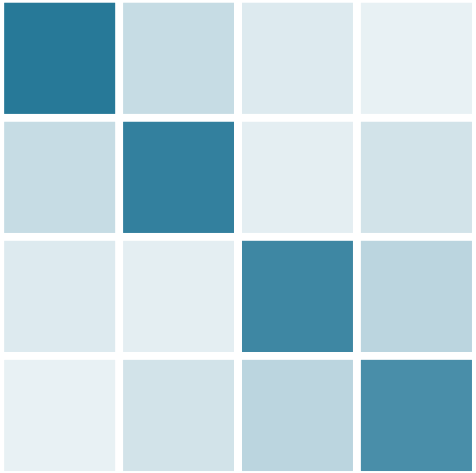
Inhomogenous random graphs: modeling the "average" graph

- Nodes carry types; **edge probability depends on endpoint types**
- The **affinity matrix** D captures community structure and degree heterogeneity
- Includes the stochastic block model; Erdős–Rényi when there is only one type

$$P_{c_1c_2} = \frac{D_{c_1c_2}}{n_{c_2}}$$

- Idea: **sharp distortion guarantees for typical graphs**, not worst case

$\boldsymbol{D}$



affinity between node types

Bollobás, Béla, Svante Janson, and Oliver Riordan. "The phase transition in inhomogeneous random graphs." *Random Structures & Algorithms* 31, no. 1 (2007): 3-122.

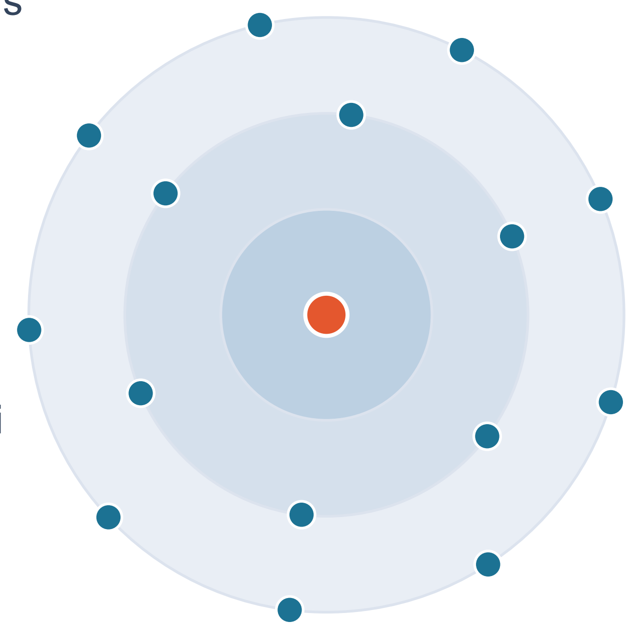
Proof idea: how neighborhoods grow

- Embeddings are computed via breadth-first search from seeds
- Local exploration behaves like a **multi-type branching process with mean matrix D**
- Neighborhoods expand exponentially at **rate λ_1, the Perron eigenvalue of D**

$$|\partial N_k(u)| = \Theta(\lambda_1^k)$$

- **Typical distances concentrate** ⇒ concentration of Bernoulli random variables

$$\frac{d(u,v)}{\log n} \to \frac{1}{\log \lambda_1}$$



λ_1 *per layer*

van der Hofstad, R. (2024). *Random Graphs and Complex Networks: Volume 2*. Cambridge University Press.
Janson, Svante, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*. John Wiley & Sons, 2011.

Distortion guarantees for the “average” graph

Theorem *(Le–Ruiz–Dhara)*

For $G \sim IHG(\vec{n}, D)$, under mild regularity, landmark embeddings attain $(1 \pm \varepsilon)$-distortion w.h.p. with embedding dimension:

$(1 - \varepsilon)$: $\Omega(n^{1-\varepsilon} \log_{\lambda_1} n)$

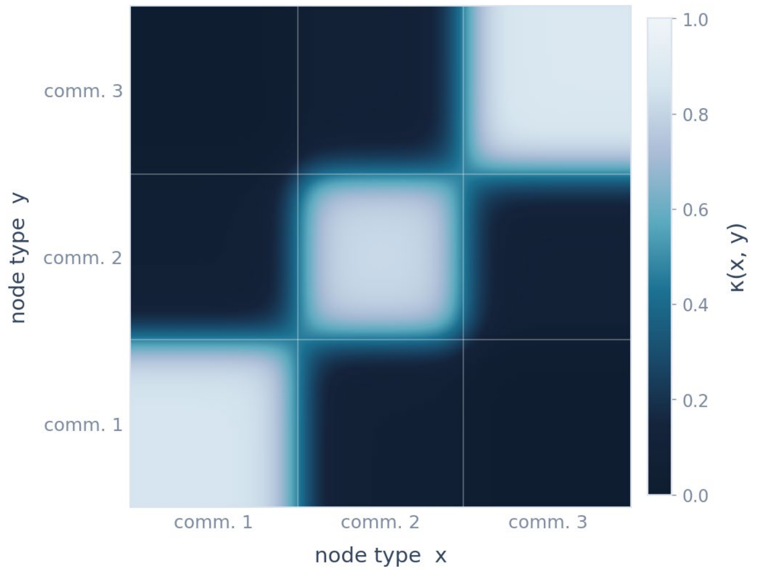
$(1 + \varepsilon)$: $\Omega\left(n^{2-2min\{\eta_n, \frac{1+\varepsilon}{2}\}} \log_{\lambda_1} n\right)$

- Governed by the **spectral radius** λ_1 and the **minimum type size exponent** η_n
- More balanced types (larger η_n) give strictly smaller embeddings
- **Polynomially smaller** than worst-case by approximately $n^{\varepsilon/2}$

Le, My, Luana **Ruiz**, and Souvik Dhara. "Landmark-Based Node Representations for Shortest Path Distance Approximations in Random Graphs." NeurIPS 2025 Workshop “New Frontiers in Graph Machine Learning”. *arXiv preprint arXiv:2504.08216* (2025).

Extending to continuous types

- Nodes may have a **continuous type space** ⇒ replace the finite set of types with $[0,1]$ (e.g., graphon-based graphs, Chung-Lu)
- Affinity becomes a **kernel** $\kappa(x,y)$ **akin to a graphon**
- Distortion bounds transfer to the kernel setting
- **Proof idea:** sandwich step-function kernels around κ, apply IHG bounds at each resolution, pass to the limit via spectral stability of T_κ

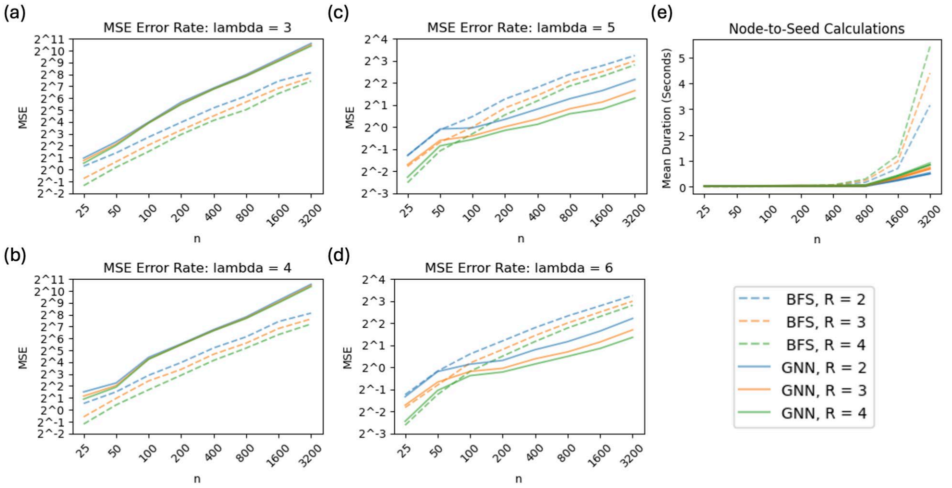


continuous kernel over type space

Bollobás, Béla, Svante Janson, and Oliver Riordan. "The phase transition in inhomogeneous random graphs." *Random Structures & Algorithms* 31, no. 1 (2007): 3-122.

Transferability via graph neural networks

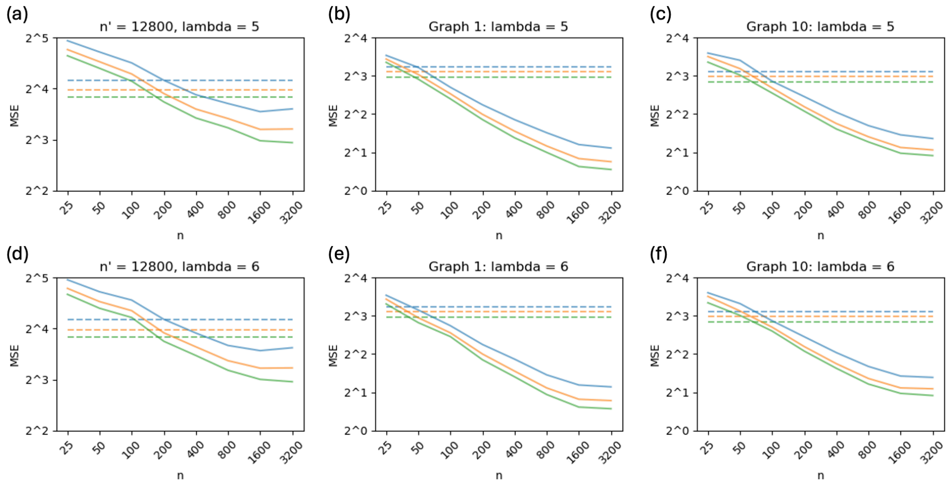
- **Replace BFS in local step with GNN** trained as shortest-path surrogate
- GNNs align with dynamic programming, naturally suited to (small) shortest-path computation
- Train on a synthetic IHG with $n = 100$ nodes
- **Transfer to graphs up to** $32\times$**larger**



Xu, Keylu, Jingling Li, Mozhi Zhang, Simon S. Du, Ken-ichi Kawarabayashi, and Stefanie Jegelka. "What Can Neural Networks Reason About?." *ICLR 2020* (2020).
Ruiz, Luana, Luiz FO Chamon, and Alejandro Ribeiro. "Transferability properties of graph neural networks." *IEEE Transactions on Signal Processing* 71 (2023): 3474-3489.

Validation on real networks

- **Real networks:** Arxiv COND-MAT collaboration network with 21,364 nodes, and GEMSEC company network with 14,113 nodes
- **GNN-based lower bounds outperform BFS-based** bounds across all networks
- MSE improves steadily with training graph size
- **Real graphs behave enough like IHGs for the theory to hold**



PART II

Exploiting Mesoscopic Structure

Cascade rewiring for structure-aware graph machine learning

GNNs and GTs struggle at mesoscopic scale

Graph neural networks

Convolutional/message-passing GNNs aggregate over K**-hop neighborhoods (small** K). The "receptive field" is local.

Graph Transformers

Graph Transformers with dense attention can capture structures beyond local, but at the cost of **lower "graph inductive bias" and $O(n^2)$ attention.**

- Multi-hop walks reveal mesoscopic structure such as **communities and bottlenecks**
- Idea: rewire the graph by promoting pairs that are **repeatedly co-activated in short multi-hop paths**

Du, Jian, John Shi, Soummya Kar, and José MF Moura. "On graph convolution for graph CNNs." In *2018 IEEE Data Science Workshop (DSW)*, pp. 1-5. IEEE, 2018.

Gama, Fernando, Antonio G. Marques, Geert Leus, and Alejandro Ribeiro. "Convolutional neural network architectures for signals supported on graphs." *IEEE Transactions on Signal Processing* 67, no. 4 (2018): 1034-1049.

Yun, Seongjun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim. "Graph transformer networks." *Advances in neural information processing systems* 32 (2019).

Contagion dynamics on graphs

- Two cascade rules for discovering mesoscopic structure: MAS and TAS
- For each seed node v, initialize an active set $S_0 = \{v\} \cup R$ where $R \subset N(v)$ is a random subset of neighbors, then propagate:

Maximum Adjacency Search (MAS)

Activate the inactive node with the most active neighbors.

The activation threshold drops as the cascade grows. Explores dense cohesive regions first but eventually crosses weak cuts.

Threshold Adjacency Search (TAS)

Activate any inactive node with $\geq \kappa$ active neighbors.

A fixed threshold κ filters weakly supported activations. Nodes need multi-neighbor reinforcement to join the cascade.

- Both rules privilege multi-hop pairs supported by many short, redundant paths (mesoscopic structure)

Chaitanya, Meher, Kshitijaa Jaglan, and Ulrik Brandes. "Adjacency Search Embeddings." *Transactions on Machine Learning Research* (2025).

Cascade-based graph rewiring

1 Ego-network cascades

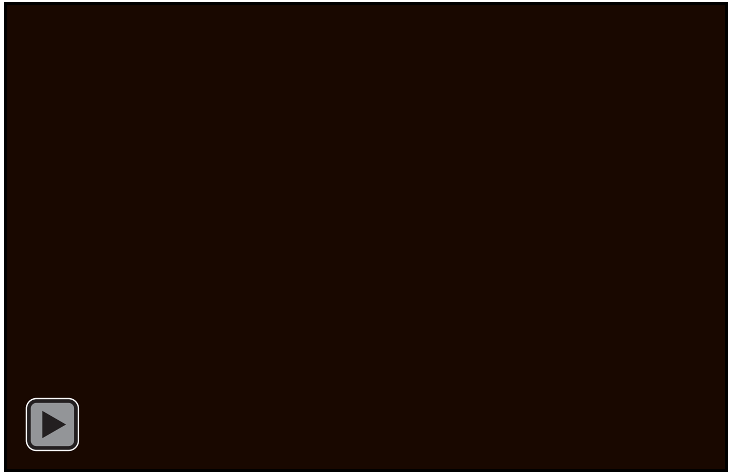
For each node v, run MAS or TAS cascades from v's ego-network. Record which nodes co-activate with v.

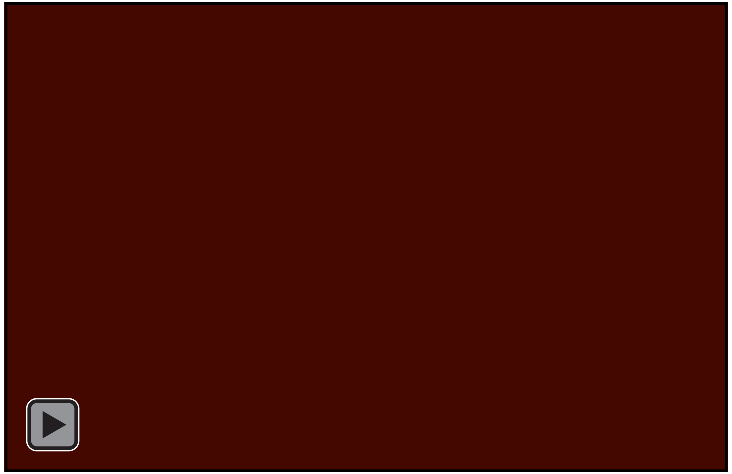
2 Top-k selection

Retain the k most frequently co-activated nodes per seed. Symmetrize and weight by co-activation frequency.

3 Auxiliary graph $G^\star$

$G^\star$ is sparse, weighted, $O(|V| + |E|)$ to build. Input to GNN or GT backbone that accepts weighted graphs.





Chaitanya, Meher, Kshitijaa Jaglan, and Ulrik Brandes. "Adjacency Search Embeddings." *Transactions on Machine Learning Research* (2025).

When does rewiring help?

For a uniformly random unordered pair of distinct nodes (u, v), define three events:

$\mathcal{L}$

label agreement

$y_u = y_v$ (same label)

$\mathcal{E}$

direct edge

$A_{uv} = 1$ (adjacent in G)

$\mathcal{R}$

reinforcement

$(A^2 + \cdots + A^L)_{uv} \geq \kappa$

The **standard edge homophily** and the **reinforcement homophily** are:

$$h_G = \mathbb{P}(\mathcal{L}|\mathcal{E}) \text{ and } h_{G^\star} = \mathbb{P}(\mathcal{L}|\mathcal{R})$$

We ask: when is $h_{G^\star} \geq h_G$**?**

Chaitanya, Meher, My Le, and Luana **Ruiz**. "Graph Cascades: Contagion-Based Mesoscopic Rewiring for Structure-Aware Graph Machine Learning." *arXiv preprint arXiv:2606.05046* (2026).

The Bayes heterophily condition

- Recall the reinforcement event $\mathcal{R} = (A^2 + \cdots + A^L)_{uv} \geq \kappa$

Proposition ***(Le–Chaitanya-Ruiz)***

Under two conditions: **"Heterophily" in G:** $\mathbb{P}(\mathcal{R}|\mathcal{L}) \geq \mathbb{P}(\mathcal{E}|\mathcal{L})$ **Reinforcement less frequent than edges:** $\mathbb{P}(\mathcal{R}) \leq \mathbb{P}(\mathcal{E})$

the reinforcement homophily satisfies $h_{G^\star} \geq h_G.$ *(strict when either condition is strict and $h_G > 0$)*

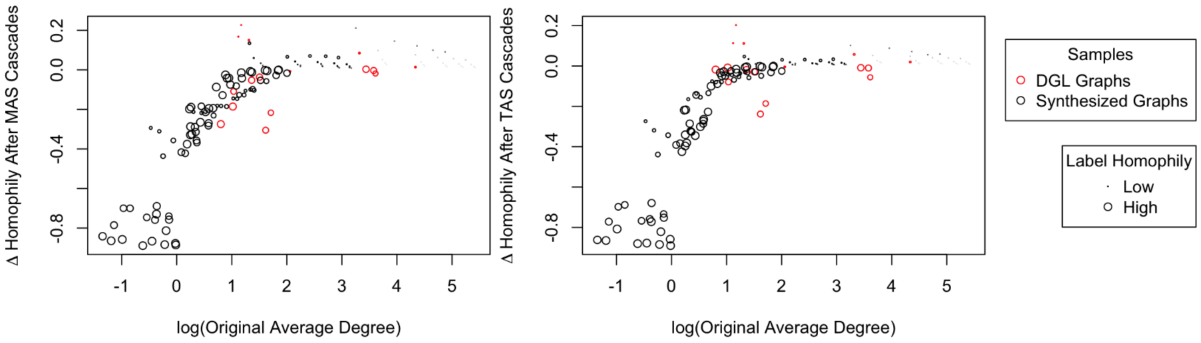


Figure 2: Effects of mesoscopic cascades on label homophily.

Walk length and effective resistance as the mesoscopic metric

- Threshold activations are backed by structural redundancy
- If a node u activates within ℓ steps under a threshold-κ cascade, then G contains κ **pairwise edge-disjoint short paths from the seed to u**

  ⇒ This directly bounds the **effective resistance**
- Place a unit current source at u and a unit sink at v

  ⇒ The effective resistance is the resulting voltage

difference $R_{eff}(u,v) = (e_u - e_v)^\top L^\dagger (e_u - e_v)$

Theorem ***(Le–Chaitanya–Ruiz)***

$$R_{eff}^{G_v}(S_0^v,u) \leq \frac{T(u)}{\kappa} \leq \frac{\ell}{\kappa}$$

where $T(u) \leq \ell$ is the activation time.

Chaitanya, Meher, My Le, and Luana **Ruiz**. "Graph Cascades: Contagion-Based Mesoscopic Rewiring for Structure-Aware Graph Machine Learning." *arXiv preprint arXiv:2606.05046* (2026).

Performance across DGL benchmarks

- Results rank within the top 5 on all 15 benchmarks tested

Test accuracy (mean, %); top-5 shaded by column

Dataset	actor	cham.	citeseer	comm.	comp.	cornell	cycle	grid	photo	pubmed	shape	sqrrl	texas	wiki	wisc.
Our accuracy (%)	36.5	63.6	75.5	78.6	90.9	75.0	61.3	62.3	94.8	88.7	89.6	47.3	82.3	83.5	81.0
Rank (of 16 methods)	2	3	4	1	2	1	5	5	5	1	1	1	1	5	2

Alignment with the theory

Regime	Heterophily	Multi-hop rarer than edges	Example datasets	Observed behavior	Δ Acc. (%)
Heterophilic	✓ holds	✓ holds	texas, cornell, wisconsin, chameleon, squirrel, actor	Largest gains	**+9.9 to +32.7**
Low-degree homophilic	× fails	✓ holds	citeseer, cora	Small or unstable	**≈ 0**
Bottlenecked / regular	× fails	× fails	cycle, grid, Roman-empire	Failure	**≤ 0**

Model-agnostic performance gains

GCN

10 / 16

graphs improved

up to +20.2%

GraphGPS

9 / 13

graphs improved

up to +9.5%

NAG

13 / 16

graphs improved

up to +34.5%

VCR

13 / 16

graphs improved

up to +32.7%

Heterophilic datasets:

actor	chameleon	cornell	squirrel	texas	wisconsin
+9.9%	+32.7%	+30.1%	+23.6%	+27.6%	+29.7%

Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*.

Rampášek, Ladislav, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. "Recipe for a general, powerful, scalable graph transformer." *Advances in Neural Information Processing Systems* 35 (2022): 14501-14515.

Chen, Jinsong, Kaiyuan Gao, Gaichao Li, and Kun He. "NAGphormer: A Tokenized Graph Transformer for Node Classification in Large Graphs." In *The Eleventh International Conference on Learning Representations*.

Fu, Dongqi, Zhigang Hua, Yan Xie, Jin Fang, Si Zhang, Kaan Sancak, Hao Wu, Andrey Malevich, Jingrui He, and Bo Long. "VCR-GRAPHORMER: A MINI-BATCH GRAPH TRANSFORMER VIA VIRTUAL CONNECTIONS." In *12th International Conference on Learning Representations, ICLR 2024*. 2024.

Structure explains performance

- Regress test accuracy on l**abel homophily, log average degree, and connectivity**
- Linear regression across 125 synthetic SBMs and 16 benchmarks

  ⇒ Random node features on synthetic graphs
- After rewiring, **structural properties of $G^\star$ explain substantially more variation in performance** than those of the original graph

M	β_0	β_1	β_2	Adjusted R^2
NAG	0.193***	0.197***	0.170**	0.454
NAG-MAS	0.143***	0.258***	0.173***	0.722
NAG-TAS	0.152***	0.265***	0.237***	0.695
VCR	0.253***	0.233***	-0.005	0.426
VCR-MAS	0.155***	0.306***	0.157***	0.704
VCR-TAS	0.172***	0.301***	0.201***	0.607
CR-MAS	0.091***	0.238***	0.261***	0.760
CR-TAS	0.132***	0.219***	0.421***	0.639

$^{***}p < 0.001$; $^{**}p < 0.01$; none: $p \geq 0.1$

Table 2: Regression coefficients for $A(M) \sim H * L(D) + H * L(D) * C$. Here H is label homophily, $L(D)$ the log average degree, and $C \in \{0, 1\}$ indicates whether the original graph is connected. $A(M)$ denotes test accuracy, computed on the original graph unless MAS/TAS specifies an auxiliary one.

Conclusions

Part I - Approximating Global Distances

- Landmark embeddings on IHGs achieve $(1 \pm \varepsilon)$-distortion of shortest path distances
- Embedding dimension governed by λ_1 and minimum type size $\Rightarrow$ polynomially smaller than worst case
- A GNN surrogate trained on small graphs transfers to graphs 32$\times$ larger and outperforms exact BFS on 16 real networks

Conclusions

Part II - Exploiting Mesoscopic Structure

- Graph Cascades builds $G^\star$ in $O(|V| + |E|)$ time by promoting pairs repeatedly co-activated in short multi-hop paths
- Under heterophily and high degree, $h_{G*} \geq h_G \Rightarrow$ Bayes condition that can predict gains before running the algorithm (up to +34.5% on heterophilic datasets)
- Threshold activation certifies κ edge-disjoint short paths to the seed, bounding the effective resistance of any selected edge $R_{eff}(u, v) \leq \ell/\kappa$

Conclusions

Convolutions/attention alone are insufficient at global or mesoscopic scales

Both scales are reached by making use of the appropriate metric

Inside the ML model: efficiently learn shortest-path embedding coordinates given a random graph prior $\Rightarrow$ followed by application of the appropriate (global) norm to compute pairwise distances

Outside the ML model: low effective resistance rewiring $\Rightarrow$ facilitate information propagation along multi-hop paths in GNNs and GTs

Future directions

- **Better landmark selection**: degree or other types of centrality, cascade-informed etc.
- Shortest-path embedding **distortion guarantees for directed graphs**
- **Signal-aware cascade selection** for graphs with reliable node features
- Combining Graph Cascades with bottleneck-targeted rewiring for graphs with both **mesoscopic structure and narrow cuts**

Thank you!

Luana Ruiz Johns Hopkins University

Joint work with My Le (JHU), Souvik Dhara (GATech) and Meher Chaitanya (KTH)

Part I: https://arxiv.org/abs/2504.08216

Part II: https://arxiv.org/pdf/2606.05046