Landmark-Based Node Representations for Shortest Path Distance Approximations in Random Graphs

My Le
Dept. of Applied Mathematics & Statistics
Johns Hopkins University
Baltimore, MD 21211
mle19@jh.edu

Luana Ruiz
Dept. of Applied Mathematics & Statistics
Johns Hopkins University
Baltimore, MD 21211
lrubini1@jh.edu

Souvik Dhara
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
sdhara@gatech.edu

Abstract

Learning node representations is a fundamental problem in graph machine learning. While existing embedding methods effectively preserve local similarity measures, they often fail to capture global functions like graph distances. Inspired by Bourgain's seminal work on Hilbert space embeddings of metric spaces [1985], we study the performance of local distance-preserving node embeddings. Known as landmark-based algorithms, these embeddings approximate pairwise distances by computing shortest paths from a small subset of reference nodes called landmarks. Our main theoretical contribution shows that random graphs, such as Erdős–Rényi random graphs, require lower dimensions in landmark-based embeddings compared to worst-case graphs. Empirically, we demonstrate that the GNN-based approximations for the distances to landmarks generalize well to larger real-world networks, offering a scalable and transferable alternative for graph representation learning.

1 Introduction

1.1 Motivations

Learning representations for network data has long been central to graph machine learning with a key objective to learn low-dimensional *node embeddings* that map structurally similar nodes to nearby points. These embeddings facilitate the application of machine learning methods to graph data, enabling a wide range of downstream tasks such as node classification, link prediction, and community detection [Hamilton et al., 2017b, Grover and Leskovec, 2016].

Traditional methods such as DeepWalk [Perozzi et al., 2014] and Node2Vec [Grover and Leskovec, 2016] use random walks to preserve local graph structures like node neighborhoods, while extensions such as GraRep [Cao et al., 2015] and PRONE [Zhang et al., 2021] capture higher-order relationships via k-hop transition factorizations. Spectral methods like Laplacian Eigenmaps [Belkin and Niyogi, 2003] preserve global geometry by embedding graphs into low-dimensional spaces that approximate their underlying manifold. Cauchy embeddings [Tang et al., 2019] further improve spectral methods by increasing their sensitivity to edge weight differences. While effective at capturing local graph

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: New Perspectives in Advancing Graph Machine Learning.

structure, these methods often fail to preserve global topology and functionals such as shortest path distances, especially in large, complex graphs [Goyal and Ferrara, 2018, Tsitsulin et al., 2018, Brunner, 2021].

In this work, we focus on the problem of learning node embeddings that preserve both local similarities and global graph distances. Motivated by Bourgain's seminal results on metric embeddings [1985], we analyze a landmark-based algorithm that approximates graph distances via shortest paths from a small set of landmarks. Our study analyzes its performance on random graphs—particularly Erdős–Rényi (ER) graphs—compared to worst-case instances. Of particular interest is the dimension of the embedding space.

1.2 Our Contributions

Theoretical Contributions. Our theoretical contribution is a detailed analysis of the dimensionality requirements for landmark-based embeddings on random graphs, in a more generalized setting than that analyzed for worst-case graphs. *This is the primary contribution of our work.*

We show that, with high probability (w.h.p.), random graphs require lower embedding dimensions: $\Omega\left(n^{\frac{1}{2c-1}+\varsigma}\theta \log n\right)$ with $\theta \in [\frac{c-1}{2c-1}, \frac{2(c-1)}{2c-1})$ for a $\frac{1}{2c-1}$-factor lower bound, and $\Omega(n^{3-2c+\varsigma})$ for a $(2c-1)$-factor upper bound for any $\varsigma > 0$, as compared to worst-case graphs with $\Omega(n^{1/c}\log n)$ for the same lower and upper bounds [Bourgain, 1985, Matoušek, 1996, Sarma et al., 2010], where $c > 1$. The proof leverages branching process approximations from the random graph literature [van der Hofstad, 2017, 2024].

Methodological Contributions. Building upon this theory, we propose a GNN-augmented variant that predicts landmark distances from graph structure. This reduces explicit shortest-path computations as GNNs can learn to approximate landmark distances in a supervised manner.

GNNs are well-suited for this task due to their alignment with dynamic programming which underpins shortest path algorithms [Xu et al., 2019b, Dudzik and Veličković, 2022]. Empirical results on ER graphs and real-world benchmarks show that GNN-based embeddings provide better global-distance lower bounds than exact landmark embeddings. Notably, GNNs trained on small ER graphs generalize effectively to larger ER graphs and real-world networks, highlighting the value of studying embedding methods in the context of random graphs.

1.3 Related Works

A rich body of theoretical works has focused on the minimum dimension k_ε required to embed worst-case graphs into $\mathbb{R}^k$ while preserving all pairwise distances up to a factor of $(1 \pm \varepsilon)$. In a seminal work, Bourgain [1985] showed $k_\varepsilon = \Omega((\log n)^2/(\log\log n)^2)$, providing a negative answer to Johnson and Lindenstrauss's Problem 3 [1984]. This was later strengthened to $k_\varepsilon = \Omega((\log n)^2)$ [Linial et al., 1995], and further to $k_\varepsilon = \Omega(n^{c/(1+\varepsilon)})$ for some universal $c > 0$ [Matoušek, 1996]. The latter was also proven recently by Naor [2016] and Naor [2021] using expanders, showing that low-distortion embeddings of graphs with strong expansion properties require polynomial dimensionality.

From an algorithmic perspective, finding embeddings with minimum distortion is NP-hard; see Sidiropoulos et al. [2019] for a survey of approximation algorithms and hardness results. Practical methods often rely on landmark-based algorithms [Goldberg and Harrelson, 2005, Sarma et al., 2010, Potamias et al., 2009, Tretyakov et al., 2011, Akiba et al., 2013, Rizi et al., 2018, Qi et al., 2020], which preselect a subset of landmark nodes and compute distances to them via local message passing (see Sommer [2014] for a review). The resulting landmark distances can be viewed as an embedding useful for approximating graph distances. Yet these methods inherit worst-case limitations of local message passing, often requiring prohibitively large dimensions for general graphs [Sarma et al., 2012, Loukas, 2020].

Notation. We let $G = (V, E)$ denote an undirected, unweighted graph, where V is the set of nodes and E is the set of edges, with $|V| = n$ and $|E| = m$, where $|X|$ denotes the cardinality of any discrete set X. We consider only one graph at a time and use $\mathcal{C}_{(i)}$ to denote the i-th largest connected component of the graph. We write $u_1 \leftrightarrow u_2$ to mean that there exists a path between u_1 and u_2 (i.e., u_1 and u_2 are in the same connected component). We often use the Bachmann–Landau asymptotic notation $o(1), O(1), \omega(1), \Omega(1), \Theta(1)$, etc. to describe the asymptotic behavior of functions. Given

a sequence of probability measures $(\mathbb{P}_n)_{n\geq 1}$, a sequence of events $(\mathcal{E}_n)_{n\geq 1}$ is said to hold *with high probability (w.h.p.)* if $\lim_{n\to\infty} \mathbb{P}_n(\mathcal{E}_n) = 1$. For a sequence of random variables $(X_n)_{n\geq 1}$, $X_n \xrightarrow{\mathbb{P}} c$ means that X_n converges to c in probability. We write statements such as $X_n = f(n)^{o(1)}$ w.h.p. to abbreviate that $\log X_n / \log f(n) \xrightarrow{\mathbb{P}} 0$. Also, we write $X_n = O(1)$ w.h.p. to mean that $\mathbb{P}(X_n \geq K) \to 0$ for a sufficiently large K.

2 The Shortest Path Problem and Landmark-Based Embeddings

Given a graph $G = (V, E)$ and nodes $u_1, u_2 \in V$, the shortest distance problem is to find the minimum number of edges connecting u_1 and u_2, i.e., the *shortest path distance* $d(u_1, u_2)$. The classical solution to this fundamental graph problem is Dijkstra's algorithm, with running time $O(n^2)$ for a single pair and $O(n^3)$ for all pairs using naive data structures, reducible to $O(m \log n)$ and $O(m + n \log n)$ with heaps and Fibonacci heaps [Schrijver, 2012]. More refined variants for single source include S-Dial ($O(m + nl_{\max})$, $l_{\max}$ is the maximum arc length), S-Heap ($O(m \log n)$ or $O(n \log n)$ in sparse graphs), and Fredman–Willard's implementation ($O(m + n \log n / \log \log n)$) [Gallo and Pallottino, 1988, Fredman and Willard, 1990]. For all pairs, Floyd–Warshall and primal sequential algorithms run in $O(n^3)$ [Gallo and Pallottino, 1988], while hidden-path achieves $O(mn + n^2 \log n)$ [Karger et al., 1993]. Despite these advances, exact computation remains costly on large graphs.

While computing exact shortest path distances is expensive, we can afford to compute local paths. Sarma et al.'s offline sketch algorithm [2010] leverages this principle in its local step to construct landmark embeddings (see Local Step in Algorithm 1). To mitigate the time and memory constraints associated with calculating shortest paths, lower and upper bounds have been used as reliable metrics for approximating shortest paths in many approaches [Bourgain, 1985, Matoušek, 1996, Sarma et al., 2010, Gubichev et al., 2010, Sommer, 2014, Akiba et al., 2014, Meng et al., 2015, Jiang et al., 2021, Awasthi et al., 2022]. As in the current setting, the resulting local embeddings can be stored in memory and later retrieved to quickly estimate $d(u_1, u_2)$ via the bounds $\underline{d}(u_1, u_2)$ and $\bar{d}(u_1, u_2)$ with a single lookup from u_1 and u_2 (see Global Step in Algorithm 1).

Algorithm 1: Landmark Algorithm Adapted From Sarma et al. [2010]

Input: Connected graph $G = (V, E)$ with $|V| = n$. Positive integer R. Set cardinalities $|S_0| = 1, |S_1|, |S_2|, ..., |S_r|$ for some positive integer r.
Output: Shortest path lower bound $\underline{d}(u, v)$ and upper bound $\bar{d}(u, v)$ for all $u, v \in V$.

for $i = 1, 2, \ldots, R$ **do** `/* LOCAL STEP */`
  for $j = 0, 1, \ldots, r$ **do**
    $S_j \leftarrow \{s_1, \ldots, s_{|S_j|} \sim \text{Uniform}(V)\}$
    for $u = 1, \ldots, n$ **do**
      $[\mathbf{x}_u]_j = \min_{s \in S_j} d(s, u)$
      $[\boldsymbol{\sigma}_u]_j = \text{argmin}_{s \in S_j} d(s, u)$
    end
  end
end
for $u = 1, \ldots, n$ **do** `/* GLOBAL STEP */`
  for $v = 1, \ldots, n$ **do**
    $\underline{d}(u, v) = \|\mathbf{x}_u - \mathbf{x}_v\|_\infty$ `/* Lower Bound */`
    $\bar{d}(u, v) = \min\{[\mathbf{x}_u]_i + [\mathbf{x}_v]_j : (i, j) \text{ s.t. } [\boldsymbol{\sigma}_u \mathbf{1}^T - \mathbf{1}\boldsymbol{\sigma}_v^T]_{ij} = 0\}$ `/* Upper Bound */`
  end
end

To show that $\underline{d}(u_1, u_2)$ is a lower bound on $d(u_1, u_2)$, without loss of generality, assume $\underline{d}(u_1, u_2) = d(u_1, S) - d(u_2, S)$ for some landmark set S with $d(u_2, S) = d(u_2, v_1)$ and $d(u_1, S) = d(u_1, v_2) \leq d(u_1, v_1)$. It then follows from triangle inequality that $\underline{d}(u_1, u_2) \leq d(u_1, v_1) - d(u_2, v_1) \leq d(u_1, u_2)$.

The proof for $d(u_1, u_2) \leq \bar{d}(u_1, u_2)$ also follows directly from the formulation of $\bar{d}(u_1, u_2)$ in Algorithm 1 and triangle inequality: $\bar{d}(u_1, u_2) = d(u_1, v) + d(u_2, v) \geq d(u_1, u_2)$ for some landmark node v. By sampling at least one landmark set of size 1, we ensure that u_1 and u_2 share a closest landmark node from such landmark sets, preventing $\bar{d}(u_1, u_2)$ from being undefined.

Since $\underline{d}(u_1, u_2)$ depends on the distance to the landmark sets but not on which landmark node is the closest, it is sufficient for the landmark embeddings to store only the closest distances from each node to the landmark sets. The trade-off for such memory reduction is that $\underline{d}(u_1, u_2)$ can be approximated only with $D = R \times (r + 1)$ dimensions, while $\bar{d}(u_1, u_2)$, which utilizes the common closest landmarks, has an approximation dimension that varies between 1 and $D \times D$, depending on how the landmark sets are sampled.

3 Lower and Upper Bound Distortions for Shortest Distance Approximations

The lower and upper bound metrics on the landmark embeddings, as described in Section 2, are only useful if we can derive guarantees on their approximation ability. For the lower bound, these have been proven by Matoušek [1996] based on Bourgain's classical embedding theorem [1985], which characterizes the distortion incurred by optimal embeddings of metric spaces into $\mathbb{R}^D$. For the upper bound, similar guarantees have been derived in Sarma et al. [2010].

Theorem 3.1 (**Lower Bound Distortion Adapted From Bourgain [1985] and Matoušek [1996]**). *Let G be a graph with $n \geq 3$ nodes and u_1, u_2 be two nodes in G. Let $c > 1$. There exist node embeddings $\mathbf{x}^*_{u_1}, \mathbf{x}^*_{u_2} \in \mathbb{R}^D$ with $D = \Omega(n^{1/c} \log n)$ for which $\underline{d}(u_1, u_2)$ as in Algorithm 1 satisfies*

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

Theorem 3.2 (**Upper Bound Distortion Adapted From Sarma et al. [2010]**). *Let G be a graph with $n \geq 3$ nodes and u_1, u_2 be two nodes in G. Let $c > 1$. There exist node embeddings $\mathbf{x}^*_{u_1}, \mathbf{x}^*_{u_2} \in \mathbb{R}^D$ with $D = \Omega(n^{1/c} \log n)$ for which $\bar{d}(u_1, u_2)$ as in Algorithm 1 satisfies*

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

In order for (1) and (2) to hold, the embeddings $\mathbf{x}^*_u$ need to be optimal. However, there is no guarantee that this can be achieved using the landmark embeddings. One way to ensure good embeddings is to control how the landmarks are sampled. Sarma et al. [2010] proposed sampling landmark sets S_i of sizes 2^i for $i = 0, 1, \ldots, r$.

For the lower bound, smaller landmark sets are beneficial since, for $\sigma_1 + \sigma_2 < 1$ with $0 < \sigma_1 < \sigma_2$, we must find at least one landmark set containing a landmark node in the $\sigma_1 d(u_1, u_2)$-hop neighborhood centered at u_1 and none in the $\sigma_2 d(u_1, u_2)$-hop neighborhood centered at u_2. For the upper bound, this strategy ensures that a landmark falls in the intersection of the $\lceil \frac{d(u_1, u_2)}{2} \rceil$-hop neighborhoods of nodes u_1 and u_2 w.h.p. Hence, having a range of landmark set cardinalities helps.

It can be shown that if $|S_i|$ is exponential in i, $R = \Omega(n^{1/c})$, and $r = \lfloor \log n \rfloor$—yielding a total embedding size of $\Theta(n^{1/c} \log n)$—then the resulting shortest path distance approximations satisfy Theorems 3.1 and 3.2 for all pairs of nodes w.h.p. for *any graph*. In Section 4, we show that both the distortions and the embedding dimensions can be improved for random graphs.

4 Lower and Upper Bound Distortions on Sparse Erdős–Rényi

In this section, we show our main results on the performance of $\underline{d}(u_1, u_2)$ and $\bar{d}(u_1, u_2)$ outputted by Algorithm 1 as shortest path approximations in sparse ER graphs, where each edge appears independently with a fixed probability. We write $G \sim \mathrm{ER}_n(\lambda/n)$ to denote this distribution over the space of all graphs on n nodes with probability λ/n, $\lambda \in [0, n]$. Based on a classical result in random graph theory [van der Hofstad, 2017, Theorems 4.4, 4.8 and Corollary 4.13], we consider $\lambda > 1$ since otherwise the giant component dies out in probability, making most pairs of nodes not connected.

4.1 Main Results on Distortions

On ER graphs, we derive the following distortion bound as a $(1 \pm \varepsilon)$-approximation of $d(u_1, u_2)$:

Theorem 4.1 (***Lower Bound Distortion on Random Graphs***). *Let $G \sim \mathrm{ER}_n(\lambda/n)$ with $\lambda > 1$. Let u_1, u_2 be chosen independently and uniformly at random with replacement from G. Fix $\varepsilon \in (0,1)$, an integer $M > 1$, $\theta \in (0, \varepsilon)$, and $r = \lfloor \frac{\theta}{\log M} \log n \rfloor$. With embedding dimension $D = \Omega\left(M n^{1-\frac{\varepsilon}{2}-\min\{\frac{\varepsilon}{2},\theta\}+\varsigma} \frac{\theta}{\log M} \log n\right)$ resulting from $R = \Omega\left(M n^{1-\frac{\varepsilon}{2}-\min\{\frac{\varepsilon}{2},\theta\}+\varsigma}\right)$ runs of the local step with set cardinalities $|S_0| = M^0, |S_1| = M^1, \ldots, |S_r| = M^r$ and any arbitrarily small $\varsigma > 0$, $\underline{d}(u_1, u_2)$ provides a $(1-\varepsilon)$-approximation of $d(u_1, u_2)$ (i.e. $d(u_1, u_2) \geq \underline{d}(u_1, u_2) \geq (1-\varepsilon) d(u_1, u_2)$) w.h.p.*

Theorem 4.2 (***Upper Bound Distortion on Random Graphs***). *Let G, λ, u_1, u_2 be as in Theorem 4.1. Fix $\varepsilon \in (0,1)$, an integer $M > 1$, $\theta \in \left(0, \frac{1-\varepsilon}{2}\right)$, and $r = \lfloor \frac{\theta}{\log M} \log n \rfloor$. With embedding dimension $D = \Omega\left(n^{1-\varepsilon+\varsigma}\right)$ resulting from $R = \Omega\left(\frac{\log M}{\theta \log n} n^{1-\varepsilon+\varsigma}\right)$ runs of the local step with set cardinalities $|S_0| = M^0, |S_1| = M^1, \ldots, |S_r| = M^r$ and any arbitrarily small $\varsigma > 0$, $\bar{d}(u_1, u_2)$ provides a $(1+\varepsilon)$-approximation of $d(u_1, u_2)$ (i.e. $d(u_1, u_2) \leq \bar{d}(u_1, u_2) \leq (1+\varepsilon) d(u_1, u_2)$) w.h.p.*

While Bourgain [1985], Matoušek [1996], and Sarma et al. [2010] showed that, in the worst case, Algorithm 1 with $M = 2$ requires embedding dimension $\Omega(n^{1/c} \log n)$ for a $\frac{1}{2c-1}$-factor lower bound and a $(2c-1)$-factor upper bound ($c > 1$), our results offer a more efficient alternative for ER graphs by loosening the dimensionality restrictions, specifically $\Omega\left(n^{\frac{1}{2c-1}+\varsigma} \theta \log n\right)$ with $\theta \in [\frac{\varepsilon}{2}, \varepsilon)$ for the same lower bound and $\Omega(n^{3-2c+\varsigma})$ for the same upper bound for any $\varsigma > 0$. Furthermore, our results pertain to a more general setting where M can be any integer greater than 1 and θ, which regulates the amount of sampling, can be ε-small for the lower bound distortion and $\left(\frac{1-\varepsilon}{2}\right)$-small for the upper bound distortion.

4.2 Idea of Proofs and Supporting Results

The proofs of Theorems 4.1 and 4.2 rely on local neighborhood expansions in ER graphs $G \sim \mathrm{ER}_n(\lambda/n)$, which can be accessed via the Poisson branching process with mean offspring λ. With $N_k(u)$ denoting the set of nodes with graph distance at most k from u and $\partial N_k(u)$ denoting the set of nodes at distance exactly k from u, the results on local neighborhood expansions are stated as follows:

Lemma 4.3. *Let G, λ, u_1, u_2 be as in Theorems 4.1 and 4.2. Let $\kappa_0 \in (0, \frac{1}{2})$, $L = \kappa_0 \log_\lambda n$, and $\varepsilon \in (0, \kappa_0)$. Let A_n denote the event that $n^{-\varepsilon}\lambda^L \leq |\partial N_L(u_1)|, |\partial N_L(u_2)| \leq n^{\varepsilon}\lambda^L$ and B_n denote the event that u_1 and u_2 are in the same connected component. Then $\mathbb{P}(A_n \setminus B_n) \to 0$ and $\mathbb{P}(B_n \setminus A_n) \to 0$ as $n \to \infty$.*

Proof. See Appendix A.1. □

Lemma 4.4. *Let $G, \lambda, u_1, u_2, k_0, L$ be as in Lemma 4.3. Let $\varepsilon \in (0, \kappa_0)$ and $\kappa \in (0, 1-\kappa_0)$. Let A_{b_m, b_M} be the event that $|\partial N_L(u_i)| \in [b_m, b_M]$ for $i \in \{1,2\}$ and $\mathcal{E}_{n,k}$ be the good event that $|\partial N_{L+k}(u_i)| \in [b_m \lambda^k, b_M \lambda^k]$ for $i \in \{1,2\}$, where $b_m = n^{-\varepsilon}\lambda^L$ and $b_M = n^{\varepsilon}\lambda^L$. Then, there exists $\delta > 0$ such that $\mathbb{P}(\cap_{l=0}^{k} \mathcal{E}_{n,l} \mid A_{b_m, b_M}) \geq 1 - 3kn^{-\delta}$ for any $k \leq \kappa \log_\lambda n$ for all sufficiently large n.*

Proof. See Appendix A.2 □

Proposition 4.5. *Let $G, \lambda, u_1, u_2, \kappa_0, \kappa, L$ be as in Lemma 4.4 and $\varepsilon > 0$. Conditionally on u_1, u_2 being in the same connected component, $|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \in \left[\frac{n^{-\varepsilon}\lambda^{k_1+k_2}}{2n}, \frac{n^{\varepsilon}\lambda^{k_1+k_2}}{n}\right]$ w.h.p. for any k_1, k_2 such that $L < k_1, k_2 \leq (\kappa_0 + \kappa) \log_\lambda n$ and $k_1 + k_2 > (1+\zeta) \log_\lambda n$ for any small $\zeta > 0$.*

Proof. See Appendix A.3 □

By Lemmas 4.3 and 4.4, $\partial N_k(u_1)$ grows as λ^k for any fixed u. By Proposition 4.5, $|\partial N_k(u_1) \cap \partial N_k(u_2)|$ grows as $\frac{\lambda^{2k}}{n}$ for any fixed u_1, u_2. These growth rates imply that, w.h.p.,

the local step selects a landmark set that intersects $N_{k_1}(u_1)$ but not the disjoint $N_{k_2}(u_2)$, with $k_2 - k_1 \geq (1-\varepsilon)d(u_1, u_2)$, yielding

$$\underline{d}(u_1, u_2) \;\geq\; (1-\varepsilon)\, d(u_1, u_2) \text{ w.h.p.}$$

For the upper bound distortion, we show that w.h.p. there is a landmark set intersecting $N_k(u_1) \cap N_k(u_2)$ but not $\big(N_k(u_1) \cup N_k(u_2)\big) \setminus \big(N_k(u_1) \cap N_k(u_2)\big)$, where $k \leq \frac{1+\varepsilon}{2} d(u_1, u_2)$. This ensures

$$\overline{d}(u_1, u_2) \;\leq\; (1+\varepsilon)\, d(u_1, u_2) \text{ w.h.p.}$$

The complete proof of Theorems 4.1 and 4.2 are provided in Appendices A.4 and A.5.

5 GNN-Based Landmark Embeddings and Experimental Results

Although Algorithm 1 outperforms traditional methods, its landmark distance calculations rely on one run of Breadth-First Search (BFS) for each landmark set, which is costly for large graphs ($O(n+m)$ per pair, $O(n(n+m))$ for all pairs [Cormen et al., 2009]). We propose replacing BFS with a GNN to approximate shortest-path distances between nodes and landmarks, which comes with three advantages: (i) embeddings are computed automatically once the GNN is trained, (ii) inference is cheaper than exact distance calculations, and (iii) the GNN's transferability [Ruiz et al., 2020, 2023] enables generalization to larger graphs from the same graphon model.

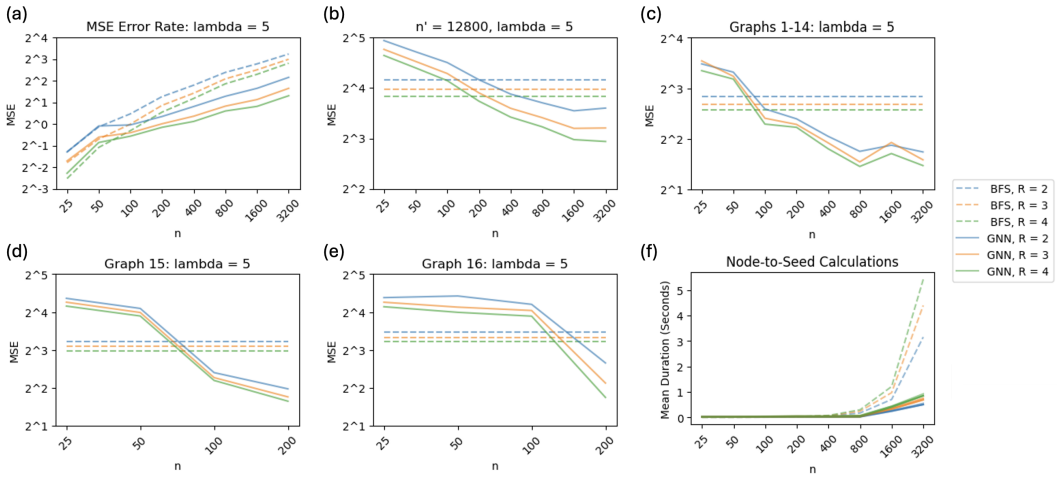

Figure 1: Error rates of BFS-based and GNN-based lower bounds on (a) test ER graphs generated from the same $\mathrm{ER}_n(\lambda/n)$ as the training graphs, (b) test ER graphs generated by $\mathrm{ER}_{n'}(\lambda/n')$ with larger graph size n', (c) real-world networks with 3,892 to 28,281 nodes, (d) Brightkite social network with 56,739 nodes, and (e) ER-AVGDEG10-100K-L2 labeled network with 99,997 nodes. (f) Duration of generating all landmark distances by NetworkX's highly optimized BFS compared with our widest and deepest GNNs—GCN, GraphSage, GAT, and GIN models were examined and are represented by solid lines of the same color for the same number of local step R. See Appendices B and C for further details and discussions on the experiments and benchmark networks.

GNNs are well-suited for this task as they align with dynamic programming strategies that are used in shortest-path algorithms [Xu et al., 2019b, Dudzik and Veličković, 2022]. As shown in Figure 1(a), the GNN achieves a substantial improvement over the vanilla lower bounds in approximating shortest path distances across all tested R, aided by better-learned embeddings and the near-certain connectivity of large graphs in this regime. Figures 1(b-e) further demonstrate the transferability of GNNs to larger ER graphs and real-world benchmark datasets. Particularly, the GNN-based embeddings achieve comparable or better performance than BFS-based embeddings, with MSE steadily improving as training graph size increases, even when learned on synthesized graphs up to 128 times smaller than the target graph. The GNN-based embeddings not only provide better distance approximations but also scale more efficiently in time and space than their BFS-based counterparts, as illustrated in Figure 1(f), making them a promising tool for large-scale graph representation learning in practical applications.

6 Conclusion

Our analysis, focused on average-case random graphs, provides a simplified framework for developing theoretical tools and insights into landmark-based embedding algorithms. Particularly, Algorithm 1 achieves $(1 \pm \varepsilon)$-factor approximations of shortest-path distances on random graphs w.h.p. even with reduced embedding dimensionality, complementing Bourgain's worst-case results [1985]. By integrating GNNs into the embedding construction, we further improve its generalizability and transferability while reducing time and space complexity, as demonstrated by experiments on ER graphs and benchmark datasets. These signal the potential of machine learning-based landmark algorithms as a solution to graph representation learning for large, complex network data.

Limitations and Future Work. While our results improve upon existing landmark-based algorithms, several limitations remain. Our theoretical analysis focuses on ER graphs, a simplified model; extending it to more complex graphs (e.g., inhomogeneous random graphs, configuration models, planar graphs) is a key direction for future work. The approach also relies on GNNs generalizing from smaller to larger graphs; further studies are needed to assess robustness across diverse graph properties. Finally, additional improvements in memory and inference efficiency may be possible with advanced GNN architectures or alternative models.

References

Takuya Akiba, Yoichi Iwata, and Yuichi Yoshida. Fast exact shortest-path distance queries on large networks by pruned landmark labeling. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 349–360, 2013.

Takuya Akiba, Yoichi Iwata, and Yuichi Yoshida. Dynamic and historical shortest-path distance queries on large evolving networks by pruned landmark labeling. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, page 237–248, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327442. doi: 10.1145/2566486.2568007. URL `https://doi.org/10.1145/2566486.2568007`.

Krishna B. Athreya and Peter E. Ney. *Branching Processes*. Springer, Berlin, Heidelberg, 1972.

Pranjal Awasthi, Abhimanyu Das, and Sreenivas Gollapudi. Beyond GNNs: An efficient architecture for graph problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6019–6027, 2022.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Charles Bordenave. *Lecture notes on random graphs and probabilistic combinatorial optimization*. Available at https://www.math.univ-toulouse.fr/ bordenave/coursRG.pdf, 2016.

J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1):46–52, 1985. doi: 10.1007/BF02776078. URL `https://doi.org/10.1007/BF02776078`.

Dustin Brunner. *Distance Preserving Graph Embedding*. PhD thesis, BS thesis, ETH Zürich, Zurich, 2021.[Online]. Available: https://pub. tik . . . , 2021.

Shaosheng Cao, Wei Lu, and Qiongkai Xu. Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM)*, pages 891–900. ACM, 2015.

Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, 3rd edition, 2009.

Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

Andrew J Dudzik and Petar Veličković. Graph neural networks are dynamic programmers. *Advances in neural information processing systems*, 35:20635–20647, 2022.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428*, 2019.

Michael L Fredman and Dan E Willard. Trans-dichotomous algorithms for minimum spanning trees and shortest paths. In *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science*, pages 719–725. IEEE, 1990.

Giorgio Gallo and Stefano Pallottino. Shortest path algorithms. *Annals of operations research*, 13(1): 1–79, 1988.

Andrew V Goldberg and Chris Harrelson. Computing the shortest path: A search meets graph theory. In *SODA*, volume 5, pages 156–165, 2005.

Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. In *Knowledge-Based Systems*, volume 151, pages 78–94. Elsevier, 2018.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 855–864. ACM, 2016.

Andrey Gubichev, Srikanta Bedathur, Stephan Seufert, and Gerhard Weikum. Fast and accurate estimation of shortest paths in large graphs. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, page 499–508, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450300995. doi: 10.1145/1871437.1871503. URL https://doi.org/10.1145/1871437.1871503.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017a.

William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. In *IEEE Data Engineering Bulletin*, volume 40, pages 52–74, 2017b.

Remco van der Hofstad. *Random Graphs and Complex Networks*, volume I. Cambridge University Press, Cambridge, 2017. doi: 10.1017/9781316779422. URL http://www.win.tue.nl/{~}rhofstad/NotesRGCN.pdf.

Remco van der Hofstad. *Random Graphs and Complex Networks*, volume II. Cambridge University Press, Cambridge, 2024.

Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random Graphs*. John Wiley & Sons, 2000.

Xiaolin Jiang, Chengshuo Xu, Xizhe Yin, Zhijia Zhao, and Rajiv Gupta. Tripoline: generalized incremental graph processing via graph triangle inequality. In *Proceedings of the Sixteenth European Conference on Computer Systems*, EuroSys '21, page 17–32, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383349. doi: 10.1145/3447786.3456226. URL https://doi.org/10.1145/3447786.3456226.

W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. In *Conference in Modern Analysis and Probability (New Haven, Conn., 1982)*, volume 26 of *Contemporary Mathematics*, pages 189–206. American Mathematical Society, Providence, RI, 1984.

David R Karger, Daphne Koller, and Steven J Phillips. Finding the hidden path: Time bounds for all-pairs shortest paths. *SIAM Journal on Computing*, 22(6):1199–1217, 1993.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *5th Int. Conf. Learning Representations*, Toulon, France, 24-26 Apr. 2017. Assoc. Comput. Linguistics.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, page 177–187, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 159593135X. doi: 10.1145/1081870.1081893. URL https://doi.org/10.1145/1081870.1081893.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):2–es, mar 2007. ISSN 1556-4681. doi: 10.1145/1217299.1217301. URL https://doi.org/10.1145/1217299.1217301.

N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

Andreas Loukas. What graph neural networks cannot learn: depth vs width. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=B1l2bp4YwS.

J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93:333–344, 1996.

Xianrui Meng, Seny Kamara, Kobbi Nissim, and George Kollios. Grecs: Graph encryption for approximate shortest distance queries. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, page 504–517, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450338325. doi: 10.1145/2810103.2813672. URL https://doi.org/10.1145/2810103.2813672.

Assaf Naor. A spectral gap precludes low-dimensional embeddings. *ArXiv*, abs/1611.08861, 2016. URL https://api.semanticscholar.org/CorpusID:2112685.

Assaf Naor. An average john theorem. *Geometry & Topology*, 25(4):1631–1717, 2021. doi: 10.2140/gt.2021.25.1631.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 701–710. ACM, 2014.

Michalis Potamias, Francesco Bonchi, Carlos Castillo, and Aristides Gionis. Fast shortest path distance estimation in large networks. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 867–876, 2009.

Jianzhong Qi, Wei Wang, Rui Zhang, and Zhuowei Zhao. A learning based approach to predict shortest-path distances. In *International Conference on Extending Database Technology*, 2020. URL https://api.semanticscholar.org/CorpusID:214613433.

Fatemeh Salehi Rizi, Joerg Schloetterer, and Michael Granitzer. Shortest path distance approximation using deep learning techniques. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1007–1014. IEEE, 2018.

Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL https://networkrepository.com.

Benedek Rozemberczki and Rik Sarkar. Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, page 1325–1334. ACM, 2020.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding, 2019a.

Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding with self clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019*, pages 65–72. ACM, 2019b.

L. Ruiz, L. F. O. Chamon, and A. Ribeiro. Graphon neural networks and the transferability of graph neural networks. In *34th Neural Inform. Process. Syst.*, Vancouver, BC (Virtual), 6-12 Dec. 2020. NeurIPS Foundation.

L. Ruiz, L. F. O. Chamon, and A. Ribeiro. Transferability properties of graph neural networks. *IEEE Transactions on Signal Processing*, 2023.

Atish Das Sarma, Sreenivas Gollapudi, Marc Najork, and Rina Panigrahy. A sketch-based distance oracle for web-scale graphs. In *Web Search and Data Mining*, 2010. URL https://api.semanticscholar.org/CorpusID:17378629.

Atish Das Sarma, Stephan Holzer, Liah Kor, Amos Korman, Danupon Nanongkai, Gopal Pandurangan, David Peleg, and Roger Wattenhofer. Distributed verification and hardness of distributed approximation. *SIAM Journal on Computing*, 41(5):1235–1265, 2012. doi: 10.1137/11085178X. URL https://doi.org/10.1137/11085178X.

Alexander Schrijver. On the history of the shortest path problem. *Documenta Mathematica*, 17(1): 155–167, 2012.

Anastasios Sidiropoulos, Mihai Badoiu, Kedar Dhamdhere, Anupam Gupta, Piotr Indyk, Yuri Rabinovich, Harald Racke, and R. Ravi. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. *SIAM Journal on Discrete Mathematics*, 33(1):454–473, 2019. doi: 10.1137/17M1113527. URL https://doi.org/10.1137/17M1113527.

Christian Sommer. Shortest-path queries in static networks. *ACM Comput. Surv.*, 46(4), March 2014. ISSN 0360-0300. doi: 10.1145/2530531. URL https://doi.org/10.1145/2530531.

Jian Tang, Yifan Lu, and Xia Zhu. Cauchy graph embedding. *Journal of Machine Learning Research*, 20(1):1–27, 2019.

David Tanny. Limit theorems for branching processes in a random environment. *The Annals of Probability*, 5(1):100 – 116, 1977. doi: 10.1214/aop/1176995894. URL https://doi.org/10.1214/aop/1176995894.

Konstantin Tretyakov, Abel Armas-Cervantes, Luciano García-Bañuelos, Jaak Vilo, and Marlon Dumas. Fast fully dynamic landmark-based estimation of shortest path distances in very large graphs. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1785–1794, 2011.

Anton Tsitsulin, Davide Mottin, Panagiotis Karras, and Emmanuel Müller. Just walk the graph: Learning node embeddings via meta paths. In *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*, pages 1306–1311. IEEE, 2018.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *Int. Conf. Learning Representations 2018*, pages 1–12, Vancouver, BC, 30 Apr.-3 May 2018. Assoc. Comput. Linguistics.

K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? In *7th Int. Conf. Learning Representations*, pages 1–17, New Orleans, LA, 6-9 May 2019a. Assoc. Comput. Linguistics.

Keyulu Xu, Jingling Li, Mozhi Zhang, Simon S Du, Ken-ichi Kawarabayashi, and Stefanie Jegelka. What can neural networks reason about? *arXiv preprint arXiv:1905.13211*, 2019b.

Wen Zhang, Yuxuan Zhang, Yansong Feng, Zheng Wang, and Jin Zhang. Prone: A scalable graph embedding method with local proximity preservation. *IEEE Transactions on Knowledge and Data Engineering*, 33(6):2500–2513, 2021.

A Proofs

A.1 Proof of Lemma 4.3

If A_n occurs but B_n does not, then $|\mathcal{C}_{(2)}| \geq n^{\kappa_0-\varepsilon}$, which occurs with probability tending to zero since $|\mathcal{C}_{(2)}| = O(\log n)$ w.h.p. On the other hand, if B_n occurs and A_n does not, then $|\partial N_L(u_1)| \notin [n^{\kappa_0-\varepsilon}, n^{\kappa_0+\varepsilon}]$ or $|\partial N_L(u_2)| \notin [n^{\kappa_0-\varepsilon}, n^{\kappa_0+\varepsilon}]$.

To bound the probabilities of these events, consider a branching process with progeny distribution Poisson(λ), and let $\mathcal{X}_l$ be the number of children at generation l. We first claim that, for any fixed node u in G,

$$\lim_{n\to\infty} \mathbb{P}(|\partial N_L(u)| = \mathcal{X}_L) = 1 \tag{3}$$

for any $\kappa_0 \in (0, \frac{1}{2})$ and $L = \kappa_0 \log_\lambda n$. Indeed, this is a consequence of Lemma 3.13 from Bordenave [2016].

Next, classical theory of branching processes shows that, on the event of survival, the growth rate of a branching process is exponential. More precisely, Theorem 5.5 (iii) from Tanny [1977], together with Theorem 2 from Athreya and Ney [1972], yields

$$\lim_{L\to\infty} \mathbb{P}\big(L(1-\varepsilon) \leq \log_\lambda \mathcal{X}_L \leq L(1+\varepsilon),\, \mathcal{X}_L > 0\big) = 1.$$

Since $\kappa_0 - \varepsilon < \kappa_0(1-\varepsilon)$ and $\kappa_0 + \varepsilon > \kappa_0(1+\varepsilon)$,

$$\lim_{L\to\infty} \mathbb{P}\big(n^{\kappa_0-\varepsilon} \leq \mathcal{X}_L \leq n^{\kappa_0+\varepsilon},\, \mathcal{X}_L > 0\big) = 1. \tag{4}$$

Combining (3) and (4), it follows that

$$\mathbb{P}(B_n \setminus A_n) \leq \mathbb{P}(|\partial N_L(u_1)| \notin [n^{\kappa_0-\varepsilon}, n^{\kappa_0+\varepsilon}]) + \mathbb{P}(|\partial N_L(u_2)| \notin [n^{\kappa_0-\varepsilon}, n^{\kappa_0+\varepsilon}]) \to 0.$$

A.2 Proof of Lemma 4.4

The proof is adapted from Section 2.6.4 in van der Hofstad [2024]. Since we need an exponential bound on the probability and L grows with n, the proof does not follow from van der Hofstad [2024].

Note that for any fixed node u and any $k \geq 1$, $|\partial N_k(u)| \leq \sum_{x\notin N_{k-1}(u)} \sum_{y\in\partial N_{k-1}(u)} I_{xy}$, where I_{xy} is the indicator random variable for the edge $\{x, y\}$ being present. Therefore, $\mathbb{E}(|\partial N_k(u)|) \leq \lambda E(|\partial N_{k-1}(u)|)$. Proceeding inductively, we have $\mathbb{E}(|\partial N_k(u)|) \leq \lambda^k$ and consequently,

$$\mathbb{E}(|N_k(u)|) \leq \frac{\lambda^{k+1}-1}{\lambda-1} = O(\lambda^k). \tag{5}$$

Then with Markov's inequality, there exists $\delta > 0$ for any $\gamma \in (\kappa_0 + \kappa, 1)$ such that

$$\mathbb{P}(|N_k(u_i)| \geq n^\gamma) \leq \frac{O(\lambda^k)}{n^\gamma} \leq \frac{O\left(n^{\kappa_0+\kappa}\right)}{n^\gamma} \leq n^{-\delta}$$

for $i = 1, 2$ and $k \leq (\kappa_0+\kappa)\log_\lambda n$ with sufficiently large n. Then for each fixed $k \leq (\kappa_0+\kappa)\log_\lambda n$,

$$\mathbb{P}(|N_k(u_i)| \leq n^\gamma : i = 1, 2) \geq 1 - \sum_{i=1,2} \mathbb{P}(|N_k(u_i)| \geq n^\gamma) \geq 1 - 2n^{-\delta}. \tag{6}$$

Let $\delta_n = n^{-\beta}$ with $0 < \beta < \frac{\kappa_0-\varepsilon'}{2}$. Also define

$$\bar{\mathcal{E}}_{n,k} = \{|\partial N_{L+k}(u_i)| \in [b'_m(1-\delta_n)^k(1-n^{\gamma-1})^k\lambda^k, b'_M(1+\delta_n)^k\lambda^k] : i = 1, 2\}$$

with $b'_m = n^{-\varepsilon'}\lambda^L$ and $b'_M = n^{\varepsilon'}\lambda^L$ for some $0 < \varepsilon' < \min\{\varepsilon, 1-\kappa_0-\kappa\}$. Conditionally on A_{b_m,b_M},

$$\begin{aligned}
\mathbb{E}(|\partial N_{L+k}(u_i)| \mid N_{L+k-1}(u_i)) &= \mathbb{E}\left(\sum_{x\notin N_{L+k-1}(u_i)} \mathbb{1}_{\{\exists y\in\partial N_{L+k-1}(u_i): I_{xy}=1\}} \mid N_{L+k-1}(u_i)\right) \\
&= (n - |N_{L+k-1}(u_i)|)\mathbb{P}(\exists y \in \partial N_{L+k-1}(u_i) : I_{xy} = 1 \mid N_{L+k-1}(u_i);\, x \notin N_{L+k-1}(u_i)) \\
&= (n - |N_{L+k-1}(u_i)|)\left(1 - \left(1 - \frac{\lambda}{n}\right)^{|\partial N_{L+k-1}(u_i)|}\right).
\end{aligned}$$

Since $\frac{\lambda}{n} \in [0,1]$,

$$1-|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n} \le \left(1-\frac{\lambda}{n}\right)^{|\partial N_{L+k-1}(u_i)|} \le 1-|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n}+\frac{|\partial N_{L+k-1}(u_i)|^2}{2}\left(\frac{\lambda}{n}\right)^2.$$

Conditionally on $\bar{\mathcal{E}}_{n,k-1}$,

$$|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n} \le n^{\varepsilon'}(1+\delta_n)^{k-1}\frac{\lambda^{L+k}}{n} \le n^{\varepsilon'-1}(1+\delta_n)^{k-1}n^{\kappa_0+\kappa} \to 0 \text{ as } n\to\infty.$$

Since $\left(|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n}\right)^2$ vanishes faster than $|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n}$, we have w.h.p. that

$$\left(1-\frac{\lambda}{n}\right)^{|\partial N_{L+k-1}(u_i)|} = 1-|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n}$$

and so

$$\mathbb{E}(|\partial N_{L+k}(u_i)| \mid N_{L+k-1}(u_i)) = (n-|N_{L+k-1}(u_i)|)|\partial N_{L+k-1}(u_i)|\frac{\lambda}{n}.$$

Conditionally on $\cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}$ and A_{b_m,b_M}, from (6) we have with probability at least $1-2n^{-\delta}$ that

$$\mathbb{E}(|\partial N_{L+k}(u_i)| \mid N_{L+k-1}(u_i)) \in \left[b'_m(1-\delta_n)^{k-1}(1-n^{\gamma-1})^k\lambda^k, b'_M(1+\delta_n)^{k-1}\lambda^k\right]$$

since $1-n^{\gamma-1} \le 1-\frac{|N_{L+k-1}(u_i)|}{n} \le 1$ for $i=1,2$ with sufficiently large n. Denote this event R_k. The fact that $\mathbb{P}(A) \le \mathbb{P}(A \mid B)+\mathbb{P}(B^c)$ implies

$$\mathbb{P}(\bar{\mathcal{E}}_{n,k}^c \mid \cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}, A_{b_m,b_M}) \le \mathbb{P}(\bar{\mathcal{E}}_{n,k}^c \mid R_k, \cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}, A_{b_m,b_M}) + 2n^{-\delta}.$$

Using union bound and Chernoff-Hoeffding bound [Dubhashi and Panconesi, 2009, Theorem 1.1],

$$\begin{aligned}
&\mathbb{P}(\bar{\mathcal{E}}_{n,k}^c \mid R_k, \cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}, A_{b_m,b_M}) \\
&\le \sum_{i=1,2}\mathbb{P}(||\partial N_{L+k}(u_i)| - \mathbb{E}(|\partial N_{L+k}(u_i)|)| \ge \delta_n\mathbb{E}(|\partial N_{L+k}(u_i)|) \mid R_k, \cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}, A_{b_m,b_M}) \\
&\le \sum_{i=1,2} 2\exp\left(-\frac{\delta_n^2}{3}n^{-\varepsilon'}(1-\delta_n)^{k-1}(1-n^{\gamma-1})^k\lambda^{L+k}\right) \\
&\le 4\exp\left(-\frac{n^{-2\beta}}{3}n^{-\varepsilon'}(1-\delta_n)^{k-1}(1-n^{\gamma-1})^k n^{\kappa_0}\right).
\end{aligned}$$

Since $(1-\delta_n)^{k-1}, (1-n^{\gamma-1})^k \to 1$ as $n\to\infty$ and $2\beta < \kappa_0-\varepsilon'$, $4\exp\left(-\frac{n^{-2\beta}}{3}n^{-\varepsilon'}(1-\delta_n)^{k-1}(1-n^{\gamma-1})^k n^{\kappa_0}\right)$ vanishes faster than $2n^{-\delta}$. Then with sufficiently large n, $\mathbb{P}(\bar{\mathcal{E}}_{n,k} \mid \cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}, A_{b_m,b_M}) \ge 1-3n^{-\delta}$. Proceed inductively,

$$\begin{aligned}
&\mathbb{P}(\cap_{l=0}^{k}\bar{\mathcal{E}}_{n,l} \mid A_{b_m,b_M}) = \mathbb{P}(\bar{\mathcal{E}}_{n,k} \mid \cap_{l=0}^{k-1}\bar{\mathcal{E}}_{n,l}, A_{b_m,b_M})\ldots\mathbb{P}(\bar{\mathcal{E}}_{n,1} \mid \bar{\mathcal{E}}_{n,0}, A_{b_m,b_M})\mathbb{P}(\bar{\mathcal{E}}_{n,0} \mid A_{b_m,b_M}) \\
&\ge (1-3n^{-\delta})\ldots(1-3n^{-\delta})\mathbb{P}(\bar{\mathcal{E}}_{n,0} \mid A_{b_m,b_M}) = (1-3n^{-\delta})^k\mathbb{P}(\bar{\mathcal{E}}_{n,0} \mid A_{b_m,b_M}).
\end{aligned}$$

Since $b'_m(1-\delta_n)^k(1-n^{\gamma-1})^k \ge b_m$ and $b'_M(1+\delta_n)^k \le b_M$, $\bar{\mathcal{E}}_{n,0} \subseteq A_{b_m,b_M}$ and $\bar{\mathcal{E}}_{n,k} \subseteq \mathcal{E}_{n,k}$ for all $k \ge 0$. Hence, $\mathbb{P}(\bar{\mathcal{E}}_{n,0} \mid A_{b_m,b_M}) = 1$ and so

$$\mathbb{P}(\cap_{l=0}^{k}\mathcal{E}_{n,l} \mid A_{b_m,b_M}) \ge \mathbb{P}(\cap_{l=0}^{k}\bar{\mathcal{E}}_{n,l} \mid A_{b_m,b_M}) \ge (1-3n^{-\delta})^k \ge 1-3kn^{-\delta}.$$

A.3 Proof of Proposition 4.5

Recall all the notation from Lemma 4.4 and its proof. Then for any k_1, k_2 such that $L < k_1, k_2 \le k = (\kappa_0+\kappa)\log_\lambda n$,

$$\begin{aligned}
&\mathbb{E}(|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \mid N_{k_1}(u_1), N_{k_2-1}(u_2)) \\
&= \mathbb{E}\left(\sum_{x\in\partial N_{k_1}(u_1)\setminus N_{k_2-1}(u_2)} \mathbb{1}_{\{\exists y\in\partial N_{k_2-1}(u_2): I_{xy}=1\}} \mid N_{k_1}(u_1), N_{k_2-1}(u_2)\right) \\
&= \left(|\partial N_{k_1}(u_1)| - \sum_{j\le k_2-1}|\partial N_{k_1}(u_1)\cap\partial N_j(u_2)|\right)\left(1-\left(1-\frac{\lambda}{n}\right)^{|\partial N_{k_2-1}(u_2)|}\right).
\end{aligned}$$

Since $\frac{\lambda}{n} \in [0,1]$,

$$1 - |\partial N_{k_2-1}(u_2)|\frac{\lambda}{n} \leq \left(1 - \frac{\lambda}{n}\right)^{|\partial N_{k_2-1}(u_2)|} \leq 1 - |\partial N_{k_2-1}(u_2)|\frac{\lambda}{n} + \frac{|\partial N_{k_2-1}(u_2)|^2}{2}\left(\frac{\lambda}{n}\right)^2.$$

Conditionally on u_1, u_2 being in the same connected component, Lemmas 4.3 and 4.4 imply that with probability at least $1 - 3n^{-\delta}(\lfloor k \rfloor - \lfloor L \rfloor) \geq 1 - 3n^{-\delta}(\kappa \log_\lambda n + 1)$ for some $\delta > 0$, $\cap_{l=0}^{k-L} \mathcal{E}_{n,l}$ occurs. Then with $\varepsilon' \in (0, \min\{\varepsilon, 1 - \kappa_0 - \kappa\})$ (ε' to be chosen later),

$$|\partial N_{k_2-1}(u_2)|\frac{\lambda}{n} \leq n^{\frac{\varepsilon'}{2}}\frac{\lambda^{k_2}}{n} \leq n^{\frac{\varepsilon'}{2}-1} n^{\kappa_0+\kappa} \to 0 \text{ as } n \to \infty.$$

Since $|\partial N_{k_2-1}(u_2)|\frac{\lambda}{n}$ vanishes and $\left(|\partial N_{k_2-1}(u_2)|\frac{\lambda}{n}\right)^2$ vanishes faster, we have w.h.p. that $\left(1 - \frac{\lambda}{n}\right)^{|\partial N_{k_2-1}(u_2)|} = 1 - |\partial N_{k_2-1}(u_2)|\frac{\lambda}{n}$, and so

$$\begin{aligned}
&\mathbb{E}\left(|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \,\middle|\, N_{k_1}(u_1), N_{k_2-1}(u_2)\right) \\
&= \left(|\partial N_{k_1}(u_1)| - \sum_{j \leq k_2 - 1} |\partial N_{k_1}(u_1) \cap \partial N_j(u_2)|\right) |\partial N_{k_2-1}(u_2)|\frac{\lambda}{n}. \qquad (7)
\end{aligned}$$

Conditionally on $\cap_{l=0}^{k-L} \mathcal{E}_{n,l}$,

$$\begin{aligned}
\mathbb{E}\left(|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \,\middle|\, N_{k_1}(u_1), N_{k_2-1}(u_2)\right) &\leq n^{\frac{\varepsilon'}{2}}\lambda^{k_1} n^{\frac{\varepsilon'}{2}}\frac{\lambda^{k_2}}{n} \qquad (8) \\
&\leq n^{\varepsilon'}\lambda^{k_1}\frac{n^{\kappa_0+\kappa}}{n} \leq \lambda^{k_1}\frac{n^{-\gamma}}{7(\lfloor k \rfloor - \lfloor L \rfloor)}
\end{aligned}$$

for all $L < k_2 \leq k$ with $0 < \gamma < \min\{\kappa_0, 1 - \kappa_0 - \kappa\}$ and sufficiently large n. Here we choose ε' small enough so that $\frac{\varepsilon'}{2} < \gamma$, $\frac{k_1}{\log_\lambda n} - \frac{\varepsilon'}{2} > \kappa_0$, and $k_1 + k_2 > (1 + \varepsilon') \log_\lambda n$.

Let A be the event that there exists $L < j \leq k_2$ such that $|\partial N_{k_1}(u_1) \cap \partial N_j(u_2)| \geq \lambda^{k_1}\frac{n^{-\gamma}}{\lfloor k \rfloor - \lfloor L \rfloor}$. Let B be the event that $\mathbb{E}(|\partial N_{k_1}(u_1) \cap \partial N_j(u_2)| \mid N_{k_1}(u_1), N_{k_2-1}(u_2)) \leq \lambda^{k_1}\frac{n^{-\gamma}}{7(\lfloor k \rfloor - \lfloor L \rfloor)}$ for all $L < j \leq k_2$. The fact that $\mathbb{P}(A) \leq \mathbb{P}(A \mid B) + \mathbb{P}(B^c)$ implies

$$\mathbb{P}(A \mid N_{k_1}(u_1), N_{k_2-1}(u_2)) \leq \mathbb{P}(A \mid B, N_{k_1}(u_1), N_{k_2-1}(u_2)) + 3n^{-\delta}(\kappa \log_\lambda n + 1).$$

By Theorem 2.8 and Corollary 2.4 from Janson et al. [2000] with union bound,

$$\begin{aligned}
\mathbb{P}(A \mid B, N_{k_1}(u_1), N_{k_2-1}(u_2)) &\leq (\lfloor k \rfloor - \lfloor L \rfloor)\exp\left(-\lambda^{k_1}\frac{n^{-\gamma}}{\lfloor k \rfloor - \lfloor L \rfloor}\right) \\
&< (\kappa \log_\lambda n + 1)\exp\left(-\frac{n^{\kappa_0-\gamma}}{\lfloor k \rfloor - \lfloor L \rfloor}\right) = (\kappa \log_\lambda n + 1)\exp\left(-n^{\gamma'}\right)
\end{aligned}$$

for some $\gamma' = \kappa_0 - \gamma > 0$. It follows that

$$\begin{aligned}
\mathbb{P}\left(A^c \mid N_{k_1}(u_1), N_{k_2-1}(u_2)\right) &\geq 1 - (\kappa \log_\lambda n + 1)\exp\left(-n^{\gamma'}\right) - 3n^{-\delta}(\kappa \log_\lambda n + 1) \\
&\geq 1 - 4n^{-\delta}(\kappa \log_\lambda n + 1) \qquad (9)
\end{aligned}$$

for sufficiently large n since $(\kappa \log_\lambda n + 1)\exp\left(-n^{\gamma'}\right)$ vanishes faster than $3n^{-\delta}(\kappa \log_\lambda n + 1)$.

Let $\gamma'' \in \left(\kappa_0, \frac{k_1}{\log_\lambda n} - \frac{\varepsilon'}{2}\right)$. Markov's inequality and Lemma 5 imply that there exists $\delta' > 0$ such that

$$\mathbb{P}(|N_L(u_2)| \geq n^{\gamma''}) \leq \frac{O(\lambda^L)}{n^{\gamma''}} \leq \frac{O\left(n^{\kappa_0}\right)}{n^{\gamma''}} \leq n^{-\delta'} \qquad (10)$$

for sufficiently large n.

Combining (7), (8), (9), (10) with Lemmas 4.3, 4.4, we have w.h.p. that

$$\begin{aligned}
&\mathbb{E}\big(|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \,\big|\, N_{k_1}(u_1), N_{k_2-1}(u_2)\big) \\
&\geq \left(|\partial N_{k_1}(u_1)| - \sum_{L<j\leq k_2-1} |\partial N_{k_1}(u_1) \cap \partial N_j(u_2)| - |N_L(u_2)|\right) |\partial N_{k_2-1}(u_2)| \frac{\lambda}{n} \\
&\geq \left(n^{-\frac{\varepsilon'}{2}}\lambda^{k_1} - (\lfloor k_2 \rfloor - 1 - \lfloor L \rfloor)\lambda^{k_1} \frac{n^{-\gamma}}{\lfloor k \rfloor - \lfloor L \rfloor} - n^{\gamma''}\right) \frac{n^{-\frac{\varepsilon'}{2}}\lambda^{k_2}}{n} > \frac{n^{-\varepsilon'}\lambda^{k_1+k_2}}{2n}
\end{aligned}$$

and

$$\mathbb{E}\big(|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \,\big|\, N_{k_1}(u_1), N_{k_2-1}(u_2)\big) \leq n^{\varepsilon'} \frac{\lambda^{k_1+k_2}}{n},$$

where the last "$\geq$" holds since $\lambda^{k_1} n^{-\gamma}$ and $n^{\gamma''}$ grow strictly slower than $n^{-\frac{\varepsilon'}{2}}\lambda^{k_1}$ as $\frac{\varepsilon'}{2} < \gamma$ and $\gamma'' < \frac{k_1}{\log_\lambda n} - \frac{\varepsilon'}{2}$. Therefore, $\mathbb{E}(|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)|) \in \left[\frac{n^{-\varepsilon'}\lambda^{k_1+k_2}}{2n}, \frac{n^{\varepsilon'}\lambda^{k_1+k_2}}{n}\right]$ and we denote this event S.

Let R denote the event that $|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \notin \left[\frac{(1-\varepsilon)n^{-\varepsilon'}\lambda^{k_1+k_2}}{2n}, \frac{(1+\varepsilon)n^{\varepsilon'}\lambda^{k_1+k_2}}{n}\right]$. Using Chernoff-Hoeffding bound [Dubhashi and Panconesi, 2009, Theorem 1.1],

$$\mathbb{P}(R) \leq \mathbb{P}(R \mid S) + \mathbb{P}(S^c) \leq 2\exp\left(-\frac{\varepsilon^2}{3}\frac{n^{-\varepsilon'}\lambda^{k_1+k_2}}{2n}\right) + \mathbb{P}(S^c).$$

Since $k_1 + k_2 > (1+\varepsilon')\log_\lambda n$ and S occurs w.h.p., $\mathbb{P}(R^c)$ converges to 1. Then w.h.p.,

$$|\partial N_{k_1}(u_1) \cap \partial N_{k_2}(u_2)| \in \left[\frac{(1-\varepsilon)n^{-\varepsilon'}\lambda^{k_1+k_2}}{2n}, \frac{(1+\varepsilon)n^{\varepsilon'}\lambda^{k_1+k_2}}{n}\right] \subseteq \left[\frac{n^{-\varepsilon}\lambda^{k_1+k_2}}{2n}, \frac{n^{\varepsilon}\lambda^{k_1+k_2}}{n}\right].$$

A.4 Proof of Theorem 4.1

Let $k_1 = \varepsilon' d(u_1,u_2)$ and $k_2 = (1-\varepsilon+\varepsilon')d(u_1,u_2)$, where $\varepsilon' = \min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\} - \varepsilon'' \in \left(0, \min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\}\right)$ ($\varepsilon'' \in \left(0, \min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\}\right)$ to be chosen later). Since $\varepsilon' < \frac{\varepsilon}{2}$, $k_1 + k_2 < d(u_1,u_2)$, and so $N_{k_1}(u_1) \cap N_{k_2}(u_2) = \varnothing$. Conditionally on u_1, u_2 being in the same connected component, Theorem 2.36 from van der Hofstad [2024] implies that $d(u_1,u_2)/\log_\lambda n \xrightarrow{\mathbb{P}} 1$. In other words, $(1-\epsilon)\log_\lambda n \leq d(u_1,u_2) \leq (1+\epsilon)\log_\lambda n$ w.h.p. for any fixed $\epsilon > 0$. With ϵ small enough so that $\varepsilon'(1+\epsilon) < 1$, $k_1 \leq \varepsilon'(1+\epsilon)\log_\lambda n < \log_\lambda n$ w.h.p., allowing us to apply Lemmas 4.3 and 4.4 on $|\partial N_{k_1}(u_1)|$.

Let S_{ij} be the landmark set of size M^i sampled in the j-th round and Z_{ij} denote the event that $S_{ij} \cap N_{k_1}(u_1) \neq \varnothing$ but $S_{ij} \cap N_{k_2}(u_2) = \varnothing$. If Z_{ij} happens for some $i \leq r$ and $j \leq R$, then $d(u_1, S_{ij}) \leq k_1$ and $d(u_2, S_{ij}) \geq k_2$, and consequently, $\underline{d}(u_1,u_2) \geq k_2 - k_1 = (1-\varepsilon)d(u_1,u_2)$. Thus, denoting $Z = \cup_{i\leq r, j\leq R} Z_{ij}$, it suffices to prove that $\mathbb{P}(Z \mid u_1 \leftrightarrow u_2) \xrightarrow{\mathbb{P}} 1$. Since $\mathbb{P}(u_1 \leftrightarrow u_2 \text{ but } u_1, u_2 \notin \mathcal{C}_{(1)}) = \frac{1}{n^2}\sum_{i\geq 2} |\mathcal{C}_{(i)}|^2 \leq \frac{|\mathcal{C}_{(2)}|}{n} \xrightarrow{\mathbb{P}} 0$, it suffices to show that $\mathbb{P}(Z \mid u_1, u_2 \in \mathcal{C}_{(1)}) \xrightarrow{\mathbb{P}} 1$ (or equivalently $\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \xrightarrow{\mathbb{P}} 0$). The fact that $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ implies, for each (i,j), that

$$\begin{aligned}
\mathbb{P}(Z_{ij} \mid u_1, u_2 \in \mathcal{C}_{(1)}) &= \mathbb{P}(S_{ij} \cap N_{k_1}(u_1) \neq \varnothing, S_{ij} \cap N_{k_2}(u_2) = \varnothing \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&= \left(1 - \frac{|N_{k_2}(u_2)|}{n}\right)^{M^i} - \left(1 - \frac{|N_{k_1}(u_1)| + |N_{k_2}(u_2)|}{n}\right)^{M^i}.
\end{aligned}$$

By independence of Z_{ij}'s,

$$
\begin{aligned}
&\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&= \left(\prod_{i=0}^{r} \left(1 - \left(1 - \frac{|N_{k_2}(u_2)|}{n} \right)^{M^i} + \left(1 - \frac{|N_{k_1}(u_1)| + |N_{k_2}(u_2)|}{n} \right)^{M^i} \right) \right)^R \\
&\leq \exp\left(- R \sum_{i=0}^{r} \left(\left(1 - \frac{|N_{k_2}(u_2)|}{n} \right)^{M^i} - \left(1 - \frac{|N_{k_1}(u_1)| + |N_{k_2}(u_2)|}{n} \right)^{M^i} \right) \right) \\
&= \exp\left(- R \sum_{i=0}^{r} \frac{|N_{k_1}(u_1)|}{n} \sum_{j=0}^{M^i-1} \left(1 - \frac{|N_{k_2}(u_2)|}{n} \right)^{M^i-1-j} \left(1 - \frac{|N_{k_1}(u_1)| + |N_{k_2}(u_2)|}{n} \right)^{j} \right) \\
&\leq \exp\left(- R \frac{|N_{k_1}(u_1)|}{n} \sum_{i=0}^{r} M^i \left(1 - \frac{|N_{k_1}(u_1)| + |N_{k_2}(u_2)|}{n} \right)^{M^i-1} \right) \\
&< \exp\left(- R \frac{|\partial N_{k_1}(u_1)|}{n} M^r \left(1 - \frac{|N_{k_1}(u_1)| + |N_{k_2}(u_2)|}{n} \right)^{M^r} \right)
\end{aligned}
$$

where the first "$\leq$" uses $1 - x \leq \exp(-x)$ and "$<$" uses $\sum_{i=0}^{r} M^i = \frac{M^{r+1}-1}{M-1} > \frac{M^{r+1}-M^r}{M-1} = M^r$.

Recall that $(1-\epsilon)\log_\lambda n \leq d(u_1, u_2) \leq (1+\epsilon)\log_\lambda n$ w.h.p. for any fixed $\epsilon > 0$. Choosing ϵ small enough so that $(1-\varepsilon+\varepsilon')(1+\epsilon) < 1-\theta$, we have that $k_2 \leq (1-\varepsilon+\varepsilon')(1+\epsilon)\log_\lambda n < (1-\theta)\log_\lambda n$ w.h.p., and so there exists $\gamma \in (0, 1-\theta)$ such that $k_1 < k_2 < \gamma \log_\lambda n$. By Markov's inequality and (5), there exists $\delta > 0$ such that $\mathbb{P}(|N_{k_i}(u_i)| \geq n^\gamma) \leq \frac{O(\lambda^{k_i})}{n^\gamma} \leq n^{-\delta}$ for $i = 1, 2$ with sufficiently large n. Therefore,

$$
\mathbb{P}(|N_{k_i}(u_i)| \leq n^\gamma : i = 1, 2) \geq 1 - \sum_{i=1,2} \mathbb{P}(|N_{k_i}(u_i)| \geq n^\gamma) \geq 1 - 2n^{-\delta},
$$

and so $|N_{k_i}(u_i)| \leq n^\gamma$ for $i = 1, 2$ w.h.p.

By Lemmas 4.3 and 4.4, $|\partial N_{k_1}(u_1)| \geq n^{-\varepsilon'''}\lambda^{k_1} \geq n^{-\varepsilon'''} n^{\varepsilon'(1-\epsilon)}$ w.h.p. for any $\varepsilon''' > 0$, and so

$$
\begin{aligned}
&\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&< \exp\left(- R \frac{n^{-\varepsilon'''} n^{\left(\min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\} - \varepsilon''\right)(1-\epsilon)}}{n} M^{\frac{\theta}{\log M}\log n - 1} \left(1 - \frac{2n^\gamma}{n} \right)^{M^{\frac{\theta}{\log M}\log n}} \right) \\
&= \exp\left(- R \frac{n^{-\varepsilon'''} n^{\min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\} - \epsilon\min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\} - \varepsilon'' + \varepsilon''\epsilon}}{nM} n^\theta \left(1 - \frac{2n^\gamma}{n} \right)^{n^\theta} \right).
\end{aligned}
$$

Since $\gamma < 1 - \theta$, $\left(1 - \frac{2n^\gamma}{n}\right)^{n^\theta} \geq 1 - \frac{2n^{\gamma+\theta}}{n} \to 1$ as $n \to \infty$. Since $\varepsilon'', \varepsilon''', \epsilon$ can be chosen small enough so that $-\varepsilon''' - \epsilon\min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\} - \varepsilon'' + \varepsilon''\epsilon < \varsigma$ for any $\varsigma > 0$, $R = \Omega\left(Mn^{1-\theta-\min\left\{\frac{\varepsilon}{2}, \varepsilon-\theta\right\}+\varsigma}\right)$ is sufficient for the final bound to tend to 0. Since $\theta \in (0, \varepsilon)$, R can be further simplified to $\Omega\left(Mn^{1-\frac{\varepsilon}{2}-\min\left\{\frac{\varepsilon}{2}, \theta\right\}+\varsigma}\right)$.

A.5 Proof of Theorem 4.2

Let $k = \varepsilon' d(u_1, u_2)$ with $\varepsilon' = \frac{1+\varepsilon}{2} - \varepsilon'' \in \left(0, \frac{1+\varepsilon}{2}\right)$ ($\varepsilon'' \in \left(0, \frac{1+\varepsilon}{2}\right)$ to be chosen later). Conditionally on u_1, u_2 being in the same connected component, Theorem 2.36 from van der Hofstad [2024] implies that $d(u_1, u_2)/\log_\lambda n \xrightarrow{\mathbb{P}} 1$. In other words, $(1-\epsilon)\log_\lambda n \leq d(u_1, u_2) \leq (1+\epsilon)\log_\lambda n$ w.h.p. for any fixed $\epsilon > 0$. With ε'', ϵ small enough so that $\varepsilon'(1+\epsilon) < 1$ and $2\left(\frac{1+\varepsilon}{2} - \varepsilon''\right)(1-\epsilon) > 1$, $k \leq \varepsilon'(1+\epsilon)\log_\lambda n < \log_\lambda n$ and $k + k \geq 2\varepsilon'(1-\epsilon)\log_\lambda n > \log_\lambda n$ w.h.p. This allows us to apply Lemmas 4.3 and 4.4 and Proposition 4.5.

Let S_{ij} be the landmark set of size M^i sampled in the j-th round and Z_{ij} be the event that S_{ij} contains at least one landmark node in $N_k(u_1) \cap N_k(u_2)$ and none in

$(N_k(u_1) \cup N_k(u_2)) \setminus (N_k(u_1) \cap N_k(u_2))$. If Z_{ij} happens for some $i \le r$ and $j \le R$, the landmarks in the intersection will be the common landmarks for calculating $\bar{d}(u_1, u_2)$, and so $\bar{d}(u_1, u_2) \le 2k \le (1+\varepsilon) d(u_1, u_2)$. Thus, denoting $Z = \cup_{i \le r, j \le R} Z_{ij}$, it suffices to prove that $\mathbb{P}(Z \mid u_1 \leftrightarrow u_2) \xrightarrow{\mathbb{P}} 1$. Since $\mathbb{P}(u_1 \leftrightarrow u_2 \text{ but } u_1, u_2 \notin \mathcal{C}_{(1)}) = \frac{1}{n^2} \sum_{i \ge 2} |\mathcal{C}_{(i)}|^2 \le \frac{|\mathcal{C}_{(2)}|}{n} \xrightarrow{\mathbb{P}} 0$, it suffices to show that $\mathbb{P}(Z \mid u_1, u_2 \in \mathcal{C}_{(1)}) \xrightarrow{\mathbb{P}} 1$ (or equivalently $\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \xrightarrow{\mathbb{P}} 0$). Note that for each (i,j),

$$\begin{aligned}
&\mathbb{P}(Z_{ij} \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&= \frac{|N_k(u_1) \cap N_k(u_2)|}{n} \left(\frac{|N_k(u_1) \cap N_k(u_2)|}{n} + 1 - \frac{|N_k(u_1) \cup N_k(u_2)|}{n} \right)^{M^i - 1}.
\end{aligned}$$

By independence of Z_{ij}'s,

$$\begin{aligned}
&\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&= \left(\prod_{i=0}^{r} \left(1 - \frac{|N_k(u_1) \cap N_k(u_2)|}{n} \left(\frac{|N_k(u_1) \cap N_k(u_2)|}{n} + 1 - \frac{|N_k(u_1) \cup N_k(u_2)|}{n} \right)^{M^i - 1} \right) \right)^R \\
&\le \exp\left(-R \sum_{i=0}^{r} \frac{|\partial N_k(u_1) \cap \partial N_k(u_2)|}{n} \left(\frac{|\partial N_k(u_1) \cap \partial N_k(u_2)|}{n} + 1 - \frac{|N_k(u_1)| + |N_k(u_2)|}{n} \right)^{M^i - 1} \right).
\end{aligned}$$

Choosing $L \in (0, \min\{k, \gamma \log_\lambda n\})$ for some $\gamma \in (0, 1-\theta)$, we obtain from Markov's inequality and Lemma 5 that $\mathbb{P}(|N_L(u_i)| \ge n^\gamma) \le \frac{O(\lambda^L)}{n^\gamma} \le n^{-\delta}$ for $i = 1, 2$ with some $\delta > 0$ and sufficiently large n. Therefore,

$$\mathbb{P}(|N_L(u_i)| \le n^\gamma : i = 1, 2) \ge 1 - \sum_{i=1,2} \mathbb{P}(|N_L(u_i)| \ge n^\gamma) \ge 1 - 2n^{-\delta},$$

and so $|N_L(u_i)| \le n^\gamma$ for $i = 1, 2$ w.h.p. Then by Lemmas 4.3 and 4.4,

$$|N_k(u_i)| = |N_L(u_i)| + \sum_{l=L+1}^{k} |\partial N_l(u_i)| \le n^\gamma + \sum_{l=L+1}^{k} n^{\varepsilon'''} \lambda^l < n^\gamma + (\log_\lambda n) n^{\varepsilon'''} \lambda^k$$

for $i = 1, 2$ w.h.p. with $0 < \varepsilon''' < 1 - \kappa_0 - \kappa$. Combining these with Proposition 4.5, we have w.h.p. that

$$\begin{aligned}
&\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&\le \exp\left(-R \frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} \sum_{i=0}^{r} \left(\frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} + 1 - \frac{2n^\gamma}{n} - \frac{2(\log_\lambda n) n^{\varepsilon'''} \lambda^k}{n} \right)^{M^i - 1} \right).
\end{aligned}$$

Since $\gamma < 1$ and $0 < \lambda^k \le n^{\kappa_0 + \kappa} < n^{1 - \varepsilon'''} < n$, $\frac{2n^\gamma}{n} + \frac{2(\log_\lambda n) n^{\varepsilon'''} \lambda^k}{n} - \frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} \in (0, 1)$ when n is large, and so

$$\begin{aligned}
&\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) \\
&\le \exp\left(-R \frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} \sum_{i=0}^{r} \left(1 - \left(\frac{2n^\gamma}{n} + \frac{2(\log_\lambda n) n^{\varepsilon'''} \lambda^k}{n} - \frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} \right) (M^i - 1) \right) \right) \\
&< \exp\left(-R \frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} \left(r - \left(\frac{2n^\gamma}{n} + \frac{2(\log_\lambda n) n^{\varepsilon'''} \lambda^k}{n} - \frac{n^{-\varepsilon'''} \lambda^{2k}}{2n^2} \right) \frac{n^\theta M}{M - 1} \right) \right)
\end{aligned}$$

since $\sum_{i=0}^{r} (M^i - 1) < \sum_{i=0}^{r} M^i = \frac{M^{r+1} - 1}{M - 1} \le \frac{n^\theta M}{M - 1}$. Since $\gamma < 1 - \theta$, $0 < \frac{2n^{\gamma + \theta}}{n} \to 0$ as $n \to \infty$. Since we can choose $\varepsilon'', \varepsilon''', \epsilon$ small enough so that $\varepsilon''' + \left(\frac{1+\varepsilon}{2} - \varepsilon'' \right)(1 + \epsilon) < 1 - \theta$, we then obtain $0 < n^{-\varepsilon'''} \frac{\lambda^{2k}}{n^2} n^\theta < n^{\varepsilon'''} \frac{\lambda^k}{n} n^\theta \le \frac{n^{\varepsilon''' + \left(\frac{1+\varepsilon}{2} - \varepsilon'' \right)(1+\epsilon) + \theta}}{n} < 1$ for sufficiently large n.

Then w.h.p.,

$$\mathbb{P}(Z^c \mid u_1, u_2 \in \mathcal{C}_{(1)}) < \exp\left(-R\frac{n^{-\varepsilon'''}n^{2\left(\frac{1+\varepsilon}{2}-\varepsilon''\right)(1-\epsilon)}}{2n^2}\left(\frac{\theta}{\log M}\log n - 1 - (1+1-0)\right)\right)$$
$$< \exp\left(-R\frac{n^{-\varepsilon'''}n^{1+\varepsilon+2\left(\varepsilon''\epsilon-\varepsilon''-\epsilon\frac{1+\varepsilon}{2}\right)}}{2n^2}\frac{\theta}{2\log M}\log n\right).$$

Since $\varepsilon'', \varepsilon''', \epsilon$ can be chosen small enough so that $-\varepsilon''' + 2\left(\varepsilon''\epsilon - \varepsilon'' - \epsilon\frac{1+\varepsilon}{2}\right) < \varsigma$ for any $\varsigma > 0$, $R = \Omega\left(\frac{\log M}{\theta \log n} n^{1-\varepsilon+\varsigma}\right)$ is sufficient for the final bound to tend to 0.

B Experimental Setup

In our experiments, we train GNNs to approximate the landmark distances in sparse, undirected, unweighted random graphs. We consider four standard GNN architectures (GCN [Kipf and Welling, 2017], GraphSAGE [Hamilton et al., 2017a], GAT [Veličković et al., 2018], and GIN [Xu et al., 2019a]) with sum aggregation, dropout, and ReLU activations. For each architecture, we test nine models with $\lfloor\sqrt{n}\rfloor$ nodes in the first and last layers and hidden layers varying in depth and width:

- Depth-6: 128-64-32-16, 64-32-16-8, 32-16-8-4
- Depth-5: 128-64-32, 64-32-16, 32-16-8
- Depth-4: 128-64, 64-32, 32-16

The training data for the GNNs are graphs generated by $\mathrm{ER}_n(\lambda/n)$ with $1 < \lambda \ll n$, which ensures sparsity and the existence of a giant component w.h.p. In particular, we consider $\lambda \in \{3, 4, 5, 6\}$ and $n \in \{25, 50, 100, 200, 400, 800, 1600, 3200\}$. Each graph is treated as a batch of nodes with a 200-50-50 train-validation-test split to generate random input signals $\mathbf{X} \in \mathbb{R}^{n\times r}$, where each column one-hot encodes a landmark node. The outputs $\mathbf{Y} \in \mathbb{R}^{n\times r}$ have the same dimensions as the inputs and represent shortest path distances between nodes u and landmarks s, i.e., $[\mathbf{Y}]_{us} = d(u, s)$. Training runs for 1000 epochs with early stopping (100 epochs), MSE loss, Adam optimizer (lr=0.01, weight decay=0.0001), and a cyclic-cosine learning rate schedule (0.001–0.1 for 10 cycles, with default cosine annealing for up to 20 iterations).

All experiments use PyTorch Geometric [Fey and Lenssen, 2019] on a Lambda Vector 1 machine (AMD Ryzen Threadripper PRO 5955WX CPU, 16 cores, 128 GB RAM, 2× NVIDIA RTX 4090 GPUs, no parallel training). Code is available at `https://github.com/ruiz-lab/shortest-path`.

C More Experimental Results

We provide additional results with more detailed explanations offering deeper insights into GNNs and their use for generating landmark embeddings in shortest path approximations. The experiments are divided into three categories: learning the predictive power of GNNs, comparing the performance of the GNN-augmented approach with the vanilla landmark-based algorithm, and evaluating the transferability of both methods to larger random graphs and real-world benchmarks.

C.1 Experiment 1: Learning the GNNs

In the first experiment, we evaluate the ability of trained GNNs to compute end-to-end shortest paths. We consider $n = 50$ and set the GNN depth to be larger than $\lceil \log_\lambda n \rceil$. Figure 2 plots the actual shortest path distances versus those predicted by our selected GNN architectures. Predictions for distances beyond the GNN depth saturate, indicating that GNNs cannot capture longer distances even with depth exceeding the expected path length. As expected, GNNs are not suitable for computing end-to-end shortest path distances, especially on sparser graphs with $\lambda \in \{3, 4\}$, which tend to exhibit longer paths.

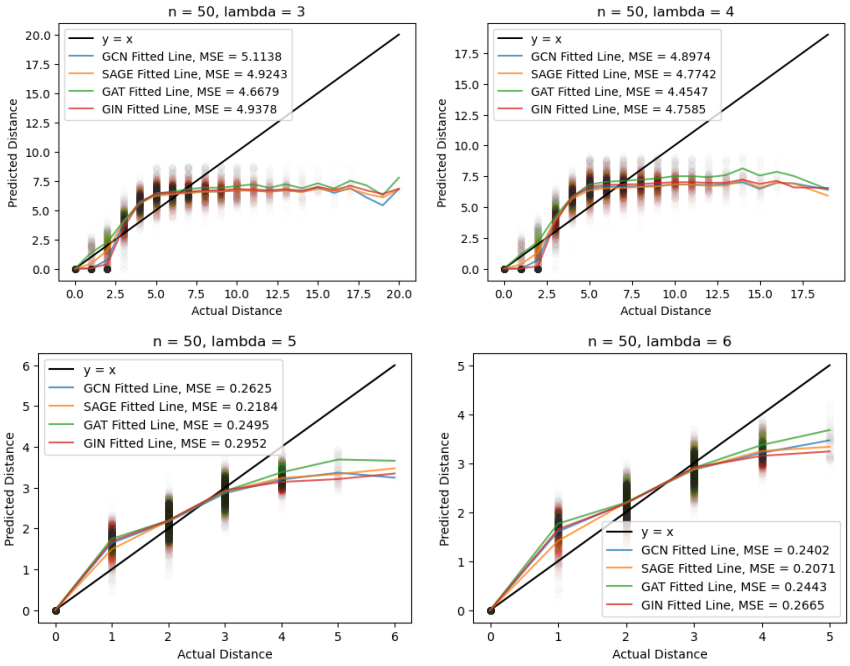

Figure 2: End-to-end shortest path distance predictions from $\lfloor\sqrt{n}\rfloor$-64-32-16-$\lfloor\sqrt{n}\rfloor$ GNNs trained on graphs generated by $\mathrm{ER}_n(\lambda/n)$. The evaluation data consists of graphs from the same model.

C.2 Experiment 2: Comparing BFS-Based and GNN-Based Landmark Embeddings

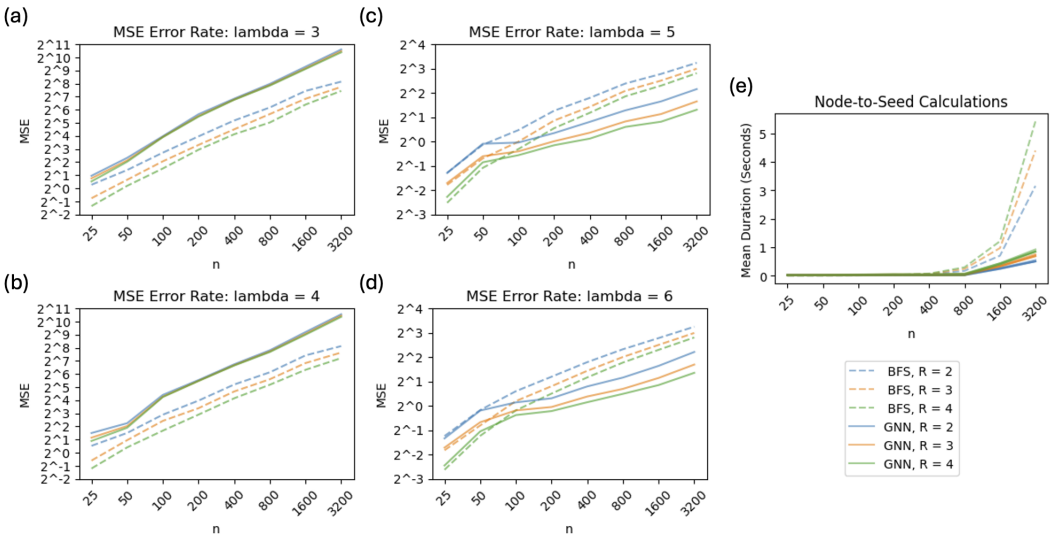

Figure 3: Error rates of BFS-based and GNN-based lower bounds on graphs generated by $\mathrm{ER}_n(\lambda/n)$, with the GNNs trained on graphs from the same model.

In this experiment, we compare the lower bounds (LBs) resulting from BFS-based and GNN-based landmark embeddings against the actual shortest path distances. *Only LBs are compared to ensure a fair evaluation, as computing the upper bounds (UBs) requires storing additional information—namely, the indices of the closest landmarks from the landmark sets to each node. Moreover, unlike in LB computations, the saturation effect inherent in GNNs cannot be mitigated*

in UB computations, making the UB an unreliable metric for shortest path approximation when calculated upon GNN-based landmark distances.

To construct the landmark embeddings, we sample $r+1$ landmark sets $S_0, S_1, \ldots, S_r$ of cardinalities $2^0, 2^1, \ldots, 2^r$ with $r = \lfloor \log n \rfloor$ for R repetitions. In Figure 3(a-d), GNN-based lower bounds underperform the vanilla lower bounds for smaller $\lambda \in \{3, 4\}$, but yield substantial improvements for larger $\lambda \in \{5, 6\}$ across all three tested values of R. Although both λ values are in the supercritical regime ($\lambda > 1$), several factors explain this difference. As shown in Figure 2, the GNN learns poorer landmark embeddings for $\lambda \in \{3, 4\}$, even on small 50-node graphs. Additionally, for large n, graphs are almost surely connected when $\lambda \in \{5, 6\}$ but not when $\lambda \in \{3, 4\}$. Finally, Figure 3(e) illustrates that GNN-based embeddings can be generated faster than BFS-based embeddings, particularly on large graphs as exact local embedding computations via BFS scale poorly with graph size.

C.3 Experiment 3: Transferability

In our last experiment, we investigate whether GNNs trained on small graphs can be transferred to compute landmark embeddings on larger networks for downstream shortest path approximation via LBs. This is motivated by Ruiz et al. [2020] and Ruiz et al. [2023], which show that GNNs are transferable as their outputs converge on convergent graph sequences. This, in turn, allows models trained on smaller graphs to generalize to similar larger graphs.

Here, we focus on $\lambda \in \{5, 6\}$ and train a sequence of eight GNNs on ER graphs ranging from $n = 25$ to $n = 3200$ nodes. These GNNs are then used to generate local node embeddings on graphs from the ER model with the same λ and $n' = 12800$ nodes. Figure 4(a,d) shows the MSE for each instance as the training graph size increases, with flat dashed lines indicating the MSE of BFS-based LBs on the n'-node graph. We observe a steady decrease in MSE as n grows, with GNN-based embeddings matching BFS-based performance when trained on graphs of $n = 100$, which is 128 times smaller than the target graph.

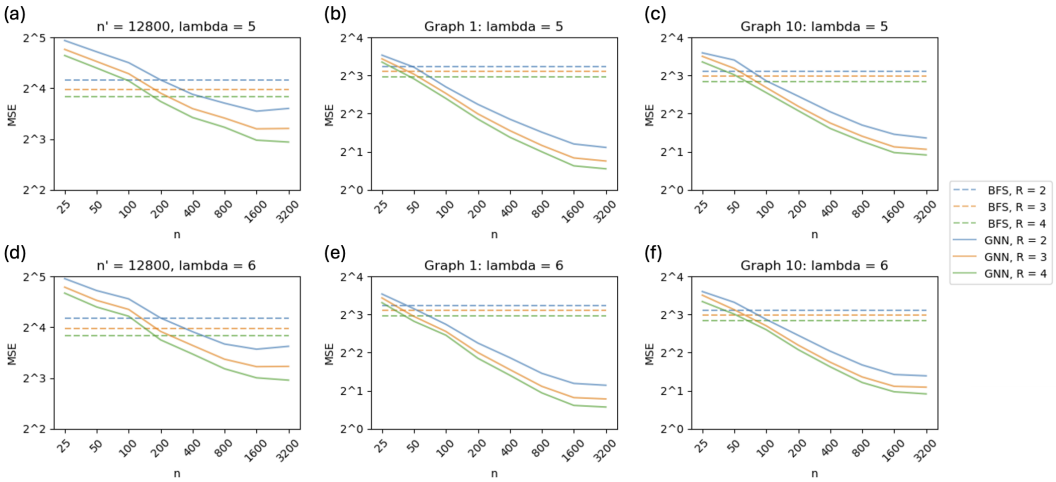

Figure 4: Error rates of BFS-based and GNN-based lower bounds on (a,d) test ER graphs generated by $\mathrm{ER}_{n'}(\lambda/n')$, (b,e) Arxiv COND-MAT collaboration network with 21,364 nodes, and (c,f) GEMSEC company network with 14,113 nodes, with the GNNs trained on graphs from $\mathrm{ER}_n(\lambda/n)$.

When examining the transferability of the same set of GNNs on sixteen real-world networks listed in Table 1, we again observe that MSE improves with training graph size and that GNN-based lower bounds outperform BFS-based lower bounds, even though the landmark embeddings are learned on much smaller graphs (see Figures 4 and 5). This can be explained as random graphs can model real-world networks in certain scenarios, and networks with similar sparsity likely exhibit similar local structures which local message-passing in GNNs can learn with sufficient training.

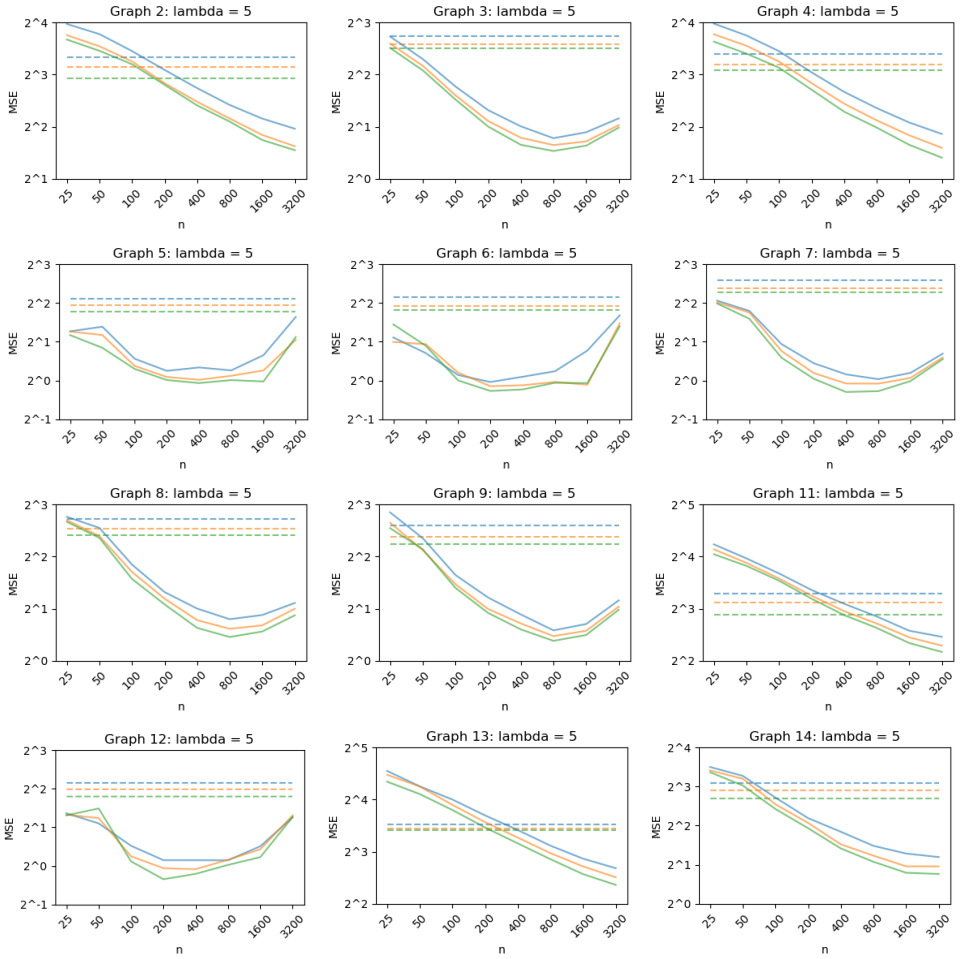

Figure 5: Additional transferability results on real networks, with the GNNs trained on graphs from $\mathrm{ER}_n(\lambda/n)$. Legend is the same as in Figure 4.

Table 1: Details on the largest connected component of selected benchmark networks.

#	Name	Category	# of Nodes	# of Edges
1	Arxiv COND-MAT [Leskovec et al., 2007]	Collaboration Network	21,364	91,315
2	Arxiv GR-QC [Leskovec et al., 2007]	Collaboration Network	4,158	13,425
3	Arxiv HEP-PH [Leskovec et al., 2007]	Collaboration Network	11,204	117,634
4	Arxiv HEP-TH [Leskovec et al., 2007]	Collaboration Network	8,638	24,817
5	Oregon Autonomous System 1 [Leskovec et al., 2005]	Autonomous System	11,174	23,409
6	Oregon Autonomous System 2 [Leskovec et al., 2005]	Autonomous System	11,461	32,730
7	GEMSEC Athletes [Rozemberczki et al., 2019b]	Social Network	13,866	86,858
8	GEMSEC Public Figures [Rozemberczki et al., 2019b]	Social Network	11,565	67,114
9	GEMSEC Politicians [Rozemberczki et al., 2019b]	Social Network	5,908	41,729
10	GEMSEC Companies [Rozemberczki et al., 2019b]	Social Network	14,113	52,310
11	GEMSEC TV Shows [Rozemberczki et al., 2019b]	Social Network	3,892	17,262
12	Twitch-EN [Rozemberczki et al., 2019a]	Social Network	7,126	35,324
13	Deezer Europe [Rozemberczki and Sarkar, 2020]	Social Network	28,281	92,752
14	LastFM Asia [Rozemberczki and Sarkar, 2020]	Social Network	7,624	27,806
15	Brightkite [Rossi and Ahmed, 2015]	Social Network	56,739	212,945
16	ER-AVGDEG10-100K-L2 [Rossi and Ahmed, 2015]	Labeled Network	99,997	499,359