The Storage Complexity of Personalized PageRank and Shortest Path Labeling Schemes

Huacheng Yu*   Hongyang Zhang†

July 3, 2018

Abstract

We introduce a new framework to prove data structure lower bounds in the communication model of labeling schemes, for all pairs personalized PageRank (PPR) and shortest paths. In a labeling scheme, each node is assigned a label vector; To answer a query between two vertices, only the labels of the two vertices are accessed. We show that for any personalized PageRank labeling schemes that can recover PPR values at least $\delta$, the total labeling size is at least $\tilde{\Omega}(n/\sqrt{\delta})$ on sparse Erdős-Rényi random graphs with high probability, where $n$ is the number of vertices in the graph. The lower bounds match the upper bound of Lofgren et al.’16. The techniques can be applied to obtain nearly tight bounds for distance labeling schemes, on sparse random graphs with power law degree distributions.

Our work is motivated by the success of labeling schemes that achieve state of the art performance for both personalized PageRank and shortest path problems in practice. We hope the techniques can also improve the understanding on other data structure problems where labeling schemes are successful for algorithm design, but obtaining tight lower bounds in the cell probe model is technically challenging.

1 Introduction

Personalized PageRank (PPR) and shortest distance are both commonly used local features in graph based recommendation and search systems [51, 31]. Instead of computing from scratch, it is more desirable to create data structures that can quickly answer queries online, by pre-processing the graph. Yet storing the data structures can be expensive, especially for very large graphs.

Recent algorithmic results have evaluated the storage requirements on social and information networks for personalized PageRank and shortest path. The synthesis of the experimental results is that low complexity data structures can be efficiently computed in practice, based on the following meta-design [41, 23, 39]: each node stores a label consisting of a list of (node, value) pairs offline during the pre-processing phase, and the query answering algorithm computes the output by taking a vectorized operation between the two labels online. Besides the simplicity and storage efficiency, such designs can be efficient wrapped with a search index, which provides much faster ranking for personalized search on social graphs [10, 39].

Despite the success found in practice, the limits of existing algorithms are far from well-understood. For personalized PageRank, the only known lower bound is that computing the value between a single pair of vertices requires $\Omega(\sqrt{n})$ time in the worst case with no pre-processing [40], where $n$ is the number of vertices. For shortest path, the seminal work of Gavioli et al. [29]
show that even for graphs with maximum degree 3, any distance labeling scheme requires $\Omega(n^{1.5})$ space for worst case instances. However, social and information networks are typically sparse expander graphs with a heavy tailed degree distribution [17], which are not always similar to the hard instances Gavoli et al. constructed.

In this work we introduce a suite of techniques for proving lower bounds for labeling schemes, closing the gap described above. A labeling scheme defines a communication model, where two vertices compute the output using just their own information/labels. While such a communication model may seem restrictive, the state of the art performance is achieved via labeling schemes for both personalized PageRank [39] and shortest path [4, 23] in practice.

We present a matching lower bound for personalized PageRank to the existing algorithms [38, 39, 40], under the communication model of labeling schemes. Our lower bound is stated in terms of the desired accuracy threshold — if one starts to care about small PPR values, then the lower bound would scale up accordingly. Our techniques also yield nearly tight bounds for shortest path labeling schemes on sparse random graphs with a power law degree distribution. The bounds interpolate between the skewness of the degree distribution and the average distance of the graph, and significantly improve over the worse case bounds via exploiting the expansion property of random graphs. We hope such theoretical analysis may find applications for proving lower bounds on other data structure problems, where the labeling methods have found algorithmic success.

1.1 Results for Personalized PageRank

Let $G = (V, E)$ be an undirected and unweighted graph. Let $n = |V|$ be the number of vertices and $m = |E|$ be the number of edges. Consider a random walk that starts at a vertex $x \in V$ with teleport probability $\alpha \in (0, 1)$. At each step, with probability $\alpha$, the random walk stops; with probability $1 - \alpha$, we jump to a uniformly random neighbor. The personalized PageRank from $x$ to a vertex $y \in V$, denoted by $\pi(x, y)$, is the probability that the random walk starting at $x$ stops at $y$. We say that an algorithm $f : V \times V \rightarrow \mathbb{R}^+$ is $(\varepsilon, \delta)$-accurate if for any $x, y \in V$:

a) if $\pi(x, y) \geq \delta$, then $\frac{1}{1+\varepsilon} \cdot \pi(x, y) \leq f(x, y) \leq (1 + \varepsilon) \cdot \pi(x, y)$.

b) if $\pi(x, y) < \delta$, then $f(x, y) \leq (1 + \varepsilon) \cdot \delta$.

Following Lofgren et al. [40], $\delta$ is assumed to be larger than $c/n$ for a certain value $c > 1$, to capture personalized PageRank values which are larger than average. It’s not hard to see that simply by sampling random walks, we obtain an $(\varepsilon, \delta)$-accurate data structure of total size $\tilde{O}(\frac{n}{\varepsilon^2 \cdot \delta})$. By combining random walks and linear algebraic methods [7], Lofgren et al. [39, 40] improved the storage complexity over the above baseline to $\tilde{O}(\frac{1}{\alpha^2} \sqrt{\frac{mn}{\delta}})$, when $\delta \geq m/n$. In this upper bound, each vertex $x \in V$ stores a set of random walks and local graph statistics as a label vector $l(x)$.

To obtain the personalized PageRank between $x$ and $y \in V$, the query answering algorithm simply computes the dot product between $l(x)$ and $l(y)$. Our main result is a matching lower bound to the above algorithm in the communication model of labeling schemes for sparse graphs.

**Theorem 1** (informal). Let $G = (V, E)$ be an Erdős-Rényi random graph where an edge is sampled independently between every vertex pair with probability $p = \log^4 n/n$, where $n = |V|$ is the number of vertices, Let $\alpha$ be the teleport probability for the random walk on $G$. With high probability over the randomness of $G$, any $(\varepsilon, \delta)$-accurate labeling data structure will output labels of total length $\Omega(n/\sqrt{\delta})$, for $\varepsilon \leq \text{poly log}(n)$ and $\frac{\log n}{\log np} \leq \frac{1}{\alpha} \leq \text{poly log } n$.

1Suppose that with probability $\alpha$, the random walk teleports to $x$ instead. The stationary distribution of this random walk is the personalized PageRank vector of $x$. 2
We remark that for graphs whose number of edges $m \gg n$, our results also imply a lower bound of $\tilde{\Omega}(\sqrt{nm}/\delta)$, under technical conditions on $\delta$. See Section 4 for details. It is worth mentioning that our results apply in the regime when $1/\alpha \lesssim \text{polylog}(n)$. This setting is critical for personalized PageRank to capture enough local graph structures [31].

Our result crucially exploits the labeling scheme communication model. In the more general cell-probe model, indeed it has been notoriously difficult to prove super-logarithmic query time lower bounds [43, 42, 34], even for non-adaptive static data structures. The communication model of labeling schemes can be viewed as a subclass of the non-adaptive data structures, where upon each query $(x,y)$, the algorithm only accesses a fixed set of memory words of form $l(x) \cup l(y)$, i.e. it only accesses two sets of memory words that depend only on the two vertices of the query. The query time, which is proportional to $|l(x)| + |l(y)|$, corresponds to the per-vertex label size in the labeling query model. Thus, within the labeling scheme communication model, we show that it becomes possible to obtain super-logarithmic query time data structure lower bounds.

Insights of the analysis: We show that each pair of labels convey a certain amount of information entropy, because the pair of labels can determine the PPR value accurately, which has certain entropy on Erdős-Rényi random graphs. To augment the entropy obtained from a single pair of vertices, we identify a maximal set of vertices, such that their PPR values are almost pairwise independent. We discover a tight connection between PPR and shortest distance on Erdős-Rényi graphs. That is, given that the random walk starting at $x$ stops at $y$, the most likely route is to walk directly along the shortest path from $x$ to $y$. While PPR is a weighted combination over different paths, the connection to shortest path allows us to extract edge information explicitly from the graph.

To obtain the “pairwise independence” of PPR values between a sufficiently large set of vertices, we describe an iterative process to grow the local neighborhood of each vertex. At every iteration of the iterative process, we grow the neighborhood of every vertex up to a certain level $d$, on the subgraph which has not been explored yet. Constructed in this way, the $d$-th level sets of every vertex are disjoint from each other. We show that based on the estimated PPR values, we can infer whether the $d$-th level sets are connected by any edge or not.

Interestingly, our analysis seems to be different from previous work [29] at a conceptual level. In the $\Omega(n^{1.5})$ lower bound obtained for distance labelings on bounded degree graphs, the hard instance consists of a set of graphs whose distance labels must all be different from each other. Hence the lower bound is obtained via a counting argument. Whereas by an entropic argument in our analysis, we show that the lower bound holds even for an average case instance from the Erdős-Rényi graph distribution.

1.2 Results for Shortest Path on Power Law Graphs

Our lower bound techniques can be extended to shortest paths labeling schemes on more general random graphs. Along the way, we also present upper bounds to complete the picture. We describe the setup and main results below. The details are deferred to Section 5.

We will focus on the Chung-Lu model, which generalizes Erdős-Rényi random graphs to general degree distributions. In the Chung-Lu model [15], each vertex $x$ has a weight $p_x$, which is the expected degree of $x$. For every pair of vertices $x$ and $y$, there is an undirected and unweighted edge between them with probability proportional to $p_x \cdot p_y$, independent of other edges. We assume that the degree distribution follows a power law distribution with exponent $\beta$: for every vertex $x$, we can extend to other random graph models as well (see the discussion in Section 5 for details).
the probability that $x$ has weight $p$ is proportional to $p^{-β}$. We point out that when the degree distribution has bounded variance, the techniques from Theorem 1 already imply the optimal bound for shortest path labeling schemes. See Section 3 for details. Hence in the following we focus on degree distributions with high variance. We characterize the storage complexity of exact distance labeling schemes, which can answer distance queries correctly for all pairs of vertices, when the degree distribution has high variance.

**Theorem 2.** Let $\mathcal{G}^n(p)$ be a sparse random power law graph model with average degree $\nu > 1$ and exponent $2 < β ≤ 3$. For a random graph $G = (V, E)$ drawn from $\mathcal{G}^n(p)$, we have that with high probability over the randomness of $G$, there exists an exact distance labeling scheme $F$ where the label size of every $x \in V$ are all bounded by $\tilde{O}(n^{1-\min(\frac{1}{β-1}, \frac{1}{4-β})})$.

Secondly, any exact distance labeling scheme will output a labeling of total length at least $\Omega(n^{5-β-o(1)})$ for $G$.

The analysis uses the fact that there are lots of very high degree vertices in the graph. However, when $β$ gets close to 2, even though the degree distribution gets more skewed towards high degree vertices, the storage complexity increases again to $Θ(n^{1.5})$. The reason is that the average distance also matters. If there exists lots of short paths that can not be compressed, then the storage complexity will increase. For approximation schemes, the amount of storage needed in the upper bound is significantly less. We also obtain a $(+1)$-stretch scheme where the label length for every vertex is $\tilde{O}(n^{(β-2)/(β-1)})$, and a $(+2)$-stretch scheme where the label length for every vertex is $\tilde{O}(n^{β/2-1})$. See Figure 1 for a summary of the results.

The lower bound builds on the insights from Erdős-Rényi graphs. However, the neighborhood growth has very high variance. To overcome the issue, we carefully construct a set of “good” path, so that with high probability, a vertex will follow a good path during the neighborhood growth. See Section 5 for details. The lower bound is nearly tight when $β$ is close to 2, and has a small gap when $β < 2.5$. It would be interesting to close the gap when $2.5 < β < 3$. We conjecture that the storage complexity should be $Ω(n^{1.5})$ when $β$ is close to 3.

**Organization:** The rest of the paper is organized as follows. In Section 2 we give a preliminary on personalized PageRank and Chung-Lu model. In Section 3 we introduce our main technical
contribution by illustrating a shortest path lower bound for $\beta > 3$. In Section 4 we prove the lower bound for personalized PageRank labeling schemes. In Section 5 we present shortest path labeling schemes for $2 < \beta < 3$. We also evaluate the performance of our algorithm on real world graphs in Appendix B. Appendix A includes missing proofs of the upper bound for $2 < \beta < 3$ and Appendix C describes relevant tools from random graph theory.

Notations: For a vertex $x$, Denote by $d_x$ the degree of $x$. For a set of vertices $S$, let $d_S = \sum_{x \in S} d_x$ denote the sum of their degrees. Denote by $x \sim y$ if there is an edge between $x, y$. For two disjoint sets $S$ and $T$, denote by $S \sim T$ if there exists an edge between $S$ and $T$, and $S \not\sim T$ if there does not exist any edge between $S$ and $T$. For a graph $G$, let dist$_G(x, y)$ denote the distance of $x$ and $y$ in $G$. When there is no ambiguity, we drop the subscript $G$ and simply denote by dist$(x, y)$ the distance between $x$ and $y$.

We use the notation $O(\cdot)$ to hide absolute multiplicative constants. Similarly, $a \lesssim b$ means that there exists an absolute constant $C > 0$ such that $a \leq Cb$. We use $\tilde{O}(\cdot)$ to hide poly-logarithmic factors.

2 Preliminaries and Related Work

Recall that $G = (V, E)$ denotes an undirected graph. Let $A$ be the adjacency matrix of $G$ and $D$ be the diagonal matrix with the degrees of every vertex in $V$. There is an equivalent algebraic definition for the personalized PageRank from $x$ to $y$:

$$\pi(x, y) = e_x^T (\text{Id} - (1 - \alpha)AD^{-1})^{-1}e_y,$$

where $\alpha$ is the teleport probability, Id is the identity matrix, $e_x$ and $e_y$ are indicator vectors.

Chung-Lu model: Recall that $p_x > 0$ denotes the weight of every vertex $x \in V$. Given the weight vector $p$ over $V$, the Chung-Lu model defines a probability distribution over the set of all graphs $\mathcal{G}^n$. Let vol$(S) = \sum_{x \in S} p_x$ denote the volume of $S$. And let vol$_2(S) := \sum_{x \in S} p_x^2$ denote the second moment of $S$. Each edge $(x, y)$ is chosen independently with probability

$$\Pr[x \sim y] = \min \left\{ \frac{p_x \cdot p_y}{\text{vol}(V)}, 1 \right\}.$$ 

Thus, $p_x$ is approximately the expected degree of $x$, and vol$(V)$ is approximately the expected number of edges (multiplied by two). Let $\mathcal{G}^n(p)$ denote such a probability distribution over $\mathcal{G}^n$, and $G \in \mathcal{G}^n(p)$ denote a sample from the distribution. The following proposition bounds the probability that two sets connect.

Proposition 3. Let $G = (V, E) \in \mathcal{G}^n(p)$ be a random graph. For any two disjoint set of vertices $S$ and $T$,

$$1 - \exp \left( -\frac{\text{vol}(S)\text{vol}(T)}{\text{vol}(V)} \right) \leq \Pr[S \sim T] \leq \frac{\text{vol}(S)\text{vol}(T)}{\text{vol}(V)}.$$ 

In particular, when $\text{vol}(S)\text{vol}(T) \leq o(\text{vol}(V))$, we have that $\Pr[S \sim T] = \Theta \left( \frac{\text{vol}(S)\text{vol}(T)}{\text{vol}(V)} \right)$. 

5
Proof. The proof follows straightforwardly from definition.

\[
\Pr[S \sim T] = 1 - \prod_{x \in S} \prod_{y \in T} \left(1 - \min\left(\frac{p_x p_y}{\text{vol}(V)}, 1\right)\right) \leq 1 - \left(1 - \sum_{x \in S} \sum_{y \in T} \min\left(\frac{p_x p_y}{\text{vol}(V)}, 1\right)\right)
\]

\[
\leq 1 - \left(1 - \sum_{x \in S} \sum_{y \in T} \frac{p_x p_y}{\text{vol}(V)}\right) = \frac{p(S) p(T)}{\text{vol}(V)}.
\]

\[
\Pr[S \sim T] = \prod_{x \in S} \prod_{y \in T} \left(1 - \min\left(\frac{p_x p_y}{\text{vol}(V)}, 1\right)\right) \leq \exp\left(-\sum_{x \in S} \sum_{y \in T} \min\left(\frac{p_x p_y}{\text{vol}(V)}, 1\right)\right)
\]

\[
\leq \exp(-\frac{p(S) p(T)}{\text{vol}(V)}).
\]

\[
\square
\]

Random power law graph: Let \( f : [x_{\text{min}}, \infty) \to \mathbb{R} \) denote the probability density function of a power law distribution with exponent \( \beta > 1 \), i.e. \( f(x) = Z x^{-\beta} \), where \( Z = (\beta - 1) \cdot x_{\text{min}}^{\beta - 1} \). The expectation of \( f(\cdot) \) exists when \( \beta > 2 \). The second moment is finite, only when \( \beta > 3 \).

In the random power law graph model, the weight of each vertex \( x \) is drawn independently from a power law distribution (with the same mean \( \nu \) and exponent \( \beta \)). Given the weight vector \( p \), we then sample a random graph according to the Chung-Lu model.

It is known that if \( \nu > 1 \), then almost surely a random graph \( G \) with weight \( p \) has a unique giant component (see e.g. Chung and Lu [16]). In this paper, we will assume that the average degree \( \nu \) is a constant greater than 1. When the degree distribution has bounded variance, we can bound the neighborhood growth rates as follows. The result is standard (see e.g. Chung and Lu [16]) – we will present a proof in Appendix C.1 for the completeness of this paper.

**Proposition 4 (Growth rates for \( \beta > 3 \)).** Let \( \mathcal{G}^n(p) \) be a random graph model with weight sequence \( p \) satisfying the following properties:

- \( \text{vol}(V) = (1 + o(1)) \nu \cdot n \) for some constant \( \nu \);
- \( \text{vol}_2(V) = (1 + o(1)) \omega \cdot n \) for some constant \( \omega \);
- \( \text{vol}_{2+\gamma}(V) = \tau \cdot n \) for some positive constant \( \gamma < 1/2 \) and \( \tau \), where \( \text{vol}_{2+\gamma}(S) := \sum_{x \in S} p_x^{2+\gamma} \);
- The growth rate \( r = \frac{\text{vol}_2(V)}{\text{vol}(V)} \) is bounded away from 1 (\( \nu > \omega \)).

Then for any vertex \( x \) with a constant weight, the set of vertices \( \Gamma_k(x) \) at distance exactly \( k \) from \( x \) satisfy that

1. \( \mathbb{E} [\text{vol}(\Gamma_k(x))] = O (r^k) \) for every \( k \leq \log_r n \);
2. \( \Pr[\text{vol}(\Gamma_k(x)) \geq \Omega (r^k)] \geq \Omega(1) \) for every positive integer \( k \leq \frac{1}{2} \log_r n \).

As a corollary, we have that \( \Pr[\text{dist}(x,y) \leq k+1] \leq O(r^k/n) \) for every \( k \leq \frac{1}{2} \log_r n \), where \( y \) is any vertex with constant weight.

\(^4\text{If} \nu < 1, \text{almost surely all connected components have at most} O(\log n) \text{ vertices.}\)
2.1 Related work

**Landmark based Labelings:** There is a rich history of study on how to preprocess a graph for answering shortest path queries \([6, 18, 13, 49]\). A commonly used algorithm is landmark based labelings \([4, 12, 19, 22, 23]\), also known as 2-hop covers \([20]\) or hub labeling \([1]\). The empirical results of Akiba et al. \([4]\) and Delling et al. \([23]\) found that only a few hundred landmarks per vertex suffices to recover all-pairs distances exactly, in a large collection of social, Web, and computer networks with tens of millions of edges. The idea is to find central landmarks that lie on the shortest paths of many sources and destinations. In a landmark based labeling, every vertex stores a set of landmarks as well as its distance to each landmark. To answer a distance query \(\text{dist}(x, y)\), we simply find a common landmark \(z\) in the landmark sets of \(x\) and \(y\) to minimize the sum of distances \(\text{dist}(x, z) + \text{dist}(z, y)\). It is NP-hard to compute the optimal landmark based labeling (or 2-hop cover), and a \(\log n\)-approximation can be obtained via a greedy algorithm \([20]\). See also the references \([8, 9, 24, 30]\) for a line of followup work. Another closely related line of work is approximate distance oracle \([3, 5, 21, 44, 45, 46, 52]\). We refer the reader to the excellent survey \([47]\) for further reading.

**Random graph models:** Existing models for social and information networks build on random graphs with a fixed degree distribution \([25, 16, 50]\). Informally, we assume that the degree sequence of our graph is given, and then we draw a “uniform” sample from graphs that have the same or very similar degree sequences. Random graphs capture the small world phenomenon \([16]\), because the average distance grows logarithmically in the number of vertices. They serve as a basic block to richer models with more realistic features, e.g. community structures \([33]\), shrinking diameters in temporal graphs \([37]\). It has been empirically observed that many social and information networks have a heavy-tailed degree distribution \([17, 26]\) — concretely, the number of vertices whose degree is \(x\), is proportional to \(x^{-\beta}\).

Previous work of Chen et al. \([14]\) presented a 3-approximate labeling scheme requiring storage \(\tilde{O}(n^{(\beta-2)/(\beta-3)})\) per vertex, on random power law graphs with \(2 < \beta < 3\). Our (+2)-stretch result improves upon this scheme in the amount of storage needed per vertex for \(2 < \beta < 2.5\), with a strictly better accuracy guarantee. Another related line of work considers compact routing schemes on random graphs. Enachescu et al. \([27]\) presented a 2-approximate compact routing scheme using space \(O(n^{1.75})\) on Erdős-Renyi graphs, and Gavoille et al. \([28]\) obtained a 5-approximate compact routing scheme on random power law graphs. Other existing mathematical models on special families of graphs related to distance queries include road networks \([2]\), planar graphs \([41]\) and graphs with doubling dimension \([32]\). However none of them can capture the expansion properties that have been observed on sub-networks of real-world social networks \([37]\). Apart from the Chung-Lu model and the configuration model that we have mentioned, the preferential attachment graph is also well-understood \([25]\). It would be interesting to see if our results extend to preferential attachment graphs as well. The Kronecker model \([35]\) allows a richer set of features by extending previous random graph models, however its mathematical properties are not as well-understood as the other three models.

3 Warm Up and Shortest Paths Lower Bounds for \(\beta > 3\)

In this section, we illustrate our main ideas by presenting a lower bound for labeling schemes that can estimate all pairs distances up to \(K \leq \log n / \log r\)\(^5\), where \(r\) is equal to \(
\frac{\text{vol}(V)}{\text{vol}(\text{vol})}\). More formally,

\(^5\)Note that the average distance of \(G\) is \(\log n / \log r\) (see e.g. Bollobás \([11]\)).
we say that a labeling scheme is $K$-accurate if for any $x, y \in V$:

a) if $\text{dist}(x, y) \leq K$, then the labeling scheme returns the exact distance $\text{dist}(x, y)$.

b) if $\text{dist}(x, y) > K$, then the labeling scheme returns “$\text{dist}(x, y) > K$”.

For any integer $1 \leq i \leq n - 1$, let $\Gamma_i(x) = \{y \in V : \text{dist}(x, y) = i\}$ denote the set of vertices whose distance from $x$ is equal to $i$. And let $N_i(x) = \{y \in V : \text{dist}(x, y) \leq i\}$ denote the set of vertices whose distance from $x$ is at most $i$. Let $d$ be an integer smaller than $K/2$.  

We may assume without loss of generality for every $x$, the label of $x$ stores the distances between $x$ and all vertices in $N_d(x)$. This is because the lower bound we are aiming at is larger than the size of $N_d(x)$, we can always afford to store them. From the labels of $x, y$, either we see a non-empty intersection between $N_d(x)$ and $N_d(y)$, which determines their distance; or the two sets are disjoint, in which case we are certain that $\text{dist}(x, y) \geq 2d + 1$. In a random graph, the event that $\text{dist}(x, y) > 2d + 1$, conditioned on $\text{dist}(x, y) \geq 2d + 1$ and $N_d(x)$ and $N_d(y)$ are disjoint, happens with probability

$$\Theta\left(\frac{\text{vol}(\Gamma_d(x)) \cdot \text{vol}(\Gamma_d(y))}{\text{vol}(V)}\right),$$

by Proposition 3 assuming that $\text{vol}(\Gamma_d(x))\text{vol}(\Gamma_d(y)) \leq o(\text{vol}(V))$. Note that this probability gives us a lower bound on the entropy of the event $1_{\text{dist}(x,y) > 2d+1}$. Since the labels of $x$ and $y$ determine their distance, if we can find a large number of pairwise independent pairs $(x, y)$ such that the entropy of $1_{\text{dist}(x,y) > 2d+1}$ is large (e.g. $1/\text{poly log}(n)$ suffices), then we obtain a lower bound on the total labeling size.

Our discussion so far suggests the following three step proof plan.

a) Pick a parameter $d$ and a maximal set of vertices $S$, such that by “growing” the local neighborhood of $S$ up to $d$, $N_d(x), N_d(y)$ are disjoint/independent and $\Gamma_d(x), \Gamma_d(x)$ have large volume, for certain pairs of $x, y \in S$.

b) Use the labels of $S$ to infer whether there are edges between $\Gamma_d(x)$ and $\Gamma_d(y)$, for certain pairs of $x, y$. Obtain a lower bound on the total label length of $S$ via entropic arguments.

c) Partition the graph into disjoint groups of size $|S|$. Apply step b) for each group.

Clearly, given any two vertices, their neighborhood growth are correlated with each other. However, one would expect that the correlation is small, so long as the volume of the neighborhood has not reached $\Theta(\sqrt{n})$. To leverage this observation, We describe an iterative process to grow the neighborhood of $S$ up to distance $d$. For simplicity, we assume that $S = \{x_1, x_2, \ldots\}$ only consists of vertices whose weight are all within $[\nu, 2\nu]$. The motivation is to find disjoint sets $L(x_i)$ for each $x_i$, such that $L(x_i)$ is almost as large as $\Gamma_d(x_i)$, and if $\text{dist}(x_i, x_j) > 2d + 1$, then there is no edge between $L(x_i)$ and $L(x_j)$.

The iterative process: Denote by $G_1 = (V_1, E_1)$, where $V_1 = V$ and $E_1 = E$. For any $i \geq 1$, define $T(x_i)$ to be the set of of vertices in $G_i$ whose distance is at most $d$ from $x_i$. Define $L(x_i)$ to be the set of vertices in $G_i$ whose distance is equal to $d$ from $x_i$. More formally,

$$T(x_i) := \begin{cases} \{y : \text{dist}_{G_i}(x_i, y) \leq d\}, & \text{if } x_i \in V_i \\ \emptyset, & \text{otherwise.} \end{cases}$$

$$L(x_i) := \{y \in T(x_i) : \text{dist}_{G_i}(x_i, y) = d\}$$

---

6We assume that $K$ is odd without loss of generality.
We then define $F_i = F_{i-1} \cup T(x_i)$ ($F_0 := \emptyset$ by default). Denote by $G_i$ the induced subgraph of $G$ on the remaining vertices $V_{i+1} = V \setminus F_i$.

We note that in the above iterative process, the neighborhood growth of $x_i$ only depends on the degree sequence of $V_i$. We show that under certain conditions, with high probability, a constant fraction of vertices $x \in S$ satisfy that $\text{vol}(L(x)) \geq \Omega(r^d)$.

**Lemma 5 (Martingale inequality).** Let $r = \frac{\text{vol}(V)}{\text{vol}(V)}$. Let $d$ be an integer and $S \subseteq V$ be a set of vertices whose weight are all within $[\nu, 2\nu]$ and $|S| \leq o\left(\frac{\text{vol}(V)}{r^d}\right)$. Assume that

1. $\Pr[\text{vol}(L(x)) \geq \nu \cdot r^d \mid \text{vol}(F_{i-1}) \leq |S| \nu \cdot r^d \log n, x_i \in V_i] \geq \Omega(1)$, for all $1 \leq i \leq |S|$;
2. $\mathbb{E}[\text{vol}(N_d(x))] \lesssim \nu \cdot r^d$, for all $x \in S$;
3. $\Pr[\text{dist}(x, y) \leq d] \lesssim \nu^2 r^{d-1} \frac{\text{vol}(V)}{V}$, for all $x, y \in S$.

Then with high probability, at least $c_1 |S|$ vertices $x \in S$ satisfy that $\text{vol}(L(x)) \geq \Omega(\nu r^d)$, for a certain fixed constant $c_1$.

**Proof.** Consider the following random variable, for any $1 \leq i \leq |S|$.

$$X_i := \begin{cases} 1 & \text{if } x_i \not\in V_i, \text{ or } \text{vol}(F_{i-1}) > |S| \nu \cdot r^d \log n, \text{or } \text{vol}(L(x_i)) \geq \Omega(r^d) \\ 0 & \text{otherwise.} \end{cases}$$

We have $\Pr[X_i = 1 \mid X_1, \ldots, X_{i-1}] \geq \Omega(1)$ by Assumption i). Thus by Azuma-Hoeffding inequality, $\sum_{i=1}^{|S|} X_i \geq \Omega(|S|)$ with high probability. We will show below that the contributions to $\sum_{i=1}^{|S|} X_i$ from the first two predicates is $o(|S|)$. Hence by taking union bound, we obtain the desired conclusion.

First, we show that the number of $x_i$ such that $x_i \not\in V_i$ is $o(|S|)$ with high probability. Note that $x_i \not\in V_i$ implies that there exists some vertex $j < i$ such that $\text{dist}(x_i, x_j) \leq d$. On the other hand, for any two vertices $x, y \in S$, $\Pr[\text{dist}(x, y) \leq d] \leq O(\nu^2 \cdot r^{d-1} / \text{vol}(V))$, by Assumption iii). Hence, the expected number of vertex pairs in $S$ whose distance is at most $d$, is $O(|S|^2 \nu^2 r^{d-1} / \text{vol}(V)) \leq o(|S|)$, by the assumption on the size of $S$. By Markov’s inequality, with high probability only $o(|S|)$ vertex pairs have distance at most $d$ in $S$. Hence there exists at most $o(|S|)$ $i$’s such that $x_i \not\in V_i$.

Secondly, for all $1 \leq i \leq |S|$, $\text{vol}(F_i) \leq |S| \nu \cdot r^d \log n$ with high probability. This is because the set of vertices $T_i$ is a subset of $N_d(x_i)$, the set of vertices within distance $d$ to $x_i$ on $G$. Thus, by Assumption ii), we have

$$\mathbb{E}[\text{vol}(T_i)] \leq \mathbb{E}[\text{vol}(N_d(x_i))] \leq O(\nu r^d).$$

And the expected volume of $F_i$ is at most

$$O(i \cdot \nu \cdot r^d) \lesssim |S| \nu \cdot r^d,$$

Hence by Markov’s inequality, the probability that $\text{vol}(N_d(S)) > |S| \nu \cdot r^d \log n$ is at most $\log^{-1} n$. This proves the lemma.

For the rest of this section, we will show how to implement the above proof plan technically, for shortest path labeling on random graphs with $\beta > 3$. To obtain the lower bound for personalized PageRank labeling, we need to show in step b) that PPR values can imply about distances also. And for shortest path labeling on random graphs with $2 < \beta < 3$, we need to deal with the fact that the local neighborhood growth has high variance in step a). We refer the reader to section 4 and section 5 for details.
3.1 The $\beta > 3$ case

In this section, we present the following result.

**Theorem 6.** Let $\mathcal{G}(p)$ be a random power law graph model with average degree $\nu > 1$ and exponent $\beta > 3$. Let $r = \frac{\text{vol}_2(V)}{\text{vol}(V)}$ and $\log \log n \lesssim K \lesssim \log n$ be a fixed integer. For a random graph $G = (V, E)$ drawn from $\mathcal{G}(p)$, we have that any $K$-accurate labeling scheme will almost surely output a labeling whose total length is $\Omega(K^{2/3})$.

We first introduce the following proposition for growing the neighborhood of vertices.

**Proposition 7** (Iterative neighborhood growth). Let $c = (3 + 1/\gamma) \log_r \log n$ and $d = K/2 - c$. Let $S$ be a set of $n/r^{K/2}$ vertices whose weight are all within $[\nu, 2\nu]$. With high probability, at least $c_1 |S|$ vertices in $S$ satisfy that $\text{vol}(L(x)) \geq \Omega(r^d)$.

**Proof.** It’s easy to verify that $|S| \cdot \nu^2 r^{d-1} \leq n/r^c \leq o(n)$. It suffices to verify the assumptions required in Lemma 3. Note that Assumption ii) and iii) simply follows from Proposition 4. Hence it suffices to verify Assumption i). Note that the subgraph $G_i$ can be viewed as a random graph sampled from Chung-Lu model over $V_i$. By setting

$$p^{(i)}_y = p_y \cdot \left(1 - \frac{\text{vol}(F_{i-1})}{\text{vol}(V)}\right), \quad \forall y \in V_i,$$

we have that $\forall y, z \in V_i$

$$\Pr[y \sim z] = \frac{p_y \cdot p_z}{\text{vol}(V)} = \frac{p_y^{(i)} \cdot p_z^{(i)}}{1 - \frac{\text{vol}(F_{i-1})}{\text{vol}(V)}} \cdot \text{vol}(V_i) = \frac{p_y^{(i)} \cdot p_z^{(i)}}{\sum_{x \in V_i} p_x^{(i)}}.$$

Hence we see that $G_i$ is equivalent to a random graph drawn from degree sequence $p^i$. Denote by $r_i := \frac{\text{vol}_2(V_i)}{\text{vol}(V)}$ the growth rate on $G_i$. When $\text{vol}(F_{i-1}) \leq n/r^{2+1/\gamma} n$, by Hölder’s inequality,

$$\text{vol}_2(F_{i-1}) \leq \text{vol}(F_{i-1})^{\frac{1}{1+\gamma}} \cdot O(n^{\frac{1}{1+\gamma}}) \leq o(n/\log n).$$

by straightforward calculation. Hence $r_i$ is a constant strictly greater than 1. By Proposition 4, with constant probability $\text{vol}(L_i) \geq \Omega(r_i^d) \geq \Omega(r^d)$, because

$$\left(\frac{\text{vol}(V)}{\text{vol}(V_i)}\right)^d \lesssim (1 + \log^{-2-\frac{1}{\gamma}} n)^{O(\log n)} \lesssim 1 + \log^{-1-\frac{1}{\gamma}} n.$$

Since the vertices at distance $d$ from $x_i$ in $G_i$ is exactly $L_i$, we have verified that Assumption i) is correct. \hfill $\Box$

Now we are ready to prove Theorem 6.

**Proof of Theorem 6.** We know that there are $\Theta(n)$ vertices whose weights are between $[\nu, 2\nu]$, by an averaging argument. Divide them into groups of size $n/r^{K/2}$. Clearly, there are $\Theta(r^{K/2})$ disjoint groups. Denote by $c_2$ a small fixed value (e.g. $1/\log \log n$ suffices). We will argue that for each group $S$,

$$\Pr[\text{The total label length of } S \leq c_2 \cdot r^{-2c}n] \leq o(1).$$  (1)
Hence by Markov’s inequality, except for \( o(n/r^{K/2}) \) groups, all the other groups will have label size at least \( \tilde{\Omega}(n) \) (recall that \( c = (3 + 1/\gamma) \log_r \log n \)). For the rest of the proof, we focus on an individual group \( S \).

Given the labels of \( S \), we can recover the pairwise distances less than \( K \) for all vertex pairs in \( S \). Let \( \text{dist}_S : S \cdot S \rightarrow \mathbb{N} \) denote the distance function restricted to all pairs in \( S \). Consider the following two cases:

a) \( \exists c_1^2 \cdot |S|^2 / 4 \) pairs \((x_i, x_j)\) such that \( \text{dist}_S(x_i, x_j) \leq 2d + 1 \). By Lemma \([4]\) we know that \( \Pr[\text{dist}(x_i, x_j) \leq 2d + 1] = O(r^{2d}/n) \), for any \( x_i, x_j \in S \). Hence the expected number of pairs with distance at most \( 2d + 1 \) in \( S \), is at most \( O(|S|^2 \cdot r^{2d}/n) \lesssim r^{-2c} \). Hence by Markov’s inequality, the probability that a random graph induces any such distance function is \( o(1) \).

b) The number of pairs such that \( \text{dist}_S(x_i, x_j) \leq 2d + 1 \) is at most \( c_1^2 \cdot |S|^2 / 4 \) in \( S \). Let

\[
A = \{(x, y) \in S \cdot S \mid \text{dist}(x, y) > 2d + 1, \text{ and } \text{vol}(L(x)), \text{vol}(L(y)) \geq \Omega(r^d)\}.
\]

By Lemma \([5]\) the size of \( A \) is at least

\[
\left( \frac{c_1 |S|}{2} \right) - c_1^2 |S|^2 / 4 \geq c_1^2 |S|^2 / 5.
\]

For any \((x, y) \in A\), \( L(x) \) and \( L(y) \) are clearly disjoint. Conditional on \( \{T(x)\} \) for all \( x \in S \), the probability of the existences of edges between \( L_i \) and \( L_j \) are unaffected.

\[
\Pr \left[ \text{dist}_S(x, y) > 2d + 1, \forall (x, y) \in A \mid \{T_i\}_{i=1}^{\lvert S \rvert} \right]
\]

\[\leq \prod_{(x, y) \in A} \Pr \left[ L(x) \not\subseteq L(y) \mid L(x) \cap L(y) = \emptyset, \text{ and } \text{vol}(L(x)), \text{vol}(L(y)) \geq \Omega(r^d) \right]\]

\[\leq \prod_{(x, y) \in A} \exp \left( - \frac{\text{vol}(L(x)) \text{vol}(L(y))}{\text{vol}(V)} \right)\]

\[\leq \exp \left( - \Omega \left( \frac{2d}{n} \right) c_1^2 |S|^2 / 5 \right)\]

\[\leq \exp(-\Omega(r^{-2c}n)).\]

Note that the number of labeling of size less than \( c_2 \cdot r^{-2c} n \) is at most \( 2^{c_2 \cdot r^{-2c} n} \). Therefore by union bound, the probability that the total label size of \( |S| \) is at most \( c_2 \cdot r^{-2c} n \) is at most:

\[
2^{c_2 \cdot r^{-2c} n} \cdot \exp(-\Omega(r^{-2c}n)) \leq o(1).
\]

By taking a union bound over cases a) and b), we have shown that Equation \([1]\) is true. Hence the proof is complete.

\[\square\]

**Remark.** It’s not hard to obtain a matching upper bound to Theorem \([6]\). To see this, in each vertex’s label set, we simply add all the vertices up to distance \( K/2 \) from the vertex. The proof uses standard arguments from the random graph literature and we omit the details.
4 Personalized PageRank Lower Bounds

In this section, we present the lower bound on the storage complexity of labeling schemes for personalized PageRank. The key intuition is based on the following lemma, which states that in an Erdős-Rényi graph, the personalized PageRank value can infer distance information.

**Lemma 8.** Let $G = (V, E)$ be an Erdős-Rényi random graph where every edge is sampled independently with probability $p \geq \frac{\log^4 n}{n}$, and $T$ be a positive integer less than $\log n$. Then almost surely, for all pair of vertices $x, y \in V$ such that $\text{dist}(x, y) \leq T$, we have $\pi(x, y) \geq \frac{\alpha}{3} \left( \frac{1 - \alpha}{np} \right)^T$.

**Proof.** By Chernoff bound, for each vertex $x$, its degree is close to $np$ with high probability:

$$\Pr[d_x \geq (1 + 1/\log n)np] \leq e^{-\frac{np}{3 \log^2 n}}.$$

Hence when $p \geq \frac{\log^4 n}{n}$, by union bound, almost surely all vertices have degree no more than $(1 + 1/\log n)np$. For any path $P = (v_0, \ldots, v_k)$, denote by $\pi(P)$ as:

$$\pi(P) := (1 - \alpha)^k \prod_{i=0}^{k-1} d_{v_i}^{-1}.$$

Therefore, almost surely for every path $P$,

$$\pi(P) \geq \left( \frac{1 - \alpha}{(1 + 1/\log n)np} \right)^{|P|}.$$

In particular, for all pairs of vertices $(x, y)$ such that $\text{dist}(x, y) \leq T$, there is a path $P_{x,y}$ from $x$ to $y$ with length at most $T$. Thus, for all such $x$ and $y$

$$\pi(x, y) \geq \alpha \cdot \pi(P_{x,y})$$

$$\geq \alpha \cdot \left( \frac{1 - \alpha}{(1 + 1/\log n)np} \right)^T$$

$$\geq \frac{\alpha}{3} \cdot \left( \frac{1 - \alpha}{np} \right)^T.$$

This proves the lemma.

Our main result is the following. The analysis will build on the insights from Section 3.

**Theorem 9.** Let $G = (V, E)$ be an Erdős-Rényi graph where every edge is sampled independently with probability from $p \geq \frac{\log^4 n}{n}$. Let $\delta \geq \frac{\log n}{n}$ be the desired accuracy threshold and $t = 1 + \varepsilon$ be the desired approximation ratio. Let $0 < \alpha < 1/2$ be the teleport probability of personalized PageRank. For any $(\varepsilon, \delta)$-accurate labeling for personalized PageRank, with high probability over the randomness of $G$, the total labeling size for $G$ is at least $\Omega \left( n \cdot (np)^{d+1} \cdot \log^{-2} n \right)$, where

$$d = \left\lfloor \frac{1}{2} \left( \log \frac{\alpha}{\delta \cdot 3} \log \frac{1 - \alpha}{np} - 1 \right) \right\rfloor.$$
To see that Theorem 1 follows from Theorem 9 let \( p = \log^4 n/n \) and \( \alpha \leq \frac{\log np}{2 \log n} \). We have that

\[
(np)^{d+1} \geq \left( \sqrt{\frac{\alpha}{3e^2 \delta}} \right)^{1+\frac{\log(1-\alpha)}{\log np - \log(1-\alpha)}}.
\]

(2)

Note that \( t \geq 1 \) and \( \alpha < 1/2 \), hence \( \left( \frac{\alpha}{3e^2 \delta} \right) \leq n \). And

\[
-\frac{\log(1-\alpha)}{\log np - \log(1-\alpha)} \leq \frac{\alpha/(1-\alpha)}{\log np - \log(1-\alpha)} \leq \frac{2\alpha}{\log np},
\]

for large enough \( n \). Hence we obtain

\[
\left( \sqrt{\frac{\alpha}{3e^2 \delta}} \right)^{1-\frac{\log(1-\alpha)}{\log np - \log(1-\alpha)}} \leq \exp \left( \log n \cdot \frac{2\alpha}{\log np} \right) \leq O(1),
\]

(3)

because \( \alpha \leq \frac{\log np}{2 \log n} \). Hence we obtain that the RHS of equation (2) is at least \( \tilde{\Omega}(\sqrt{\frac{\alpha}{3e^2 \delta}}) \).

As a remark, when \( \frac{\alpha}{3e^2 \delta} \) is an odd integer power of \( \frac{np}{1-\alpha} \), the above lower bound simplifies to

\[
\tilde{\Omega} \left( \sqrt{\frac{nm \cdot \left( \frac{\alpha}{3e^2 \delta} \right)^{1-\frac{\log(1-\alpha)}{\log np - \log(1-\alpha)}}}} \right) \geq \tilde{\Omega} \left( \sqrt{\frac{nm}{\delta}} \right),
\]

for \( t \lesssim \text{poly log}(n) \) and \( 1/\text{poly log}(n) \lesssim \alpha \leq \frac{\log np}{2 \log n} \) by equation (3). In the following we present the proof of Theorem 9 based on Lemma 8 and Lemma 5.

**Proof of Theorem 9**. Divide \( V \) to groups of size \( \frac{n}{(np)^{d \log^2 n}} \). We will show that for each group \( S \),

\[
\Pr[\text{The total label size of } S \leq c_2 |S| \cdot (np)^{d+1} \log^{-2} n] \leq o(1).
\]

(4)

where \( c_2 \) is a certain constant to be specified later. Hence by Markov’s inequality, almost surly there are \( \Omega((np)^{d \log^2 n}) \) groups with total label size \( c_2 |S| \cdot (np)^{d+1} \log^{-2} n \). And the total label size of \( G \) will be at least \( \Omega((np)^{d+1} \cdot n \cdot \log^{-2} n) \). For the rest of the proof, we focus on proving equation (4).

We apply the iterative process in Section 3 to generate boundary sets \( L(x) \subseteq V \) for each \( x \in S \). Note that Erdős-Rényi graphs can be generated from the Chung-Lu model where each vertex has weight \( np \). Hence \( \nu = np \), \( \text{vol}(V) = n^2p \) and \( r = \frac{\text{vol}(V)}{\text{vol}(V)} = np \). And we obtain that

\[
|S| \leq o \left( \frac{\text{vol}(V)}{p \log n} \right).
\]

We also verify the assumptions required to in Lemma 5. Note that the second and third assumption follows from straightforward calculation. For the first assumption, conditional on \( \text{vol}(F_i-1) \leq |S| \cdot r^d \log n \), we have that the size of \( F_i-1 \) is at most \( |S| \cdot r^d \log n \leq n/\log n \). Thus the growth rate of \( x_i \) on \( G_i \) is \( np(1 - O \left( \frac{1}{\log n} \right) ) \) with high probability because \( p \geq \log^4 n/n \). Since \( d \lesssim \log n \), we obtain that the first assumption holds.

Thus by Lemma 5, almost surely there exists \( S' \subseteq S \) with size at least \( c_1 \cdot |S| \), such that for all \( x \in S' \) we have \( |L(x)| \geq \Omega((np)^d) \), for a certain fixed constant \( c_1 < 1 \). Consider the following two cases:

a) There exists \( c_2^2 |S'|^2/4 \) pairs \((x,y) \in S \times S \) such that \( \pi(x,y) \geq \delta \): We show that such an event happens on \( G \) with very small probability. By definition, \( \sum_{y \in V} \pi(x,y) = 1 \). Observe that an Erdős-Rényi random graph is symmetric in \( V \). Thus for fixed \( x,y \), \( E_G[\pi(x,y)] = 1/(n-1) \). By Markov’s inequality, \( \Pr[\pi(x,y) > \delta] \leq \frac{1}{(n-1)\delta} \). Hence the expected number of pairs \((x,y) \in S \times S \) such that \( \pi(x,y) > \delta \) is at most \( |S|^2/(n\delta) \leq |S|^2/\log n \), because \( \delta \geq \log n/n \). The desired conclusion then follows by Markov’s inequality.
b) The number of vertex pairs \((x, y) \in S \times S\) such that \(\pi(x, y) \geq \delta\) is at most \(c_1^2 |S|/4\): Denote by \(\tilde{\pi}(x, y)\) the output of the labeling scheme, for each query \((x, y) \in V \times V\). Let \(B = \{(x, y) \in S' \times S' \mid \tilde{\pi}(x, y) < t\delta\}\). We first show that the size of \(B\) is at least \(c_1^2 |S|/4\). To see this, note that there are at least \(\left(\binom{|S'|}{2} \right) - \frac{c_1^2 |S|^2}{4} \geq \frac{c_1^2 |S|^2}{5}\) vertex pairs \((x, y) \in S' \times S'\) where \(\pi(x, y) \leq \delta\). For any such pair, by the \((\varepsilon, \delta)\)-accuracy guarantee, we have that \(\tilde{\pi}(x, y) \leq t\delta\). Hence the the size of \(B\)'s at least \(c_1^2 |S|^2/5\).

Next we show that for each \((x, y) \in B\), \(\text{dist}(x, y) \geq 2d + 2\). Assume that \(\text{dist}(x, y) \leq 2d + 1\). By Lemma 3 with high probability for all vertex pairs where \(\text{dist}(x, y) \leq 2d + 1\), we have that \(\pi(x, y) \leq \frac{\alpha}{3} \left(\frac{1-\alpha}{np}\right)^{2d+1}\) holds. On the other hand,

\[
\pi(x, y) \leq t \cdot \max(\delta, \tilde{\pi}(x, y)) < t^2 \delta \leq \frac{\alpha}{3} \cdot \left(\frac{1-\alpha}{np}\right)^{2d+1},
\]

where the first step is because of the \((\varepsilon, \delta)\)-accuracy, the second step is because \((x, y) \in B\), and the last step is because of the setting of \(d\). Hence we have arrived at a contradiction.

Lastly, since for each \((x, y) \in B\), \(\text{dist}(x, y) > 2d + 1\), we infer that there are no edges between \(L(x)\) and \(L(y)\). However, the size of \(L(x)\) and \(L(y)\) are both at least \(\Omega((np)^d)\). And

\[
\Pr[L(x) \sim L(y) \forall (x, y) \in B \mid \{L(x)\}_{x \in S}] \leq \left(1 - p\right)^{\Omega((np)^{2d})} \frac{c_2 |S|^2}{5} \leq \exp\left(-\Theta\left(p \cdot (np)^{2d} |S|^2\right)\right).
\]

Let \(c_2\) be a small enough constant. By taking a union bound over the set of all possible labels for \(S\) whose total size is less than \(c_2 \cdot p(n(p)^{2d} |S|^2) = c_2 |S| \cdot (np)^{d+1} \log^{-2} n\), we obtain that equation 4 is true for case b).

By a union bound over cases a) and b), we proved that equation 4 is true. Hence the proof is complete. \(\square\)

5 Shortest Path Labeling Schemes for \(2 < \beta < 3\)

In this section we consider random power law graphs with degree exponent \(2 < \beta < 3\). Because the degree sequence has high variance, the branching process grows doubly exponentially near the boundary. To resolve this issue, we will carefully follow the growth of the high degree vertices. We first state the upper bound result.

**Theorem 10.** Let \(\mathcal{G}^n(p)\) be a random power law graph model with average degree \(\nu > 1\) and exponent \(2 < \beta \leq 3\). For a random graph \(G = (V, E)\) drawn from \(\mathcal{G}^n(p)\), with high probability there exists an exact distance labeling scheme \(F\) such that \(|F(x)| \leq n^{1-\min\left(\frac{1}{2\beta}, \frac{1}{1+\beta}\right)} \log^3 n\) for all \(x \in V\).

Given the labeling \(F\) (and the corresponding distances), the query algorithm for \(\text{dist}(x, y)\) for any \(x, y \in V\) is given by

\[
\min_{z \in F(x) \cap F(y)} \text{dist}(x, z) + \text{dist}(z, y).
\]

If no common vertex is found between \(F(x)\) and \(F(y)\), we return that \(x, y\) are disconnected. Clearly, each query takes no more than \(O(\log |F(x)| + |F(y)|)\) time.

We use the fact that \(G\) contains a heavy vertex whose weight is approximately \(n^{\frac{1}{\beta-1}}\). We first add all such high degree vertices (and their distances) to the labeling set of every vertex. Then we
Algorithm 1 ALGSkewDegree

Input: An undirected graph $G = (V, E)$; Parameter $K = \sqrt{n}$ if $2.5 \leq \beta \leq 3$, or $n^{(1+\beta)/(2-\beta)}$ if $2 < \beta < 2.5$.
1. Let $H = \{x \in V : d_x \geq K\}$
2. for $x \in V$ do
3. $(F(x), l(x)) = \text{ALGBfs}(x, 4\nu \log^2 n \times n^{\frac{\beta-2}{2-\beta}})$ (add closest $\tilde{O}(n^{\frac{\beta-2}{2-\beta}})$ vertices)
4. $F(x) = F(x) \cup H$ (add all high degree vertices)
5. for $y \in \Gamma_{l(x)-1}(x)$ do
6. if $d_y \leq K$ then
7. $F(y) = F(x) \cup \{z \in N(y) : d_y \leq d_z \leq K\}$ (add all neighbors of $y$ with a higher degree)
8. end if
9. end for
10. end for
11.
12. procedure ALGBfs($x, t$)
13. $S = \{x\}$
14. $\alpha_0(x) = d_x; k = 0$
15. while $\alpha_k(x) \leq t \land |\Gamma_{k+1}(x)| > 0$ do
16. $S = S \cup \Gamma_k(x)$
17. $Y = \{(y, z) \in E : y \in \Gamma_k(x), z \in \Gamma_{k+1}(x)\}$
18. $k = k + 1$
19. $\alpha_k(x) = \sum_{y \in \Gamma_k(x)} d_y - |Y|$ (number of edges between $\Gamma_k(x)$ and $V \setminus N_{k-1}(x)$)
20. end while
21. return $(S, k)$ ($k$ is the first integer that satisfies: $\alpha_{k-1}(x) > t$ or $\Gamma_k(x) = \emptyset$)
22. end procedure

One can also obtain a $(+2)$-stretch labeling by setting $t = \tilde{O}(n^{\beta/2-1})$ in ALGBfs. To see this, for two vertices $x, y$, once the bottom layers of $x, y$ have size at least $K$, they are at most three hops away from each other. This is because with high probability the bottom layer will connect to a vertex with weight $\Omega(\sqrt{n})$ in the next layer. The maximum label size for all $x \in V$ is $\tilde{O}(K)$ — the proof is similar to the proof of Theorem [10] and we omit the proof.

In Section [12] we test our algorithm on real world graphs. We found that our algorithm achieves fairly accurate results — the 80%-percentile multiplicative error is less than 0.25 in our experiment. In addition, the algorithm is scalable to preprocess graphs with millions of edges in several minutes.

Discussion. We now describe a few extensions of Theorem [10]. First, the Chung-Lu model has a natural extension to directed graphs. Consider two power law distributions $f^{in}(x)$ and $f^{out}(x)$ with mean value bigger than 1, representing the indegree and outdegree distributions, respectively. Each node $v$ is associated with two parameters $p^{in}_v \sim f^{in}(\cdot)$ and $p^{out}_v \sim f^{out}(\cdot)$. For any two nodes $u$ and $v$, there is a directed edge from $u$ to $v$ with probability $\frac{p^{out}_v p^{in}_u}{M}$, where $M$ is a normalization term. It’s not hard to see that $O(\sqrt{n})$ storage per node suffices to recover all pairs distances.
we do a breadth-first search forward from every node $x$ to include $\tilde{O}(\sqrt{n})$ landmarks for $x$ as well as a backward BFS from $x$ to include $\tilde{O}(\sqrt{n})$ landmarks, then for every pair of nodes $x$ and $y$, the forward frontier of $x$ and the reverse frontier of $y$ will intersect with high probability. A second possible extension is to consider configuration models with a power law degree distribution. We believe all of our proofs can be extended to configuration models, since our technical tools only involve bounding the growth of branching processes from every node; We leave the details to future work.

The high level intuition behind our algorithmic result is that as long as the breadth-first search process of the graph grows neither too fast nor too last, but rather at a uniform rate, then an efficient labeling scheme can be obtained. It would be interesting to come up with a deterministic graph model when the degree distribution has high variance.

5.1 Lower bound for $2 < \beta < 3$

When the degree distribution has unbounded variance which grows with $n$, we no longer have the growth rates obtained for the case $\beta > 3$. For a vertex with weight $p$ that is large enough, we expect that $p$ is connected to a vertex with weight $p^{1/(\beta-2)}$ with high probability. To formulate such a growth pattern, we introduce a set of ranges increasing by an exponential rate, such that under certain initial conditions, with high probability the growth rates will follow in the specified ranges. Let $d = \log \log \log n / \log \frac{1}{\beta-2}$, $\varepsilon = 1 / \log \log n$, and $w = \log \log n$. For all $0 \leq i \leq d + 1$, let

$$
\mu_i := n^{1/(\beta - 1 - \varepsilon)(\beta - 2)^{d-i}}, \\
\sigma_i := w \cdot (\beta - 2)^{1/(\beta - 2)}
$$

$\mu_i$ can be thought of as the “expected” volume at distance $i$ from a vertex $x \in S$. Denote by $a_i = \mu_i / \sigma_i$ and $b_i = \mu_i \sigma_i$. If the volume at distance $i$ from $x$ always stays inside $[a_i, b_i]$, then we say $x$ follows a “good” path.

**Proposition 11** (Growth rates for $2 < \beta < 3$). Let $G_n(p)$ be a random graph with average degree $\nu > 1$. Suppose that $p$ satisfy that for a set of fixed values $S \subseteq [1, n^{3/2-1}]$ of size at most $2d$, for all $t \in S$, the following holds:

$$
\sum_{y : p_y \geq t} p_y \gtrsim nt^{2-\beta}, \\
\sum_{y : p_y \leq t} p_y^2 \lesssim nt^{3-\beta}, \\
\sum_{y : p_y \geq t} p_y \lesssim nt^{2-\beta}.
$$

Then the following statements are true:

a) Following a good path: Let $x$ be a fixed vertex and $1 \leq k \leq d + 1$. Suppose that $\text{vol}(\Gamma_i(x)) \in [a_i, b_i]$ for any $1 \leq i < k$, then $\text{vol}(\Gamma_k(x)) \in [a_k, b_k]$, with probability at least $1 - O(1/w^{\beta-2})$, where $w = \log \log n$;

b) Average distance: let $x, y$ be two vertices such that $p_x, p_y \in [a_0, b_0]$, then $\Pr[\text{dist}(x, y) \leq 2d + 3] = o(1)$.

We leave the proof to Appendix C.2. Next we state the lower bound for shortest path labeling.
Theorem 12. Let \( G = (V, E) \in \mathcal{G}(p) \) be a random power law graph with average distance \( \nu > 1 \) and exponent \( 2 < \beta < 3 \). With high probability over the randomness of \( G \), any exact distance labeling scheme will output a labeling whose total length is at least \( \Omega(n^{\frac{2-\beta}{2}} - o(1)) \).

The proof consists of three parts. In the first part, we apply an iterative process for growing the neighborhood of a set of vertices. While the high level idea is similar to Section 3, there are important differences due to the fact that the degree distribution has unbounded variance.

Iterative neighborhood growth. Let \( S = \{x_1, x_2, \ldots \} \) be an arbitrary vertex set of size \( n^{\frac{2-\beta}{2}} \) such that all \( x_i \) have weights between \( a_0 \) and \( b_0 \). Denote by \( G_1 = (V_1, E_1) \) where \( V_1 = V \) and \( E_1 = E \). For \( 1 \leq i \leq |S| \), we consider the following inductive process:

1. If \( x_i \in V_i \), let \( 1 \leq \lambda_i \leq d \) be the maximum \( k \) that still satisfy \( \text{vol}(\Gamma_k(x_i)) \in [a_k, b_k] \) in graph \( G_i \), where \( \Gamma_k(x_i) \) is the set of vertices at distance exactly \( k \) from \( x_i \) in \( G_i \);

2. Denote by \( T_i \) the set of vertices within distance \( \min\{d, \lambda_i + 1\} \) from \( x_i \) in \( G_i \);

3. If \( \lambda_i = d \), let \( L(x_i) = \Gamma_d(x_i) \); otherwise, let \( L(x_i) = \emptyset \).

The difference between the above process and the one in Section 3 is that we terminate the growth as soon as it falls out from the good path. Define \( F_i = F_{i-1} \cup T_i \) (\( F_0 = \emptyset \) by default). Let \( G_{i+1} \) be the subgraph of \( G_i \) on remaining vertices \( V_{i+1} = V \setminus T_i \). If we reach distance \( d \), then \( L(x_i) \) is the set of vertices at distance \( d \) from \( x_i \). We make the following crucial observation.

Proposition 13 (Martingale inequality for \( 2 < \beta < 3 \)). In the setting of this subsection, with high probability at least \( \Theta(n^{\frac{2-\beta}{2}}) \) vertices \( x_i \) in \( S \) satisfy that \( \text{vol}(L(x_i)) \in [a_d, b_d] \).

Proof. Consider the following random variables.

\[
X_i = \begin{cases} 
1 & \text{vol}(L(x_i)) \in [a_d, b_d], \\
0 & \text{otherwise}.
\end{cases}
\]

We show that \( X_i = 1 \) with high probability for all \( 1 \leq i \leq |S| \). We first verify that \( \Pr[x_i \notin V_i] = \Pr[x_i \in F_{i-1}] \leq o(1) \). Consider any vertex \( z \in V \) and \( 1 \leq j \leq i - 1 \), we have that

\[
\Pr[z \in T_j] \leq \sum_{l=0}^{d-1} \frac{p_z \cdot b_l}{\text{vol}(V_j)} \leq p_z \cdot n^{(\beta/2-1-\epsilon)(\beta-2)-1} \cdot O(\log^2 \log n).
\]

Thus, by union bound from \( 1 \leq j \leq i - 1 \), we have

\[
\Pr[z \in F_{i-1}] \leq p_z \cdot n^{\frac{1}{2}(\beta^2 - 5\beta + 5) - \epsilon(\beta - 2) + o(\epsilon)} = p_z \cdot n^\lambda,
\]

i.e. denote the exponent by \( \lambda \) above. Next, we verify that the weight sequence of \( G_i \) satisfies the premises of Proposition 11 with high probability. It suffices to verify the first premise – the second and the third hold because \( V_i \) is a subset of \( V \). It’s not hard to see that the initial weight sequence of \( G_1 = G \) satisfies all the premises by Chernoff bound (details omitted). It suffices to show that
$F_{i-1}$ has small volume, i.e. we only remove a small volume from $G$ in total. By Equation \((5)\), we have:

$$E[\text{vol}(F_{i-1})] \leq \sum_{z \in V} p_z \cdot \min\{1, p_z \cdot n^\lambda\} = \sum_{z: p_z \geq n^{-\lambda}} p_z + \sum_{z: p_z < n^{-\lambda}} p_z^2 \cdot n^\lambda \lesssim n^{1-\lambda(2-\beta)} + n^{1-\lambda(3-\beta)+\lambda} \lesssim n^{1+\lambda(\beta-2)}.$$ 

The second inequality above is by applying the Assumption of Proposition \(11\) with $t = n^{-\lambda}$. The last step is because $1 + \lambda(\beta - 2) \leq 1 - \frac{1}{2}(\beta - 2)^2 - \Theta(\varepsilon)$. Having bounded the expected volume of $F_{i-1}$, we obtain that with high probability only a total volume of $o(n^{1-(\beta-2)^2/2})$ is from $V$ in $V_i$. Thus, we obtain the first premise of Proposition \(11\) because $nt^{2-\beta} = \Omega(n^{1-(\beta-2)^2/2})$ for any $t \leq n^{3/2-1}$.

Now we can apply Proposition \(11\) to $G_i$ to obtain that $\Pr[X_i = 0] = o(1)$. By Markov’s inequality, $\sum_i (1 - X_i) \lesssim n^{1/2 - \beta}$ with high probability. Hence at least $\Theta(n^{-\beta})$ vertices in $S$ satisfy that $\text{vol}(L(x_i)) \in [a_d, b_d]$ with high probability.

**Connecting to heavy vertices.** In this part, we show how $L(x_i)$ and certain set of high degree vertices are connected. Let $A = \{x \in V : p_x \in [(\frac{n}{w})^{1/(\beta-2)}, 2(\frac{n}{w})^{1/(\beta-2)}] \setminus F_{|S|}. \text{ We show that there is a constant fraction of the vertices } x \text{ in } S_1 \text{ such that each } L(x) \text{ is connected to a different vertex in } A.\)

**Proposition 14.** In the setting of this subsection, with high probability, there exists a set $S' \subseteq S$ and a injective function $h : S' \to A$ such that $|S'| \gtrsim n^{3-\beta}$, and for every $x \in S_2$, $h(x)$ connects to some vertex in $L(x)$.

**Proof.** We first show that $A$ has a large volume. By Equation \((5)\), any vertex with weight at most $2(\frac{n}{w})^{1/(\beta-2)} < o(n^{1/2})$ belongs to $F_{|S|}$ with probability at most $n^{\frac{1}{2}(\beta-2)(\beta-3)} \leq o(1)$. Thus, the volume of $A$ satisfies that $\text{vol}(A) \gtrsim nw/a_d$ with high probability. This implies that $|A| \gtrsim \frac{nw}{a_d(\frac{n}{w})^{1/(\beta-2)}} \gtrsim n^{(3-\beta)/2} \cdot w$.

By Proposition \(13\) there exists a set $S_1 \subseteq S$ of size $\Theta(|S|)$ such that for all $x \in S_1$, $\text{vol}(L(x)) \in [a_d, b_d]$. We will construct the set $S'$ and the function $h$ as follows. Initially, all vertices in $A$ are marked as “unused”. For each $x$ in $S_1$: if $L(x)$ has a neighbor $y$ in $A$ that is “unused”, add $x$ to $S'$ and set $h(x)$ to $y$; Mark $y$ as “used” and continue to the next vertex in $S_1$. We claim that this procedure will generate a set $S'$ of size at least $\Theta(n^{3-\beta})$ with high probability. This is because $\text{vol}(L(x)) \geq a_d$ by Proposition \(13\) for all $x \in S_1$, hence

$$\Pr[L(x) \sim y \mid \forall y \in A, \text{ s.t. } y \text{ is “unused”}] \leq e^{-\frac{\text{vol}(L(x)) \cdot \frac{nw}{w}}{\text{vol}(V)}} \leq o(1), \quad (6)$$

because the volume of the “used” vertices is at most $n^{1-\beta/2} \cdot (\frac{n}{w})^{1/(\beta-2)} \leq o(nw/a_d) \leq o(\text{vol}(A))$. Given equation \((6)\), it’s not hard to see that the conclusion follows by Markov’s inequality (details omitted).

In the last part, we bound the entropy information obtained from the labeling and use the entropy to get a lower bound on the total labeling length.
Proof of Theorem 14. Consider the set of vertices with weights between \(a_0\) and \(b_0\), and divide them into groups of size \(n^{-\beta}\). The number of groups is \(n^{\beta-1}/O(e)\) with high probability, because there are \(n^{1-O(e)}\) such vertices. Recall that \(e = 1/\log \log n\). We will show that for each group \(S\),

\[
\Pr[\text{The total label size of } S \leq n^{3-\beta}] \leq o(1).
\] (7)

Hence by Markov’s inequality, with high probability, at least a constant fraction of the groups have label sizes at least \(n^{3-\beta-O(e)}\). This implies that the total label size of \(G\) is at least \(n^{\frac{5}{2}-\beta-O(e)}\). For the rest of the proof we focus on Equation (7).

Consider the distance function on \(G\) restricted to all vertex pairs in \(S\), \(d_{st} : S \times S \to \mathbb{N}\). Clearly, \(d_{st}\) can be determined from the labels of \(S\). By Proposition 14, almost surely there exists \(S' \subseteq S\) of size \(\Theta(|S|)\) such that for every \(x \in S'\), \(L(x)\) is connected to a unique high degree vertex \(h(x)\) whose weight is at least \((\frac{w}{d})^{1/(\beta-2)}\). Consider the following two cases:

a) If there exists \(0.01 \times |S'|^2\) vertex pairs from \(S\) whose distance is at most \(2d + 3\), we claim that the probability that \(G\) induces any such distance function is at most \(o(1)\). By Proposition 11, the probability that two vertices have distance at most \(2d + 3\) is \(o(1)\), hence the expected number of vertex pairs in \(S\) within distance \(2d + 3\) is \(o(n^{3-\beta})\). The claim then follows by Markov’s inequality.

b) If the number of vertex pairs from \(S\) within distance \(2d + 3\) is at most \(0.01 \cdot |S'|^2\), then we infer that there are \(\Theta(|S'|^2) = \Theta(|S|^2)\) vertex pairs in \(S'\) whose pairwise distance is at least \(2d + 4\). Let \(B \subseteq S' \times S'\) denote the set of such pairs. Since \(d_{st}(x, y) \geq 2d + 4\) for any \(x, y \in B\), we infer that \(h(x)\) and \(h(y)\) are not connected. Hence we have

\[
\Pr\left[\text{dist}(x, y) \geq 2d + 4, \forall (x, y) \in B \mid \{T_i\}_{i=1}^{\lfloor S \rfloor}\right]
\]

\[
\leq \prod_{(x, y) \in B} \Pr[h(x) \neq h(y)]
\]

\[
\leq \left(1 - \frac{1}{n} \times \Theta\left(\frac{a_d}{w} \frac{n^{2}}{\log n}\right)\right)^{\Theta(n^{3-\beta})}
\]

\[
\leq \exp\left(-\Theta\left(\frac{a_d}{w} \frac{n^{2}}{\log n}\right) \times n^{\beta-4}\right)
\]

\[
= \exp\left(-\Theta(n^{\beta-3-2\varepsilon/(\beta-2)} \times w^{-\frac{2}{\beta-2}} - \frac{2}{\beta-2} \log^2 \log n)\right) = \exp\left(-\Theta(n^{\beta-3-O(e)})\right).
\]

where the last few steps follows by straightforward calculations. To summarize, the distance function induced on \(S\) only occurs with probability \(\exp\left(-n^{\beta-3-O(e)}\right)\). Then by union bound over all possible labels for \(S\) whose length is less than \(n^{3-\beta}\), we obtain that with high probability, the total label size of \(S\) should be at least \(n^{3-\beta}\).

By taking a union bound over cases a) and b), we have proved equation (7). Hence the proof is complete.

6 Conclusions

In this work, we studied data structure lower bounds under the labeling query model. We introduced a general technique based on random graph distributions. We showed a matching lower bound for personalized PageRank. The same techniques are also used to obtain nearly tight bounds for
shortest path labeling schemes on sparse random graphs with power law degree distributions. We hope that our results may be useful to obtain lower bounds for other data structure problems, where the labeling query model have been successfully used for algorithm design. It would be interesting to further understand the query-space trade-off under the labeling query model. We believe that random graphs could be a hard instance for studying the storage complexity of set intersections [48].

Acknowledgements. The authors would like to thank Ashish Goel, Fan Chung Graham, Tim Roughgarden and D. Sivakumar for useful feedbacks and suggestions at various stages of this work.

References


In this section, we fill in the proof of Theorem 10. We assume that \( p \) satisfies all properties in Proposition 22. We will show that Algorithm 1 is correct in Lemma 15 and bound its output size in Lemma 18.
Lemma 15. Algorithm [I] with parameter $\delta$ and $K$ finds a labeling $F$ that can recover the distances for all pairs of vertices with high probability.

Proof. Consider the random variable $l(x)$ that is computed in the Algorithm for each node $x$. Let $\Omega_S$ denote the set of graphs that satisfies

$$\Gamma_{l(x)}(x) = \emptyset$$

or

$$\text{dist}(v^*, x) \leq l(x), \forall x \in V$$

where $v^*$ is the node with the maximum weight. We argue that Algorithm [I] finds a 2-hop cover for any $G \in \Omega_S$, and

$$1 - \Pr[\Omega_S] \leq 2/n$$

This would imply that Algorithm [I] succeeds with probability at least $1 - 2/n$.

We first argue that Algorithm [I] is correct if $G \in \Omega_S$. Let $x$ and $y$ be two different vertices in $V$. If $x$ and $y$ are not reachable from each other, then clearly $F(x) \cap F(y) = \emptyset$. If $x$ and $y$ are reachable from each other, consider their distance $\text{dist}(x, y)$. Note that when $\Gamma_{l(x)}(x)$ (or $\Gamma_{l(y)}(y)$) is empty, then $F(x)$ (or $F(y)$) includes the entire connected component that contains $x$ (or $y$). Therefore, $y \in F(x)$, vice versa. When none of them are empty, we know that $\text{dist}(x, v^*) \leq l(x)$ and $\text{dist}(y, v^*) \leq l(y)$ since $G \in \Omega_S$. We consider three cases:

- If $\text{dist}(x, y) \leq l(x) + l(y) - 2$, then there exists a node $z$ such that $\text{dist}(x, z) \leq l(x) - 1$ and $\text{dist}(y, z) \leq l(y) - 1$. By our construction, $z$ is in $F(x)$ and $F(y)$.

- If $\text{dist}(x, y) = l(x) + l(y) - 1$, then consider the two nodes $z$ and $z'$ on one of the shortest path from $x$ to $y$, with $\text{dist}(x, z) = l(x) - 1$ and $\text{dist}(y, z) = l(y)$. If either $d_z$ or $d_{z'}$ is at least $K$, then they have been added as a landmark to every node in $V$. Otherwise, assume without loss of generality that $d_z \geq d_{z'}$. Then our construction adds $z$ into $F(y)$ and clearly $z$ is also in $F(x)$, hence $z$ is a common landmark for $x$ and $y$.

- If $\text{dist}(x, y) = l(x) + l(y)$, then clearly $v^*$ is a common landmark for $x$ and $y$.

We now bound $1 - \Pr[\Omega_S]$. Clearly,

$$1 - \Pr[\Omega_S] \leq \sum_{x \in V} \Pr[\Gamma_{l(x)}(x) \neq \emptyset, \text{dist}(v^*, x) > l(x)]$$

$$= \sum_{x \in V} \sum_{k=0}^{n-1} \Pr[l(x) = k + 1, \Gamma_{k+1}(x) \neq \emptyset, \text{dist}(v^*, x) > k + 1]$$

Note that $l(x) = k + 1$ and $\Gamma_{k+1}(x) \neq \emptyset$ is the same as the event that:

- $\alpha_i(x) \leq \delta n^{1 - \frac{1}{\beta-1}}$, for $i = 0, \ldots, k$;

- $\alpha_k(x) > \delta n^{1 - \frac{1}{\beta-1}}$.

Hence,

$$\Pr[l(x) = k + 1, \Gamma_{k+1}(x) \neq \emptyset, \text{dist}(v^*, x) > k + 1]$$

$$\leq \Pr[\alpha_k(x) > \delta n^{1 - \frac{1}{\beta-1}}, \text{dist}(v^*, x) > k + 1]$$

$$\leq \Pr[\alpha_k(x) > \delta n^{1 - \frac{1}{\beta-1}}, \text{vol}(\Gamma_k(x)) \leq \frac{\delta n^{1 - \frac{1}{\beta-1}}}{3}]$$

$$+ \Pr[\text{vol}(\Gamma_k(x)) > \frac{\delta n^{1 - \frac{1}{\beta-1}}}{3}, \text{dist}(v^*, u) > k + 1]$$

(9)
For Equation (8), consider how $\alpha_k(x)$ is discovered when we do the level set expansion from node $x$. Conditioned on $a = \text{vol}(\Gamma_k(x)) \leq \delta n^{1-\frac{1}{\beta-1}}/3$, $\alpha_k(x)$ is the sum of $0$-$1$ independent random variables, with expected value less than $\delta n^{1-\frac{1}{\beta-1}}/3$. Hence by Chernoff bound, Equation (8) is at most $\exp(-\delta n^{1-\frac{1}{\beta-1}}/6) \sim o(n^{-3})$. For Equation (9), conditioned on $\text{vol}(\Gamma_k(x)) \geq \delta n^{1-\frac{1}{\beta-1}}/2$ and $v^* \notin N_k(x)$,

$$\Pr[v^* \sim \Gamma_k(x)] \leq \exp(-\frac{\delta n^{1-\frac{1}{\beta-1}}p_{v^*}}{2\text{vol}(V)}) \sim o(n^{-3})$$

The first inequality is because of Proposition 3. The second inequality is because $\text{vol}(V) \sim \nu n \pm o(n)$ by Proposition 22. In summary, $1 - \Pr[\Omega_S] \leq 2/n$.}

We now consider the size of our landmark scheme. There are three parts in each landmark set: (1) the heavy nodes whose degree is at least $K$; (2) all the level sets before the last layer; (3) the last layer that we carefully constructed. It’s not hard to bound the first part, since the degree of a node is concentrated near its weight, and the number of nodes whose weight is $\Omega(K)$ is $O(nK^{1-\beta})$. The second part can be bounded by the maximum number of layers, hence the diameter of $G$, which is $O(\log n)$. For the third part, the idea is that before adding all the nodes on the boundary layer, we already have a $(+1)$-stretch scheme. Therefore, for a given vertex $x$, it is enough if we only add neighbors whose degree is bigger than $d_x$ — this reduces the amount of vertices from $d_x$ to $O(d_x^{\beta-\beta})$.

We first show that the volume of all the level sets is at most $O(\delta n^{1-\frac{1}{\beta-1}})$ before the boundary layer. For the rest of the section, let $\alpha_k = \alpha_k(x)$ for any $0 \leq k \leq n-1$, unless there is any ambiguity on the vertex we are considering. Recall that $\alpha_k$ denotes the number of edges between $\Gamma_k(x)$ and $V \setminus N_{k-1}(x)$.

**Lemma 16.** Let $x$ be a fixed node. Let $k$ be an integer less than or equal to $O(\log n)$. Let $\Omega_k$ denote the set of graphs such that

$$\text{vol}(\Gamma_i(x)) < 4\delta n^{1-\frac{1}{\beta-1}}, \text{ for any } 0 \leq i \leq k-1,$$

and

$$\text{vol}(\Gamma_k(x)) > 4\delta n^{1-\frac{1}{\beta-1}}.$$

Then $\Pr[\alpha_k \leq \delta n^{1-\frac{1}{\beta-1}} \mid \Omega_k] \leq n^{-2}$.

**Proof.** Let $a = \text{vol}(\Gamma_k(x))$ and $b = \text{vol}(N_{k-1}(x))$. Conditioned on $\Omega_k$,

$$a > 4\delta n^{1-\frac{1}{\beta-1}} \text{ and } b \leq 4k\delta n^{1-\frac{1}{\beta-1}}.$$

Clearly, the random variable $\alpha_k$ is the sum of independent $0$-$1$ random variables. Let $\mu$ denote its expected value. For each $y \in \Gamma_k(x)$, we know that $p_y \leq a = O(\sqrt{n})$. Let $\mu_y$ denote the expected number of edges between $y$ and $V \setminus N_{k-1}(x)$, then

$$\mu_y = \sum_{z: z \neq y \land z \notin N_{k-1}(x)} \min(\frac{p_y p_z}{\text{vol}(V)}, 1)1_{p_z \leq \sqrt{n}} \geq p_y(1 - \frac{b + \sum_{z \in V} p_z 1_{p_z \geq \sqrt{n}}}{\text{vol}(V)}) = p_y(1 - \kappa(n))$$
because of Proposition \[22\] And $\mu = \sum_{y \in \Gamma_k(x)} \mu_y = (1 - o(1))a$. Let $c = \frac{\mu}{\delta n^{1-\frac{1}{\beta - 1}}} \geq 2 - o(1)$. By Chernoff bound,

$$\Pr[\alpha \leq \delta n^{1-\frac{1}{\beta - 1}} | \Omega_k] \leq \exp\left(-\frac{(c-1)^2 \delta n^{1-\frac{1}{\beta - 1}}}{4}\right) \sim o(n^{-2})$$

\[\Box\]

**Lemma 17.** Let $x$ be a fixed vertex. Let $0 \leq k \leq O(\log n)$. Denote by $\Omega_k^*$ the set of graphs such that

$$\alpha_i \leq \delta n^{1-\frac{1}{\beta - 1}}, \text{ for any } 0 \leq i \leq k$$

Then $\Pr[\text{vol}(\Gamma_k(x)) > 4\delta n^{1-\frac{1}{\beta - 1}}, \Omega_k^*] \leq (k+1)n^{-2}$

**Proof.** When $k = 0$, the claim is proved by Lemma \[16\]. When $k \geq 1$, we will repeatedly apply Lemma \[16\] to prove the statement. For any values of $i$ smaller than or equal to $k$, let $S_i \subset \Omega_k^*$ denote the set of graphs that also satisfy: (1) $\text{vol}(\Gamma_i(x)) \leq 4\delta n^{1-\frac{1}{\beta - 1}}$, for any $0 \leq j \leq i - 1$; (2) $\text{vol}(\Gamma_k(x)) > 4\delta n^{1-\frac{1}{\beta - 1}}$. We show that $\Pr[S_i] - \Pr[S_{i+1}] \leq n^{-2}$ if $0 \leq i \leq k - 1$, and $\Pr[S_k] \leq n^{-2}$. Our Lemma follows from the two claims.

For the first part,

$$\Pr[S_i] - \Pr[S_{i+1}] = \Pr[\text{vol}(\Gamma_i(x)) > 4\delta n^{1-\frac{1}{\beta - 1}}, S_i]$$

$$\leq \Pr[\Omega_i, \alpha_i \leq \delta n^{1-\frac{1}{\beta - 1}}] \leq n^{-2}$$

The first inequality is because if $G \in S_i$ and $G$ satisfies $\text{vol}(\Gamma_i(x)) > 4\delta n^{1-\frac{1}{\beta - 1}}$, then $G \in \Omega_i$. Also $\alpha_i \leq \delta n^{1-\frac{1}{\beta - 1}}$ since $G \in S_i \subset \Omega_k^*$. The second inequality is because of Lemma \[16\] The other part can be proved similarly and we omit the details. \[\Box\]

Now we are ready to bound the size of our landmark scheme.

**Lemma 18.** The following holds almost surely

- $|F(x)| \lesssim O(n^{1 - \min\left(\frac{1}{\beta - 1}, \frac{1}{4 - \gamma}\right)} \cdot \log^3 n)$ for all $x \in V$;
- The algorithm terminates in time $O(n^{2 - \min\left(\frac{1}{\beta - 1}, \frac{1}{4 - \gamma}\right)} \cdot \log^3 n)$.

**Remark** To implement Line \[7\] one can first sort $N(x)$ for each $x \in V$, in descending order on their degrees, and then create a separate list that truncates the nodes whose degree is at least $K$. Given this list, one can find the set of neighbors of $x$ whose degree is between $[d_x, K]$. The amount of time it takes to sort $N(x)$ is $O(d_x \log d_x) \sim O(d_x \log n)$. Hence the total amount of time it takes to sort all the adjacency lists is $O(|E| \log n) = O(n \log n)$.

We will use the following lemma for technical reasons — the proof is deferred to the end of the section.

**Lemma 19.** Let $x$ be a fixed node with weight $p_x \leq 2K$. Denote by

$$S_x = \{y \in N(x) : d_x \leq d_y \text{ and } d_y \leq K\}$$

and let $\hat{d}_x = |S_x|$. Then

$$\Pr[\hat{d}_x \geq \max(c_1 p_x^{3-\beta}, c_2 \log n)] \leq n^{-3}$$

where $c_1 = \frac{192Z}{\beta(\beta - 2)}$ and $c_2 = 130$. 

26
Proof of Lemma 18: We first bound the number of nodes in \( H \). By Proposition 21 with probability \( 1 - n^{-1} \)

\[
|H| = O(nK^{1-\beta}) = O(n^{1-\min(\frac{1}{2(3-\beta)}, \frac{1}{3})})
\]

Secondly, we bound the number of landmarks added before reaching the boundary layer. For any vertex \( x \), with \( i = 0, \ldots, l(x) - 2 \), \( |\Gamma_i(x)| \leq \alpha_i(x) = O(\delta n^{1-\frac{1}{1-\beta}}) \). Since \( l(x) \leq O(\log n) \), the total landmarks for these layers are at most \( O(n^{1-\frac{1}{1-\beta}} \log^3 n) \). The rest of the proof will bound the number of landmarks on the boundary layer with depth \( l(x) - 1 \).

Denote by

\[
\pi_k(x) = \sum_{y \in \Gamma_k(x)} d_y \mathbb{1}_{d_y \leq K} \quad \text{for } x \in V, 0 \leq k \leq n - 1
\]

Hence \( \pi_{l(x)-1}(x) \) gives the number of landmarks added on the boundary layer.

Set \( c = \frac{3Z}{[2\beta - 1]} \max(x_{\min}^{-5-2\beta}, 1) \), \( \psi = 12c_3\delta n^{1-\min(\frac{1}{1-\beta}, \frac{1}{3})} \), and \( \Delta = \max(c_1\psi, c_2\delta n^{1-\frac{1}{1-\beta}} \log n) \), where \( c_1 \) and \( c_2 \) are defined in Lemma 19. We show that \( \pi_{l(x)-1}(x) \leq \Delta \) with probability \( 1 - n^{-2} \) for the rest of the proof — our conclusion follows by taking union bound over \( x \in V \) and \( 1 \leq l(x) \leq O(\log n) \).

When \( l(x) = 1 \), \( \pi_0(x) = d_x \leq K \leq \Delta \). When \( l(x) = k + 1 \geq 2 \), we know that \( G \in \Omega_{k-1}^* \). Hence by Lemma 17, \( \operatorname{vol}(\Gamma_{k-1}(x)) \leq 4\delta n^{1-\frac{1}{1-\beta}} \) with high probability. More concretely,

\[
\Pr[l(x) = k + 1, \pi_k(x) \geq \Delta] \\
\leq (k + 1)n^{-2} + \Pr[l(x) = k + 1, \operatorname{vol}(\Gamma_{k-1}(x)) \leq 4\delta n^{1-\frac{1}{1-\beta}}, \pi_k(x) \geq \Delta]
\tag{10}
\]

Denote by

\[
w_k = \sum_{y \in \Gamma_k(x)} p_y^{3-\beta} \mathbb{1}_{p_y \leq 2K}.
\]

Conditional on \( a = \operatorname{vol}(\Gamma_{k-1}(x)) \leq 4\delta n^{1-\frac{1}{1-\beta}} \), we show that \( w_k \leq \psi \) with high probability. Denote by \( \Omega_w \) the set of graphs satisfying \( a \leq 4\delta n^{1-\frac{1}{1-\beta}} \). Conditioned on \( \Omega_w \), \( w_k \) is the sum of independent random variables that are all bounded in \([0, (2K)^{3-\beta}]\). Hence

\[
\mathbb{E}[w_k] = \sum_{y \in \Gamma_{k-1}(x)} \Pr[y \sim N_{k-1}(x)] p_y^{3-\beta} \mathbb{1}_{p_y \leq 2K} \\
\leq \frac{a}{\operatorname{vol}(V)} \left( \sum_{y \notin N_{k-1}(x)} p_y^{4-\beta} \mathbb{1}_{p_y \leq 2K} \right) \quad \text{(by Proposition 3)} \\
\leq \frac{a}{\operatorname{vol}(V)} \left( \sum_{y \in V} p_y^{4-\beta} \mathbb{1}_{p_y \leq 2K} \right) \\
\leq \frac{a\phi(K)n}{\operatorname{vol}(V)} \quad \text{(by Proposition 22)} \\
\sim \frac{a\phi(K)}{\nu} \quad \text{((\nu \operatorname{vol}(V) = \nu n \pm o(n) by Proposition 22)} \\
\leq \frac{\psi}{3}.
\]

The last line follows by \( a \leq 4\delta n^{1-\frac{1}{1-\beta}} \) and \( \phi(K)n^{1-\frac{1}{1-\beta}} \leq c_2 n^{1-\min(\frac{1}{3}, \frac{1}{3})} \). Now we apply Chernoff bound on \( w_k \),

\[
\Pr[w_k > \psi | \Omega_w] \leq \exp\left(-\frac{\psi}{4(2K)^{3-\beta}}\right) \sim o(n^{-2})
\]
because when $2.5 \leq \beta \leq 3$,
\[
\psi
\frac{1}{K^{3-\beta}} = \Theta(n^{1-\frac{1}{3-\beta} - \frac{3-\beta}{2}}) = \Theta(n^{(\frac{\beta-1)^2 - 2}{2(\beta-1)}})
\]
And when $2 < \beta < 2.5$,
\[
\psi
\frac{1}{K^{3-\beta}} = \Theta(n^{\frac{(3-\beta)(\beta-2)}{2(\beta-1)}})
\]
Hence the second part in Equation (10) is bounded by $o(n^{-2})$ plus
\[
\Pr[l(x) = k + 1, \text{vol}(\Gamma_k(x)) \leq 4\delta n^{1-\frac{1}{\beta-1}}, w_k \leq \psi, \pi_k(x) \geq \Delta] \\
\leq \Pr[w_k \leq \psi, \alpha_{k-1} \leq \delta n^{1-\frac{1}{\beta-1}}, \pi_k(x) \geq \Delta] \\
\leq \Pr[w_k \leq \psi, |\Gamma_k(x)| \leq \delta n^{1-\frac{1}{\beta-1}}, \pi_k(x) \geq \Delta]
\]
In the reminder of the proof we show the above Equation is at most $n^{-2}$. Denote by
\[
\pi'_k(x) = \sum_{y\in\Gamma_k(x)} \hat{d}_y^k p_y \leq 2K
\]
By Proposition 20, $\Pr[d_y \leq K \mid p_y > 2K] \leq \exp(-K/8) \sim o(n^{-3})$ for any $y \in V$. Hence $\pi'_k(x) = \pi_k(x)$ with probability at least $1 - o(n^{-2})$. Lastly, we have
\[
\Pr[w_k \leq \psi, |\Gamma_k(x)| \leq \delta n^{1-\frac{1}{\beta-1}}, \pi'_k(x) \geq \Delta] \leq n^{-2}
\]
Otherwise, there exists a vertex $y \in \Gamma_k(x)$ such that $p_y \leq 2K$ and $\hat{d}_y \geq \max(c_1p_y^{3-\beta}, c_2 \log n)$, because $\Delta \geq \max(c_1\psi, c_2\delta n^{1-\frac{1}{\beta-1}} \log n)$. This happens with probability at most $n^{-2}$, by taking union bound over every vertex with Lemma 19.

**Proof of Lemma 19** When $p_x \leq c_2 \log n/2$,
\[
\Pr[\hat{d}_x \geq c_2 \log n] \leq \Pr[\hat{d}_x \geq c_2 \log n] \leq o(n^{-3})
\]
Now suppose that $p_x > c_2 \log n/2$. Consider any vertex $y$ whose weight is at most $p_x/8$. Then
\[
\Pr[d_y \geq d_x] \leq \Pr[d_y \geq p_x/4] + \Pr[d_x < p_x/4] \sim o(n^{-4})
\]
The second inequality is because of Proposition 20. Hence $y$ is not in $S_x$.
Now if $p_y \geq 2K$, then $\Pr[d_y \leq K] \sim o(n^{-4})$. Hence $y$ is also not in $S_x$. Lastly, let $X$ denote the set of vertices whose weight is between $[\frac{p_x}{8}, 2K]$ and who is connected to $x$. We have
\[
\mathbb{E}[X] = \sum_{y \in V \setminus \{x\}; p_y/8 \leq p_y \leq 2K} \frac{p_xp_y}{\text{vol}(V)} \\
\leq \frac{4p_x}{\text{vol}(V)} \max(8Z, np_x^{2-\beta}, \sqrt{n} \log n) \\
\leq \max(c_1p_x^{3-\beta}, c_2 \log n)/3
\]
The first inequality is because of Proposition 22. The second inequality is because $\text{vol}(V) = \nu n + o(n)$ by Proposition 22 and $p_x \leq 2K \leq 2\sqrt{n}$. From here it is not hard to obtain that $\Pr[|X| \leq \max(c_1p_x^{3-\beta}, c_2 \log n)] \sim o(n^{-3})$.  

28
B Experiment

In this section, we evaluate our algorithms on a collection of large networks. We compare with the algorithm of Akiba et al.'s [4] and the Thorup-Zwick distance oracle [49, 14]. The first algorithm produces an exact landmark labeling via recursively pruning during breadth first search over all vertices – we will refer to it as PrunedLabel later. The second algorithm adapts the 3-approximate distance oracle of Thorup and Zwick [49], via picking high degree vertices as global landmarks – we refer to it as BallGrow. In Table 1 we list the graphs used in our experiment. More details are available at Stanford Large Network Dataset Collection [36].

<table>
<thead>
<tr>
<th>graph</th>
<th># nodes</th>
<th># edges</th>
<th>category</th>
<th>90% effective diameter</th>
<th>average distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>81,306</td>
<td>1,768,149</td>
<td>Social</td>
<td>4.5</td>
<td>3.8</td>
</tr>
<tr>
<td>Stanford</td>
<td>281,903</td>
<td>2,312,497</td>
<td>Web</td>
<td>9.7</td>
<td>5.2</td>
</tr>
<tr>
<td>Google</td>
<td>875,713</td>
<td>5,105,039</td>
<td>Web</td>
<td>8.1</td>
<td>6.0</td>
</tr>
<tr>
<td>BerkStan</td>
<td>685,230</td>
<td>7,600,595</td>
<td>Web</td>
<td>9.9</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Table 1: Basic statistics of graphs in experiments.

Implementation We implemented all three algorithms in Scala. The graph library we used is available at https://github.com/teapot-co/tempest. We run the experiments on Amazon EC2 m4.4xlarge instance, with 64GB of RAM and 16 Intel Xeon 2.3GHz CPUs. We used a variant of AlgSkewDegree in the experiments [1]. We hand tune the two parameters used in the algorithm. For PrunedLabel, a vertex ordering is required: we simply sort all vertices by indegree plus outdegree. For BallGrow, it is necessary to specify the number of global landmarks; we handtune this parameter and choose the number of high degree vertices as global landmarks accordingly.

We measure accuracy over 2000 randomly sampled pairs of source/destination vertices. We look at the 80 and 90-percentile multiplicative error (|estimated-distance / true-distance − 1|).

Algorithm 2 A description of our algorithm in experiments.

Input: A directed graph $G = (V, E)$; Parameters $d$ and $K$.

1: $\sigma = \text{vertices ordered by (indegree + outdegree)}$
2: for $i \leq n$ do
3:   if $i \leq K$ then
4:     \text{COMPUTEGLOBALLM($\sigma_i$)}
5:   else
6:     \text{COMPUTELOCALLM($\sigma_i$)}
7:   end if
8: end for
9: \text{procedure COMPUTEGLOBALLM($x$)}
10: $\{(y, \text{dist}(x, y)), \forall y \in V\} = \text{Run a forward BFS}$
11: $\{(y, \text{dist}(y, x)), \forall y \in V\} = \text{Run a backward BFS}$
12: end procedure
13: \text{procedure COMPUTELOCALLM($x$)}
14: Run a forward BFS from $x$ up to distance $d$, prune any node from $\{\sigma_i\}_{i=1}^K$.
15: end procedure

Results Table 2 compares the landmark size and running time of the three tested algorithms. Table 3 compares the accuracy. Looking at accuracy, we found that both our algorithm and
BallGrow are fairly accurate on the three Web graphs. However, our algorithm does slightly worse for the first test cases. From the performance comparison, we found that both our algorithm and BallGrow are more scalable compared to PrunedLabel. This is to be expected, since PrunedLabel is designed to guarantee exact distances. Our algorithm found smaller landmark sets compared to BallGrow and PrunedLabel in three out of four tests, and runs faster than PrunedLabel on the two largest instance.

<table>
<thead>
<tr>
<th></th>
<th>Landmark size per node</th>
<th>Running time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ours</td>
<td>PrunedLabel</td>
</tr>
<tr>
<td>Twitter</td>
<td>227</td>
<td>261</td>
</tr>
<tr>
<td>Stanford</td>
<td>82</td>
<td>95</td>
</tr>
<tr>
<td>Google</td>
<td>215</td>
<td>285</td>
</tr>
<tr>
<td>BerkStan</td>
<td>63</td>
<td>155</td>
</tr>
</tbody>
</table>

Table 2: Comparison of performances over our algorithm, PrunedLabel and BallGrow. The landmark size is equal to the total number of forward and backward landmarks stored, divided by the total number of vertices.

<table>
<thead>
<tr>
<th></th>
<th>90% error</th>
<th>80% error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ours</td>
<td>BallGrow</td>
</tr>
<tr>
<td>Twitter</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Stanford</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Google</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>BerkStan</td>
<td>0.125</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 3: Comparison of accuracy. The accuracy of PrunedLabel is not listed because it is guaranteed to output exact distances.

Remark. We have only presented experiments on social and information networks, since our algorithm is designed for graphs with small average distance and a heavy tailed distribution.

C Random Graph Toolbox

The following Lemma characterizes the probability that a vertex’s actual degree deviates from its weight.

Proposition 20. Let $G = (V, E) \in \mathcal{G}^n(p)$ be a random graph. Let $x$ be a fixed vertex with weight $p_x$ and degree $d_x$ in $G$. Then

1. If $c \geq 3$, then
   \[
   \Pr[d_x \geq cp_x] \leq \exp\left(-\frac{(c - 1)p_x}{2}\right)
   \]

2. If $0 < c < 1$, then
   \[
   \Pr[d_x \leq cp_x] \leq \exp\left(-\frac{(1 - c)^2p_x}{8}\right)
   \]
Proof. Let $\mu = \mathbb{E}[d_x]$. First,

\[
\mu = \sum_{y \in V \setminus \{x\}} \min\left(\frac{p_x p_y}{\text{vol}(V)}, 1\right) \leq \sum_{y \in V} \frac{p_x p_y}{\text{vol}(V)} = p_x
\]

By Chernoff bound, for any $c \geq 3$,

\[
\Pr[d_x \geq cp_x] \leq \exp\left(-\frac{cp_x - \mu}{2}\right) \leq \exp\left(-\frac{(c-1)p_x}{2}\right)
\]

since $cp_x - \mu \geq 2\mu$.

On the other hand, let $t = \frac{\nu}{2\log(n)} n^{1-\frac{1}{\beta-1}}$, then for any $y \in V$ where $p_y \leq t$, we know that $p_x p_y \leq \text{vol}(V)$ by Proposition 22. Hence

\[
\mu \geq p_x \left(1 - \frac{p_x}{\text{vol}(V)} - \sum_{y \in V} \frac{p_y \mathbb{I}[p_y \geq t]}{\text{vol}(V)}\right)
\]

By Proposition 22,

\[
\sum_{y \in V} p_y \mathbb{I}[p_y \geq t] \sim o(n)
\]

Since $p_x \sim o(n)$ and $\text{vol}(V) = \nu n + o(n)$, we conclude that $\mu = p_x (1 - o(1))$. By Chernoff bound, for any $0 < c < 1$,

\[
\Pr[d_x \leq cp_x] \leq \exp\left(-\frac{(cp_x - \mu)^2}{4\mu}\right) \leq \exp\left(-\frac{p_x(1-c)^2}{8}\right)
\]

for large enough $n$.

The following Proposition characterizes the number of nodes whose degree is at least $K$.

**Proposition 21.** Let $G = (V, E) \in \mathcal{G}_n(p)$ be a random graph. Let $8 \log n \leq K \leq \sqrt{n}$ denote a fixed value and $S = \{x \in V : d_x \geq K\}$. With probability at least $1 - n^{-1}$, $|S| \leq 3 \max\left(\frac{Z_3^{\beta-1}}{\beta-1}nK^{1-\beta}, \log n\right)$.

**Proof.** Let $Y_1 = \{x \in V : p_x \geq \frac{K}{3}\}$ and $Y_2 = \{x \in V : p_x < \frac{K}{3} \text{ and } K \leq d_x\}$. Clearly, $S \subseteq Y_1 \cup Y_2$.

We first show that $Y_2$ is empty with probability at least $1 - n^{-1}$. Consider a fixed node $x \in V$ with weight $p_x \leq K/3$. By Proposition 20,

\[
\Pr[d_x \geq K] \leq \exp(-K) \sim o(n^{-2})
\]

Hence $\Pr[Y_2 \neq \emptyset] = o(n^{-1})$ by union bound.

We then bound the size of $Y_1$. The expected value of $Y_1$ is $\frac{Z_3^{\beta-1}}{\beta-1}nK^{1-\beta}$. Then by Chernoff bound, it’s not hard to obtain the desired conclusion (details omitted).

**Proposition 22.** Let $f$ denote the probability density function of a power law distribution with mean value $\nu > 1$ and exponent $2 < \beta \leq 3$. Let $p$ denote $n$ independent samples from $f(\cdot)$. Let $\log n \leq d \leq 2\sqrt{n}$ be any fixed value and let $\epsilon(n)$ be a function that goes to 0 when $n$ goes to infinity. Then almost surely the following holds:
i) The maximum weight $\max \mathbf{p} \geq \varepsilon(n)n^{\frac{1}{\beta}}$.

ii) The sum of weights beyond $d$ is $\sum_{x \in V} p_x \mathbb{1}_{p_x \geq d} \sim o(n)$.

iii) The volume of $V$ is $\text{vol}(V) = vn \pm o(n)$.

iv) Let $\log n < K \leq 2\sqrt{n}$ be a fixed value. Set

$$c(K) = \begin{cases} 3Z_{\beta}^{\beta - 2\beta} & \text{if } 2.5 \leq \beta \leq 3 \\ \frac{3Z_{\beta}}{\sqrt{\beta - 2\beta}} K^{5 - 2\beta} & \text{if } 2 < \beta < 2.5 \end{cases}$$

Then

$$\sum_{x \in V} p_x^{4 - \beta} \mathbb{1}_{p_x \leq K} \leq c(K)n.$$ 

v) Let $c > 1$ denote a fixed constant value. For any vertex $x \in V$,

$$\sum_{y \in V} p_y \mathbb{1}_{p_y \leq 2\sqrt{n}} \leq 6 \max\left(\frac{c^{\beta - 2\beta}Z_{\beta}}{\beta - 2\beta} npu^{\beta - 2\beta}, \sqrt{n} \log n\right).$$

The proof is via standard concentration inequality (details omitted).

### C.1 Proof of Growth Lemma 4: $\beta > 3$

We first show an upper bound for the expected volume for each level $\Gamma_k(x)$.

**Proof of Part 1.** Let us first fix $\Gamma_k(x)$, and consider the set $\Gamma_{k+1}(x)$. For a vertex $y$, the probability that it is in $\Gamma_{k+1}(x)$ is at most

$$p_y \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)}$$

by Proposition 3. Thus, the expected volume of $\Gamma_{k+1}(x)$ conditioned on $\text{vol}(\Gamma_k(x))$ is at most

$$\mathbb{E}[\text{vol}(\Gamma_{k+1}(x)) \mid \text{vol}(\Gamma_k(x))]$$

$$\leq \sum_{y \not\in N_k(x)} p_y^2 \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)}$$

$$\leq \text{vol}(\Gamma_k(x)) \cdot \frac{\text{vol}(V)}{\text{vol}(V)}$$

$$= \text{vol}(\Gamma_k(x)) \cdot r.$$

On the other hand, $\text{vol}(\Gamma_0(x)) = O(1)$. Thus, we have $\mathbb{E}[\text{vol}(\Gamma_k(x))] = O(r^k)$. $\square$

Next we present the proof for part 2.

**Proof of Part 2.** The proof is split into three steps.
Two fixed constant-weight vertices are close with very low probability

Fix two vertices \( x, y \in S \). By Item 1,
\[
\mathbb{E}[\text{vol}(N_k(x))] \leq O(r^k).
\]
However, for each \( i \), the probability that \( y \) is at distance \( i \) from \( x \) conditioned on \( N_{i-1}(x) \) is at most
\[
\Pr[y \in \Gamma_i(x) \mid N_{i-1}(x)] \leq p_y \cdot \frac{\text{vol}(\Gamma_{i-1}(x))}{\text{vol}(V)}.
\]
The probability \( y \) is within distance \( k+1 \) from \( x \) is at most
\[
\Pr[y \in N_{k+1}(x)] \leq \sum_{i=1}^{k+1} \Pr[y \in \Gamma_i(x)] \leq \frac{p_y}{\text{vol}(V)} \cdot \sum_{i=1}^{k+1} \mathbb{E}[\text{vol}(\Gamma_{i-1}(x))] \leq p_y \cdot \frac{N_k(x)}{\text{vol}(V)} \leq O(r^k).
\]

With large probability, \( \Gamma_{k+1}(x) \) has volume not much smaller than \( \Gamma_k(x) \cdot \frac{\text{vol}_2(V \setminus N_k(x))}{\text{vol}(V)} \).

Conditioned on \( \Gamma_k(x) \), the probability that a vertex \( y \notin N_k(x) \) is in \( \Gamma_{k+1}(x) \) is at least
\[
1 - e^{-p_y \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)}},
\]
by Proposition \( 6 \).

For any \( T > 0 \), we have
\[
\sum_{y : p_y > T} p_y^2 \leq T^{-\gamma} \cdot \sum_{y : p_y > T} p_y^{2+\gamma} \leq \tau \cdot n \cdot T^{-\gamma}.
\]
We also have
\[
\sum_{y : p_y \leq T} p_y^3 \leq T^{1-\gamma} \cdot \sum_{y : p_y \leq T} p_y^{2+\gamma} \leq \tau \cdot n \cdot T^{1-\gamma}.
\]
Let us focus on all \( y \)'s with weight at most \( T \). By that fact that \( 1 - e^{-x} \geq x - x^2/2 \) when \( x \geq 0 \), the expected volume of \( \Gamma_{k+1}(x) \cap \{y : p_y \leq T\} \) conditioned on \( N_k(x) \) is at least
\[
\sum_{y \notin N_k(x) : p_y \leq T} p_y \left( 1 - e^{-p_y \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)}} \right)
\geq \sum_{y \notin N_k(x) : p_y \leq T} p_y^2 \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} - \frac{1}{2} \sum_{y \notin N_k(x) : p_y \leq T} p_y^3 \left( \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} \right)^2
\geq (\text{vol}_2(V \setminus N_k(x))) \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} - \sum_{y : p_y \geq T} p_y^2 \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} - \frac{1}{2} \sum_{y : p_y \leq T} p_y^3 \left( \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} \right)^2
\geq \text{vol}(\Gamma_k(x)) \cdot \frac{\text{vol}_2(V \setminus N_k(x))}{\text{vol}(V)} - \tau \cdot n \cdot T^{-\gamma} \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} - \tau \cdot n \cdot T^{1-\gamma} \cdot \left( \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} \right)^2
\geq \text{vol}(\Gamma_k(x)) \cdot \left( \frac{\text{vol}_2(V \setminus N_k(x))}{\text{vol}(V)} - \tau \cdot \left( T^{-\gamma} + T^{1-\gamma} \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)} \right) \right).
Note that “$y \in \Gamma_{k+1}(x)$” are independent events conditioned on $N_k(x)$ for different $y \notin N_k(x)$. Now we apply Chernoff Bound to lower bound the probability that the volume of $\Gamma_{k+1}(x)$ is too small. The above inequality holds for every $T > 0$. In the following, we set $T = \text{vol}(\Gamma_k(x))^{1/2}$.

When $\text{vol}(\Gamma_k(x)) \leq \text{vol}(V)^{2/3}$, $T^{-\gamma} \geq T^{1-\gamma} \cdot \frac{\text{vol}(\Gamma_k(x))}{\text{vol}(V)}$, the expected volume of $\Gamma_{k+1}(x)$ conditioned on $N_k(x)$ is at least:

$$\mathbb{E}[\text{vol}(\Gamma_{k+1}(x) \cap \{ y : p_y \leq T \}) \mid N_k(x)] \geq \text{vol}(\Gamma_k(x)) \cdot \left( \frac{\text{vol}(V) \setminus N_k(x)}{\text{vol}(V)} - 2\tau \cdot \text{vol}(\Gamma_k(x))^{-\gamma/2} \right).$$

Since each $p_y \leq T = \text{vol}(\Gamma_k(x))^{1/2}$, by Chernoff bound, we have

$$\Pr \left[ \text{vol}(\Gamma_{k+1}(x)) \leq \text{vol}(\Gamma_k(x)) \cdot \left( \frac{\text{vol}(V) \setminus N_k(x)}{\text{vol}(V)} - \text{vol}(\Gamma_k(x))^{-\gamma/3} \right) \mid N_k(x) \right] \leq 2^{-\Theta(\text{vol}(\Gamma_k(x))^{1/2-2\gamma/3})},$$

as long as $\text{vol}(\Gamma_k(x)) = O(n^{2/3})$ and $\text{vol}(\Gamma_k(x))$ sufficiently large.

**With constant probability, $\Gamma_k(x)$ has volume at least $\Omega(r^k)$** Fix a sufficiently large constant $C$, denote by $\mathcal{E}_0$ the event that $x$ has a neighborhood of volume at least $C$. Then it is not hard to verify that for any constant $C$, the probability of $\mathcal{E}_0$ is at least a constant:

$$\Pr[\text{vol}(\Gamma_1(x)) \geq C] \geq \Omega(1).$$

Moreover, for $i \geq 1$, denote by $\mathcal{E}_i$ the event that either

$$\text{vol}(\Gamma_{i+1}(x)) > \text{vol}(\Gamma_i(x)) \cdot \left( \frac{\text{vol}(V) \setminus N_i(x)}{\text{vol}(V)} - \text{vol}(\Gamma_i(x))^{-\gamma/3} \right)$$

or

$$\text{vol}(\Gamma_i(x)) \geq n^{2/3}.$$

By the argument above,

$$\Pr[\mathcal{E}_i \mid N_i(x)] \leq 2^{-\Theta(\text{vol}(\Gamma_i(x))^{1/2-2\gamma/3}).$$

We claim that these events have the following properties.

**Claim 23.** When $\mathcal{E}_i$ occurs for all $0 \leq i < k$, we must have either $\text{vol}(\Gamma_k(x)) \geq \Omega(r^k)$ or $\text{vol}(N_k(x)) \geq n^{2/3}$ for sufficiently large $C$.

**Claim 24.** All events $\mathcal{E}_i$’s ($0 \leq i < k$) occur simultaneously with constant probability.

Before proving the two claims, let us first show that they together imply that $\Pr[\text{vol}(\Gamma_k(x)) \geq \Omega(r^k)] \geq \Omega(1)$.

By Markov’s inequality, the first inequality in the lemma statement and $k \leq \frac{1}{2} \log_r n$, we have

$$\Pr[\text{vol}(N_k(x)) \geq n^{2/3}] \leq O(\sqrt{n}/n^{2/3}) = o(1).$$

34
Therefore, we have the lower bound
\[
\Pr[\text{vol}(\Gamma_k(x)) \geq \Omega(r^k)] \\
\geq \Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \cdot \Pr[\text{vol}(\Gamma_k(x)) \geq \Omega(r^k) \mid \mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \\
\geq \Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \cdot (1 - \Pr[\text{vol}(N_k(x)) \geq n^{2/3} \mid \mathcal{E}_0, \ldots, \mathcal{E}_{k-1}]) \\
\geq \Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \cdot (1 - \Pr[\text{vol}(N_k(x)) \geq n^{2/3}]) / \Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \\
= \Omega(1).
\]
This concludes the proof. \hfill \square

**Proof of Claim 23.** Assume $\mathcal{E}_i$ occurs for all $0 \leq i < k$ and $\text{vol}(N_k(x)) < n^{2/3}$. The goal is to show that in this case, we must have $\text{vol}(\Gamma_k(x)) \geq \Omega(r^k)$.

In particular, $\text{vol}(N_k(x)) < n^{2/3}$ implies that $\text{vol}(N_i(x)) < n^{2/3}$ and $\text{vol}(\Gamma_i(x)) \leq n^{2/3}$ for every $i \leq k$. By Hölder’s inequality, we also have
\[
\text{vol}_2(N_i(x)) \leq \text{vol}_{2+\gamma}(N_i(x))^{1/1+\gamma} \cdot \text{vol}(N_i(x))^{1/3(1+\gamma)} \\
\leq \tau^{1/1+\gamma} \cdot n^{1-\frac{\gamma}{3(1+\gamma)}}.
\]
Thus, the event $\mathcal{E}_i$ ($i > 0$) implies
\[
\text{vol}(\Gamma_{i+1}(x)) > \text{vol}(\Gamma_i(x)) \cdot \left( r - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - \text{vol}(\Gamma_i(x))^{-\gamma/3} \right).
\]
Let $\hat{r} = r - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - C^{-\gamma/3}$. For sufficiently large $C$, $\hat{r} > \sqrt{r} > 1$. First we can show $\text{vol}(\Gamma_i(x)) \geq C \cdot r^{i-1/2}$ for $i \geq 1$ inductively:

- By the definition of $\mathcal{E}_0$, $\text{vol}(\Gamma_1(x)) \geq C$;
- If $\text{vol}(\Gamma_i(x)) \geq C \cdot \hat{r}^{i-1}$, then we have
  \[
  \text{vol}(\Gamma_{i+1}(x)) \geq \text{vol}(\Gamma_i(x)) \cdot \left( r - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - \text{vol}(\Gamma_i(x))^{-\gamma/3} \right) \\
  \geq \text{vol}(\Gamma_i(x)) \cdot \left( r - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - C^{-\gamma/3} \right) \\
  \geq \text{vol}(\Gamma_i(x)) \cdot \hat{r}.
  \]

Thus, we have $\text{vol}(\Gamma_i(x)) \geq \Omega(\hat{r}^i) \geq \Omega(r^{i/2})$. By Equation 12 again, we have
\[
\text{vol}(\Gamma_k(x)) > \text{vol}(\Gamma_{k-1}(x)) \cdot \left( r - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - r^{-(k-1)\gamma/6} \right) \\
\geq \text{vol}(\Gamma_1(x)) \cdot \left( \prod_{i=1}^{k-1} \left( r - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - r^{-i\gamma/6} \right) \right) \\
\geq r^{k-1} \cdot C \cdot \left( \prod_{i=1}^{k-1} \left( 1 - \tau^{1/1+\gamma} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} \cdot r^{-1} - r^{-i\gamma/6-1} \right) \right) \\
= r^k \cdot \alpha_k,
\]
where $\alpha_k$ is decreasing as $k$ increases. $\alpha_{\frac{1}{2} \log_r n}$ is lower bounded by a constant $\alpha$. Thus, $\text{vol}(\Gamma_k(x)) \geq \alpha \cdot r^k \geq \Omega(r^k)$. This proves the claim. \hfill \square
Proof of Claim 24. By Lemma 23, conditioned on $E_0, \ldots, E_{i-1}$, we have either

$$\text{vol}(\Gamma_i(x)) \geq \alpha \cdot r^i$$

or

$$\text{vol}(N_i(x)) \geq n^{2/3}.$$

Thus, by Equation (11), we have

$$\Pr[E_i \mid E_0, \ldots, E_{i-1}] \leq 2^{-\Theta(r^{O(i)})}.$$

Since $\Pr[E_0] = \Omega(1)$, we may lower bound the probability that all events happen simultaneously

$$\Pr[E_0, \ldots, E_{k-1}] \geq \Omega \left( \prod_{i=0}^{k-1} \left( 1 - 2^{-\Theta(r^{O(i)})} \right) \right)$$

$$\geq \Omega(1).$$

\[\square\]

C.2 Growth Lemma for $2 < \beta < 3$

Proof. Let us first upper bound the volume of $k$-neighborhood of any vertex $x$.

Upper bounding $\text{vol}(\Gamma_k(x))$. We will upper bound $\text{vol}(\Gamma_k(x))$ in terms of $\text{vol}(\Gamma_{k-1}(x))$. First, we have $\text{vol}(\Gamma_0(x)) = \text{vol}(x)$ by definition. For any vertex $y$, the probability that it is connected to some vertex in $\Gamma_{k-1}(x)$ is at most

$$\sum_{z \in \Gamma_{k-1}(x)} p_{yz} \cdot \frac{P_z}{\text{vol}(V)} = p_y \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}.$$

Thus, the probability that $\Gamma_k(x)$ contains any vertex with very high weight is low:

$$\Pr \left[ \exists y, p_y \geq (\text{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)} \cdot w, y \in \Gamma_k(x) \mid \text{vol}(\Gamma_{k-1}(x)) \right]$$

$$\leq \sum_{y : p_y \geq (\text{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)} \cdot w} p_y \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}$$

$$\leq O \left( n \cdot \frac{1}{\text{vol}(\Gamma_{k-1}(x))} \cdot w^{\beta-2} \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)} \right)$$

$$= O(1/w^{\beta-2}).$$

That is, the highest weight in $\Gamma_k(x)$ is at most

$$w \cdot (\text{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)}$$

with probability at least $1 - O(1/w^{\beta-2})$. Denote this event by $E_k$. We have

$$\mathbb{E}[\text{vol}(\Gamma_k(x)) \mid E_k, \text{vol}(\Gamma_{k-1}(x))] \leq \sum_{y : p_y \leq (\text{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)} \cdot w} p_y^2 \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}$$

$$\leq O \left( n \cdot \text{vol}(\Gamma_{k-1}(x))^{(3-\beta)/(\beta-2)} \cdot w^{3-\beta} \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)} \right)$$

$$\leq O \left( \text{vol}(\Gamma_{k-1}(x))^{1/(\beta-2)} \cdot w^{3-\beta} \right).$$
By Markov’s inequality, we have

$$\Pr[\text{vol}(\Gamma_k(x)) \geq \text{vol}(\Gamma_{k-1}(x))^{1/(\beta-2) : w} | E_k] \leq O(1/w^{\beta-2}).$$

Since $E_k$ occurs with high probability, by union bound, we have

$$\Pr[\text{vol}(\Gamma_k(x)) \geq \text{vol}(\Gamma_{k-1}(x))^{1/(\beta-2) : w}] \leq O(1/w^{\beta-2}).$$

On the other hand, we have

$$\text{vol}(\Gamma_{k-1}(x))^{1/(\beta-2) : w} \leq b_k^{-1/(\beta-2) : w}$$

$$= (c_{k-1} \cdot w^{1/(\beta-2)(2k-1)}/(3-\beta))^{1/(\beta-2) : w}$$

$$= c_k \cdot w^{1/(\beta-2)2k-1(3-\beta)+1}$$

$$\leq b_k.$$

Thus, we have $\Pr[\text{vol}(\Gamma_k(x)) > b_k] \leq O(1/w^{\beta-2}).$

**Lower Bounding $\text{vol}(\Gamma_k(x))$** For any vertex $y$, if $p_y \cdot p_z \geq \text{vol}(V)$ for some $z \in \Gamma_{k-1}(x)$, then $y$ must be connected to $\Gamma_{k-1}(x)$, otherwise the probability that $y$ does not connect to $\Gamma_{k-1}(x)$ is at most

$$\prod_{z \in \Gamma_{k-1}(x)} \left(1 - \frac{p_y \cdot p_z}{\text{vol}(V)}\right) \leq e^{-\sum_{z \in \Gamma_{k-1}(x)} \frac{p_y \cdot p_z}{\text{vol}(V)}}$$

$$= e^{-p_y \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}}.$$

That is, in either case, if $y \notin N_{k-1}(x)$, then $\Pr[y \notin \Gamma_k(x)] \leq e^{-p_y \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}}$. When $n$ is large enough, we have $b_{k-1} < a_k$, and $N_{k-1}(x)$ does not contain high weight vertex by the premises of the lemma. Therefore, the probability that $\Gamma_k(x)$ contains no high weight vertex is low:

$$\Pr[\forall y, \text{ s.t. } p_y \geq a_k, y \notin \Gamma_k(x)] \leq \prod_{y: p_y \geq a_k} e^{-p_y \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}}$$

$$= e^{-\sum_{y: p_y \geq a_k} p_y \cdot \frac{\text{vol}(\Gamma_{k-1}(x))}{\text{vol}(V)}}$$

$$\leq e^{-\Omega(a_k^{2-\beta} - a_{k-1})}$$

$$= e^{-\Omega(w^{1/(\beta-2)2k-1(3-\beta)-1/(3-\beta)^{2k-2(3-\beta)}})}$$

$$= e^{-\Omega(w^{1/(\beta-2)2k-1})}$$

$$\leq e^{-w} = 1/\log n.$$

In particular, it implies that the probability that $\text{vol}(\Gamma_k(x)) < a_k$ is at most $1/\log n$.

Finally, by union bound, the probability that $\text{vol}(\Gamma_k(x)) \in [a_k, b_k]$ is at least $1 - O(1/w^{\beta-2})$ as the lemma states.
Lower Bounding $\text{dist}(x, y)$ By the above argument, we have

$$\Pr[\text{dist}(x, y) \leq 2d + 3] \leq \Pr[\exists 0 \leq i, j \leq d + 1, \exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v]$$

$$\leq \sum_{i,j=0}^{d+1} \Pr[\exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v]$$

$$= \sum_{i,j=0}^{d+1} \mathbb{E}_{\Gamma_i(x), \Gamma_j(y)} \left[ \Pr[\exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v | \Gamma_i(x), \Gamma_j(y)] \right]$$

$$\leq \sum_{i,j=0}^{d+1} \left( \Pr[\exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v | \Gamma_i(x) \in [a_i, b_i], \Gamma_j(y) \in [a_j, b_j] \right)$$

$$+ \Pr[\Gamma_i(x) \notin [a_i, b_i]] + \Pr[\Gamma_j(y) \notin [a_j, b_j]]$$

$$\leq \sum_{i,j=0}^{d+1} \left( b_i \cdot b_j / \text{vol}(V) + O(1/w^{\beta-2}) \right)$$

$$\leq O(b_{d+1}^2/n + d^2/w^{\beta-2})$$

$$= O(n^{\varepsilon/(\beta-2)} + \log^2 \log \log n / \log^{\beta-2} \log n) = o(1).$$

This concludes the proof. \qed