The Space Complexity of Personalized PageRank and Shortest Path Labeling Schemes

Huacheng Yu* Hongyang Zhang†

July 12, 2018

Abstract

We present tight space lower bounds for static personalized PageRank (PPR) labeling schemes. In a labeling scheme, each vertex is assigned a *label vector*; To answer a query between two vertices, only the labels of the two vertices may be accessed. We show that for any personalized PageRank labeling schemes which can approximately recover PPR values that are at least δ, the total labeling size must be at least $\tilde{\Omega}(n/\sqrt{\delta})$ on sparse Erdös-Rényi random graphs with high probability, where n is the number of vertices in the graph. The result matches the upper bound of Lofgren et al.'16.

We also study shortest path labeling schemes on sparse random graphs with a power law degree distribution of exponent $\beta \in (2, 3)$. We obtain upper and lower bounds which are parameterized by β, and are nearly tight for $\beta \in (2, 2.5)$.

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

*Harvard University, yuhch123@gmail.com

†Stanford University, hongyang@cs.stanford.edu

1 Introduction

Personalized PageRank (PPR) and shortest distance are both commonly used local features in graph based recommendation and search systems [31, 52]. 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 surprisingly low complexity data structures can be efficiently computed in practice, based on the following meta-design [4, 23, 40]: 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 have provided much faster ranking results for personalized search on social graphs [10, 40].

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 a computational lower bound, which says that computing the PPR value between a single pair of vertices requires $\Omega(\sqrt{n})$ time in the worst case (without any pre-processing) [41], where n is the number of vertices. For shortest path, the seminal work of Gavioli et al. [29] shows that for the family of graphs with maximum degree 3, any distance labeling scheme requires $\Omega(n^{1.5})$ space in the worst case. 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 Gavioli et al. constructed. This raises the question whether similar results can be obtained for average case instances.

In this work we introduce a general framework for proving lower bounds for labeling schemes, closing the gap described above. A labeling scheme defines a data access model, where two vertices compute the output using just their own information/labels. While such an access model may seem restrictive, the state of the art performance is achieved via labeling schemes for both personalized PageRank [40] and shortest paths [4, 23] in practice.

We present a tight lower bound for personalized PageRank in the data access model of labeling schemes. Our lower bound matches the existing algorithms [39, 40, 41], and is stated in terms of the desired accuracy threshold — if one starts to care about smaller PPR values, then the lower bound would scale up accordingly.

We then study distance labeling schemes on sparse random graphs with a power law degree distribution. Our techniques yield upper and lower bounds which depend on the power law exponent, and nearly tight in certain regimes. The bounds 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 α, the random walk stops; with probability $1 - \alpha$, it moves 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.[1] We say that an algorithm $f : V \times V \to \mathbb{R}^+$ is (ε, δ)-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.[41], δ is assumed to be larger than c/n for a certain value $c > 1$, to capture PPR values which are above average. And α is assumed to be at least $\Omega(1/\text{polylog}(n))$, so that the expected length of a random walk is not too large. This setting is critical for personalized PageRank to capture enough local graph structures [31].

It's not hard to see that simply by sampling at most $\tilde{O}(\frac{1}{\varepsilon^2 \cdot \delta})$ random walks per vertex, we obtain an (ε, δ)-accurate data structure of total size $\tilde{O}(\frac{n}{\varepsilon^2 \cdot \delta})$. By combining random walk and linear algebraic methods [7], Lofgren et al. [40, 41] improved the storage complexity over the above baseline to $\tilde{O}(\frac{1}{\varepsilon^2}\sqrt{\frac{nm}{\delta}})$, when $\delta \leq n/m$. 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 algorithm simply computes the dot product between $l(x)$ and $l(y)$. The total length of the labeling is simply $\sum_{x \in V} |l(x)|$. Our main result is a matching lower bound to the above algorithm for labeling schemes on sparse graphs, even if the algorithm is only required to approximate the answer within a factor of $\text{poly}\log(n)$.

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

We remark that for graphs where the number of edges $m \gg n$, our results also imply a lower bound of $\tilde{\Omega}(\sqrt{nm/\delta})$, under technical conditions on δ (see Section 4 for details). Since labeling schemes are only a special family of data structures, it would be interesting to see whether one can obtain stronger lower bounds. However, in the more general cell-probe model, it has been notoriously difficult to prove super-logarithmic query time lower bounds [34, 43, 44], even for *non-adaptive* static data structures. Thus we view our work as the first step towards closing the gap between the upper and lower bounds of static personalized PageRank data structures.

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 Appendix C.

We will focus on the Chung-Lu model, which generalizes Erdös-Rényi random graphs to general degree distributions.[2] 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 for each $x \in V$, p_x is drawn independently from a power law distribution with exponent β, e.g. the probability that $p_x = q$ is proportional to $q^{-\beta}$.

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

[1] Equivalently, suppose that with probability α, the random walk teleports to x instead. Then the stationary distribution of this random walk is also equal the personalized PageRank vector of x (see e.g. Chapter 1 in [38]).

[2] Our results can be extended to other random graph models as well (see the discussion in Appendix C for details).

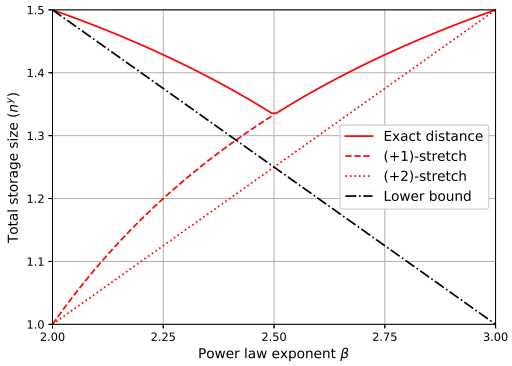

Figure 1: An illustration of the results for shortest path labeling schemes: The x-axis is the exponent of the power law degree distribution and for each value in the y-axis it means that the amount of storage is $\tilde{\mathrm{O}}(n^y)$. The lower bound curve is for exact distances.

details. This lower bound bears resemblance to the seminal work of Gavoille et al. [29], and extends their analysis to obtain average case lower bounds for natural distributions. In the $\Omega(n^{1.5})$ lower bound obtained for distance labeling 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 for the worst case instance via a counting argument. Whereas in our result, we show that the lower bound holds *even for an average case instance* from the Erdös-Rényi random graph distribution with high probability.

For degree distributions with high variance, we present upper and lower bounds on the space complexity of exact distance labeling schemes, which can answer distance queries correctly for all pairs of vertices.

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

Secondly, any exact distance labeling scheme will output a labeling of total length at least $\tilde{\Omega}(n^{\frac{5-\beta}{2}-o(1)})$ for G.

The bounds are nearly tight when β is close to 2, and has a small gap when $\beta < 2.5$. It would be interesting to close the gap when $2.5 < \beta < 3$. We mention that if the labeling scheme is only required to output approximately correct distances, then the amount of space needed can be significantly reduced. Our techniques can also be used to obtain a $(+1)$-stretch scheme where the maximum label length among all the vertices is at most $\tilde{O}(n^{(\beta-2)/(\beta-1)})$, and a $(+2)$-stretch scheme where the maximum label length among all the vertices is at most $\tilde{O}(n^{\beta/2-1})$. See Figure 1 for a summary of the results.

1.3 Intuitions for The Analyses

We show that on an Erdös-Rényi random graph G, each pair of labels convey a certain amount of information about G, as the pair of labels determines their PPR value. To augment the entropy obtained from pairs of vertices, we identify a maximal set of vertices, such that their pairwise PPR values are almost independent — as the number of such vertices k increases, their total label size, which upper bounds the amount of information they can possibly convey, grows linearly in

k, whereas the total entropy of pairwise PPR values grows quadratically. We also discover a tight connection between PPR and shortest distance. 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 "independence" of PPR values between a sufficiently large set of vertices S, we describe an iterative process to "grow" the local neighborhood of each vertex in S. At every iteration of the process, we grow the neighborhood of one 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.

For Theorem 2, the upper bound 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 space complexity increases again to $\tilde{\Theta}(n^{1.5})$. The reason is that the average distance also matters. If there exists lots of short paths that can not be compressed, then the space complexity will increase. The lower bound builds on the insights from Erdös-Rényi random 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.[3] See Appendix C for details.

Organization: The rest of the paper is organized as follows. In Section 2 we give a review of 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. Appendix A and B contain missing proofs from earlier sections. In Appendix C we present distance labeling schemes for $2 < \beta < 3$. Then we present the analysis of the algorithm in Appendix D and also evaluate it experimentally in Appendix E. Appendix F 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 $S \nsim T$ if there does not exist any edge between S and T. For a graph G, let $\mathrm{dist}_G(x, y)$ denote the distance of x and y in G. When there is no ambiguity, we drop the subscript G and simply denote by $\mathrm{dist}(x, y)$ the distance between x and y.

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

2 Preliminaries and Related Work

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

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

[3] The idea is inspired from Chapter 3 in Van Der Hofstad [51].

Thus, p_x is approximately the expected degree of x, and $\mathrm{vol}(V)$ is approximately the expected number of edges (multiplied by two). Let $\mathscr{G}^n(\mathbf{p})$ denote such a probability distribution over $\mathscr{G}^n$, and $G \in \mathscr{G}^n(\mathbf{p})$ denote a sample from the distribution. The following proposition bounds the probability that two sets connect. The proof can be found in Appendix F.1.

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

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

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

Random power law graph: Let $f : [x_{\min}, \infty) \to \mathbb{R}$ denote the probability density function of a power law distribution with exponent $\beta > 1$, i.e. $f(x) = Zx^{-\beta}$, where $Z = (\beta - 1) \cdot x_{\min}^{\beta-1}$ [17]. 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 ν and exponent β). Given the weight vector $\mathbf{p}$, 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 $\mathbf{p}$ has a unique giant component (see e.g. Chung and Lu [16]).[4] In this paper, we will assume that the average degree ν is a constant greater than 1.

2.1 Related work

Landmark based labeling: There is a rich history of study on how to preprocess a graph for answering shortest path queries [6, 18, 13, 50]. 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 $\mathrm{dist}(x, y)$, we simply find a common landmark z in the landmark sets of x and y to minimize the sum of distances $\mathrm{dist}(x, z) + \mathrm{dist}(z, y)$. 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, 45, 46, 47, 53]. We refer the reader to the excellent survey [48] for further reading.

Random graph models: Existing models for social and information networks build on random graphs with a fixed degree distribution [25, 16, 51]. 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

[4] If $\nu < 1$, almost surely all connected components have at most $O(\log n)$ vertices.

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 Erdös-Rényi random 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 [42] 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 $\frac{\mathrm{vol}_2(V)}{\mathrm{vol}(V)}$. More formally, we say that a labeling scheme is K-accurate if for any $x, y \in V$:

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

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

For any integer $1 \leq i \leq n-1$, let $\Gamma_i(x) = \{y \in V : \mathrm{dist}(x, y) = i\}$ denote the set of vertices whose distance from x is equal to i. And let $N_i(x) = \{y \in V : \mathrm{dist}(x, y) \leq i\}$ denote the set of vertices whose distance from x is at most i. Let d be an integer smaller than $K/2$. [6]

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 for is larger than the size of $N_d(x)$, and hence 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 $\mathrm{dist}(x, y) \geq 2d + 1$. In a random graph, the event that $\mathrm{dist}(x, y) > 2d+1$, conditioned on $\mathrm{dist}(x, y) \geq 2d+1$ and $N_d(x)$ and $N_d(y)$ are disjoint, happens with probability

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

by Proposition 3, assuming that $\mathrm{vol}(\Gamma_d(x))\mathrm{vol}(\Gamma_d(y)) \leq o(\mathrm{vol}(V))$. Note that this probability gives us a lower bound on the entropy of the event $\mathbf{1}_{\mathrm{dist}(x,y)>2d+1}$. 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 $\mathbf{1}_{\mathrm{dist}(x,y)>2d+1}$ is large (e.g. $1/\mathrm{poly}\log(n)$ suffices), then we obtain a lower bound on the total labeling size.

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

[5] Note that the average distance of G is $\log n / \log r$ (see e.g. Bollobás [11]).

[6] We assume that K is odd without loss of generality.

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 a large number of (x, y) pairs from S.

b) Use the labels of S to infer whether there are edges between $\Gamma_d(x)$ and $\Gamma_d(y)$, for a large number of (x, y) pairs from S. 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 the first two steps 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. Let $S = \{x_1, x_2, \ldots\}$ be a set of vertices whose weights are roughly close to expected degree of G. 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 $\mathrm{dist}(x_i, x_j) > 2d + 1$, then there is no edge between $L(x_i)$ and $L(x_j)$.

The iterative process: We grow the neighborhood of each vertex in S by an arbitrary but fixed order, up to level d. 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,

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

We then define $F_i = F_{i-1} \cup T(x_i)$ ($F_0 := \varnothing$ by default). Denote by G_{i+1} to be the induced subgraph of G_i on the remaining vertices $V_{i+1} = V \backslash F_i$.

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 $\mathrm{vol}(L(x)) \geq \Omega(r^d)$. We leave the proof to Appendix A.

Lemma 4 (Martingale inequality)**.** *Let $r = \frac{\mathrm{vol}_2(V)}{\mathrm{vol}(V)}$. Let d be an integer and $S \subseteq V$ be a set of vertices whose weight are all within $[\nu, 2\nu]$ and $|S| \leq o(\frac{\mathrm{vol}(V)}{\nu^2 r^{d-1}})$. Assume that*

i) $\Pr[\mathrm{vol}(L(x_i)) \geq \nu \cdot r^d \mid \mathrm{vol}(F_{i-1}) \leq |S|\, \nu \cdot r^d \log n, x_i \in V_i] \geq \Omega(1)$*, for all* $1 \leq i \leq |S|$*;*

ii) $\mathbb{E}[\mathrm{vol}(N_d(x))] \lesssim \nu \cdot r^d$*, for all* $x \in S$*;*

iii) $\Pr[\mathrm{dist}(x, y) \leq d] \lesssim \frac{\nu^2 \cdot r^{d-1}}{\mathrm{vol}(V)}$*, for all* $x, y \in S$*.*

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

For the rest of this section, we show how to implement the three step proof plan, for distance labeling on random graphs with $\beta > 3$. For personalized PageRank, we need to show in step b) that PPR values can infer distance information. And for distance 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 Appendix C for details.

3.1 The $\beta > 3$ case

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

Proposition 5 (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 $\mathrm{vol}(L(x)) \geq \Omega(r^d)$.*

The proof is left to Appendix A. Now we state the main result of this section.

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

Proof. 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 argue that for each group S,

$$\Pr[\text{The total label length of } S \leq c_2 \cdot r^{-2c} n] \leq o(1). \tag{1}$$

Hence by Markov's inequality, except for $o(n/r^{K/2})$ groups, all the other groups will have label size at least $\tilde{\Omega}(n)$. For the rest of the proof, we focus on an individual group S.

Given the labels of S, we can recover all pairwise distances which are less than K in S. Let $\mathrm{dist}_S : S \times S \to \mathbb{N}$ denote the distance function restricted to S. Consider the following two cases:

a) $\exists c_1^2 \cdot |S|^2/4$ pairs (x_i, x_j) such that $\mathrm{dist}_S(x_i, x_j) \leq 2d + 1$. By Lemma 9, we know that $\Pr[\mathrm{dist}(x_i, x_j) \leq 2d + 1] = O(r^{2d}/n)$, 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 $\mathrm{dist}_S(x_i, x_j) \leq 2d + 1$ is at most $c_1^2 \cdot |S|^2/4$ in S. Let

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

By Lemma 4, the size of A is at least $\binom{c_1|S|}{2} - c_1^2 |S|^2/4 \geq c_1^2 |S|^2/5$. 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.

$$\begin{aligned}
&\Pr\left[\mathrm{dist}_S(x, y) > 2d + 1,\ \forall (x, y) \in A \mid \{T_i\}_{i=1}^{|S|}\right] \\
&\leq \prod_{(x,y)\in A} \Pr\left[L(x) \not\sim L(y) \mid L(x) \cap L(y) = \varnothing, \text{ and } \mathrm{vol}(L(x)), \mathrm{vol}(L(y)) \geq \Omega(r^d)\right] \\
&\leq \prod_{(x,y)\in A} \exp\left(-\frac{\mathrm{vol}(L(x))\mathrm{vol}(L(y))}{\mathrm{vol}(V)}\right) \qquad \text{(by Proposition 3)} \\
&\leq \exp\left(-\Omega\left(\frac{r^{2d}}{n}\right)\right)^{c_1^2|S|^2/5} \leq \exp(-\Omega(r^{-2c}n)).
\end{aligned}$$

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 combining the two cases, we have shown that Equation (1) is true. Hence the proof is complete. □

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 space complexity of labeling schemes for personalized PageRank. The key intuition is based on the following lemma, which states that in an Erdös-Rényi random graph, personalized PageRank values are closely related to distances.

Lemma 7. *Let $G = (V, E)$ be an Erdös-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 surly, for all pair of vertices $x, y \in V$ such that $\mathrm{dist}(x, y) \leq T$, we have $\pi(x, y) \geq \frac{\alpha}{3}\left(\frac{1-\alpha}{np}\right)^T$.*

The proof can be found at Appendix B. Next we describe our main result. The analysis builds on Lemma 7 and the insights from Section 3, and can be found in Appendix B.

Theorem 8. *Let $G = (V, E)$ be an Erdös-Rényi random graph where every edge is sampled independently with probability $p \geq \frac{\log^4 n}{n}$. 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 the personalized PageRank random walk on G. For any (ε, δ)-accurate labeling data structures 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*

$$d = \left\lfloor \frac{1}{2}\left(\frac{\log \frac{\alpha}{3t^2\delta}}{\log \frac{np}{1-\alpha}} - 1\right)\right\rfloor.$$

To see that Theorem 1 follows from Theorem 8, let $p = \log^4 n/n$. We have that

$$\begin{aligned} d + 1 &\geq \frac{1}{2} \cdot \frac{\log \frac{\alpha}{3t^2\delta}}{\log np}\left(1 + \frac{\log(1-\alpha)}{\log np - \log(1-\alpha)}\right) - \frac{1}{2} \\ &\geq \frac{1}{2} \cdot \frac{\log \frac{\alpha}{3t^2\delta}}{\log np} \cdot \left(1 - \frac{2\alpha}{\log np}\right) - \frac{1}{2}, \end{aligned} \tag{2}$$

for large enough n and $\alpha < 1/2$. Since $1 \leq t$ and $\delta \leq \frac{\log n}{n}$, we have $\frac{\alpha}{3t^2\delta} \leq n$. Then

$$(np)^{\frac{\log \frac{\alpha}{3t^2\delta}}{\log np} \cdot \frac{\alpha}{\log np}} \leq \exp\left(\log n \cdot \frac{\alpha}{\log np}\right) \leq O(1), \tag{3}$$

where the second inequality is because $\alpha \leq \frac{\log np}{2 \log n}$. Hence by combining Equation (2) and (3), we obtain that

$$n \cdot (np)^{d+1} \geq \tilde{\Omega}\left(n \cdot \sqrt{\frac{\alpha}{3t^2\delta}}\right) \geq \tilde{\Omega}\left(n/\sqrt{\delta}\right),$$

since $t, \alpha^{-1} \lesssim \mathrm{poly}\log(n)$. As another remark, when $\frac{\alpha}{3t^2\delta}$ is an odd integer power of $\frac{np}{1-\alpha}$, the calculations from Equation 2 and 3 also shows that

$$n \cdot (np)^{d+1} \geq \tilde{\Omega}\left(n \cdot \sqrt{np \cdot \left(\frac{\alpha}{3t^2\delta}\right)^{1 - \frac{\log 1-\alpha}{\log np - \log 1-\alpha}}}\right) \geq \tilde{\Omega}\left(\sqrt{\frac{nm}{\delta} \cdot \frac{\alpha}{3t^2}}\right) \geq \tilde{\Omega}\left(\sqrt{nm/\delta}\right),$$

since $m = \Theta(n^2 p)$ with high probability. Hence for general number of edges m, we obtain a lower bound which also matches the upper bound of Lofgren et al. [40].

5 Conclusions

In this work, we presented a general framework to show lower bounds for labeling schemes. The hard instance is based on random graph distributions. Our lower bound for personalized PageRank labeling schemes matches the previous upper bound. 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 the techniques may shed light on obtaining lower bounds for other data structure problems, where labeling schemes have achieved successful performance. It would be interesting to further understand the time-space trade-off for labeling schemes. Another direction is to consider random graphs as a hard instance for studying other problems such as the space complexity of set intersections [49].

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. Hongyang Zhang is supported by NSF grant 1447697.

References

[1] Ittai Abraham, Daniel Delling, Andrew V Goldberg, and Renato F Werneck. A hub-based labeling algorithm for shortest paths in road networks. In *International Symposium on Experimental Algorithms*, pages 230–241. Springer, 2011.

[2] Ittai Abraham, Amos Fiat, Andrew V Goldberg, and Renato F Werneck. Highway dimension, shortest paths, and provably efficient algorithms. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 782–793. Society for Industrial and Applied Mathematics, 2010.

[3] Ittai Abraham and Cyril Gavoille. On approximate distance labels and routing schemes with affine stretch. In *International Symposium on Distributed Computing*, pages 404–415. Springer, 2011.

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

[5] Stephen Alstrup, Søren Dahlgaard, Mathias Bæk Tejs Knudsen, and Ely Porat. Sublinear distance labeling. *arXiv preprint arXiv:1507.02618*, 2015.

[6] Ingo Althöfer, Gautam Das, David Dobkin, Deborah Joseph, and José Soares. On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 9(1):81–100, 1993.

[7] Reid Andersen, Christian Borgs, Jennifer Chayes, John Hopcraft, Vahab S Mirrokni, and Shang-Hua Teng. Local computation of pagerank contributions. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 150–165. Springer, 2007.

[8] Haris Angelidakis, Yury Makarychev, and Vsevolod Oparin. Algorithmic and hardness results for the hub labeling problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1442–1461. Society for Industrial and Applied Mathematics, 2017.

[9] Maxim Babenko, Andrew V Goldberg, Haim Kaplan, Ruslan Savchenko, and Mathias Weller. On the complexity of hub labeling. In *International Symposium on Mathematical Foundations of Computer Science*, pages 62–74. Springer, 2015.

[10] Bahman Bahmani and Ashish Goel. Partitioned multi-indexing: bringing order to social search. In *Proceedings of the 21st international conference on World Wide Web*, pages 399–408. ACM, 2012.

[11] Béla Bollobás. Random graphs. In *Modern Graph Theory*, pages 215–252. Springer, 1998.

[12] Michele Borassi, Pierluigi Crescenzi, and Luca Trevisan. An axiomatic and an average-case analysis of algorithms and heuristics for metric properties of graphs. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 920–939. SIAM, 2017.

[13] Shiri Chechik. Approximate distance oracles with improved bounds. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 1–10. ACM, 2015.

[14] Wei Chen, Christian Sommer, Shang-Hua Teng, and Yajun Wang. Compact routing in power-law graphs. In *International Symposium on Distributed Computing*, pages 379–391. Springer, 2009.

[15] Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.

[16] Fan RK Chung and Linyuan Lu. *Complex graphs and networks*, volume 107. American mathematical society Providence, 2006.

[17] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[18] Edith Cohen. Size-estimation framework with applications to transitive closure and reachability. *Journal of Computer and System Sciences*, 55(3):441–453, 1997.

[19] Edith Cohen, Daniel Delling, Fabian Fuchs, Andrew V Goldberg, Moises Goldszmidt, and Renato F Werneck. Scalable similarity estimation in social networks: Closeness, node labels, and random edge lengths. In *Proceedings of the first ACM conference on Online social networks*, pages 131–142. ACM, 2013.

[20] Edith Cohen, Eran Halperin, Haim Kaplan, and Uri Zwick. Reachability and distance queries via 2-hop labels. *SIAM Journal on Computing*, 32(5):1338–1355, 2003.

[21] Hagai Cohen and Ely Porat. On the hardness of distance oracle for sparse graph. *arXiv preprint arXiv:1006.1117*, 2010.

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

[23] Daniel Delling, Andrew V Goldberg, Thomas Pajor, and Renato F Werneck. Robust distance queries on massive networks. In *European Symposium on Algorithms*, pages 321–333. Springer, 2014.

[24] Daniel Delling, Andrew V Goldberg, Ruslan Savchenko, and Renato F Werneck. Hub labels: Theory and practice. In *International Symposium on Experimental Algorithms*, pages 259–270. Springer, 2014.

[25] Richard Durrett. *Random graph dynamics*, volume 200. Citeseer, 2007.

[26] Nicole Eikmeier and David F Gleich. Revisiting power-law distributions in spectra of real world networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 817–826. ACM, 2017.

[27] Mihaela Enachescu, Mei Wang, and Ashish Goel. Reducing maximum stretch in compact routing. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008.

[28] Cyril Gavoille, Christian Glacet, Nicolas Hanusse, and David Ilcinkas. Brief announcement: Routing the internet with very few entries. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing*, pages 33–35. ACM, 2015.

[29] Cyril Gavoille, David Peleg, Stéphane Pérennes, and Ran Raz. Distance labeling in graphs. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 210–219. Society for Industrial and Applied Mathematics, 2001.

[30] Andrew V Goldberg, Ilya Razenshteyn, and Ruslan Savchenko. Separating hierarchical and general hub labelings. In *International Symposium on Mathematical Foundations of Computer Science*, pages 469–479. Springer, 2013.

[31] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. Wtf: The who to follow service at twitter. In *Proceedings of the 22nd international conference on World Wide Web*, pages 505–514. ACM, 2013.

[32] Ken-ichi Kawarabayashi, Philip N Klein, and Christian Sommer. Linear-space approximate distance oracles for planar, bounded-genus and minor-free graphs. In *International Colloquium on Automata, Languages, and Programming*, pages 135–146. Springer, 2011.

[33] Tamara G Kolda, Ali Pinar, Todd Plantenga, and Comandur Seshadhri. A scalable generative graph model with community structure. *SIAM Journal on Scientific Computing*, 36(5):C424–C452, 2014.

[34] Kasper Green Larsen. Higher cell probe lower bounds for evaluating polynomials. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 293–301, 2012.

[35] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research*, 11(Feb):985–1042, 2010.

[36] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

[37] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. Statistical properties of community structure in large social and information networks. In *Proceedings of the 17th international conference on World Wide Web*, pages 695–704. ACM, 2008.

[38] Peter Lofgren. Efficient algorithms for personalized pagerank. *arXiv preprint arXiv:1512.04633*, 2015.

[39] Peter Lofgren, Siddhartha Banerjee, and Ashish Goel. Bidirectional pagerank estimation: From average-case to worst-case. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 164–176. Springer, 2015.

[40] Peter Lofgren, Siddhartha Banerjee, and Ashish Goel. Personalized pagerank estimation and search: A bidirectional approach. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 163–172. ACM, 2016.

[41] Peter A Lofgren, Siddhartha Banerjee, Ashish Goel, and C Seshadhri. Fast-ppr: scaling personalized pagerank estimation for large graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1436–1445. ACM, 2014.

[42] Shay Mozes and Christian Sommer. Exact distance oracles for planar graphs. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 209–222. SIAM, 2012.

[43] Rina Panigrahy, Kunal Talwar, and Udi Wieder. Lower bounds on near neighbor search via metric expansion. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 805–814, 2010.

[44] Mihai Patrascu. Lower bounds for 2-dimensional range counting. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 40–46, 2007.

[45] Mihai Patrascu and Liam Roditty. Distance oracles beyond the thorup-zwick bound. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 815–823. IEEE, 2010.

[46] Mihai Patrascu, Liam Roditty, and Mikkel Thorup. A new infinity of distance oracles for sparse graphs. In *Foundations of Computer Science (focs), 2012 Ieee 53rd Annual Symposium on*, pages 738–747. IEEE, 2012.

[47] Ely Porat and Liam Roditty. Preprocess, set, query! *Algorithmica*, 67(4):516–528, 2013.

[48] Christian Sommer. Shortest-path queries in static networks. *ACM Computing Surveys (CSUR)*, 46(4):45, 2014.

[49] Christian Sommer, Elad Verbin, and Wei Yu. Distance oracles for sparse graphs. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 703–712. IEEE, 2009.

[50] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *Journal of the ACM (JACM)*, 52(1):1–24, 2005.

[51] Remco Van Der Hofstad. Random graphs and complex networks. *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf*, page 11, 2009.

[52] Akhil Wable. Intro to facebook search. `https://code.facebook.com/posts/422334894553532/intro-to-facebook-search/`.

[53] Christian Wulff-Nilsen. Approximate distance oracles with improved preprocessing time. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 202–208. SIAM, 2012.

A Missing Proofs from Section 3

In this section we present the missing proofs from Section 3. 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 Section F.2 for the completeness of this paper.

Proposition 9 (Growth rates for $\beta > 3$)**.** *Let $\mathscr{G}^n(\mathbf{p})$ be a random graph model with weight sequence $\mathbf{p}$ satisfying the following properties:*

- $\mathrm{vol}(V) = (1 + o(1))\nu \cdot n$ *for some constant* ν*;*
- $\mathrm{vol}_2(V) = (1 + o(1))\omega \cdot n$ *for some constant* ω*;*
- $\mathrm{vol}_{2+\gamma}(V) = \tau \cdot n$ *for some positive constant* $\gamma < 1/2$ *and* τ*, where* $\mathrm{vol}_{2+\gamma}(S) := \sum_{x \in S} p_x^{2+\gamma}$*;*
- *The growth rate* $r = \frac{\mathrm{vol}_2(V)}{\mathrm{vol}(V)}$ *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. $\mathbb{E}[\mathrm{vol}(\Gamma_k(x))] = O\left(r^k\right)$ *for every* $k \le \log_r n$*;*
2. $\Pr[\mathrm{vol}(\Gamma_k(x)) \ge \Omega\left(r^k\right)] \ge \Omega(1)$ *for every positive integer* $k \le \frac{1}{2}\log_r n$*.*

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

Based on the above proposition, we present the proof of Lemma 4, which is a martingale inquality for the iterative neighborhood growth.

Proof of Lemma 4. Consider the following random variable, for any $1 \le i \le |S|$.

$$X_i := \begin{cases} 1 & \text{if } x_i \notin V_i, \text{ or } \mathrm{vol}(F_{i-1}) > |S|\,\nu \cdot r^d \log n, \text{or } \mathrm{vol}(L(x_i)) \ge \Omega(r^d) \\ 0 & \text{otherwise.} \end{cases}$$

We have $\Pr[X_i = 1 \mid X_1, \ldots, X_{i-1}] \ge \Omega(1)$ by Assumption i). Thus by Azuma-Hoeffding inequality, $\sum_{i=1}^{|S|} X_i \ge \Omega(|S|)$ 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 $x_i \notin V_i$ is $o(|S|)$ with high probability. Note that $x_i \notin V_i$ implies that there exists some vertex $j < i$ such that $\mathrm{dist}(x_i, x_j) \le d$. On the other hand, for any two vertices $x, y \in S$, $\Pr[\mathrm{dist}(x, y) \le d] \le O(\nu^2 \cdot r^{d-1}/\mathrm{vol}(V))$, by Assumption iii). Hence, the expected number of vertex pairs in S whose distance is at most d, is $O(|S|^2\,\nu^2 \cdot r^{d-1}/\mathrm{vol}(V)) \le o(|S|)$, 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 $x_i \notin V_i$.

Secondly, for all $1 \le i \le |S|$, $\mathrm{vol}(F_i) \le |S|\,\nu \cdot r^d \log n$ 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

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

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 $\mathrm{vol}(N_d(S)) > |S|\,\nu \cdot r^d \log n$ is at most $\log^{-1} n$. This proves the lemma. □

Next we present the proof of Proposition 5, which instantiates the martingale inequality towards proving the distance labeling lower bound for $\beta > 3$.

Proof of Proposition5. It's easy to verify that $|S| \cdot \nu^2 r^{d-1} \le n/r^c \le o(n)$. It suffices to verify the assumptions required in Lemma 4. Note that Assumption ii) and iii) simply follows from Proposition 9. 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

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

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

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

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

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

by straightforward calculation. Hence r_i is a constant strictly greater than 1. By Proposition 9, with constant probability $\mathrm{vol}(L_i) \ge \Omega(r_i^d) \ge \Omega(r^d)$, because

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

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

B Missing Proofs from Section 4

In this section, we present the missing proofs from Section 4. First, we present the proof of Lemma 7, which relates PPR values to shortest distance on Erdös-Rényi random graphs.

Proof of Lemma 7. By Chernoff bound, for each vertex x, its degree is close to np with high probability:

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

Hence when $p \geq \log^4 / n$, by union bound, almost surly 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 surly 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 $\mathrm{dist}(x, y) \leq T$, there is a path $P_{x,y}$ from x to y with length at most T. Thus, for all such x and y, by the random walk definition of personalized PageRank,

$$\begin{aligned}
\pi(x, y) &\geq \alpha \cdot \pi(P_{x,y}) \\
&\geq \alpha \cdot \left(\frac{1-\alpha}{(1+1/\log n)np}\right)^T \\
&\geq \frac{\alpha}{3} \cdot \left(\frac{1-\alpha}{np}\right)^T.
\end{aligned}$$

This proves the lemma. □

Based on the above Lemma, next we present the analysis for our main result.

Proof of Theorem 8. Divide V to groups of size $\frac{n}{(np)^d \log^2 n}$. We will show that for each group S,

$$\Pr[\text{The total label size of } S \leq c_2 \,|S| \cdot (np)^{d+1} \log^{-2} n] \leq o(1). \tag{4}$$

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 at least $c_2 \,|S| \cdot (np)^{d+1} \log^{-2} n$. And this implies 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 Erdös-Rényi random graphs can be generated from the Chung-Lu model where each vertex has weight np. Hence the average weight is equal to $\nu = np$, $\mathrm{vol}(V) = n^2 p$ and the growth rate r is $\frac{\mathrm{vol}_2(V)}{\mathrm{vol}(V)} = np$. And we obtain that $|S| \leq o\left(\frac{\mathrm{vol}(V)}{\nu^2 r^{d-1}}\right)$. We also verify the assumptions required to in Lemma 4. Note that the second and third assumption follows from straightforward calculation, hence we omit the details. For the first assumption, conditional on $\mathrm{vol}(F_{i-1}) \leq |S|\, \nu \cdot r^d \log n$, we have that the size of F_{i-1} is at most $|S|\, r^d \log n \leq n/\log n$, because every vertex of F_{i-1} has weight equal to np. Thus the number of vertices in V_i is at least $n(1 - O(\frac{1}{\log n}))$, and the growth rate of x_i on G_i is $r_i = np(1 - O(\frac{1}{\log n}))$. By Chernoff bound, the neighborhood growth on G_i concentrates at rate r_i with high probability because $r_i \geq \log^4(n)(1 - O(\frac{1}{\log n}))$. Hence we obtain that the size of $L(x_i)$ is at least $\Theta(r_i^d) \geq \Theta\left(r^d \cdot \left(1 - O(\frac{1}{\log n})\right)^d\right) \gtrsim \Theta(r^d)$ with high probability, since $d \leq \log n$.

Thus by Lemma 4, 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 $c_1^2|S|^2/4$ 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 Erdös-Rényi random graph is symmetric in V. Thus for fixed x, y, $\mathbb{E}_G[\pi(x,y)] = 1/(n-1)$. By Markov's inequality, $\Pr[\pi(x,y) > \delta] < \frac{1}{(n-1)\delta}$. Hence the expected number of pairs $(x,y) \in S \times S$ such that $\pi(x,y) > \delta$ is at most $\frac{|S|^2}{(n-1)\delta} \lesssim |S|^2 / \log n$, 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 $c_1^2 |S| /5$. To see this, note that there are at least $\binom{|S'|}{2} - c_1^2 |S|^2 /4 \geq c_1^2 |S|^2 /5$ vertex pairs $(x,y) \in S' \times S'$ where $\pi(x,y) \leq \delta$. For any such pair, by the (ε, δ)-accuracy guarantee, we have that $\tilde{\pi}(x,y) \leq t\delta$. Hence the the size of B is at least $c_1^2 |S|^2 /5$.

Next we show that for each $(x,y) \in B$, $\mathrm{dist}(x,y) \geq 2d+2$. Assume that $\mathrm{dist}(x,y) \leq 2d+1$. By Lemma 7, with high probability for all vertex pairs where $\mathrm{dist}(x,y) \leq 2d+1$, we have that $\pi(x,y) \geq \frac{\alpha}{3} \left(\frac{1-\alpha}{np}\right)^{2d+1}$ holds. On the other hand,

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

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

Lastly, since for each $(x,y) \in B$, $\mathrm{dist}(x,y) \geq 2d+2$, 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

$$\begin{aligned}&\Pr[L(x) \nsim L(y)\ \forall (x,y) \in B \mid \{L(x)\}_{x \in S}] \\ &\leq \left((1-p)^{\Omega((np)^{2d})}\right)^{c_1^2|S|^2/5} \leq \exp\left(-\Theta\left(p \cdot (np)^{2d}|S|^2\right)\right).\end{aligned}$$

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 $c_2 \cdot p(np)^{2d}|S|^2 = c_2 |S| \cdot (np)^{d+1} \log^{-2} n$, we obtain that Equation (4) is true for case b).

By taking a union bound over the two cases, we proved that Equation (4) is true. Hence the proof is complete. □

C 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 [15]. 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 $\mathscr{G}^n(\mathbf{p})$ 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 $\mathscr{G}^n(\mathbf{p})$, with high probability there exists an exact distance labeling scheme F such that $|F(x)| \lesssim n^{1-\min(\frac{1}{\beta-1}, \frac{1}{4-\beta})} \log^3 n$ for all $x \in V$.*

Algorithm 1 AlgSkewDegree

Input: An undirected graph $G = (V, E)$; Parameter $K = \sqrt{n}$ if $2.5 \le \beta \le 3$, or $n^{\frac{1}{(4-\beta)(\beta-1)}}$ if $2 < \beta < 2.5$.

Let $H = \{x \in V : d_x \ge K\}$
for $x \in V$ **do**
  $(F(x), l(x)) =$ AlgBfs(x, $4\nu \log^2 n \times n^{\frac{\beta-2}{\beta-1}}$) (add closest $\tilde{O}(n^{\frac{\beta-2}{\beta-1}})$ vertices)
  $F(x) = F(x) \cup H$ (add all high degree vertices)
  for $y \in \Gamma_{l(x)-1}(x)$ **do**
    if $d_y \le K$ **then**
      $F(x) = F(x) \cup \{z \in N(y) : d_y \le d_z \le K\}$ (add all neighbors of y with a higher degree)
    end if
  end for
end for

procedure AlgBfs(x, t)
  $S = \{x\}$
  $\alpha_0(x) = d_x$; $k = 0$
  while $\alpha_k(x) \le t \wedge |\Gamma_{k+1}(x)| > 0$ **do**
    $S = S \cup \Gamma_k(x)$
    $Y = \{(y, z) \in E : y \in \Gamma_k(x), z \in \Gamma_{k+1}(x)\}$
    $k = k + 1$
    $\alpha_k(x) = \sum_{y \in \Gamma_k(x)} d_y - |Y|$ (number of edges between $\Gamma_k(x)$ and $V \setminus N_{k-1}(x)$)
  end while
  return (S, k) (k is the first integer that satisfies: $\alpha_{k-1}(x) > t$ or $\Gamma_k(x) = \varnothing$)
end procedure

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

$$\min_{z \in F(x) \cap F(y)} \mathrm{dist}(x, z) + \mathrm{dist}(z, y).$$

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 $O(|F(x)| + |F(y)|)$ 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 do a breadth-first search, but stop right before the boundary size exceeding $\tilde{\Theta}(n^{\frac{\beta-2}{\beta-1}})$, and put all vertices that we have explored in the landmark set. We claim that this is a $(+1)$-sketch labeling.

To see this, for two vertices x, y, if their landmark sets intersect with each other, then we can already compute their distances correctly from their labels. Otherwise, the bottom layer of x and y have distance at most two (through the heavy vertex) with high probability. Therefore, we only have to check whether they have distance one. To resolve the $(+1)$-stretch, for each vertex on the boundary, we add all of its neighbors with a higher degree to the landmark set. Clearly, this fixes the $(+1)$-stretch, if there is an edge connecting the two boundaries. To bound the landmark size, we refer the reader to Appendix D. Algorithm 1 describes the whole 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 E, we test our algorithm on real world graphs. We found that our algorithm achieves accurate results — the 80%-percentitle 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 $f^{\text{IN}}(x)$ and $f^{\text{OUT}}(x)$ with mean value bigger than 1, representing the indegree and outdegree distributions, respectively. Each node v is associated with two parameters $p_v^{\text{IN}} \sim f^{\text{IN}}(\cdot)$ and $p_v^{\text{OUT}} \sim f^{\text{OUT}}(\cdot)$. For any two nodes u and v, there is a directed edge from u to v with probability $\frac{p_u^{\text{OUT}} \cdot p_v^{\text{IN}}}{M}$, 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. If 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 distance labeling scheme can be obtained. It would be interesting to come up with a deterministic graph model when the degree distribution has high variance, and then characterize the distance labeling on such a graph model.

C.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, 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 of $(\beta-2)^{-1}$, such that with high probability the neighborhood growth will follow the specified ranges. Let $d = \frac{\log\log\log n}{\log \frac{1}{\beta-2}}$, $\varepsilon = 1/\log\log n$ and $w = \log\log n$. For all $0 \le i \le d+1$, let

$$\mu_i := n^{(\frac{\beta}{2}-1-\varepsilon)(\beta-2)^{d-i}}, \text{and}$$
$$\sigma_i := w^{\frac{1}{(\beta-2)^{2i}(3-\beta)}},$$

μ_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. Initially, μ_0 is on the order of $n^{\Theta(\frac{1}{\log\log n})}$. At the end of the process, μ_{d+1} is on the order of $n^{\frac{1}{2}-O(\varepsilon)}$.

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

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

Then the following statements are true:

a) Following a good path: Let x be a fixed vertex and $1 \le k \le d+1$. Suppose that $\mathrm{vol}(\Gamma_i(x)) \in [a_i, b_i]$ for any $1 \le i < k$, then $\mathrm{vol}(\Gamma_k(x)) \in [a_k, b_k]$, with probability at least $1 - O(w^{2-\beta})$, 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 $\Pr[\mathrm{dist}(x,y) \le 2d+3] = o(1)$.

We leave the proof to Appendix F.3. Next we state the lower bound for shortest path labeling.

Theorem 12. *Let $G = (V, E) \in \mathscr{G}^n(\mathbf{p})$ be a sparse 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 $\Omega(n^{\frac{5-\beta}{2}-o(1)})$.*

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 $S = \{x_1, x_2, \dots\}$ be an arbitrary vertex set of size $n^{\frac{3-\beta}{2}}$ 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 $1 \le i \le |S|$, we consider the following inductive process:

1. If $x_i \in V_i$, let $1 \le \lambda_i \le d$ be the maximum k that still satisfy $\mathrm{vol}(\Gamma_k(x_i)) \in [a_k, b_k]$ 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 $L(x_i) = \varnothing$;

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 $F_i = F_{i-1} \cup T_i$ ($F_0 = \varnothing$ by default). Let G_{i+1} be the subgraph of G_i on remaining vertices $V_{i+1} = V_i \backslash T_i$. 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 $\Theta(n^{\frac{3-\beta}{2}})$ vertices x_i in S satisfy that $\mathrm{vol}(L(x_i)) \in [a_d, b_d]$.*

Proof. Consider the following random variables.

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

We show that $X_i = 1$ with high probability for all $1 \le i \le |S|$. We first verify that $\Pr[x_i \notin V_i] = \Pr[x_i \in F_{i-1}] \le o(1)$. Consider any vertex $z \in V$ and $1 \le j \le i-1$, by Proposition 3 we have that

$$\begin{aligned} \Pr[z \in T_j] &\le \sum_{l=0}^{d-1} \frac{p_z \cdot b_l}{\mathrm{vol}(V_j)} \\ &\le p_z \cdot n^{(\beta/2-1-\varepsilon)(\beta-2)-1} \cdot w^{O(\log^2 \log n)}. \end{aligned}$$

Thus, by union bound over $1 \le j \le i-1$, we have

$$\Pr[z \in F_{i-1}] \le \sum_{j=1}^{i-1} \Pr[z \in T_j] \le p_z \cdot n^{\frac{1}{2}(\beta^2-5\beta+5)-\varepsilon(\beta-2)+o(\varepsilon)} = p_z \cdot n^{\lambda}, \tag{5}$$

i.e. denote the exponent by λ 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, and the initial weight sequence of $V_1 = V$ 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:

$$\begin{aligned} E[\mathrm{vol}(F_{i-1})] &\le \sum_{z \in V} p_z \cdot \min\{1, p_z \cdot n^{\lambda}\} \\ &= \sum_{z: p_z \ge n^{-\lambda}} p_z + \sum_{z: p_z < n^{-\lambda}} p_z^2 \cdot n^{\lambda} \\ &\lesssim n^{1-\lambda(2-\beta)} + n^{1-\lambda(3-\beta)+\lambda} \lesssim n^{1-\frac{1}{2}(\beta-2)^2-\Theta(\varepsilon)}, \end{aligned}$$

where the second inequality is by applying the Assumption of Proposition 11 with $t = n^{-\lambda}$, and the last step is because $1 + \lambda(\beta-2) \le 1 - \frac{1}{2}(\beta-2)^2 - \Theta(\varepsilon)$. 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 $t \le n^{\beta/2-1}$.

Now we can apply Proposition 11 on G_i to obtain that $\Pr[X_i = 0] \lesssim w^{2-\beta} \le o(1)$, for any $1 \le i \le |S|$. By Markov's inequality, $\sum_{i=1}^{|S|}(1 - X_i) \lesssim n^{\frac{3-\beta}{2}}$ with high probability. Hence at least $\Theta(n^{\frac{3-\beta}{2}})$ vertices in S satisfy that $\mathrm{vol}(L(x_i)) \in [a_d, b_d]$ 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 $A = \{x \in V : p_x \in [(\frac{a_d}{w})^{1/(\beta-2)}, 2(\frac{a_d}{w})^{1/(\beta-2)}]\} \setminus F_{|S|}$. We show that there exists a constant fraction of vertices x in S such that $L(x)$ is connected to a *different* vertex in A for each $x \in V$.

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 $|S'| \gtrsim n^{\frac{3-\beta}{2}}$, and for every $x \in S'$, $h(x)$ 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 $2(\frac{a_d}{w})^{1/(\beta-2)} < o(n^{1/2})$ belongs to $F_{|S|}$ with probability at most $O(n^{\frac{1}{2}(\beta-2)(\beta-3)}) \le o(1)$. Thus, the volume of A satisfies that $\mathrm{vol}(A) \gtrsim nw/a_d$ with high probability. This implies that

$$|A| \gtrsim \frac{nw}{a_d \cdot (\frac{a_d}{w})^{1/(\beta-2)}} \ge \frac{n}{a_d^{1/(\beta-2)}} \cdot w^{(\beta-1)/(\beta-2)} \gtrsim n^{(3-\beta)/2} \cdot w.$$

By Proposition 13, there exists a set $S_1 \subseteq S$ of size $\Theta(|S|)$ such that for all $x \in S_1$, $\mathrm{vol}(L(x)) \in [a_d, b_d]$. 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 $\Theta(n^{\frac{3-\beta}{2}})$ with high probability. This is because $\mathrm{vol}(L(x)) \geq a_d$ by Proposition 13 for all $x \in S_1$, hence by Proposition 3

$$\Pr[L(x) \nsim y \mid \forall y \in A, \text{ s.t. } y \text{ is "unused"}] \lesssim \exp\left(-\frac{\mathrm{vol}(L(x)) \cdot \frac{na_d}{w}}{\mathrm{vol}(V)}\right) \leq o(1), \tag{6}$$

because the volume of the "used" vertices is at most

$$n^{\frac{3-\beta}{2}} \cdot \left(\frac{a_d}{w}\right)^{1/(\beta-2)} \leq o\left(n \cdot \frac{w}{a_d}\right) \leq o(\mathrm{vol}(A)).$$

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 12. Consider the set of vertices with weights between a_0 and b_0, and divide them into groups of size $n^{\frac{3-\beta}{2}}$. It's not hard to see that the number of groups is $n^{\frac{\beta-1}{2}-O(\varepsilon)}$ with high probability, because there are $n^{1-O(\varepsilon)}$ such vertices. Recall that $\varepsilon = 1/\log\log n$. We will show that for each group S,

$$\Pr[\text{The total label size of } S \leq c_2 \cdot n^{3-\beta-c_3}] \leq o(1), \tag{7}$$

for two fixed constants c_2 and c_3. Hence by Markov's inequality, with high probability, at least a constant fraction of the groups have label sizes at least $n^{3-\beta-O(\varepsilon)}$. This implies that the total label size of G is at least $n^{\frac{5-\beta}{2}-O(\varepsilon)}$. For the rest of the proof we focus on Equation (7).

Consider the distance function on G restricted to all vertex pairs in S, $\mathrm{dist}_S : S \times S \to \mathbb{N}$. Clearly, dist_S 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 distinct high degree vertex $h(x)$ whose weight is at least $(\frac{a_d}{w})^{1/(\beta-2)}$. Consider the following two cases:

a) If there exists $c_1^2 |S|^2 / 4$ 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 $c_1^2 |S|^2 / 5$, then we infer that there are at least $c_1^2 |S|^2 / 5$ 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 $\mathrm{dist}_S(x,y) \geq 2d+4$ for any $x, y \in B$, we infer that $h(x)$ and $h(y)$ are not connected by any edge. Hence we have

$$\begin{aligned}
&\Pr\left[\mathrm{dist}(x,y) \geq 2d+4, \ \forall (x,y) \in B \mid \{T_i\}_{i=1}^{|S|}\right] \\
&\leq \prod_{(x,y)\in B} \Pr[h(x) \nsim h(y)] \\
&\leq \left(1 - \Theta\left(\frac{1}{\mathrm{vol}(V)} \cdot \left(\frac{a_d}{w}\right)^{\frac{2}{\beta-2}}\right)\right)^{\Theta(n^{\beta-3})} \\
&\leq \exp\left(-\Theta\left(\left(\frac{a_d}{w}\right)^{\frac{2}{\beta-2}} \times n^{\beta-4}\right)\right) \\
&= \exp\left(-\Theta(n^{\beta-3-2\varepsilon/(\beta-2)} \times w^{-\frac{2}{\beta-2}-\frac{1}{3-\beta}\log^2\log n})\right) = \exp\left(-\Theta(n^{\beta-3-O(\varepsilon)})\right).
\end{aligned}$$

where the last few steps follows by straightforward calculations. To summarize, the distance function induced on S only occurs with probability at most $\exp\left(-c_2 \cdot n^{\beta-3-c_3\varepsilon}\right)$. Then by union bound over all possible labels for S whose length is less than $c_2 n^{3-\beta-c_3\varepsilon}$, we have shown that Equation (7) is true.

By taking a union bound over the two cases, we have proved Equation (7). Hence the proof is complete. □

D Proof of Theorem 10

In this section, we fill in the proof of Theorem 10. We assume that $\mathbf{p}$ 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. *In the setting of this section, Algorithm 1 returns an exact distance labeling scheme F with high probability.*

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

$$\Gamma_{l(x)}(x) = \varnothing \text{ or } \mathrm{dist}(v^*, x) \leq l(x), \forall x \in V$$

where v^* is the node with the maximum weight. We argue that Algorithm 1 finds a 2-hop cover for any $G \in \Omega_S$, and

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

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

We first argue that Algorithm 1 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 $F(x) \cap F(y) = \varnothing$. If x and y are reachable from each other, consider their distance $\mathrm{dist}(x, y)$. 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 $\mathrm{dist}(x, v^*) \leq l(x)$ and $\mathrm{dist}(y, v^*) \leq l(y)$ since $G \in \Omega_S$. We consider three cases:

- If $\mathrm{dist}(x, y) \leq l(x) + l(y) - 2$, then there exists a node z such that $\mathrm{dist}(x, z) \leq l(x) - 1$ and $\mathrm{dist}(y, z) \leq l(y) - 1$. By our construction, z is in $F(x)$ and $F(y)$.
- If $\mathrm{dist}(x, y) = l(x) + l(y) - 1$, then consider the two nodes z and z' on one of the shortest path from x to y, with $\mathrm{dist}(x, z) = l(x) - 1$ and $\mathrm{dist}(y, z) = l(y)$. 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 $\mathrm{dist}(x, y) = l(x) + l(y)$, then clearly v^* is a common landmark for x and y.

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

$$\begin{aligned}
& 1 - \Pr[\Omega_S] \\
\leq & \sum_{x \in V} \Pr[\Gamma_{l(x)}(x) \neq \varnothing, \mathrm{dist}(v^*, x) > l(x)] \\
= & \sum_{x \in V} \sum_{k=0}^{n-1} \Pr[l(x) = k+1, \Gamma_{k+1}(x) \neq \varnothing, \mathrm{dist}(v^*, x) > k+1]
\end{aligned}$$

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

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

Hence,

$$\begin{aligned}
&\Pr[l(x) = k+1, \Gamma_{k+1}(x) \neq \varnothing, \mathrm{dist}(v^*, x) > k+1] \\
\leq\ &\Pr[\alpha_k(x) > \delta n^{1-\frac{1}{\beta-1}}, \mathrm{dist}(v^*, x) > k+1] \\
\leq\ &\Pr[\alpha_k(x) > \delta n^{1-\frac{1}{\beta-1}}, \mathrm{vol}(\Gamma_k(x)) \leq \frac{\delta n^{1-\frac{1}{\beta-1}}}{3}] \qquad (8) \\
+\ &\Pr[\mathrm{vol}(\Gamma_k(x)) > \frac{\delta n^{1-\frac{1}{\beta-1}}}{3}, \mathrm{dist}(v^*, u) > k+1] \qquad (9)
\end{aligned}$$

For Equation (8), consider how $\alpha_k(x)$ is discovered when we do the level set expansion from node x. Conditioned on $a = \mathrm{vol}(\Gamma_k(x)) \leq \delta n^{1-\frac{1}{\beta-1}}/3$, $\alpha_k(x)$ is the sum of 0-1-2 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 $\mathrm{vol}(\Gamma_k(x)) \geq \delta n^{1-\frac{1}{\beta-1}}/2$ and $v^* \notin N_k(x)$,

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

The first inequality is because of Proposition 3. The second inequality is because $\mathrm{vol}(V) \sim \nu n \pm o(n)$ 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 $\mathrm{O}(nK^{1-\beta})$. 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 $\mathrm{O}(d_x^{3-\beta})$.

We first show that the volume of all the level sets is at most $\mathrm{O}(\delta n^{1-\frac{1}{\beta-1}})$ 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 α_k denotes the number of edges between $\Gamma_k(x)$ and $V \backslash N_{k-1}(x)$.

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

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

and

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

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

Proof. Let $a = \mathrm{vol}(\Gamma_k(x))$ and $b = \mathrm{vol}(N_{k-1}(x))$. Conditioned on Ω_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 α_k is the sum of independent 0-1 random variables. Let μ denote its expected value. For each $y \in \Gamma_k(x)$, we know that $p_y \leq a = O(\sqrt{n})$. Let μ_y denote the expected number of edges between y and $V \backslash N_{k-1}(x)$, then

$$\begin{aligned}\mu_y &= \sum_{z: z\neq y \wedge z \notin N_{k-1}(x)} \min(\frac{p_y p_z}{\mathrm{vol}(V)}, 1)\mathbb{1}_{p_z \leq \sqrt{n}} \\ &\geq p_y(1 - \frac{b + \sum_{z\in V} p_z \mathbb{1}_{p_z \geq \sqrt{n}}}{\mathrm{vol}(V)}) = p_y(1-\kappa(n))\end{aligned}$$

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,

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

□

Lemma 17. *Let x be a fixed vertex. Let $0 \leq k \leq O(\log n)$. Denote by Ω_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 $\Pr[\mathrm{vol}(\Gamma_k(x)) > 4\delta n^{1-\frac{1}{\beta-1}}, \Omega_k^*] \leq (k+1)n^{-2}$.

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) $\mathrm{vol}(\Gamma_j(x)) \leq 4\delta n^{1-\frac{1}{\beta-1}}$, for any $0 \leq j \leq i-1$; (2) $\mathrm{vol}(\Gamma_k(x)) > 4\delta n^{1-\frac{1}{\beta-1}}$. 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,

$$\begin{aligned}\Pr[S_i] - \Pr[S_{i+1}] &= \Pr[\mathrm{vol}(\Gamma_i(x)) > 4\delta n^{1-\frac{1}{\beta-1}}, S_i] \\ &\leq \Pr[\Omega_i, \alpha_i \leq \delta n^{1-\frac{1}{\beta-1}}] \leq n^{-2}\end{aligned}$$

The first inequality is because if $G \in S_i$ and G satisfies $\mathrm{vol}(\Gamma_i(x)) > 4\delta n^{1-\frac{1}{\beta-1}}$, 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. □

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

Lemma 18. *In the setting of this section, we have that the following holds almost surely*

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

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 $\mathrm{O}(d_x \log d_x) \sim \mathrm{O}(d_x \log n)$. Hence the total amount of time it takes to sort all the adjacency lists is $\mathrm{O}(|E| \log n) = \mathrm{O}(n \log n)$.

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

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

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

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

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

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

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

$$|H| = O(nK^{1-\beta}) = O(n^{1-\min(\frac{1}{\beta-1}, \frac{1}{4-\beta})})$$

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

$$\pi_k(x) = \sum_{y \in \Gamma_k(x)} \hat{d}_y \mathbb{1}_{d_y \le K} \quad \text{for } x \in V, 0 \le k \le n-1$$

Hence $\pi_{l(x)-1}(x)$ gives the number of landmarks added on the boundary layer.
Set $c_3 = \frac{3Z}{|2\beta-5|} \max(x_{\min}^{5-2\beta}, 1)$, $\psi = 12 c_3 \delta n^{1-\min(\frac{1}{\beta-1}, \frac{1}{4-\beta})}$, and $\Delta = \max(c_1\psi, c_2 \delta n^{1-\frac{1}{\beta-1}} \log n)$, where c_1 and c_2 are defined in Proposition 19. We show that $\pi_{l(x)-1}(x) \le \Delta$ with probability $1 - n^{-2}$ for the rest of the proof — our conclusion follows by taking union bound over $x \in V$ and $1 \le l(x) \le O(\log n)$.

When $l(x) = 1$, $\pi_0(x) = d_x \le K \le \Delta$. When $l(x) = k + 1 \ge 2$, we know that $G \in \Omega^*_{k-1}$. Hence by Lemma 17, $\mathrm{vol}(\Gamma_{k-1}(x)) \le 4\delta n^{1-\frac{1}{\beta-1}}$ with high probability. More concretly,

$$\begin{aligned} &\Pr[l(x) = k+1, \pi_k(x) \ge \Delta] \\ \le\ &(k+1)n^{-2} + \Pr[l(x) = k+1, \mathrm{vol}(\Gamma_{k-1}(x)) \le 4\delta n^{1-\frac{1}{\beta-1}}, \pi_k(x) \ge \Delta] \end{aligned} \tag{10}$$

Denote by

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

Conditional on $a = \mathrm{vol}(\Gamma_{k-1}(x)) \le 4\delta n^{1-\frac{1}{\beta-1}}$, we show that $w_k \le \psi$ with high probability. Denote by Ω_w the set of graphs satisfying $a \le 4\delta n^{1-\frac{1}{\beta-1}}$. Conditioned on Ω_w, w_k is the sum of independent

random variables that are all bounded in $[0, (2K)^{3-\beta}]$. Hence

$$\begin{aligned}
\mathbb{E}[w_k] &= \sum_{y \notin N_{k-1}(x)} \Pr[y \sim N_{k-1}(x)] {p_y}^{3-\beta} \mathbb{1}_{p_y \le 2K} \\
&\le \frac{a}{\mathrm{vol}(V)} \Big(\sum_{y \notin N_{k-1}(x)} {p_y}^{4-\beta} \mathbb{1}_{p_y \le 2K} \Big) && \text{(by Proposition 3)} \\
&\le \frac{a}{\mathrm{vol}(V)} \Big(\sum_{y \in V} {p_y}^{4-\beta} \mathbb{1}_{p_y \le 2K} \Big) \\
&\le \frac{a\phi(K)n}{\mathrm{vol}(V)} && \text{(by Proposition 22)} \\
&\sim \frac{a\phi(K)}{\nu} && (\mathrm{vol}(V) = \nu n \pm o(n) \text{ by Proposition 22}) \\
&\le \frac{\psi}{3}.
\end{aligned}$$

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

$$\Pr[w_k > \psi \mid \Omega_w] \le \exp(-\frac{\psi}{4(2K)^{3-\beta}}) \sim o(n^{-2})$$

because when $2.5 \le \beta \le 3$,

$$\frac{\psi}{K^{3-\beta}} = \Theta(n^{1-\frac{1}{\beta-1}-\frac{3-\beta}{2}}) = \Theta(n^{\frac{(\beta-1)^2-2}{2(\beta-1)}})$$

And when $2 < \beta < 2.5$,

$$\frac{\psi}{K^{3-\beta}} = \Theta(n^{\frac{(3-\beta)(\beta-2)}{(4-\beta)(\beta-1)}})$$

Hence the second part in Equation (10) is bounded by $o(n^{-2})$ plus

$$\begin{aligned}
&\Pr[l(x) = k+1, \mathrm{vol}(\Gamma_{k-1}(x)) \le 4\delta n^{1-\frac{1}{\beta-1}}, w_k \le \psi, \pi_k(x) \ge \Delta] \\
\le\ &\Pr[w_k \le \psi, \alpha_{k-1} \le \delta n^{1-\frac{1}{\beta-1}}, \pi_k(x) \ge \Delta] \\
\le\ &\Pr[w_k \le \psi, |\Gamma_k(x)| \le \delta n^{1-\frac{1}{\beta-1}}, \pi_k(x) \ge \Delta]
\end{aligned}$$

In the reminder of the proof we show the above Equation is at most n^{-2}. Denote by

$$\pi'_k(x) = \sum_{y \in \Gamma_k(x)} \hat{d}_y \mathbb{1}_{p_y \le 2K}$$

By Proposition 20, $\Pr[d_y \le K \mid p_y > 2K] \le \exp(-K/8) \sim o(n^{-3})$ for any $y \in V$. Hence $\pi'_k(x) = \pi_k(x)$ with probability at least $1 - o(n^{-2})$. Lastly, we have

$$\Pr[w_k \le \psi, |\Gamma_k(x)| \le \delta n^{1-\frac{1}{\beta-1}}, \pi'_k(x) \ge \Delta] \le n^{-2}$$

Otherwise, there exists a vertex $y \in \Gamma_k(x)$ such that $p_y \le 2K$ and $\hat{d}_y \ge \max(c_1 {p_y}^{3-\beta}, c_2 \log n)$, because $\Delta \ge \max(c_1\psi, c_2\delta n^{1-\frac{1}{\beta-1}} \log n)$. This happens with probability at most n^{-2}, by taking union bound over every vertex with Proposition 19.

Proof of Proposition 19: When $p_x \le c_2 \log n/2$,

$$\Pr[\hat{d}_x \ge c_2 \log n] \le \Pr[d_x \ge c_2 \log n] \le o(n^{-3})$$

Now suppose that $p_x > c_2 \log n/2$. Consider any vertex y whose weight is at most $p_x/8$. Then

$$\Pr[d_y \ge d_x] \le \Pr[d_y \ge p_x/4] + \Pr[d_x < p_x/4] \sim o(n^{-4})$$

The second inequality is because of Proposition 20. Hence y is not in S_x.
Now if $p_y \ge 2K$, then $\Pr[d_y \le K] \le\sim o(n^{-4})$. 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

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{y \in V \setminus \{x\}: p_x/8 \le p_y \le 2K} \frac{p_x p_y}{\mathrm{vol}(V)} \\
&\le \frac{4p_x}{\mathrm{vol}(V)} \max(\frac{8Z}{\beta - 1} n {p_x}^{2-\beta}, \sqrt{n} \log n) \\
&\le \max(c_1 {p_x}^{3-\beta}, c_2 \log n)/3
\end{aligned}$$

The first inequality is because of Proposition 22. The second inequality is because $\mathrm{vol}(V) = \nu n + o(n)$ by Proposition 22, and $p_x \le 2K \le 2\sqrt{n}$. From here it is not hard to obtain that $\Pr[|X| \le \max(c_1 {p_x}^{3-\beta}, c_2 \log n)] \sim o(n^{-3})$.

E 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 [50, 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 [50], 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].

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

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 E. 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 numer 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.

σ = vertices ordered by (indegree + outdegree)
for $i \leq n$ **do**
  if $i \leq K$ **then**
    computeGlobalLm(σ_i)
  else
    computeLocalLm(σ_i)
  end if
end for
procedure computeGlobalLm(x)
  $\{(y, \text{dist}(x, y)), \forall\, y \in V\}$ = Run a forward BFS
  $\{(y, \text{dist}(y, x)), \forall\, y \in V\}$ = Run a backward BFS
end procedure
procedure computeLocalLm(x)
  Run a forward BFS from x up to distance d, prune any node from $\{\sigma_i\}_{i=1}^{K}$.
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 gaurantee 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.

	Landmark size per node			Running time (min)		
	Ours	PrunedLabel	BallGrow	Ours	PrunedLabel	BallGrow
Twitter	227	261	637	< 1	10.5	1.8
Stanford	82	95	367	< 1	< 1	2.0
Google	215	285	276	7.2	84	8.2
BerkStan	63	155	742	1.0	46.9	8.6

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.

	90% error		80% error	
	Ours	BallGrow	Ours	BallGrow
Twitter	0.5	0.0	0.25	0.125
Stanford	0.07	0.08	0.0	0.0
Google	0.0	0.0	0.0	0.0
BerkStan	0.125	0.1	0.0	0.0

Table 3: Comparion of accuracy. The accuracy of PrunedLabel is not listed becaue it is gauranteed 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,

F Random Graph Toolbox

We first present a helper proposition which characterizes the probability that a vertex's actual degree deviates from its weight.

Proposition 20. *Let $G = (V, E) \in \mathscr{G}^n(\mathbf{p})$ 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*
$$\Pr[d_x \geq cp_x] \leq \exp(-\frac{(c-1)p_x}{2})$$
2. *If $0 < c < 1$, then*
$$\Pr[d_x \leq cp_x] \leq \exp(-\frac{(1-c)^2 p_x}{8})$$

Proof. Let $\mu = \mathbb{E}[d_x]$. First,

$$\begin{aligned}\mu &= \sum_{y \in V \setminus \{x\}} \min(\frac{p_x p_y}{\mathrm{vol}(V)}, 1) \\ &\leq \sum_{y \in V} \frac{p_x p_y}{\mathrm{vol}(V)} = p_x\end{aligned}$$

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

$$\begin{aligned}\Pr[d_x \geq cp_x] &\leq \exp(-\frac{cp_x - \mu}{2}) \\ &\leq \exp(-\frac{(c-1)p_x}{2})\end{aligned}$$

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

$$\mu \geq p_x(1 - \frac{p_x}{\mathrm{vol}(V)} - \frac{\sum_{y \in V} p_y \mathbb{1}_{p_y \geq t}}{\mathrm{vol}(V)})$$

By Proposition 22,

$$\sum_{y \in V} p_y \mathbb{1}_{p_y \geq t} \sim o(n)$$

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

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

for large enough n. □

The nexxt proposition helps us to characterize the number of nodes whose degree is at least K, for a certain value K.

Proposition 21. *Let $G = (V, E) \in \mathscr{G}^n(\mathbf{p})$ be a random graph. Let $8 \log n \le K \le \sqrt{n}$ denote a fixed value and $S = \{x \in V : d_x \ge K\}$. With probability at least $1 - n^{-1}$, $|S| \le 3 \max(\frac{Z 3^{\beta-1}}{\beta-1} n K^{1-\beta}, \log n)$.*

Proof. Let $Y_1 = \{x \in V : p_x \ge \frac{K}{3}\}$ and $Y_2 = \{x \in V : p_x < \frac{K}{3} \text{ and } K \le d_x\}$. Clearly, $S \subset Y_1 \cup Y_2$. We first show that Y_2 is empty with probability at least $1 - n^{-1}$. Consider a fixed node $x \in V$ with weight $p_x \le K/3$. By Proposition 20,

$$\Pr[d_x \ge K] \le \exp(-K) \sim o(n^{-2})$$

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

We then bound the size of Y_1. The expected value of Y_1 is $\frac{Z 3^{\beta-1}}{\beta-1} n K^{1-\beta}$. Then by Chernoff bound, it's not hard to obtain the desired conclusion (details omitted). □

We will need the following basic properties of the degree sequence of a Chung-Lu model.

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

i) The maximum weight $\max \mathbf{p} \ge \varepsilon(n) n^{\frac{1}{\beta-1}}$.

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

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

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

$$c(K) = \begin{cases} \frac{3 Z x_{\min}^{5-2\beta}}{2\beta-5} & \text{if } 2.5 \le \beta \le 3 \\ \frac{3Z}{5-2\beta} K^{5-2\beta} & \text{if } 2 < \beta < 2.5 \end{cases}$$

Then

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

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

$$\sum_{y \in V} p_y \mathbb{1}_{\frac{p_y}{c} \le p_x \le 2\sqrt{n}} \le 6 \max(\frac{c^{\beta-2} Z}{\beta - 2} n {p_u}^{2-\beta}, \sqrt{n} \log n).$$

The proof is via standard concentration inequality (details omitted). For the rest of this section, we fill in the missing proofs stated in previous sections regarding random graphs.

F.1 Proof of Proposition 3

Proof. The proof follows straightforwardly from definition.

$$\begin{aligned}
\Pr[S \sim T] &= 1 - \prod_{x\in S}\prod_{y\in T}\left(1 - \min(\frac{p_x p_y}{\mathrm{vol}(V)}, 1)\right) \le 1 - \left(1 - \sum_{x\in S}\sum_{y\in T}\min(\frac{p_x p_y}{\mathrm{vol}(V)}, 1)\right) \\
&\le 1 - \left(1 - \sum_{x\in S}\sum_{y\in T}\frac{p_x p_y}{\mathrm{vol}(V)}\right) = \frac{p(S)p(T)}{\mathrm{vol}(V)}. \\
\Pr[S \nsim T] &= \prod_{x\in S}\prod_{y\in T}\left(1 - \min(\frac{p_x p_y}{\mathrm{vol}(V)}, 1)\right) \le \exp\left(-\sum_{x\in S}\sum_{y\in T}\min(\frac{p_x p_y}{\mathrm{vol}(V)}, 1)\right) \\
&\le \exp(-\frac{p(S)p(T)}{\mathrm{vol}(V)}).
\end{aligned}$$

□

F.2 Proof of Growth Lemma 9: $\beta > 3$

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

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

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

$$\begin{aligned}
&\mathbb{E}[\mathrm{vol}(\Gamma_{k+1}(x)) \mid \mathrm{vol}(\Gamma_k(x))] \\
\le &\sum_{y\notin N_k(x)} p_y^2 \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)} \\
\le &\ \mathrm{vol}(\Gamma_k(x)) \cdot \frac{\mathrm{vol}_2(V)}{\mathrm{vol}(V)} \\
= &\ \mathrm{vol}(\Gamma_k(x)) \cdot r.
\end{aligned}$$

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

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,

$$\mathbb{E}[\mathrm{vol}(N_k(x))] \le O(r^k).$$

However, for each i, the probability that y is at distance i from x conditioned on $N_{i-1}(x)$ is at most

$$\Pr[y \in \Gamma_i(x) \mid N_{i-1}(x)] \le p_y \cdot \frac{\mathrm{vol}(\Gamma_{i-1}(x))}{\mathrm{vol}(V)}.$$

The probability y is within distance $k+1$ from x is at most

$$\begin{aligned}\Pr[y \in N_{k+1}(x)] &\le \sum_{i=1}^{k+1} \Pr[y \in \Gamma_i(x)] \\ &\le \frac{p_y}{\mathrm{vol}(V)} \cdot \sum_{i=1}^{k+1} \mathbb{E}[\mathrm{vol}(\Gamma_{i-1}(x))] \\ &= p_y \cdot \frac{N_k(x)}{\mathrm{vol}(V)} \\ &\le O(r^k).\end{aligned}$$

With large probability, $\Gamma_{k+1}(x)$ has volume not much smaller than $\Gamma_k(x) \cdot \frac{\mathrm{vol}_2(V \setminus N_k(x))}{\mathrm{vol}(V)}$ Conditioned on $\Gamma_k(x)$, the probability that a vertex $y \notin N_k(x)$ is in $\Gamma_{k+1}(x)$ is at least

$$1 - e^{-p_y \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)}}$$

by Proposition 3.

For any $T > 0$, we have

$$\sum_{y:p_y>T} p_y^2 \le T^{-\gamma} \cdot \sum_{y:p_y>T} p_y^{2+\gamma} \le \tau \cdot n \cdot T^{-\gamma}.$$

We also have

$$\sum_{y:p_y\le T} p_y^3 \le T^{1-\gamma} \cdot \sum_{y:p_y\le T} p_y^{2+\gamma} \le \tau \cdot n \cdot T^{1-\gamma}.$$

Let us focus on all y's with weight at most T. By that fact that $1 - e^{-x} \ge x - x^2/2$ when $x \ge 0$, the expected volume of $\Gamma_{k+1}(x) \cap \{y : p_y \le T\}$ conditioned on $N_k(x)$ is at least

$$\begin{aligned}&\sum_{y \notin N_k(x): p_y \le T} p_y \left(1 - e^{-p_y \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)}}\right) \\ \ge\ &\sum_{y \notin N_k(x): p_y \le T} p_y^2 \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)} - \frac{1}{2} \sum_{y \notin N_k(x): p_y \le T} p_y^3 \cdot \left(\frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)}\right)^2 \\ \ge\ &(\mathrm{vol}_2(V \setminus N_k(x))) \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)} - \sum_{y:p_y \ge T} p_y^2 \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)} - \frac{1}{2} \sum_{y:p_y \le T} p_y^3 \cdot \left(\frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)}\right)^2 \\ \ge\ &\mathrm{vol}(\Gamma_k(x)) \cdot \frac{\mathrm{vol}_2(V \setminus N_k(x))}{\mathrm{vol}(V)} - \tau \cdot n \cdot T^{-\gamma} \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)} - \tau \cdot n \cdot T^{1-\gamma} \cdot \left(\frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)}\right)^2 \\ \ge\ &\mathrm{vol}(\Gamma_k(x)) \cdot \left(\frac{\mathrm{vol}_2(V \setminus N_k(x))}{\mathrm{vol}(V)} - \tau \cdot \left(T^{-\gamma} + T^{1-\gamma} \cdot \frac{\mathrm{vol}(\Gamma_k(x))}{\mathrm{vol}(V)}\right)\right).\end{aligned}$$

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 $T = \mathrm{vol}(\Gamma_k(x))^{1/2}$.

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

$$\begin{aligned}&\mathbb{E}[\mathrm{vol}(\Gamma_{k+1}(x) \cap \{y : p_y \le T\}) \mid N_k(x)]\\ \ge\ &\mathrm{vol}(\Gamma_k(x)) \cdot \left(\frac{\mathrm{vol}_2(V \setminus N_k(x))}{\mathrm{vol}(V)} - 2\tau \cdot \mathrm{vol}(\Gamma_k(x))^{-\gamma/2}\right).\end{aligned}$$

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

$$\begin{aligned}&\Pr\left[\mathrm{vol}(\Gamma_{k+1}(x)) \le \mathrm{vol}(\Gamma_k(x)) \cdot \left(\frac{\mathrm{vol}_2(V \setminus N_k(x))}{\mathrm{vol}(V)} - \mathrm{vol}(\Gamma_k(x))^{-\gamma/3}\right) \,\middle|\, N_k(x)\right]\\ &\le 2^{-\Theta\left(\mathrm{vol}(\Gamma_k(x))^{1/2-2\gamma/3}\right)},\end{aligned} \tag{11}$$

as long as $\mathrm{vol}(\Gamma_k(x)) = O(n^{2/3})$ and $\mathrm{vol}(\Gamma_k(x))$ 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:

$$\Pr[\mathrm{vol}(\Gamma_1(x)) \ge C] \ge \Omega_C(1).$$

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

$$\mathrm{vol}(\Gamma_{i+1}(x)) > \mathrm{vol}(\Gamma_i(x)) \cdot \left(\frac{\mathrm{vol}_2(V \setminus N_i(x))}{\mathrm{vol}(V)} - \mathrm{vol}(\Gamma_i(x))^{-\gamma/3}\right)$$

or

$$\mathrm{vol}(\Gamma_i(x)) \ge n^{2/3}.$$

By the argument above,

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

We claim that these events have the following properties.

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

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

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

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

$$\Pr[\mathrm{vol}(N_k(x)) \ge n^{2/3}] \le O(\sqrt{n}/n^{2/3}) = o(1).$$

Therefore, we have the lower bound

$$\begin{aligned}&\Pr[\mathrm{vol}(\Gamma_k(x)) \ge \Omega(r^k)]\\ \ge\ &\Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \cdot \Pr[\mathrm{vol}(\Gamma_k(x)) \ge \Omega(r^k) \mid \mathcal{E}_0, \ldots, \mathcal{E}_{k-1}]\\ \ge\ &\Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \cdot (1 - \Pr[\mathrm{vol}(N_k(x)) \ge n^{2/3} \mid \mathcal{E}_0, \ldots, \mathcal{E}_{k-1}])\\ \ge\ &\Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] \cdot (1 - \Pr[\mathrm{vol}(N_k(x)) \ge n^{2/3}] / \Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}])\\ =\ &\Omega(1).\end{aligned}$$

This concludes the proof. □

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

In particular, $\mathrm{vol}(N_k(x)) < n^{2/3}$ implies that $\mathrm{vol}(N_i(x)) < n^{2/3}$ and $\mathrm{vol}(\Gamma_i(x)) \le n^{2/3}$ for every $i \le k$. By Hölder's inequality, we also have

$$\begin{aligned}\mathrm{vol}_2(N_i(x)) &\le \mathrm{vol}_{2+\gamma}(N_i(x))^{\frac{1}{1+\gamma}} \cdot \mathrm{vol}(N_i(x))^{\frac{\gamma}{1+\gamma}} \\ &\le \tau^{\frac{1}{1+\gamma}} \cdot n^{1-\frac{\gamma}{3(1+\gamma)}}.\end{aligned}$$

Thus, the event $\mathcal{E}_i$ $(i > 0)$ implies

$$\mathrm{vol}(\Gamma_{i+1}(x)) > \mathrm{vol}(\Gamma_i(x)) \cdot \left(r - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - \mathrm{vol}(\Gamma_i(x))^{-\gamma/3}\right). \tag{12}$$

Let $\hat{r} = r - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - C^{-\gamma/3}$. For sufficiently large C, $\hat{r} > \sqrt{r} > 1$. First we can show $\mathrm{vol}(\Gamma_i(x)) \ge C \cdot \hat{r}^{i-1} \ge C \cdot r^{(i-1)/2}$ for $i \ge 1$ inductively:

- By the definition of $\mathcal{E}_0$, $\mathrm{vol}(\Gamma_1(x)) \ge C$;
- If $\mathrm{vol}(\Gamma_i(x)) \ge C \cdot \hat{r}^{i-1}$, then we have

$$\begin{aligned}\mathrm{vol}(\Gamma_{i+1}(x)) &\ge \mathrm{vol}(\Gamma_i(x)) \cdot \left(r - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - \mathrm{vol}(\Gamma_i(x))^{-\gamma/3}\right) \\ &\ge \mathrm{vol}(\Gamma_i(x)) \cdot \left(r - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - C^{-\gamma/3}\right) \\ &\ge \mathrm{vol}(\Gamma_i(x)) \cdot \hat{r}.\end{aligned}$$

Thus, we have $\mathrm{vol}(\Gamma_i(x)) \ge \Omega(\hat{r}^i) \ge \Omega(r^{i/2})$. By Equation (12) again, we have

$$\begin{aligned}\mathrm{vol}(\Gamma_k(x)) &> \mathrm{vol}(\Gamma_{k-1}(x)) \cdot (r - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - r^{-(k-1)\gamma/6}) \\ &\ge \mathrm{vol}(\Gamma_1(x)) \cdot \left(\prod_{i=1}^{k-1}(r - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} - r^{-i\gamma/6})\right) \\ &\ge r^{k-1} \cdot C \cdot \left(\prod_{i=1}^{k-1}(1 - \tau^{\frac{1}{1+\gamma}} \cdot n^{-\frac{\gamma}{3(1+\gamma)}} \cdot r^{-1} - r^{-i\gamma/6-1})\right) \\ &= r^k \cdot \alpha_k,\end{aligned}$$

where α_k is decreasing as k increases. $\alpha_{\frac{1}{2}\log_r n}$ is lower bounded by a constant α. Thus, $\mathrm{vol}(\Gamma_k(x)) \ge \alpha \cdot r^k \ge \Omega(r^k)$. This proves the claim. □

Proof of Claim 24. By Lemma 23, conditioned on $\mathcal{E}_0, \ldots \mathcal{E}_{i-1}$, we have either

$$\mathrm{vol}(\Gamma_i(x)) \ge \alpha \cdot r^i$$

or

$$\mathrm{vol}(N_i(x)) \ge n^{2/3}.$$

Thus, by Equation (11), we have

$$\Pr[\overline{\mathcal{E}_i} \mid \mathcal{E}_0, \ldots, \mathcal{E}_{i-1}] \le 2^{-\Theta(r^{\Omega(i)})}.$$

Since $\Pr[\mathcal{E}_0] = \Omega(1)$, we may lower bound the probability that all events happen simultaneously

$$\begin{aligned}\Pr[\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}] &\geq \Omega\left(\prod_{i=0}^{k-1}\left(1 - 2^{-\Theta(r^{\Omega(i)})}\right)\right)\\ &\geq \Omega(1).\end{aligned}$$

□

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

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

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

$$\sum_{z \in \Gamma_{k-1}(x)} p_y \cdot \frac{p_z}{\mathrm{vol}(V)} = p_y \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}.$$

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

$$\begin{aligned}&\Pr\left[\exists y, p_y \geq (\mathrm{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)} \cdot w, y \in \Gamma_k(x) \mid \mathrm{vol}(\Gamma_{k-1}(x))\right]\\ &\leq \sum_{y: p_y \geq (\mathrm{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)} \cdot w} p_y \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}\\ &\leq O\left(n \cdot \frac{1}{\mathrm{vol}(\Gamma_{k-1}(x)) \cdot w^{\beta-2}} \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}\right)\\ &= O(1/w^{\beta-2}).\end{aligned}$$

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

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

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

$$\begin{aligned}\mathbb{E}[\mathrm{vol}(\Gamma_k(x)) \mid E_k, \mathrm{vol}(\Gamma_{k-1}(x))] &\leq \sum_{y: p_y \leq (\mathrm{vol}(\Gamma_{k-1}(x)))^{1/(\beta-2)} \cdot w} p_y^2 \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}\\ &\leq O\left(n \cdot \mathrm{vol}(\Gamma_{k-1}(x))^{(3-\beta)/(\beta-2)} \cdot w^{3-\beta} \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}\right)\\ &\leq O\left(\mathrm{vol}(\Gamma_{k-1}(x))^{1/(\beta-2)} \cdot w^{3-\beta}\right).\end{aligned}$$

By Markov's inequality, we have

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

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

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

On the other hand, we have

$$\begin{aligned}\mathrm{vol}(\Gamma_{k-1}(x))^{1/(\beta-2)} \cdot w &\le b_{k-1}^{1/(\beta-2)} \cdot w \\ &= (c_{k-1} \cdot w^{1/(\beta-2)^{2(k-1)}(3-\beta)})^{1/(\beta-2)} \cdot w \\ &= c_k \cdot w^{1/(\beta-2)^{2k-1}(3-\beta)+1} \\ &= c_k \cdot w^{(1/(\beta-2)^{2k}(3-\beta))\cdot(\beta-2+(3-\beta)(\beta-2)^{2k})} \\ &\le b_k.\end{aligned}$$

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

Lower Bounding $\mathrm{vol}(\Gamma_k(x))$ For any vertex y, if $p_y \cdot p_z \ge \mathrm{vol}(V)$ 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

$$\begin{aligned}\prod_{z\in\Gamma_{k-1}(x)} \left(1 - \frac{p_y \cdot p_z}{\mathrm{vol}(V)}\right) &\le e^{-\sum_{z\in\Gamma_{k-1}(x)} \frac{p_y \cdot p_z}{\mathrm{vol}(V)}} \\ &= e^{-p_y \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}}.\end{aligned}$$

That is, in either case, if $y \notin N_{k-1}(x)$, then $\Pr[y \notin \Gamma_k(x)] \le e^{-p_y \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}}$. 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:

$$\begin{aligned}\Pr[\forall y, \text{ s.t. } p_y \ge a_k, y \notin \Gamma_k(x)] &\le \prod_{y:p_y\ge a_k} e^{-p_y \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}} \\ &= e^{-\sum_{y:p_y\ge a_k} p_y \cdot \frac{\mathrm{vol}(\Gamma_{k-1}(x))}{\mathrm{vol}(V)}} \\ &\le e^{-\Omega(a_k^{2-\beta} \cdot a_{k-1})} \\ &= e^{-\Omega(w^{1/(\beta-2)^{2k-1}(3-\beta)-1/(\beta-2)^{2k-2}(3-\beta)})} \\ &= e^{-\Omega(w^{1/(\beta-2)^{2k-1}})} \\ &\le e^{-w} = 1/\log n.\end{aligned}$$

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

Finally, by union bound, the probability that $\mathrm{vol}(\Gamma_k(x)) \in [a_k, b_k]$ is at least $1 - O(1/w^{\beta-2})$ as the lemma states.

Lower Bounding $\mathrm{dist}(x,y)$ By the above argument, we have

$$\begin{aligned}
&\Pr[\mathrm{dist}(x,y) \le 2d+3] \\
\le\ &\Pr[\exists 0 \le i,j \le d+1, \exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v] \\
\le\ &\sum_{i,j=0}^{d+1} \Pr[\exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v] \\
=\ &\sum_{i,j=0}^{d+1} \mathop{\mathbb{E}}_{\Gamma_i(x),\Gamma_j(y)} [\Pr[\exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v \mid \Gamma_i(x), \Gamma_j(y)]] \\
\le\ &\sum_{i,j=0}^{d+1} (\Pr[\exists u \in \Gamma_i(x), v \in \Gamma_j(y), u \sim v \mid \Gamma_i(x) \in [a_i,b_i], \Gamma_j(y) \in [a_j,b_j]] \\
+\ &\Pr[\Gamma_i(x) \notin [a_i,b_i]] + \Pr[\Gamma_j(y) \notin [a_j,b_j]]) \\
\le\ &\sum_{i,j=0}^{d+1} \Big(b_i \cdot b_j / \mathrm{vol}(V) + O(1/w^{\beta-2}) \Big) \\
\le\ &O(b_{d+1}^2/n + d^2/w^{\beta-2}) \\
=\ &O(n^{-\varepsilon/(\beta-2)} + \log^2 \log \log n / \log^{\beta-2} \log n) = o(1).
\end{aligned}$$

This concludes the proof. □