Ultra-Fast Load Balancing on Scale-Free Networks

Karl Bringmann[1], Tobias Friedrich[2], Martin Hoefer[3], Ralf Rothenberger[2] and Thomas Sauerwald[4]

[1] Institute of Theoretical Computer Science, ETH Zurich, Switzerland
[2] Hasso Plattner Institute, Potsdam, Germany
[3] Max Planck Institute for Informatics, Saarbrücken, Germany
[4] University of Cambridge, United Kingdom

Abstract. The performance of large distributed systems crucially depends on efficiently balancing their load. This has motivated a large amount of theoretical research how an imbalanced load vector can be smoothed with local algorithms. For technical reasons, the vast majority of previous work focuses on regular (or almost regular) graphs including symmetric topologies such as grids and hypercubes, and ignores the fact that large networks are often highly heterogeneous.

We model large scale-free networks by Chung-Lu random graphs and analyze a simple local algorithm for iterative load balancing. On n-node graphs our distributed algorithm balances the load within $\mathcal{O}((\log\log n)^2)$ steps. It does not need to know the exponent $\beta \in (2,3)$ of the power-law degree distribution or the weights w_i of the graph model. To the best of our knowledge, this is the first result which shows that load-balancing can be done in double-logarithmic time on realistic graph classes.

1 Introduction

Load balancing. Complex computational problems are typically solved on large parallel networks. An important prerequisite for their efficient usage is to balance the work load efficiently. Load balancing is also known to have applications to scheduling [19], routing [7], numerical computation such as solving partial differential equations [18, 21], and finite element computations [14]. In the standard abstract formulation of load balancing, processors are represented by nodes of a graph, while links are represented by edges. The objective is to balance the load by allowing nodes to exchange loads with their neighbors via the incident edges. Particularly popular are decentralized, round-based iterative algorithms where a processor knows only its current load and that of the neighboring processors. We focus on *diffusive* load balancing strategies, where each processor decides how many jobs should be sent and balances its load with its neighbors in each round. As the degrees of the topologies of many networks follow heavy tailed statistics, our main interest lies on *scale-free* networks.

Diffusion. On networks with n nodes, our balancing model works as follows: At the beginning, each node i has some work load $x_i^{(0)}$. The goal is to obtain (a good approximation of) the balanced work load $\overline{x} := \sum_{i=1}^{n} x_i^{(0)}/n$ on all nodes. On heterogeneous graphs with largely varying node degrees it is natural to consider a multiplicative quality measure: We want to find an algorithm which achieves $\max_i x_i^{(t)} = \mathcal{O}(\overline{x})$ at the earliest time t possible. Load-balancing is typically considered *fast* if this can be achieved in time logarithmic in the number of nodes. We aim at double-logarithmic time, which we call *ultra-fast* (following the common use of the superlative "ultra" for double-logarithmic bounds [5, 11, 20]).

The diffusion model was first studied by Cybenko [7] and, independently, Boillat [2]. The standard implementation is the *first order scheme* (FOS), where the load vector is multiplied with a diffusion matrix $\mathbf{P}$ in each step. For regular graphs with degree d, a common choice is $\mathbf{P}_{ij} = 1/(d+1)$ if $\{i,j\} \in E$. Already Cybenko [7] in 1989 shows for regular graphs a tight connection between the convergence rate of the diffusion algorithm and the absolute value of the second largest eigenvalue $\lambda_{\max}$ of the diffusion matrix $\mathbf{P}$. While FOS can be defined for non-regular graphs, its convergence is significantly affected by the loops which are induced by the degree discrepancies. Regardless of how the damping factor is chosen, FOS requires $\Omega(\log n)$ rounds on a broad class of non-regular graphs (cf. Lemma 3.1 and the discussion in Appendix 3).

Scale-free networks. Many real-world graphs have a power law degree distribution, meaning that the number of vertices with degree k is proportional to $k^{-\beta}$, where β is a constant intrinsic to the network. Such networks are synonymously called scale-free networks and have been widely studied. As a model for large scale-free networks we use the *Chung-Lu random graph model* with a power-law degree distribution with exponent $\beta \in (2,3)$. (See Section 4 for a formal definition.) This range of β's is typically studied as many scale-free networks (e.g. co-actors, protein interactions, internet, peer-to-peer [17]) have a power law exponent with $2 < \beta < 3$. It is known that the diameter of this graph model is $\Theta(\log n)$ while the average distance between two vertices is $\Theta(\log\log n)$ [4].

Results. Scale-free networks are omnipresent, but surprisingly few rigorous insights are known about their ability to efficiently balance load. Most results and developed techniques for theoretically studying load balancing only apply to regular (or almost-regular) graphs. In fact, we cannot hope for ultra-fast balancing on almost-regular graphs: Even for expander graphs of maximum degree d, there is a general lower bound of $\Omega(\log n/\log d)$ iterations for *any* distributed load balancing algorithms (cf. Lemma 3.2). Our main result (cf. Theorem 4.1) shows that within $\mathcal{O}((\log\log n)^2)$ steps, our simple local balancing algorithm (cf. Algorithm 1) can balance the load on a scale-free graph with high probability. The algorithm assumes that the initial load is only distributed on nodes with degree $\Omega(\operatorname{polylog} n)$ (cf. Theorem 4.3), which appears to be a natural assumption in typical load balancing applications. As the diameter of the graph is $\Theta(\log n)$, ultra-fast balancing is impossible if the initial load is allowed on arbitrary vertices. As standard FOS requires $\Omega(\log n)$ rounds (cf. Lemma 3.1), our algorithm uses a different, novel approach to overcome these restrictions.

Algorithm. The protocol proceeds in waves, and each wave (roughly) proceeds as follows. First, the remaining load is balanced within a core of high-degree nodes. These nodes are known to compose a structure very similar to a dense Erdős-Rényi random graph and thereby allow very fast balancing. Afterwards, the load is disseminated into the network from high- to low-degree nodes. Each node absorbs some load and forwards the remaining to lower-degree neighbors. If there are no such neighbors, the excess load is routed back to nodes it was received from. In this way, the load moves like a wave over the graph in decreasing order of degree and then swaps back into the core. We will show that each wave needs $\mathcal{O}(\log\log n)$ rounds. The algorithm keeps initiating waves until all load is absorbed, and we will show that only $\mathcal{O}(\log\log n)$ waves are necessary.

Techniques. There are a number of technical challenges in our analysis, mostly coming from the random graph model, and we have to develop new techniques to cope with them. For example, in scale-free random graphs there exist large sparse areas with many nodes of small degree that result in a high diameter. A challenge is to avoid that waves get lost by pushing too much load deep into these periphery areas. This is done by a partition of nodes into layers with significantly different degrees and waves that proceed only to neighboring layers. To derive the layer structure, we classify nodes based on their realized degrees. However, this degree might be different from the expected degree corresponding to the weights w_i of the network model, which is unknown to the algorithm. This implies that nodes might not play their intended role in the graph and the analysis (cf. Definition 6.1). This can lead to poor spread and the emersion of a few, large single loads during every wave. Here we show that several types of "wrong-degree" events causing this problem are sufficiently rare, or, more precisely, they tend to happen frequently only in parts of the graph that turn out not to be critical for the result. At the core, our analysis adjusts and applies fundamental probabilistic tools to derive concentration bounds, such as a variant of the method of bounded variances (cf. Section 2).

2 Probabilistic Tail Bounds

We frequently apply the following Chernoff-type bound.

Theorem 2.1. *If X is a sum of independent Bernoulli trials, then for $\delta < 2e-1$ we have*

$$\mathbf{Pr}\left[\, |X - \mathbf{E}\left[\, X \,\right]| \geqslant \delta \mathbf{E}\left[\, X \,\right] \,\right] < \exp(-\mathbf{E}\left[\, X \,\right] \delta^2/4) \enspace .$$

In addition, for $\delta > 0$ it holds

$$\mathbf{Pr}\left[\, X \geqslant (1+\delta)\mathbf{E}\left[\, X \,\right] \,\right] < \exp(-\mathbf{E}\left[\, X \,\right] \cdot \min\{\delta, \delta^2\}/4) \enspace .$$

Our proof of the main result also requires a variant of the method of bounded variances [9], which seems to be unpublished so far.

For random variables $X_1, \ldots, X_n$ we write $\mathbf{X}_i := (X_1, \ldots, X_i)$.

Theorem 2.2. *Let $X_1, \ldots, X_n$ be independent random variables taking values in $\{0, 1\}$, and set $\mu := \mathbf{E}\left[\, \sum_{i=1}^{n} X_i \,\right]$. Let $f := f(X_1, \ldots, X_n)$ be a function with* *finite* $\mathbf{E}[f]$ *satisfying for all* $i \in [n]$ *and all* $\mathbf{X}_{i-1}$

$$\left|\mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 0] - \mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 1]\right| \leqslant c,$$

for some $c > 0$. *Then for any* $0 \leqslant t \leqslant c\mu$ *we have*

$$\Pr[|f - \mathbf{E}[f]| > t] \leqslant 2\exp(-\tfrac{t^2}{3c^2\mu}).$$

Note that the tail bound is in terms of μ instead of n. In particular, it is completely independent of n.

Proof. We closely follow the proof of the method of bounded variances as can be found in [9]. Consider the Doob martingale $Y_i := \mathbf{E}[f \mid \mathbf{X}_i]$ and the difference $\Delta_i := Y_i - Y_{i-1}$. Note that $Y_0 = \mathbf{E}[f]$ and $Y_n = f$. Thus,

$$p_{\text{err}} := \Pr[f - \mathbf{E}[f] > t] = \Pr[Y_n - Y_0 > t] = \Pr\Big[\sum_{i=1}^{n} \Delta_i > t\Big].$$

For any $\lambda > 0$, using Markov's inequality we obtain

$$p_{\text{err}} = \Pr\left[e^{\lambda \sum_{i=1}^n \Delta_i} > e^{\lambda t}\right] \leqslant e^{-\lambda t}\mathbf{E}\left[e^{\lambda \sum_{i=1}^n \Delta_i}\right].$$

Applying basic identities of expectations we can rewrite this as

$$\begin{aligned} p_{\text{err}} &\leqslant e^{-\lambda t}\mathbf{E}\left[\mathbf{E}\left[e^{\lambda \sum_{i=1}^n \Delta_i} \mid \mathbf{X}_{n-1}\right]\right] \\ &= e^{-\lambda t}\mathbf{E}\left[\mathbf{E}\left[e^{\lambda Y_{n-1}} \cdot e^{\lambda \Delta_n} \mid \mathbf{X}_{n-1}\right]\right] \\ &= e^{-\lambda t}\mathbf{E}\left[e^{\lambda Y_{n-1}}\mathbf{E}\left[e^{\lambda \Delta_n} \mid \mathbf{X}_{n-1}\right]\right]. \end{aligned}$$

We continue by establishing an upper bound $U_n = U_n(\lambda, t, c, \mu)$ for $\mathbf{E}[[]\, e^{\lambda \Delta_n} \mid \mathbf{X}_{n-1}]$. Then by induction we obtain

$$\begin{aligned} p_{\text{err}} &\leqslant e^{-\lambda t}\mathbf{E}\left[e^{\lambda Y_{n-1}}\mathbf{E}\left[e^{\lambda \Delta_n} \mid \mathbf{X}_{n-1}\right]\right] \\ &\leqslant e^{-\lambda t}\mathbf{E}\left[e^{\lambda Y_{n-1}}\right] \cdot U_n \\ &\leqslant e^{-\lambda t}\prod_{i=1}^{n} U_i. \end{aligned} \tag{2.1}$$

To obtain such upper bounds U_i we use arguments specific to our situation as follows. By independence of $X_1, \ldots, X_n$,

$$\begin{aligned} \mathbf{E}[f \mid \mathbf{X}_{i-1}] &= \Pr[X_i = 0] \cdot \mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 0] \\ &\quad + \Pr[X_i = 1] \cdot \mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 1] \\ &= \mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 0] + p_i z_i, \end{aligned} \tag{2.2}$$

where $p_i := \Pr[X_i = 1]$ and

$$z_i := \mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 1] - \mathbf{E}[f \mid \mathbf{X}_{i-1}, X_i = 0].$$

Recall that $\Delta_i = Y_i - Y_{i-1} = \mathbf{E}\left[\, f \mid \mathbf{X}_i \,\right] - \mathbf{E}\left[\, f \mid \mathbf{X}_{i-1} \,\right]$. Note that the randomness of Δ_i depends only on $\mathbf{X}_i$. Hence, $(\Delta_i \mid \mathbf{X}_i)$ is a deterministic value, which we compute by plugging equation (2.2) into the definition of Δ_i, obtaining

$$\begin{aligned}(\Delta_i \mid \mathbf{X}_{i-1}, X_i = 0) &= -p_i z_i,\\ (\Delta_i \mid \mathbf{X}_{i-1}, X_i = 1) &= (1-p_i) z_i.\end{aligned}$$

These equations allow to compute

$$\begin{aligned}\mathbf{E}\left[\, \Delta_i^k \mid \mathbf{X}_{i-1} \,\right] =& \Pr[X_i = 0] \cdot \mathbf{E}\left[\, \Delta_i^k \mid \mathbf{X}_{i-1}, X_i = 0 \,\right]\\ &+ \Pr[X_i = 1] \cdot \mathbf{E}\left[\, \Delta_i^k \mid \mathbf{X}_{i-1}, X_i = 1 \,\right]\\ =& (1-p_i)(-p_i z_i)^k + p_i((1-p_i) z_i)^k.\end{aligned}$$

This evaluates to 1 for $k = 0$ and to 0 for $k = 1$. Since $|z_i| \leqslant c$ by the assumption on f, for $k \geqslant 2$ we have

$$\begin{aligned}\mathbf{E}\left[\, \Delta_i^k \mid \mathbf{X}_{i-1} \,\right] &\leqslant (1-p_i)(p_i c)^k + p_i((1-p_i)c)^k\\ &\leqslant p_i \cdot p_i c^k + p_i \cdot (1-p_i) c^k\\ &= p_i c^k.\end{aligned}$$

Hence, we have

$$\begin{aligned}\mathbf{E}\left[\, e^{\lambda \Delta_i} \mid \mathbf{X}_{i-1} \,\right] = \sum_{k \geqslant 0} \mathbf{E}\left[\, \lambda^k \Delta_i^k / k! \mid \mathbf{X}_{i-1} \,\right] &\leqslant 1 + \sum_{k \geqslant 2} p_i (\lambda c)^k / k!\\ &= 1 + p_i(e^{\lambda c} - 1 - \lambda c)\\ &\leqslant \exp(p_i(e^{\lambda c} - 1 - \lambda c)),\end{aligned}$$

since $1 + x \leqslant e^x$ for any x. Now we plug this upper bound into equation (2.1) to obtain

$$\begin{aligned}p_{\text{err}} &\leqslant \exp(-\lambda t + \sum_{i=1}^{n} p_i(e^{\lambda c} - 1 - \lambda c))\\ &= \exp(-\lambda t + \mu(e^{\lambda c} - 1 - \lambda c)).\end{aligned}$$

Putting $\lambda := \frac{1}{c} \ln(1 + \frac{t}{c\mu})$ yields

$$\begin{aligned}p_{\text{err}} &\leqslant \exp(-(\tfrac{t}{c} + \mu) \ln(1 + \tfrac{t}{c\mu}) + \tfrac{t}{c})\\ &= \exp(-\mu g(\tfrac{t}{c\mu})),\end{aligned}$$

where $g(x) = (1+x)\ln(1+x) - x$. We note that $g(x) \leqslant \frac{x^2}{3}$ for $x \leqslant 1.81$ to prove

$$\Pr[f - \mathbf{E}\left[\, f \,\right] > t] \leqslant \exp(-\tfrac{t^2}{3c^2\mu}).$$

Using this inequality on f and $-f$ yields the desired

$$\Pr[|f - \mathbf{E}\left[\, f \,\right]| > t] \leqslant 2\exp(-\tfrac{t^2}{3c^2\mu}). \qquad \square$$

We generalize the above concentration result to include an error event.

Theorem 2.3. *Let $X_1, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$, and set $\mu := \mathbf{E}\left[\sum_{i=1}^n X_i\right]$. Let $f := f(X_1, \ldots, X_n)$ be a function satisfying*

$$|f| \leqslant M,$$

and consider an error event $\mathcal{B}$ such that for every $\mathbf{X}_n \in \overline{\mathcal{B}}$

$$|f(\mathbf{X}_n) - f(\mathbf{X}'_n)| \leqslant c$$

for every $\mathbf{X}'_n$ that differs in only one position X_i from $\mathbf{X}_n$, and for some $c > 0$. Then for any $0 \leqslant t \leqslant c\mu$ we have

$$\Pr\left[\left|f - \mathbf{E}\left[\, f\,\right]\right| > t + \tfrac{(2M)^2}{c}\Pr[\mathcal{B}]\right] \leqslant \tfrac{2M}{c}\Pr[\mathcal{B}] + 2\exp\left(-\tfrac{t^2}{16c^2\mu}\right).$$

Proof. We construct a function $g = g(X_1, \ldots, X_n)$ as follows. Iteratively sample $X_1, \ldots, X_n$. If after sampling X_{i-1}, $i \in [n]$, we have

$$\left|\mathbf{E}\left[\, f \mid X_i = 1, \mathbf{X}_{i-1}\right] - \mathbf{E}\left[\, f \mid X_i = 0, \mathbf{X}_{i-1}\right]\right| > 2c, \tag{2.3}$$

then set $g(X_1, \ldots, X_n) := \mathbf{E}\left[\, f \mid \mathbf{X}_{i-1}\right]$ for all $X_i, \ldots, X_n$. If this event never occurs, then we indeed sample $X_1, \ldots, X_n$ and simply set $g(X_1, \ldots, X_n) := f(X_1, \ldots, X_n)$.

Claim. Fix any $\mathbf{X}_n$ and $i \in [n]$. If (2.3) does not hold for any $i' < i$, then

$$\mathbf{E}\left[\, g \mid \mathbf{X}_{i-1}\right] = \mathbf{E}\left[\, f \mid \mathbf{X}_{i-1}\right].$$

Proof. If (2.3) holds for i, then we set $g(X_1, \ldots, X_n) = \mathbf{E}\left[\, f \mid \mathbf{X}_{i-1}\right]$, so the claim holds trivially. If $i = n$ and (2.3) does not hold for i then we set $g(X_1, \ldots, X_n) = f(X_1, \ldots, X_n)$, so the claim holds trivially. If $i < n$ and (2.3) does not hold for i, then we have

$$\mathbf{E}\left[\, g \mid \mathbf{X}_{i-1}\right] = \Pr[X_i = 1]\cdot\mathbf{E}\left[\, g \mid X_i = 1, \mathbf{X}_{i-1}\right] + \Pr[X_i = 0]\cdot\mathbf{E}\left[\, g \mid X_i = 0, \mathbf{X}_{i-1}\right].$$

Inductively, this equals

$$\begin{aligned}\mathbf{E}\left[\, g \mid \mathbf{X}_{i-1}\right] &= \Pr[X_i = 1]\cdot\mathbf{E}\left[\, f \mid X_i = 1, \mathbf{X}_{i-1}\right] + \Pr[X_i = 0]\cdot\mathbf{E}\left[\, f \mid X_i = 0, \mathbf{X}_{i-1}\right] \\ &= \mathbf{E}\left[\, f \mid \mathbf{X}_{i-1}\right],\end{aligned}$$

which finishes the proof of the claim. □

Claim. For all $i \in [n]$ and $\mathbf{X}_{i-1}$ we have

$$\left|\mathbf{E}\left[\, g \mid X_i = 0, \mathbf{X}_{i-1}\right] - \mathbf{E}\left[\, g \mid X_i = 1, \mathbf{X}_{i-1}\right]\right| \leqslant 2c.$$

Proof. If (2.3) holds for some $i' \leqslant i$ then both $(g \mid X_i = 0, \mathbf{X}_{i-1})$ and $(g \mid X_i = 1, \mathbf{X}_{i-1})$ are identically $\mathbf{E}\left[\, g \mid \mathbf{X}_{i'-1}\right]$. Otherwise (2.3) does not hold for i, and since by the last claim we have $\mathbf{E}\left[\, f \mid X_i = k, \mathbf{X}_{i-1}\right] = \mathbf{E}\left[\, g \mid X_i = k, \mathbf{X}_{i-1}\right]$ for any $k \in \{0,1\}$, the statement follows. □

Thus, we can use Theorem 2.2 on g, which yields

$$\Pr[|g - \mathbf{E}[g]| > t] \leqslant 2\exp(-\tfrac{t^2}{16c^2\mu}).$$

In the remainder, we show the upper bound

$$\Pr[f \neq g] \leqslant \tfrac{2M}{c}\Pr[\mathcal{B}] =: U. \tag{2.4}$$

This then implies $|\mathbf{E}[g] - \mathbf{E}[f]| \leqslant 2M \cdot U$, since $|f|, |g| \leqslant M$. Finally, we have

$$\begin{aligned}\Pr[|f - \mathbf{E}[f]| > t + 2M \cdot U] &\leqslant \Pr[|f - \mathbf{E}[g]|| > t] \\ &\leqslant U + \Pr[|g - \mathbf{E}[g]| > t] \\ &\leqslant U + 2\exp(-\tfrac{t^2}{16c^2\mu}).\end{aligned}$$

Consider the set S of all outcomes $\mathbf{X}_{i-1}$ for any $i \in [n]$ such that (2.3) holds for i but not for any $i' < i$. Note that

$$\Pr[f \neq g] \leqslant \sum_{\mathbf{X}_{i-1} \in S} \Pr[\mathbf{X}_{i-1}],$$

since only when (2.3) holds for the first time we set g to a value that possibly differs from f. Recall that for any $\mathbf{X}_n$ in $\overline{\mathcal{B}}$ we have $|f(\mathbf{X}_n) - f(\mathbf{X}_n^{(i)})| \leqslant c$, where $\mathbf{X}_n^{(i)}$ stems from $\mathbf{X}_n$ by flipping X_i. Using this, inequality (2.3), and $|f| \leqslant M$, for any $\mathbf{X}_{i-1}$ in S, $i \in [n]$, we have

$$\Pr[\mathbf{X}_n \in \mathcal{B} \mid \mathbf{X}_{i-1}] \geqslant \tfrac{c}{2M}.$$

Since the sequences $\mathbf{X}_{i-1}$ in S correspond to disjoint events, we have

$$\begin{aligned}\Pr[\mathcal{B}] \geqslant \sum_{\mathbf{X}_{i-1} \in S} \Pr[\mathbf{X}_{i-1}] \cdot \Pr[\mathbf{X}_n \in \mathcal{B} \mid \mathbf{X}_{i-1}] &\geqslant \tfrac{c}{2M} \sum_{\mathbf{X}_{i-1} \in S} \Pr[\mathbf{X}_{i-1}] \\ &\geqslant \tfrac{c}{2M}\Pr[f \neq g].\end{aligned}$$

This proves inequality (2.4), as desired. □

3 Lower Bounds for other Load Balancing Algorithms and Networks

First notice that scale-free networks are not only heterogeneous, but also sparse in the sense of $|E| = \Theta(n)$. One of the motivations of our work is the fact that double-logarithmic load balancing time is a rare feature in sparse graphs. To make this observation more tangible, this section presents two logarithmic lower bounds: First, we show that the standard diffusion protocol is not able to achieve double logarithmic balancing time in heterogeneous graphs. Second, we show that heterogeneity is needed to achieve double logarithmic balancing time for sparse graphs.

Recall that FOS diffusion is defined by the diffusion matrix $\mathbf{P} = \mathbf{I} - \alpha \cdot \mathbf{L}$, where $\mathbf{I}$ is the n by n identity matrix, $\mathbf{L}$ is the Laplace matrix, and $\alpha < \frac{1}{\operatorname{maxdeg}(G)}$ is the damping factor (the smaller α, the more load is retained at a node). The load vector $x^{(t+1)}$ satisfies $x^{(t+1)} = \mathbf{P} \cdot x^{(t)}$, and thus converges towards the uniform distribution for any connected, possibly non-regular, graph. By $\overline{x}$, we denote the average load.

Lemma 3.1. *Let $G = (V, E)$ be a non-regular graph with $\frac{\operatorname{mindeg}(G)}{\operatorname{maxdeg}(G)} < 1 - c'$ for a constant $c' > 0$ and consider the execution of FOS with matrix $\mathbf{P}$. Let u be any node with $\deg(u) \leqslant (1-c)\operatorname{maxdeg}(G)$ for a constant $c > 0$. Then if the initial load vector is one at u and zero elsewhere it holds that*

$$x_u^{(t)} \geqslant c^t.$$

In particular, the time until u has a load $\mathcal{O}(\overline{x}) = \mathcal{O}(1/n)$ is at least $\Omega(\log n)$.

Proof. It follows by the definition of FOS that $P_{u,u} = 1 - \alpha \cdot \deg(u) \geqslant 1 - \frac{1}{\operatorname{maxdeg}(G)} \cdot \deg(u) \geqslant c$ and therefore

$$x_u^{(t)} \geqslant P_{u,u} \cdot x_u^{(t-1)} \geqslant c \cdot x_u^{(t-1)},$$

and hence $x_u^{(t)} \geqslant c^t$. $\square$

While this choice of the diffusion matrix $\mathbf{P}$ with all non-zero entries $P_{u,v}$ being identical is the most common one [16], there are also other possibilities. For instance, we could define $\mathbf{P} = \mathbf{D}^{-1} \cdot \mathbf{A}$ corresponding to the way we balance the load in the core (see Section 5). However, this would converge towards a load distribution where load is proportional to the degree. Furthermore, by using a similar argument as in the proof of Lemma 3.1 it follows that whenever load is initially on a node u with degree $\mathcal{O}(\log n)$ and this node has at least one neighbor with degree $\mathcal{O}(\log n)$, then it takes at least $\Omega(\log n / \log\log n)$ rounds before the load distribution is balanced.

We could see that the high heterogeneity of scale-free networks doesn't allow us to use simple FOS diffusion with matrices defined as before if we want to achieve double logarithmic balancing time. The following Lemma shows that this heterogeneity is essential for achieving double logarithmic runtime on sparse graphs.

Lemma 3.2. *Let $G = (V, E)$ be any d-regular graph ($d \geqslant 2$ constant), and consider any iterative load balancing protocol which is being run for τ steps. Then for any initial load vector which is one at a vertex and zero elsewhere, there exists a node u with $x_u^{(\tau)} = \Omega(d^{-\tau})$. In particular, it takes at least $\Omega(\log n)$ rounds until the maximum load is bounded by $\mathcal{O}(\overline{x})$.*

Proof. Consider the number of nodes which can be reached by a path of length τ from v, where v is the node with $x_v^{(0)} = 1$. This number of nodes is at most $d + d^2 + d^3 + \cdots + d^\tau \leqslant 2d^\tau$. Since only these nodes can have a non-zero load, it follows by the pigeonhole principle that there must be a node u with $x_u^{(t)} \geqslant \frac{x_u^0}{2d^\tau} = \frac{1}{2d^\tau}$. $\square$

4 Model, Algorithms, and Formal Result

Chung-Lu random graph model. We consider random graphs $G = (V, E)$ as defined by Chung and Lu [4]. Every vertex $i \in V = \{1, \ldots, n\}$ has a weight w_i with

$$w_i := \frac{\beta - 2}{\beta - 1} d n^{1/(\beta-1)} i^{-1/(\beta-1)}$$

for $i = 1, 2 \ldots, n$. The probability for placing an edge $\{i, j\} \in E$ is then set to $\min\{w_i\, w_j / W, 1\}$ with $W := \sum_{i=1}^{n} w_i$. This creates a random graph where the expected degrees follow a power-law distribution with exponent $\beta \in (2, 3)$, the maximum expected node degree is $\frac{\beta-2}{\beta-1} d n^{1/(\beta-1)}$ and the parameter d can be used to scale the average expected node degree [4]. The graph has a *core* of densely connected nodes which we define as

$$C := \left\{ i \in V \colon\ \deg_i \geqslant n^{1/2} - \sqrt{n^{1/2} \cdot (c+1) \ln n} \right\}.$$

Distributing the load in waves. Our main algorithm is presented in Algorithm 1. It assumes that an initial total load of m resides exclusively on the core C of the network. The first rounds are spent on simple diffusion on the core with diffusion matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{A}$ is the adjacency matrix and $\mathbf{D}$ is the degree matrix. Afterwards, the algorithm pushes the load to all other nodes in waves from the large to the small degree nodes and the other way around. To define the direction of the waves, the algorithm partitions the nodes into layers, where on layer k we have all nodes v of degree $\deg_v \in (\omega_k, \omega_{k-1}]$, where $\omega_0 = n^{1/2} - \sqrt{n^{1/2} \cdot (c+1) \ln n}$ and $\omega_{k+1} = \omega_k^{1-\varepsilon}$ for a constant

$$0 < \varepsilon < \min \left\{ \tfrac{(3-\beta)}{(\beta-1)},\ \tfrac{\beta-2}{3},\ \tfrac{1}{2}\left(1 - \sqrt{\tfrac{3}{\beta+1}}\right) \right\}.$$

For every layer k we have $\omega_k > 2^{\frac{1}{\varepsilon(\beta-1)}}$. The last layer ℓ is the first, for which $\omega_\ell \leqslant 2^{\frac{1}{\varepsilon(\beta-1)}}$ holds. In this case, we define the interval simply to include all nodes with degree less than $\omega_{\ell-1}$. Note that in total we obtain at most

$$L := \tfrac{1}{\log(1/(1-\varepsilon))} \left(\log\log n + \log \tfrac{\varepsilon(\beta-1)}{2} \right)$$

layers.

To choose an appropriate ε, we have to know lower and upper bounds on β. These bounds are either known or can be chosen as constants arbitrarily close to 2 and 3. The algorithm therefore does not need to know the precise β. Our main result is then as follows.

Theorem 4.1. *Let $G = (V, E)$ be a Chung-Lu random graph as defined above. For any load vector $x^{(0)} \in \mathbb{R}^n_{\geqslant 0}$ with support only on the core C of the graph, there is a $\tau = \mathcal{O}((\log\log n)^2)$ such that for all steps $t \geqslant \tau$ of Algorithm 1, the resulting load vector $x^{(t)}$ fulfills $x_u^{(t)} = \mathcal{O}(\overline{x})$ for all $u \in V$ w. h. p.*[1]

[1] w. h. p. is short for "with high probability". We use w.h.p. to describe events that hold with probability $1 - n^{-c}$ for an arbitrary large constant c.

Algorithm 1: Balance load in waves from core to all other nodes

```
repeat
    for phase t ← 1 to log log n do
        for 32/(3−β) rounds do                                   // 1. diffusion on the core
            Nodes v with deg(v) ⩾ ω_0 perform diffusion with P = D^{-1}A
        for L rounds do                                          // 2. downward propagation
            Every node absorbs at most m/nt^2 load
            All remaining load is forwarded in equal shares to neighbors on
            the next lower layer
        for L rounds do                                          // 3. upward propagation
            All nodes send their load back to the the next higher layer over
            the edges they received it from
until terminated;
```

In the analysis of our algorithm we need the following definitions and observations. We define $V_k = \{v \mid w_v \in (\omega_k, \omega_{k-1}]\}$ as the set of nodes on layer k and $n_k = |V_k|$. Let $W_k = \sum_{v \in V_k} w_v$ be the total weight of nodes in layer k. For the sake of simplification, we also define the following value $\gamma := \frac{1}{2}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}$. From the given weight sequence and the requirements $\omega_k > 2^{\frac{1}{\varepsilon(\beta-1)}}$ and $\omega_k < n^{1/(\beta-1)}$, we can easily derive the following bounds. For all $0 \leqslant k < \ell$ it holds that

$$\frac{\gamma}{2} \cdot n\omega_k^{1-\beta} \leqslant n_k \leqslant 4\gamma \cdot n\omega_k^{1-\beta}. \tag{4.1}$$

This implies

$$W_k \geqslant \frac{\gamma}{2} \cdot n\omega_k^{2-\beta}. \tag{4.2}$$

Finally, let $\overline{d} = \frac{W}{n}$ the expected average degree. After stating the main result, we will now turn to some probabilistic tail bounds which will be used extensively throughout the rest of the paper.

Reaching the core. Algorithm 1 and Theorem 4.1 above require that the initial total load resides exclusively on the core C of the network. As the diameter of the network is $\Theta(\log n)$ [4], we cannot hope to achieve a double-logarithmic balancing time if all the initial load starts at an arbitrary small and remote vertex. However, we can allow initial load on all nodes with at least some polylogarithmic degree and run the following simple Algorithm 2 to bring the load to the core.

Algorithm 2: Push load to large nodes in core

```
for L rounds do
    Every node sends all load to any node in next higher layer;
```

In Algorithm 2 all nodes send all their load to an arbitrary neighbor on the next-highest layer. This local routing algorithm succeeds if all nodes have at least one neighbor on the next-highest layer. The following lemma states that this is the case for all nodes of degree at least $\Omega\left((\ln n)^3 + (\ln n)^{2/(3-\beta)}\right)$.

Lemma 4.2. *For $\varepsilon < \min\left\{1/3,\ \frac{3-\beta}{\beta-1}\right\}$ all nodes with degree at least*

$$\max\left\{128\,(c\ln n)^3 + 3c\ln n,\ 24^{\frac{3(1-\varepsilon)}{1-3\varepsilon}},\ 72^{\frac{3(1-\varepsilon)}{1+3\varepsilon(\beta-1)}},\right.$$
$$\left.2\left(\left(8c\ln n + \frac{3}{\overline{d}}\right)8\overline{d}\left(d\frac{\beta-2}{\beta-1}\right)^{1-\beta}\right)^{\frac{2}{3-\beta}} + 3c\ln n\right\}$$

have at least one neighbor on the next-highest layer with probability at least $1 - 4n^{-c}$.

Proof. First we note that due to Theorem 2.1

$$\Pr(|\deg_v - w_v| > w_v^{2/3}) \leqslant \exp(-w_v^{1/3}/4)$$

for all $v \in V$. This means that all nodes with $w_v > 64(c\ln n)^3$ fulfill $|\deg_v - w_v| \leqslant w_v^{2/3}$ with probability at least $1 - n^{-c}$. Especially, nodes $v \in V_k$ with $w_v \in (\omega_k + \omega_{k-1}^{2/3},\ \omega_{k-1} - \omega_{k-1}^{2/3})$ do not leave layer k. Let V_k' the set of these nodes and $W_k' = \sum_{v\in V_k'} w_v$ their total weight. We now want to show that each node on layer k has at least one neighbor in V_k'. Due to equations (6.4) and (6.6) it holds that

$$\begin{aligned}
|V_k'| &\geqslant \left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1} n\left(\frac{1}{2}\omega_k^{1-\beta} - 3\omega_k^{1-\beta+\frac{3\varepsilon-1}{3(1-\varepsilon)}} - 6\omega_{k-1}^{1-\beta-\frac{1}{3}}\right) - 3\\
&= \left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1} n\cdot\omega_k^{1-\beta}\left(\frac{1}{2} - 3\omega_k^{\frac{3\varepsilon-1}{3(1-\varepsilon)}} - 6\omega_k^{\frac{3\varepsilon(1-\beta)-1}{3(1-\varepsilon)}}\right) - 3.
\end{aligned}$$

The requirement $\deg_v \geqslant \max\left\{24^{\frac{3(1-\varepsilon)}{1-3\varepsilon}},\ 72^{\frac{3(1-\varepsilon)}{1+3\varepsilon(\beta-1)}}\right\}$ ensures

$$|V_k'| \geqslant \frac{1}{4}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1} n\cdot\omega_k^{1-\beta} - 3$$

and therefore

$$W_k' \geqslant \omega_k\cdot|V_k'| \geqslant \frac{1}{4}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1} n\cdot\omega_k^{2-\beta} - 3\omega_k.$$

Let D'_v the number of edges node $v \in V_k$ has to V'_{k-1}. For $v \in V_k$ it holds that

$$\begin{aligned}\mathbf{E}\left[\, D'_v \,\right] &= w_v \frac{W'_{k-1}}{W} \\ &\geqslant \omega_k \omega_{k-1}^{2-\beta} \frac{\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}}{4\overline{d}} - \frac{3\omega_k\omega_{k-1}}{\overline{d}n} \\ &\geqslant \omega_k^{\frac{3-\beta}{2}} \frac{\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}}{4\overline{d}} - \frac{3}{\overline{d}},\end{aligned}$$

where we used $\omega_k \leqslant n^{1/2}$ and $\omega_k\omega_{k-1}^{2-\beta} \geqslant \omega_k^{\frac{3-\beta}{2}}$, which holds for $\varepsilon \leqslant \frac{3-\beta}{\beta-1}$ as in the proof of Lemma 6.3. Again, we apply a Chernoff Bound and show

$$\Pr(D'_v < \frac{1}{2}\mathbf{E}\left[\, D'_v \,\right]) \leqslant \exp\left(-\frac{\mathbf{E}\left[\, D'_v \,\right]}{8}\right).$$

For $\omega_k \geqslant \left(\left(8c\ln n + \frac{3}{\overline{d}}\right) 4\overline{d} \left(d\frac{\beta-2}{\beta-1}\right)^{1-\beta}\right)^{\frac{2}{3-\beta}}$ it holds that $\mathbf{E}\left[\, D'_v \,\right] \geqslant 8c\ln n$. This implies $D'_v \geqslant 4c\ln n > 1$ with probability at least $1 - n^{-c}$.

For nodes $v \in V_k \setminus V'_k$ we have to show that they have enough edges into both layers they might end up in. Nodes $v \in V_k$ with $w_v \in \left[\omega_k,\ \omega_k + \omega_{k-1}^{2/3}\right]$ can end up in layer k or $k+1$. So we have to show, that they have enough edges to V'_k as well. The expected number of these edges is

$$w_v \frac{W'_k}{W} \geqslant \omega_k \omega_k^{2-\beta} \frac{\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}}{4\overline{d}} - \frac{3\omega_k\omega_k}{\overline{d}n}.$$

Since this lower bound is even bigger than the one for $\mathbf{E}\left[\, D'_v \,\right]$, the same bounds hold in this case.

Nodes $v \in V_k$ with $w_v \in \left[\omega_{k-1} - \omega_{k-1}^{2/3},\ \omega_{k-1}^{2/3}\right]$ can end up in layer k or $k-1$. The expected number of edges from such a node v to V'_{k-2} is

$$w_v \frac{W'_{k-2}}{W} \geqslant \omega_{k-1}\left(1 - \omega_{k-1}^{-1/3}\right) \omega_{k-2}^{2-\beta} \frac{\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}}{4\overline{d}} - \frac{3\omega_{k-1}\left(1 - \omega_{k-1}^{-1/3}\right)\omega_{k-2}}{\overline{d}n}.$$

For $\omega_k \geqslant 8$ it holds that $\omega_{k-1}\left(1 - \omega_{k-1}^{-1/3}\right) \geqslant \frac{1}{2}\omega_{k-1}$. Again we can use $\varepsilon \leqslant \frac{3-\beta}{\beta-1}$ to show $\omega_{k-1}\omega_{k-2}^{2-\beta} \geqslant \omega_{k-1}^{\frac{3-\beta}{2}}$, which gives us the same results as above if $\omega_{k-1} \geqslant \left(\left(8c\ln n + \frac{3}{\overline{d}}\right) 8\overline{d} \left(d\frac{\beta-2}{\beta-1}\right)^{1-\beta}\right)^{\frac{2}{3-\beta}}$.

A last thing we need to ensure is that the nodes with the required minimum degree also have the minimum weight they need. This part is easily shown by a Chernoff Bound, which states

$$\Pr\left(\deg_v > 2w_v + 3c\ln n\right) < \exp\left(-\frac{3c\ln n}{3}\right) = n^{-c}$$

for all $v \in V$. This means, each node $v \in V$ has a weight of at least $\frac{\deg_v - 3c \ln n}{2}$. Together with our requirements for ω_k this establishes the requirements of the lemma. At last we utilize a union bound to collect all failure probabilities and establish the result as desired. $\square$

The following statement is a simple corollary of the above lemma.

Theorem 4.3. *Let $G = (V, E)$ be a Chung-Lu random graph as defined in section 4. For any load vector $x^{(0)} \in \mathbb{R}^n_{\geqslant 0}$ with support only on nodes with degree $\Omega((\log n)^{\max(3,2/(3-\beta))})$, Algorithm 2 reaches after $L = \Theta(\log \log n)$ steps a load vector $x^{(L)}$ such that $x_u^{(L)}$ has support only on the core C w. h. p.*

The rest of the paper proves Theorem 4.1.

5 Analysis of Load Balancing on the Core

We start our analysis of Algorithm 1 with its first step, the diffusion on the core. Recall the definition of the core C of the network and consider the core subgraph $\widetilde{G} = (\widetilde{V}, \widetilde{E})$ induced by C. The following lemma provides an upper bound on the spectral gap of $\widetilde{G}$'s normalized Laplacian, which will be used in Lemma 5.2 to bound the convergence rate of the diffusion process in the core.

Lemma 5.1. *The core subgraph $\widetilde{G}$ of G fulfills*

$$|1 - \lambda_k(\mathbf{L})| \leqslant \Theta\left(\frac{\sqrt{(c+1)\ln(4n)}}{n^{(3-\beta)/4}} + \frac{(2(c+1)\ln n)^{1/4}}{n^{1/8}}\right)$$

for all eigenvalues $\lambda_k(\mathbf{L}) > \lambda_{\min}(\mathbf{L})$ of the normalized Laplacian $\mathbf{L}(\widetilde{G})$ w. h. p.

Proof. First of all, we note that ω_0 is chosen in such a way, that, according to a Chernoff bound, all nodes with weight at least $n^{1/2}$ remain in the core with probability at least $1 - 1/n^{c+1}$. We will henceforth denote the set of these nodes by C. Additionally, with probability at least $1 - 1/n^{c+1}$ only nodes with weight bigger than or equal to $n^{1/2} - \sqrt{n^{1/2} \cdot 2(c+1)\ln n} = n^{1/2}\left(1 - \frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)$ can possibly join the core. Let

$$C' = \left\{ v \in V \mid w_v \in \left[n^{1/2}\left(1 - \frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right), n^{1/2}\right)\right\}$$

the set of potential additional core nodes. To bound the number of these additional nodes, we observe that the number of nodes with weight at least $n^{1/2}\left(1 - \frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)$ is at most

$$\left(\frac{\beta-2}{\beta-1}d\right)^{\beta-1} \cdot n^{(3-\beta)/2}\left(1 - \frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)^{1-\beta}.$$

The number of nodes with weight at least $n^{1/2}$ is at least

$$\left(\frac{\beta-2}{\beta-1}d\right)^{\beta-1}\cdot n^{(3-\beta)/2}-1.$$

This gives us

$$|C'|\leqslant\left(\tfrac{\beta-2}{\beta-1}d\right)^{\beta-1}\cdot n^{(3-\beta)/2}\left(\left(1-\tfrac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)^{1-\beta}-1\right)+1.$$

We can now bound

$$\begin{aligned}
&\left(\left(1-\tfrac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)^{1-\beta}-1\right)\\
=&\quad\left(\left(1+\tfrac{\sqrt{2(c+1)\ln n}}{n^{1/4}-\sqrt{2(c+1)\ln n}}\right)^{\beta-1}-1\right)\\
<&\quad\left(1+2\tfrac{\sqrt{2(c+1)\ln n}}{n^{1/4}-\sqrt{2(c+1)\ln n}}+\tfrac{2(c+1)\ln n}{\left(n^{1/4}-\sqrt{2(c+1)\ln n}\right)^2}-1\right)\\
\leqslant&\quad 6\tfrac{\sqrt{2(c+1)\ln n}}{n^{1/4}},
\end{aligned}$$

where the second last line follows with $\beta<3$ and the last line holds for sufficiently large n.

For each possible set $S\subseteq C\cup C'$ of core nodes, we look at the Chung-Lu graph induced by those nodes. These graphs can be interpreted as Chung-Lu random graphs of their own if we scale the node weights by a factor of $\frac{W_S}{W}$, where W_S denotes the total weight of the nodes from S. By doing so, all edge probabilities remain the same. We already know that $W_S\geqslant n^{1/2}\left(\left(\frac{\beta-2}{\beta-1}d\right)^{\beta-1}\cdot n^{(3-\beta)/2}-1\right)$ and $W=\overline{d}\cdot n$. This means, that the minimum rescaled weight of a node from S is at least

$$w_{\min}:=\frac{1}{\overline{\overline{d}}}\left(1-\frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)\left(\left(\frac{\beta-2}{\beta-1}d\right)^{\beta-1}\cdot n^{(3-\beta)/2}-1\right)=\Theta\left(n^{(3-\beta)/2}\right).$$

Now we can use Theorem 2 from [13] with the failure probability set to $\varepsilon:=2^{-|C'|}\cdot n^{-c}$. This gives us

$$\begin{aligned}
|1-\lambda_k(\mathbf{L})|\leqslant 2\sqrt{\frac{3\ln(4n/\varepsilon)}{w_{\min}}}&\\
=\Theta\left(\sqrt{\frac{\ln(4n^{c+1})+n^{(3-\beta)/2}\frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\ln 2}{n^{(3-\beta)/2}}}\right)&\\
=\Theta\left(\frac{\sqrt{(c+1)\ln(4n)}}{n^{(3-\beta)/4}}+\frac{(2(c+1)\ln n)^{1/4}}{n^{1/8}}\right)&\qquad(5.1)
\end{aligned}$$

for the eigenvalues $\lambda_k(\mathbf{L}) > \lambda_{\min}(\mathbf{L})$ of the normalized Laplacian $\mathbf{L}$ of the induced subgraph on S. As we consider only $2^{|C'|}$ many subsets, we can utilize a union bound to show that with probability at most $2^{|C'|}\varepsilon = n^{-c}$ one of these subgraphs does not fulfill bound 5.1. Collecting the n failure probabilities for the node degrees and the one for the eigenvalues gives us the desired result with probability at least $1 - \frac{2}{n^c}$. □

The following lemma states that after only a constant number of diffusion rounds in $\widetilde{G}$, the load of node $v \in C$ is more or less equal to $m \cdot w_v / W_0$.

Lemma 5.2. *After $\frac{32}{3-\beta}$ rounds of diffusion with $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$ in the core subgraph $\widetilde{G}$, each node $v \in C$ has a load of at most $\mathcal{O}\left(\frac{w_y}{\sum_{x\in C} w_x} m\right)$ w. h. p.*

Proof. Recall that we perform diffusion with a diffusion matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{A}$ is the adjacency matrix and $\mathbf{D}$ is the degree matrix. We now show that the eigenvalues of $\mathbf{D}^{-1}\mathbf{A}$ and $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ are the same. Indeed, let v be an eigenvector with eigenvalue λ of $\mathbf{D}^{-1}\mathbf{A}$. Then $\mathbf{D}^{1/2} \cdot v$ is an eigenvector with eigenvalue λ of $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, since

$$\left(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\right) \cdot \mathbf{D}^{1/2} \cdot v = \mathbf{D}^{1/2}\mathbf{D}^{-1}\mathbf{A} \cdot v = \mathbf{D}^{1/2}\lambda v = \lambda \mathbf{D}^{1/2} v.$$

Since the diffusion matrix $\mathbf{P}$ is reversible, we can use the following result (see book by Peres [15] eq. 12.11):

$$\left|\frac{\mathbf{P}^t_{x,y}}{\pi_y} - 1\right| \leqslant \frac{\lambda^t_{\max}}{\pi_{\min}}$$

for $\lambda_{\max} = \max\{|\lambda_k| : \lambda_k \text{ eigenvalue of P}, \lambda_k \neq 1\}$. Since the normalized Laplacian is defined as $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, we know $\lambda_{\max} = \Theta\left(\frac{\sqrt{(c+1)\ln(4n)}}{n^{(3-\beta)/4}} + \frac{(2c\ln n)^{1/4}}{n^{1/8}}\right)$ due to Lemma 5.1. Now we choose $t := \frac{32}{3-\beta}$ to obtain that

$$\mathbf{P}^t_{x,y} = \pi_y \pm n^{-4}.$$

It now holds that

$$\pi_y = \deg_{\widetilde{G}}(y) / \sum_{x\in C} \deg_{\widetilde{G}}(x)$$

and

$$\deg_{\widetilde{G}}(y) \leqslant 2 \cdot \frac{w_v}{W} W_{\widetilde{G}} = \mathcal{O}\left(\frac{w_v}{W} n \left(n^{1/2}\left(1 - \frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)\right)^{2-\beta}\right) = \mathcal{O}\left(w_v \cdot n^{\frac{2-\beta}{2}}\right)$$

due to Theorem 2.1, the fact that by not allowing self-loops and using $\min\left\{\frac{w_i w_j}{W}, 1\right\}$ the edge probabilities only decrease and the observation from Lemma 5.1 that only nodes of weight at least $n^{1/2}\left(1 - \frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)$ can join

the core w.h.p.
Furthermore it holds that

$$\sum_{x\in\widetilde{G}} \deg_{\widetilde{G}}(x) \geqslant \Theta\left(\left(n\cdot n^{\frac{1-\beta}{2}}\right)^2\right) = \Theta\left(n^{3-\beta}\right)$$

since the core contains a complete sub graph of nodes with weights at least $\sqrt{W} = \Theta(n^{1/2})$. We also know that

$$\begin{aligned} W_0 &\leqslant \sum_{i=1}^{\left(\frac{\beta-2}{\beta-1}d\right)^{\beta-1} n\left(n^{1/2}\left(1-\frac{\sqrt{2(c+1)\ln n}}{n^{1/4}}\right)\right)^{1-\beta}} \frac{\beta-2}{\beta-1} d\left(\frac{n}{i}\right)^{\frac{1}{\beta-1}} \\ &= \mathcal{O}\left(n\cdot n^{\frac{2-\beta}{2}}\right). \end{aligned}$$

It now holds that

$$\begin{aligned} \pi_y &= \deg_{\widetilde{G}}(y)/\sum_{x\in C}\deg_{\widetilde{G}}(x) \\ &= \mathcal{O}\left(w_v\cdot\frac{n^{\frac{2-\beta}{2}}}{n^{3-\beta}}\right) \\ &= \mathcal{O}\left(w_v\cdot n\cdot n^{\frac{2-\beta}{2}}\right) \\ &= \mathcal{O}\left(\frac{w_y}{W_0}\right). \end{aligned}$$

Since also $\pi_y = \deg_{\widetilde{G}}(y)/\sum_{x\in C}\deg_{\widetilde{G}}(x) = \mathcal{O}\left(w_y/\sum_{x\in C} w_x\right)$ for all $y\in C$ with probability at least $1-1/n^c$, it follows that

$$\mathbf{P}^t_{x,y} = \mathcal{O}\left(w_y/\sum_{x\in C} w_x\right) \pm o(n^{-1}) = \mathcal{O}\left(w_y/\sum_{x\in C} w_x\right)$$

after $\frac{32}{3-\beta}$ rounds. $\square$

An implication of the lemma is, that there is a constant $\varepsilon_0 > 0$ such that each node $v\in C$ has a load of at most $(1+\varepsilon_0)\frac{w_v}{W_0}m$ after the first phase of Algorithm 1.

6 Analysis of Top-Down Propagation

We continue our analysis of Algorithm 1. This section studies the downward/upward propagation.

First note that our algorithm deals with a random graph and therefore it might happen that some of the nodes' neighborhoods look significantly different from what one would expect by looking at the expected values. We call these nodes *dead-ends* as they can not be utilized to effectively forward load. This definition will be made precise in Definition 6.1 below.

Only for the sake of analysis we assume that dead-ends do not push load to neighbors on the next lowest layer, but instead keep all of it. In reality the algorithm does *not* differentiate between nodes which are dead-ends and nodes which are no dead-ends. We also assume in this section that nodes do not consume any load during the top-down distribution.

The main goal of this section is twofold. We first show that no node which is not a dead-end, gets too much load. Then we show that the total load on all dead-ends from the core down to a layer with nodes of a certain constant degree is at most a constant fraction of the total load. The converse means that at least a constant fraction of load reaches the nodes of the last layer we are considering.

For a node $v \in V_k$ we consider two partial degrees. Let D_v^h be the number of edges to nodes in the higher layer $k-1$, and D_v^ℓ is the number of edges to nodes in the lower layer $k+1$. Note that D_v^h and D_v^ℓ are random variables, composed of sums of independent Bernoulli trials:

$$D_v^h = \sum_{u \in V_{k-1}} \mathsf{Ber}\left(\frac{w_v \cdot w_u}{W}\right) \quad \text{and} \quad D_v^\ell = \sum_{u \in V_{k+1}} \mathsf{Ber}\left(\frac{w_v \cdot w_u}{W}\right) \quad .$$

In our proofs we will apply several well-known Chernoff bounds which use the fact that partial degrees are sums of independent Bernoulli trials.

We now define four properties which will be used throughout the analysis.

Definition 6.1. *A node $v \in V$ is a* dead-end *if one of the following holds:*

⟨D1⟩ In-/Out-degree: *A node $v \in V_k$ has this property if either $|D_v^h - \mathbf{E}\left[D_v^h\right]| \geqslant \mathbf{E}\left[D_v^h\right]^{2/3}$ or $|D_v^\ell - \mathbf{E}\left[D_v^\ell\right]| \geqslant \mathbf{E}\left[D_v^\ell\right]^{2/3}$.*

⟨D2⟩ Wrong layer: *A node $v \in V_k$ has this property if it has a degree that deviates by at least $w_v^{2/3}$ from its expected degree.*

⟨D3⟩ Border: *A node $v \in V_k$ has this property if it does not fulfill property ⟨D2⟩ and if it is of weight at least $\omega_{k-1} - \omega_{k-1}^{2/3}$ or at most $\omega_k + \omega_{k-1}^{2/3}$ and if it is assigned to the wrong layer.*

⟨D4⟩ Induced Out-degree: *A node $v \in V_k$ has this property if it fulfills none of the properties ⟨D1⟩ – ⟨D3⟩ and if it has at least $(\omega_k W_{k+1}/W)^{2/3}$ many lower-layer neighbors with properties ⟨D2⟩ or ⟨D3⟩.*

The next lemma shows that for a non-dead-end node $v \in V_k$ the received load x_v in phase k is almost proportional to the "layer-average load" $m \cdot w_v / W_k$. For dead-ends, the received load can be higher, but the probability to receive significantly higher load is small.

Lemma 6.2. *For $v_k \in V_k$ and the received load x_v in phase k the following holds. If v is not a dead-end,*

$$x_v \leqslant (1+\varepsilon_k) \cdot m \cdot \frac{w_v}{W_k} \quad ,$$

where for every layer k the error term ε_k is given by

$$(1+\varepsilon_k) = (1+\varepsilon_{k-1}) \cdot (1+\mathcal{O}(\omega_k^{-1+\beta/3})) \cdot (1+\mathcal{O}(\omega_k^{-(3-\beta)/6})) \ ,$$

so $\varepsilon_k \leqslant \varepsilon_{k+1}$ and $\varepsilon_k = \mathcal{O}(1)$.

Proof. Every node $u \in V_{k-1}$ that sends load to v is not a dead-end. This means, that u has at least $\mathbf{E}\left[D_u^\ell \right] - \mathbf{E}\left[D_u^\ell \right]^{2/3}$ many edges to V_k as it does not fulfill $\langle$D1$\rangle$. Due to $\langle$D2$\rangle$ and $\langle$D3$\rangle$ we know, that u is assigned to the correct layer. Now it is possible that u's neighbors in V_k are assigned to layers different than k, but due to the fact that u does not fulfill $\langle$D4$\rangle$, there can be at most $\left(\omega_{k-1}\frac{W_k}{W}\right)^{2/3} \leqslant \mathbf{E}\left[D_u^\ell \right]^{2/3}$ of those nodes. This means, that u has at least $\mathbf{E}\left[D_u^\ell \right] - 2\mathbf{E}\left[D_u^\ell \right]^{2/3}$ many edges to nodes which are correctly assigned to V_k and a load of at most $(1+\varepsilon_{k-1}) \cdot m \cdot w_u/W_{k-1}$.

Then the load on $v \in V_k$ becomes at most

$$\begin{aligned}
&\sum_{u \in V'_{k-1}} \mathsf{Ber}\left(\frac{w_u w_v}{W}\right) \cdot (1+\varepsilon_{k-1}) \cdot m \cdot \frac{w_u}{W_{k-1}} \cdot \frac{1}{\mathbf{E}\left[D_u^\ell \right] - \mathbf{E}\left[D_u^\ell \right]^{2/3}} \\
&= \sum_{u \in V'_{k-1}} \mathsf{Ber}\left(\frac{w_u w_v}{W}\right) \cdot (1+\varepsilon_{k-1}) \cdot m \cdot \frac{w_u}{W_{k-1}} \cdot \frac{1}{\mathbf{E}\left[D_u^\ell \right]} \cdot \left(1 + \frac{1}{\mathbf{E}\left[D_u^\ell \right]^{1/3} - 1}\right) \\
&= \sum_{u \in V'_{k-1}} \mathsf{Ber}\left(\frac{w_u w_v}{W}\right) \cdot (1+\varepsilon_{k-1}) \cdot m \cdot \frac{w_u}{W_{k-1}} \cdot \frac{W}{w_u \cdot W_k} \cdot \left(1 + \frac{1}{\mathbf{E}\left[D_u^\ell \right]^{1/3} - 1}\right) \\
&\leqslant \sum_{u \in V'_{k-1}} \mathsf{Ber}\left(\frac{w_u w_v}{W}\right) \cdot (1+\varepsilon_{k-1}) \cdot m \cdot \frac{W}{W_k \cdot W_{k-1}} \cdot \left(1 + \frac{1}{\left(\frac{\gamma}{2d}\omega_k^{3-\beta}\right)^{1/3} - 1}\right) \ ,
\end{aligned}$$

where the last equality uses the expressions for $\mathbf{E}\left[D_u^\ell \right]$ derived in the proof of Lemma 6.3. Using the assumption that v is not a dead-end, the number of incoming edges from the higher layer which actually forward any load is at most $\mathbf{E}\left[D_v^h \right] + \mathbf{E}\left[D_v^h \right]^{2/3}$. Note that there can not be any additional edges by wrongly assigned nodes as these nodes would be dead-ends and therefore would not forward any load. Thus we can simplify

$$\begin{aligned}
&\frac{W}{W_k \cdot W_{k-1}} \cdot \sum_{u \in V'_{k-1}} \mathsf{Ber}\left(\frac{w_u w_v}{W}\right) \qquad (6.1)\\
&\leqslant \frac{W}{W_k \cdot W_{k-1}} \cdot \left(\mathbf{E}\left[\, D_v^h \,\right] + \mathbf{E}\left[\, D_v^h \,\right]^{2/3}\right)\\
&= \frac{W}{W_k \cdot W_{k-1}} \cdot \mathbf{E}\left[\, D_v^h \,\right]\left(1 + \mathbf{E}\left[\, D_v^h \,\right]^{-1/3}\right)\\
&= \frac{W}{W_k \cdot W_{k-1}} \cdot \frac{w_v \cdot W_{k-1}}{W} \cdot \left(1 + \mathbf{E}\left[\, D_v^h \,\right]^{-1/3}\right)\\
&= \left(1 + \mathbf{E}\left[\, D_v^h \,\right]^{-1/3}\right) \cdot \frac{w_v}{W_k}\\
&\leqslant \left(1 + \frac{1}{\left(\frac{\gamma}{2d}\omega_k^{(3-\beta)/2}\right)^{1/3}}\right) \cdot \frac{w_v}{W_k} \quad ,
\end{aligned}$$

where the last line follows with $\varepsilon \leqslant \frac{3-\beta}{\beta-1}$.

Finally, we prove that the ε_k values do not grow arbitrarily large for the final layer $k = \mathcal{O}(\log\log n)$. We define a constant $\nu = 2^{1/\varepsilon(\beta-1)}$ and a constant $r = 1/(1-\varepsilon) > 1$. Note that for the last layer of interest $\omega_{k-1} \geqslant \nu$, and, more generally, $\omega_i = \omega_{i+1}^r$. Hence, by reversing the index, we observe that

$$\begin{aligned}
(1+\varepsilon_k) \leqslant (1+\varepsilon_0) \cdot \prod_{i=1}^{k-1} &\left(1 + \frac{1}{\left(\frac{\gamma}{2d}\right)^{1/3} \left(\nu^{r^{i-k}}\right)^{(3-\beta)/3} - 1}\right)\\
&\cdot \left(1 + \frac{1}{\left(\frac{\gamma}{2d}\right)^{1/3} \left(\nu^{r^{i-k}}\right)^{(3-\beta)/6}}\right)
\end{aligned}$$

To bound the product, we apply the natural logarithm and note that

$$\begin{aligned}
&\sum_{i=1}^{k-1} \ln\left(1 + \frac{1}{\left(\frac{\gamma}{2d}\right)^{1/3} \left(\nu^{r^{i-k}}\right)^{(3-\beta)/3} - 1}\right) + \ln\left(1 + \frac{1}{\left(\frac{\gamma}{2d}\right)^{1/3} \left(\nu^{r^{i-k}}\right)^{(3-\beta)/6}}\right)\\
&\leqslant \frac{2}{\left(\frac{\gamma}{2d}\right)^{1/3}} \cdot \sum_{i=1}^{k-1} (\nu^{(3-\beta)/6})^{-r^i}\\
&< \frac{2}{\left(\frac{\gamma}{2d}\right)^{1/3}} \cdot \sum_{i=1}^{\infty} (\nu^{(3-\beta)/6})^{-r^i}.
\end{aligned}$$

This series converges to a constant value, and thus the overall product is also bounded by a constant. We obtain

$$(1+\varepsilon_k) \leqslant (1+\varepsilon_0) \cdot (1+\mathcal{O}(1)) = 1+\mathcal{O}(1) \ ,$$

as ε_0 measures the relative load differences in the core subgraph $\widetilde{G}$, for which $\varepsilon_0 = o(1)$. □

Now we want to show that on each layer with sufficiently large constant weight at most a small fraction of the total load remains on dead-ends. To do so, we show that for each property ⟨D1⟩ – ⟨D4⟩ the nodes with these properties only contribute a small enough fraction to the total dead-end load of each layer. We begin by bounding the contribution of ⟨D1⟩-nodes to the total dead-end load.

Lemma 6.3. *If $\varepsilon \leqslant (3-\beta)/(\beta-1)$ and $\omega_k > \left(\frac{2\bar{d}}{\gamma}\left(\frac{1}{2e-1}\right)^3\right)^{2/(3-\beta)}$, the probability that a node $v \in V_k$ is a ⟨D1⟩-node is at most*

$$2\exp(-c \cdot \omega_k^{(3-\beta)/6}),$$

for $c = \frac{1}{4}\left(\frac{\gamma}{2\bar{d}}\right)^{1/3}$.

Proof. We apply the Chernoff bound from Theorem 2.1 and choose $\delta = \mathbf{E}\left[\, D_v^h \,\right]^{-1/3}$. This yields

$$\mathbf{Pr}\left[\left|D_v^h - \mathbf{E}\left[\, D_v^h \,\right]\right| \geqslant \mathbf{E}\left[\, D_v^h \,\right]^{2/3}\right] < \exp\left(-\frac{1}{4} \cdot \mathbf{E}\left[\, D_v^h \,\right]^{1/3}\right)$$

and

$$\mathbf{Pr}\left[\left|D_v^\ell - \mathbf{E}\left[\, D_v^\ell \,\right]\right| \geqslant \mathbf{E}\left[\, D_v^\ell \,\right]^{2/3}\right] < \exp\left(-\frac{1}{4} \cdot \mathbf{E}\left[\, D_v^\ell \,\right]^{1/3}\right)$$

as long as $\mathbf{E}\left[\, D_v^h \,\right]^{-1/3}, \mathbf{E}\left[\, D_v^\ell \,\right]^{-1/3} < 2e-1$.For the expected degrees $\mathbf{E}\left[\, D_v^h \,\right]$ and $\mathbf{E}\left[\, D_v^\ell \,\right]$ we have

$$\mathbf{E}\left[\, D_v^h \,\right] = w_v \cdot \frac{W_{k-1}}{W} \quad \text{and} \quad \mathbf{E}\left[\, D_v^\ell \,\right] = w_v \cdot \frac{W_{k+1}}{W} \ .$$

From inequality 4.2 we get

$$\mathbf{E}\left[\, D_v^h \,\right] \geqslant w_v \cdot \frac{\gamma}{2}\frac{n\omega_{k-1}^{2-\beta}}{\bar{d} \cdot n} = w_v \cdot \frac{\gamma}{2\bar{d}}\omega_{k-1}^{2-\beta} \ .$$

With a similar derivation, we obtain that

$$\mathbf{E}\left[\, D_v^\ell \,\right] \geqslant w_v \cdot \frac{\gamma}{2\bar{d}}\omega_{k-1}^{(1-\varepsilon)^2(2-\beta)} \ .$$

As $\beta > 2$, we see that the lower bound for $\mathbf{E}\left[\, D_v^h \,\right]$ dominates. Thus, by using a union bound and observing that $w_v > \omega_k$, the probability for a node v becoming a dead end is at most

$$2\exp\left(-\frac{1}{4}\cdot\left(\frac{\gamma}{2\bar{d}}\omega_{k-1}^{2-\beta}\omega_k\right)^{1/3}\right) \quad .$$

Note that $\omega_{k-1}^{2-\beta}\cdot\omega_k \geqslant \omega_k^{3-\beta+\varepsilon(2-\beta)/(1-\varepsilon)}$ with equality for all but the last layer. The latter exponent is at least $(3-\beta)/2$ if we pick a sufficiently small $\varepsilon \leqslant (3-\beta)/(\beta-1)$. In this case, the overall probability becomes at most $2\exp(-c\cdot\omega_k^{(3-\beta)/6})$for the constant $c = \frac{1}{4}\cdot\left(\frac{\gamma}{2\bar{d}}\right)^{1/3}$ as desired. We also see now, that $\mathbf{E}\left[\, D_v^h \,\right]^{-1/3} \leqslant 2e-1$ is fulfilled, if $\omega_k > \left(\frac{2\bar{d}}{\gamma}\left(\frac{1}{2e-1}\right)^3\right)^{2/(3-\beta)}$. □

An implication of the former lemma is that there are no ⟨D1⟩-nodes on layers with weight at least polylog(n). Now that we have an understanding of which layers actually contain ⟨D1⟩-nodes, we can start to derive high probability upper bounds on the total load that is left on these nodes throughout the top-down phase.

Lemma 6.4. *If $v \in V_k$ is a* ⟨D1⟩*-node, then*

$$\mathbf{Pr}\left[x_v \geqslant \alpha\cdot\frac{m\cdot w_v}{W_k}\right] < \exp\left(-\Omega(\omega_k^{(3-\beta)/2}\cdot\min\{\alpha-1,(\alpha-1)^2\})\right) \quad .$$

Proof. To prove the upper bound on the probability for ⟨D1⟩-nodes, we apply a Chernoff bound. In particular, given that the load for every node $v \in V_k$ is upper bounded by (6.1), a ⟨D1⟩-node $v \in V_k$ can receive a total load of more than $\alpha\cdot m\cdot\frac{w_v}{W_k}$ only if this number is larger than the bound in (6.1), or, equivalently

$$\begin{aligned}
D_v^h &\geqslant \sum_{u\in V'_{k-1}} \mathsf{Ber}\left(\frac{w_u w_v}{W}\right)\\
&\geqslant \frac{\alpha}{(1+\varepsilon_{k-1})(1+\mathcal{O}(\omega_k^{-1+\beta/3}))}\cdot w_v\cdot\frac{W_{k-1}}{W}\\
&= \frac{\alpha}{(1+\varepsilon_{k-1})(1+\mathcal{O}(\omega_k^{-1+\beta/3}))}\cdot\mathbf{E}\left[\, D_v^h \,\right]\\
&> \frac{\alpha}{c}\cdot\mathbf{E}\left[\, D_v^h \,\right] \quad ,
\end{aligned}$$

for an appropriately chosen constant c. By the Chernoff bound above, we have

$$\begin{aligned}
\mathbf{Pr}\left[x_v \geqslant \alpha\cdot\frac{m\cdot w_v}{W_k}\right] &< \exp\left(-\mathbf{E}\left[\, D_v^h \,\right]\cdot\min\left\{\frac{\alpha}{c}-1,\left(\frac{\alpha}{c}-1\right)^2\right\}/4\right)\\
&= \exp\left(-\Omega(\omega_k^{(3-\beta)/2}\cdot\min\{\alpha-1,(\alpha-1)^2\})\right). \qquad \square
\end{aligned}$$

Now we use the tail bound from Lemma 6.4 and overestimate the load distribution of ⟨D1⟩-nodes with an exponential distribution. In particular, for each node $v \in V_k$ we introduce the following variable that measures the "⟨D1⟩-load" of this node, i.e. the load that each ⟨D1⟩-node keeps.

Definition 6.5. *The load of a node $v \in V_k$ under the condition that it has property ⟨D1⟩ is upper-bounded by the following random variable*

$$X_v = \begin{cases} 0 & \text{with prob. } 1 - p_v \\ s_v D_v^h & \text{with prob. } p_v \end{cases},$$

where p_v is an upper bound for the probability that v fulfills ⟨D1⟩ and $s_v = (1+\varepsilon_k)mW/(W_k W_{k-1})$ is an upper bound on the load per edge between layers $k-1$ and k.

We can now show that for each node $v \in V_k$ the following random variable stochastically dominates the ⟨D1⟩-load X_v.

Definition 6.6. *For a node $v \in V_k$ let*

$$\widehat{X}_v = \begin{cases} 0 & \text{with prob. } 1 - \widehat{p}_v \\ \ell_v \left(1 + Exp(\lambda_v) + \mathbf{E}\left[\, D_v^h \,\right]^{-2/3}\right) & \text{with prob. } \widehat{p}_v \end{cases},$$

where $\widehat{p}_v = 2\exp\left(-\frac{\mathbf{E}\left[\, D_v^h \,\right]^{1/3}}{4}\right)$ is an upper bound for the probability p_v, $\lambda_v = \frac{1}{4}\mathbf{E}\left[\, D_v^h \,\right]$ and $\ell_v = 2(1+\varepsilon_k)m\frac{w_v}{W_k}$.

Note that our ⟨D1⟩-load overestimates the contribution of v to the total load left on ⟨D1⟩-nodes during the top-down phase. In particular, if v is not a ⟨D1⟩-node, then no ⟨D1⟩-load is left on v and consequently the contribution is 0. Otherwise, we use the tail bound from Lemma 6.4 as follows. We overestimate the load by assuming that at least twice the layer-average load is present on v. For the additional load, we can apply the tail bound under the condition $\alpha \geqslant 2$, which implies that this excess load is upper bounded by an exponentially distributed random variable with a parameter $\lambda_v = \frac{1}{4}\mathbf{E}\left[\, D_v^h \,\right]$.

We first obtain a high probability bound on the total load left on ⟨D1⟩-nodes in each layer k during the top-down phase.

Lemma 6.7. *For every constant $c > 0$ and any k the total load left on ⟨D1⟩-nodes in layer k is at most*

$$4(1+\varepsilon_k)m\frac{\omega_{k-1}}{W_k}c\ln n + 40(1+\varepsilon_k)m\frac{\omega_{k-1}}{W_k}n_k \exp\left(-\frac{1}{4}\left(\omega_k \frac{W_{k-1}}{W}\right)^{1/3}\right)$$

with probability at least $1 - n^{-c}$.

Proof. We begin by observing that

$$\begin{aligned}
\mathbf{Pr}\left[\sum_{v\in V_k} X_v > t\right] &= \mathbf{Pr}\left[\exp\left(\alpha \sum_{v\in V_k} X_v\right) > \exp(\alpha t)\right] \\
&\leqslant \mathbf{E}\left[\exp\left(\alpha \sum_{v\in V_k} X_v\right) / \exp(\alpha t)\right] \\
&= \frac{1}{\exp(\alpha t)} \prod_{v\in V_k} \mathbf{E}\left[\, \exp(\alpha X_v)\,\right] \\
&\leqslant \frac{1}{\exp(\alpha t)} \prod_{v\in V_k} \mathbf{E}\left[\, \exp(\alpha \widehat{X}_v)\,\right]
\end{aligned}$$

for every $\alpha > 0$. The last equality was shown by using independence. The next stop is now bounding the expectation of $\exp(\alpha\widehat{X}_v)$, which amounts to

$$\begin{aligned}
&\mathbf{E}\left[\exp(\alpha\widehat{X}_v)\right] = \\
&\quad (1-\widehat{p}_v) + \widehat{p}_v\left(\exp\left(\alpha\ell_v\left(1+\mathbf{E}\left[\, D_v^h \,\right]^{-2/3}\right)\right)\cdot \mathbf{E}\left[\, \exp(\alpha\ell_v \cdot \mathsf{Exp}(\lambda_v))\,\right]\right)
\end{aligned}$$

We analyze the remaining expectation in the right-hand side by explicitly looking at the definition of the exponential distribution. This yields

$$\begin{aligned}
\mathbf{E}\left[\, \exp(\alpha\ell_v \cdot \mathsf{Exp}(\lambda_v))\,\right] &= \int_0^\infty \exp(\alpha\ell_v x)\cdot\lambda_v\exp(-\lambda_v x)dx \\
&= \frac{\lambda_v}{\lambda_v - \alpha\ell_v}\cdot\left[-\exp((\alpha\ell_v - \lambda_v)x)\right]_0^\infty \\
&= \frac{\lambda_v}{\lambda_v - \alpha\ell_v} \quad .
\end{aligned}$$

It has to be noted that for this expectation to hold, we need $0 < \alpha\ell_v < \lambda_v$. Using this in our expression above, we see

$$\mathbf{E}\left[\exp(\alpha\widehat{X}_v)\right] = (1-\widehat{p}_v) + \widehat{p}_v\left(\exp\left(\alpha\ell_v\left(1+\mathbf{E}\left[\, D_v^h \,\right]^{-2/3}\right)\right)\cdot\frac{\lambda_v}{\lambda_v - \alpha\ell_v}\right)$$

We can further bound this expression by choosing α small enough such that $\alpha\ell_v$ and α/λ_v become small as well. In particular, we denote by $\ell = 2(1+\varepsilon_k)m\omega_{k-1}/W_k$ and observe $\ell_v \leqslant \ell$ for all $v \in V_k$. We choose $\alpha = 1/2\ell$, which implies that $\alpha\ell_v \leqslant 1/2 < \alpha_v$ for all v with $\mathbf{E}\left[\, D_v^h \,\right] > 2$. We can safely assume, that $\mathbf{E}\left[\, D_v^h \,\right] \geqslant 4$ if w_v is a big enough constant. In this case, we can bound $\exp(\alpha\ell_v\left(1+\mathbf{E}\left[\, D_v^h \,\right]^{-2/3}\right)) \leqslant \exp\left(\frac{1}{2}\left(1+4^{-2/3}\right)\right) < \exp(1)$ and $\frac{\lambda_v}{\lambda_v - \alpha\ell_v} \leqslant 2$. Using these bounds, we arrive at the following expression for our expectation

$$\begin{aligned}
\mathbf{E}\left[\exp(\alpha\widehat{X}_v)\right] &< (1-\widehat{p}_v) + \widehat{p}_v\cdot 2\exp(1) \\
&\leqslant 1 + 5\cdot\widehat{p}_v \\
&\leqslant \exp(5\cdot\widehat{p}_v) \quad .
\end{aligned}$$

This expression is now a suitable bound for the sum of dead-end loads. We apply this bound as follows

$$\begin{aligned}\mathbf{Pr}\left[\sum_{v\in V_k} X_v > t\right] &\leqslant \exp\left(5\sum_{v\in V_k}\widehat{p}_v - \alpha t\right)\\ &\leqslant \exp\left(10 n_k \exp\left(-\frac{1}{4}\left(\omega_k \frac{W_{k-1}}{W}\right)^{1/3}\right) - \alpha t\right)\end{aligned}$$

As we chose $\alpha = 1/2\,(2(1+\varepsilon_k)m\omega_{k-1}/W_k)$, we can see, that

$$t = 4(1+\varepsilon_k)m\frac{\omega_{k-1}}{W_k}c\ln n + 40(1+\varepsilon_k)m\frac{\omega_{k-1}}{W_k}n_k\exp\left(-\frac{1}{4}\left(\omega_k\frac{W_{k-1}}{W}\right)^{1/3}\right)$$

is sufficient to obtain the high probability bound. □

Now we take a closer look at nodes with property ⟨D2⟩. We can employ a Chernoff Bound to show that nodes with polylogarithmically large weights do not deviate by $w_v^{2/3}$ from their expected degree with high probability. This means that none of these nodes fulfills property ⟨D2⟩ with high probability. In the following analysis we can therefore concentrate on nodes with weight at most polylog(n). This observation is crucial for the proof of Lemma 6.8.

Lemma 6.8. *For any k all nodes $v \in V_k$ with property ⟨D2⟩ contribute at most*

$$\mathcal{O}\left(\left(1+\omega_k^{-2/3}\right)^3 \omega_k^{\frac{4-\beta+\varepsilon(\beta-1)}{1-\varepsilon}} \cdot \exp\left(-\omega_k^{1/3}/4\right) m\right) + \mathcal{O}\left(\frac{\mathrm{polylog}(n)}{\sqrt{n}}m\right)$$

to the total dead-end load of all layers with probability at most $1-\frac{3}{n^C}$, for a constant $C > 1 + (\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)$.

Proof. We use Theorem 2.3 to show this result. In the context of the theorem, the X_e are random variables indicating the existence of edge e in the graph. This means $\mu = dn/2$ since the expected average degree is d. The function f is defined as the total load all nodes from V_k with property ⟨D2⟩ get. Then we can bound

$$\begin{aligned}|f| &\leqslant n_k\frac{(1+\varepsilon_k)\,\overline{d}}{\gamma^2}n^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)}\frac{m}{n}\\ &\leqslant \frac{4\gamma\cdot(1+\varepsilon_k)\cdot\overline{d}}{\gamma^2}\omega_k^{1-\beta}n^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)}m\\ &=: M\end{aligned}$$

due to equation (6.2) and the trivial bound $\deg_v \leqslant n$.

As the error event $\mathcal{B}$ we define that any of the nodes $v \in V_k$ fulfills $\deg_v > w_v\left(1+\frac{c_{\mathcal{B}}\ln n}{w_v}\right)$ for $w_v < c_{\mathcal{B}}\ln n$ or $\deg_v > 2\cdot w_v$ for $w_v \geqslant c_{\mathcal{B}}\ln n$. To unify this bound, we can assume $\deg_v < 2 \cdot w_v + c_{\mathcal{B}} \ln n$. A Chernoff Bound from [1] states, that this is fulfilled with probability at most $n^{-c_{\mathcal{B}}/3}$ for every $v \in V_k$. By applying a union bound we get $\Pr[\mathcal{B}] \leqslant n^{-c_{\mathcal{B}}/3+1}$.

To bound the differences $|f(\mathbf{X}_n) - f(\mathbf{X}'_n)|$ for every $\mathbf{X}'_n$ that differs in only one position X_e from $\mathbf{X}_n$, we observe the following: By changing one edge, only two nodes can change their degree by one. We can now trivially upper bound $|f(\mathbf{X}_n) - f(\mathbf{X}'_n)|$ by two times the maximum dead-end load a ⟨D2⟩-node can get while $\mathcal{B}$ does not hold. This gives us

$$c := 2\frac{(1+\varepsilon_k)\,\overline{d}}{\gamma^2}\left(2 \cdot \omega_{k-1} + c_{\mathcal{B}} \ln n\right)^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)} \frac{m}{n}.$$

If we set $t = 4c\sqrt{\mu \cdot C \ln n}$ for an arbitrary constant $C > 0$ we get the following result from Theorem 2.3

$$\Pr\left[\left|f - \mathbf{E}\left[\,f\,\right]\right| > 4c\sqrt{\mu \cdot C \ln n} + \frac{(2M)^2}{c}\Pr[\mathcal{B}]\right] \leqslant \frac{2M}{c}\Pr[\mathcal{B}] + 2n^{-C}.$$

We can now conclude

$$\frac{2M}{c}\Pr[\mathcal{B}] \leqslant \frac{4\gamma}{2}\omega_k^{1-\beta} n \left(\frac{n}{2 \cdot \omega_{k-1} + c_{\mathcal{B}} \ln n}\right)^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)} n^{-c_{\mathcal{B}}/3+1}.$$

For $c_{\mathcal{B}} \geqslant 3C + 9 + 3(\beta-2)(1+\frac{1}{1-\varepsilon})$ and sufficiently large n this is at most n^{-C}. We obtain an upper bound of $\mathbf{E}\left[\,f\,\right] + 4c\sqrt{\mu \cdot C \ln n} + \frac{(2M)^2}{c}\Pr[\mathcal{B}]$ on the dead-end load of nodes $v \in V_k$. We can see that

$$\begin{aligned} 4c\sqrt{\mu \cdot C \ln n} &\leqslant 2\frac{(1+\varepsilon_k)\,\overline{d}}{\gamma^2}\left(2 \cdot \omega_{k-1} + c_{\mathcal{B}} \ln n\right)^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)} \frac{m}{n} \cdot \sqrt{d \cdot n \cdot C \ln n} \\ &= \Theta\left(\frac{\operatorname{polylog}(n)}{\sqrt{n}} m\right) \end{aligned}$$

for $\omega_k = \mathcal{O}\left(\operatorname{polylog}(n)\right)$. This is acceptable since nodes with property ⟨D2⟩ can only appear for polylogarithmically small ω_k. We can also see that

$$\begin{aligned} \frac{(2M)^2}{c}\Pr[\mathcal{B}] &\leqslant 2Mn^{-C} \\ &\leqslant 2\frac{4\gamma \cdot (1+\varepsilon_k) \cdot \overline{d}}{\gamma^2}\omega_k^{1-\beta} n^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)} m \cdot n^{-C}, \end{aligned}$$

which is arbitrarily small for $C > 1 + (\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)$. The only thing missing now is an upper bound on $\mathbf{E}\left[\,f\,\right]$, the expected total dead-end load of nodes $v \in V_k$ with property ⟨D2⟩.

Let $L_v^{(2)}$ the load of node v with property ⟨D2⟩. Suppose the node $v \in V_k$ lands on layer k. Then it gets a load of at most

$$(1+\varepsilon_k)\,\frac{W}{W_k W_{k-1}} m \leqslant \frac{\overline{d}n}{\gamma n \omega_k^{2-\beta} \gamma n \omega_{k-1}^{2-\beta}} m = \frac{(1+\varepsilon_k)\,\overline{d}}{\gamma^2}\omega_k^{\beta-2}\omega_{k-1}^{\beta-2}\frac{m}{n}$$

from each neighbor on layer $k-1$ according to Lemma 6.2. We observe that $\omega_k \leqslant \deg_v$ and $\omega_{k-1} \leqslant \deg_v^{\frac{1}{1-\varepsilon}}$. If we trivially bound the number of v's neighbors on layer $k-1$ by $\deg_v$, we get an upper bound of

$$L_v^{(2)} \leqslant \frac{(1+\varepsilon_k)\,\overline{d}}{\gamma^2} \deg_v^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)} \frac{m}{n} \tag{6.2}$$

if $\deg_v$ deviates by at least $w_v^{2/3}$ from w_v. The probability of this is at most

$$\exp\left(-\frac{w_v^{1/3}}{4}\right)$$

by using the Chernoff Bound 2.1.

We can see now that $\deg_v^{(2)}$, the degree of v if it has property $\langle \mathrm{D2} \rangle$, is stochastically dominated by the following random variable

$$\widehat{\deg}_v^{(2)} = \begin{cases} 0 & \text{with prob. } 1-\widehat{p}_v \\ 2w_v\left(1+Exp(\frac{w_v}{4})+w_v^{-2/3}\right) & \text{with prob. } \widehat{p}_v\ , \end{cases}$$

where $\widehat{p}_v = \exp\left(-\frac{w_v^{1/3}}{4}\right)$. This means that $L_v^{(2)}$ is stochastically dominated by

$$\widehat{L}_v^{(2)} := \frac{(1+\varepsilon_k)\,\overline{d}}{\gamma^2}\left(\widehat{\deg}_v^{(2)}\right)^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)} \frac{m}{n}.$$

From our proof of lemma 6.3 we know that $\varepsilon \leqslant (3-\beta)/(\beta-1)$. This gives us

$$1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right) \leqslant \frac{3}{2}(\beta-1) < 3.$$

Since each value of $\widehat{\deg}_v^{(2)}$ is either 0 or at least 1, we conclude that

$$\mathbf{E}\left[\left(\widehat{\deg}_v^{(2)}\right)^{1+(\beta-2)\left(1+\frac{1}{1-\varepsilon}\right)}\right] \leqslant \mathbf{E}\left[\left(\widehat{\deg}_v^{(2)}\right)^3\right] = \Theta\left(\left(1+\varepsilon_k^{(2)^2}\right)^3 w_v^3 \cdot \widehat{p}_v\right).$$

This gives us

$$\mathbf{E}\left[L_v^{(2)}\right] \leqslant \Theta\left(\frac{m}{n}\left(1+w_v^{-2/3}\right)^3 w_v^3 \cdot \exp\left(-w_v^{1/3}/4\right)\right).$$

Now let $\mathbf{E}\left[L_k^{(2)}\right]$ the expected total dead-end load of nodes $v \in V_k$ with property $\langle \mathrm{D2} \rangle$. We can now use the result above to show

$$\begin{aligned} \mathbf{E}\left[L_k^{(2)}\right] &\leqslant \sum_{v \in V_k} \Theta\left(\frac{m}{n}\left(1+w_v^{-2/3}\right)^3 w_v^3 \cdot \exp\left(-w_v^{1/3}/4\right)\right) \\ &\leqslant 4\gamma \cdot \omega_k^{1-\beta} \cdot n \cdot \Theta\left(\frac{m}{n}\left(1+\omega_k^{-2/3}\right)^3 \omega_{k-1}^3 \cdot \exp\left(-\omega_k^{1/3}/4\right)\right) \\ &\leqslant \Theta\left(m\left(1+\omega_k^{-2/3}\right)^3 \omega_k^{\frac{4-\beta+\varepsilon(\beta-1)}{1-\varepsilon}} \cdot \exp\left(-\omega_k^{1/3}/4\right)\right). \end{aligned} \tag{6.3}$$

Putting all upper bounds together gives us the desired result. $\square$

After successfully bounding the contribution of nodes with properties $\langle\mathrm{D1}\rangle$ and $\langle\mathrm{D2}\rangle$ to dead-end load, we will now turn to the border nodes with property $\langle\mathrm{D3}\rangle$. We already know that these nodes cannot deviate too much from their expected degrees, because they do not fulfill property $\langle\mathrm{D2}\rangle$ by definition. Therefore they can only be on one of two layers. We still have to differ between nodes in the upper half of a border and those in the lower half. The following lemma bounds the contribution of nodes in the upper half of a border.

Lemma 6.9. *For any k all nodes $v \in V_k$ with property $\langle\mathrm{D3}\rangle$ and $\omega_k \leqslant w_v \leqslant \omega_k + \omega_{k-1}^{2/3}$ contribute at most*

$$\Theta\left(\omega_k^{-\varepsilon(\beta-2)} + \frac{\omega_k^{\beta-2}\omega_{k+1}^{\beta-2}\cdot c\ln n}{n}\right) m$$

to the total dead-end load of layer $k+1$ w. h. p.

Proof. In the worst case, all considered nodes change layer. Then we simply have to bound the number of edges these nodes have into layer k. Let B_k^u the total weight of nodes from the upper half of the border around ω_k. The expected number of edges these nodes have to other nodes from V_k is at most $B_k^u \frac{W_k}{W}$ and at least $B_k^u \frac{W_k - B_k^u}{W}$. By utilizing a Chernoff Bound, we get that the number of edges in consideration is at most $B_k^u \frac{W_k}{W} + c\ln n$ (if the expected value is asymptotically smaller than $c\ln n$) or $2B_k^u \frac{W_k}{W}$ (if the expected value is bigger) w. h. p.

The load per edge from layer k is at most $(1+\varepsilon_{k+1})\frac{W}{W_k W_{k+1}} m$. This gives us an upper bound of

$$\begin{aligned}
&(1+\varepsilon_{k+1})\frac{W}{W_k W_{k+1}}\left(B_k^u \frac{W_k}{W} + c\ln n\right) m \\
&\leqslant (1+\varepsilon_{k+1})\left(\frac{B_k^u}{W_{k+1}} + \frac{d\cdot n\cdot c\ln n}{\left(\frac{\gamma}{2}\right)^2 n^2 \omega_k^{2-\beta}\omega_{k+1}^{2-\beta}}\right) m
\end{aligned}$$

on the total dead-end load of nodes with property $\langle\mathrm{D3}\rangle$ from the upper half of the border around ω_k. As B_k^u is trivially bounded by W_k this amounts to at most $\Theta\left(\left(\frac{\omega_k}{\omega_{k+1}}\right)^{2-\beta} + \frac{\omega_k^{\beta-2}\omega_{k+1}^{\beta-2}\cdot c\ln n}{n}\right) m = \Theta\left(\omega_k^{-\varepsilon(\beta-2)} + \frac{\omega_k^{\beta-2}\omega_{k+1}^{\beta-2}\cdot c\ln n}{n}\right) m.$ $\square$

The following lemma about the contribution of nodes in the lower half of a border uses the smoothness of the weight distribution.

Lemma 6.10. *For any k all nodes $v \in V_{k+1}$ with property $\langle\mathrm{D3}\rangle$ and $\omega_k - \omega_k^{2/3} \leqslant w_v \leqslant \omega_k$ contribute at most*

$$(1+\varepsilon_k)\left(\frac{6\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}}{\frac{\gamma}{2}}\omega_k^{-1/3} + \frac{2\cdot\omega_k^{\beta-1}}{\frac{\gamma}{2}n} + \frac{\overline{d}\cdot\omega_k^{\beta-2}\omega_{k+1}^{\beta-2}\cdot c\ln n}{\left(\frac{\gamma}{2}\right)^2 n}\right) m$$

to the total dead-end load of layer k w. h. p.

Proof. Again, all nodes change layer in the worst case. Let B_k^l denote the total weight of nodes from the lower half of the border around ω_k. The expected number of edges these nodes have into V_{k-1} is exactly $B_k^l \frac{W_{k-1}}{W}$. With the same Chernoff bound as in the proof of lemma 6.9, we get that the number of these edges is at most $B_k^l \frac{W_{k-1}}{W} + c \ln n$ with high probability. The load per edge from layer $k-1$ is at most $(1+\varepsilon_k) \frac{W}{W_{k-1} W_k} m$. This gives us an upper bound of

$$\begin{aligned} &(1+\varepsilon_k) \frac{W}{W_{k-1} W_k} \left(B_k^l \frac{W_{k-1}}{W} + c \ln n \right) m \\ &\quad \leqslant (1+\varepsilon_k) \left(\frac{B_k^l}{W_k} + \frac{\overline{d} \cdot n \cdot c \ln n}{\left(\frac{\gamma}{2}\right)^2 n^2 \omega_k^{2-\beta} \omega_{k+1}^{2-\beta}} \right) m \end{aligned}$$

on the total dead-end load of nodes with property $\langle \mathrm{D3} \rangle$ from the lower half of the border around ω_k. Now we can bound B_k^l as follows. A node is of weight at least ω_k iff

$$i \leqslant \left(d \frac{\beta-2}{\beta-1} \right)^{\beta-1} \cdot n \cdot \omega_k^{1-\beta}$$

and of weight at most $\omega_k - \omega_k^{2/3} = \omega_k \left(1 - \omega_k^{-1/3}\right)$ iff

$$i > \left(d \frac{\beta-2}{\beta-1} \right)^{\beta-1} \cdot n \cdot \omega_k^{1-\beta} \left(1 - \omega_k^{-1/3}\right)^{1-\beta}$$

This means the number of nodes in the lower half of the border can be at most

$$\left(d \frac{\beta-2}{\beta-1} \right)^{\beta-1} \cdot n \cdot \omega_k^{1-\beta} \left(\left(1 - \omega_k^{-1/3}\right)^{1-\beta} - 1 \right) + 2.$$

We can see $\left(1 - \omega_k^{-1/3}\right)^{1-\beta} = \left(1 + \frac{1}{\omega_k^{1/3} - 1}\right)^{\beta-1}$. Together with the fact, that $\beta < 3$, we get that the number of nodes in the lower half of the border is at most

$$\begin{aligned} &\left(d \frac{\beta-2}{\beta-1} \right)^{\beta-1} \cdot n \cdot \omega_k^{1-\beta} \left(\frac{2}{\omega_k^{1/3} - 1} + \frac{1}{\left(\omega_k^{1/3} - 1\right)^2} \right) + 2 \\ \leqslant &\left(d \frac{\beta-2}{\beta-1} \right)^{\beta-1} \cdot n \cdot \omega_k^{1-\beta} \frac{6}{\omega_k^{1/3}} + 2 \end{aligned} \tag{6.4}$$

for $\omega_k \geqslant 8$. Since the maximum weight of these nodes is ω_k, it holds that

$$B_k^l \leqslant 6 \left(d \frac{\beta-2}{\beta-1} \right)^{\beta-1} \cdot n \frac{\omega_k^{2-\beta}}{\omega_k^{1/3}} + 2 \cdot \omega_k, \tag{6.5}$$

which implies

$$\frac{B_k^l}{W_k} \leqslant \frac{6\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}}{\frac{\gamma}{2}}\omega_k^{-1/3} + \frac{2\cdot\omega_k^{\beta-1}}{\frac{\gamma}{2}n},$$

giving us the result. □

At last we have to show that the dead-end load of nodes with property $\langle$D4$\rangle$ is properly bounded. We already know that each of these nodes obeys the upper bound from Lemma 6.2. Therefore it is sufficient to bound the number of these nodes. To bound the number of $\langle$D4$\rangle$-nodes in V_k, we simply have to bound the total number of edges lost between nodes from V_k and nodes with properties $\langle$D2$\rangle$ or $\langle$D3$\rangle$ from V_{k+1}. Then we divide this total number of edges by the minimum number of edges a node $v \in V_k$ has to lose to obtain property $\langle$D4$\rangle$. The following two lemmas give upper bounds on the the total number of edges between nodes from V_k and nodes with properties $\langle$D2$\rangle$ and $\langle$D3$\rangle$ from V_{k+1} respectively.

Lemma 6.11. *For $\varepsilon < 1/3$ the number of edges from nodes $v \in V_{k+1}$ with property $\langle$D3$\rangle$ to nodes from V_k is at most*

$$\left(\frac{12}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\frac{\omega_k^{2-\beta}}{\omega_k^{1/3}} + \frac{4\cdot\omega_k}{d\cdot n} + 2\frac{\omega_{k+1}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)}{d\cdot n}\right.$$
$$\left.+\frac{6}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\omega_{k+1}^{2-\beta}\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)\right)W_k \qquad \textit{w. h. p.}$$

Proof. We consider border nodes $v \in V_{k+1}$ from the border around ω_k and ω_{k+1} independently. Let B_k^l the total weight of nodes $v \in V_{k+1}$ from the border around ω_k. The expected number of edges between this border and V_k is $B_k^l\frac{W_k}{W}$. We already know $W = d\cdot n$, $W_k = \Theta\left(n\omega_k^{2-\beta}\right)$ and, from the proof of Lemma 6.10, $B_k^l = \Theta\left(n\frac{\omega_k^{2-\beta}}{\omega_k^{1/3}}\right)$. Therefore, the expected value is at least $c\ln n$ as long as ω_k is small enough. This allows us to apply a Chernoff Bound to show that the number of edges between this border and V_k is at most $2B_k^l\frac{W_k}{W}$ with probability at least $1-n^{-c/3}$. Otherwise, the number of edges is at most $B_k^l\frac{W_k}{W}+c\ln n$ with probability at least $1-n^{-c/3}$. From equation (6.5) we know

$$B_k^l \leqslant 6\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\cdot n\frac{\omega_k^{2-\beta}}{\omega_k^{1/3}} + 2\cdot\omega_k,$$

which gives us the first term of our result.

Now let B_{k+1}^u the total weight of nodes $v \in V_{k+1}$ from the border around ω_{k+1}. Again, the expected number of edges between this border and V_k is $B_{k+1}^u\frac{W_k}{W}$. Using a Chernoff Bound, we get at most $2B_k^l\frac{W_k}{W}$ edges with probability at least $1-n^{-c/3}$ or at most $B_k^l\frac{W_k}{W}+c\ln n$ edges with probability at least

$1-n^{-c/3}$. Note that in this border there are all nodes with weight at least ω_{k+1} and at most $\omega_{k+1}+\omega_k^{2/3}=\omega_{k+1}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)$. As in the proof of Lemma 6.10 we can bound the number n_{k+1}^u of these nodes to

$$\left|n_{k+1}^u-\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1} n\cdot\omega_{k+1}^{1-\beta}\left(1-\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)^{1-\beta}\right)\right|\leqslant 1. \tag{6.6}$$

Some simple calculations and the observation $\beta<3$ give us

$$\begin{aligned}\left(1-\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)^{1-\beta}\right) &\leqslant \left(\frac{2\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}+\omega_{k+1}^{2\frac{3\varepsilon-1}{3(1-\varepsilon)}}}{1+2\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}+\omega_{k+1}^{2\frac{3\varepsilon-1}{3(1-\varepsilon)}}}\right)\\ &\leqslant 2\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}+\omega_{k+1}^{2\frac{3\varepsilon-1}{3(1-\varepsilon)}}\\ &\leqslant 3\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}},\end{aligned}$$

where in the last line we used the assumptions $\varepsilon<1/3$, which gave us $\frac{3\varepsilon-1}{3(1-\varepsilon)}<0$, and $\omega_{k+1}\geqslant 1$. Together with our bound of $\omega_{k+1}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)$ on the maximum weight of these nodes this results in

$$B_{k+1}^u\leqslant\frac{3}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\cdot\omega_{k+1}^{2-\beta}\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)+\frac{\omega_{k+1}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)}{d\cdot n}$$

and therefore establishes the second term of our result. □

Lemma 6.12. *For all k the number of edges from nodes $v\in V_{k+1}$ with property $\langle\mathrm{D2}\rangle$ to nodes from V_k is at most*

$$6\exp\left(-\omega_{k+1}^{1/3}/4\right)W_{k+1}+\left(4\omega_k\frac{W_k}{W}+6(c'+3)\ln n\right)\frac{d\sqrt{c'}}{\sqrt{2}}\sqrt{n\ln n}+n^{-1}$$

with probability at least $1-3\cdot n^{-c'}$ for some constant $c'>3$.

Proof. As in the proof of Lemma 6.8 we are overestimating the number of edges of a node $v\in V_{k+1}$ with property $\langle\mathrm{D2}\rangle$ to nodes from V_k by $\deg_v^{(2)}$. We already know from that Lemma, that $\deg_v^{(2)}$ is stochastically dominated by the following random variable

$$\widehat{\deg}_v^{(2)}=\begin{cases}0 & \text{with prob. } 1-\widehat{p}_v\\ 2w_v\left(1+Exp(\frac{w_v}{4})+w_v^{-2/3}\right) & \text{with prob. } \widehat{p}_v\end{cases},$$

where $\widehat{p}_v = \exp\left(-\frac{w_v^{1/3}}{4}\right)$. This means

$$\begin{aligned}\mathbf{E}\left[\, D_v^h \mid v \text{ fulfills } \langle \text{D2} \rangle \,\right] &\leqslant \mathbf{E}\left[\, deg_v^{(2)} \,\right] \\ &\leqslant \left(2w_v + 8 + 2w_v^{1/3}\right) \exp\left(-\frac{w_v^{1/3}}{4}\right) \\ &\leqslant 6w_v \exp\left(-\frac{w_v^{1/3}}{4}\right)\end{aligned}$$

for $w_v \geqslant 4$. As $w_v \geqslant \omega_{k+1}$, this gives us

$$6 \exp\left(-\omega_{k+1}^{1/3}/4\right) W_{k+1}$$

as an upper bound on the expected number of edges from $\langle$D2$\rangle$-nodes on layer $k+1$ to nodes on layer k.

Now we want to show concentration using Theorem 2.3. In the context of the Theorem, the X_e are random variables indicating the existence of edge e in the graph. This means

$$\mu = dn/2,$$

since the expected average degree is d. The function f is defined as the total number of edges from all nodes from V_{k+1} with property $\langle$D2$\rangle$ to nodes from V_k. Then we can bound

$$|f| \leqslant n_k \cdot n_{k+1} := M,$$

since n_k and n_{k+1} are the number of nodes on layers k and $k+1$ respectively. As the error event $\mathcal{B}$ we define the event, that any node $v \in V_{k+1}$ has more than $\max\left\{2\mathbf{E}\left[\, D_v^h \,\right], \mathbf{E}\left[\, D_v^h \,\right] + c_{\mathcal{B}} \ln n\right\}$ edges to V_k. This results in

$$\Pr(\mathcal{B}) \leqslant n^{-c_{\mathcal{B}}/3+1}.$$

We can now assume, that each $v \in V_{k+1}$ with property $\langle$D2$\rangle$ has at most $2\mathbf{E}\left[\, D_v^h \,\right] + c_{\mathcal{B}} \ln n$ many edges to V_k. Especially, this means that for every $\mathbf{X}_n \in \overline{\mathcal{B}}$

$$|f(\mathbf{X}_n) - f(\mathbf{X}_n')| \leqslant 4\omega_k \frac{W_k}{W} + 2c_{\mathcal{B}} \ln n =: c$$

for every $\mathbf{X}_n'$ that differs in only one position X_e from $\mathbf{X}_n$. The former inequality holds, because in the worst case changing one edge results in two new nodes with property $\langle$D2$\rangle$. Each of these nodes can only have $2\omega_k \frac{W_k}{W} + c_{\mathcal{B}} \ln n$ edges to V_k as we have seen before. If we choose $t = 4c\sqrt{\mu \cdot c' \ln n} \ll c\mu$ for a constant $c' > 0$, we get the following result from Theorem 2.3

$$\Pr\left[\left|f - \mathbf{E}\left[\, f \,\right]\right| > c\sqrt{\mu \cdot c' \ln n} + \tfrac{(2M)^2}{c} \Pr[\mathcal{B}]\right] \leqslant \tfrac{2M}{c} n^{-c_{\mathcal{B}}/3+1} + 2n^{-c'}.$$

Let $c_{\mathcal{B}} = 3(c'+3)$. As $M \ll n^2$ and $c > 1$ this gives us that the number of edges between nodes $v \in V_k$ with property $\langle$D2$\rangle$ and nodes in V_{k+1} is upper-bounded by

$$\mathbf{E}\left[\, f \,\right] + c\sqrt{\mu \cdot c' \ln n} + \tfrac{(2M)^2}{c} \Pr[\mathcal{B}]$$

with probability at least $1-3\cdot n^{-c'}$. With $c'>3$ it holds that $\frac{(2M)^2}{c}\Pr[\mathcal{B}]<n^{-1}$. We also know that

$$c\sqrt{\mu\cdot c'\ln n}\leqslant\left(4\omega_k\frac{W_k}{W}+6(c'+3)\ln n\right)\frac{d\sqrt{c'}}{\sqrt{2}}\sqrt{n\ln n}.$$

Together with the upper bound on the expected number of edges from earlier in this proof, we get the result as desired. □

Lemma 6.13. *Let* $\varepsilon<\min\left\{\frac{\beta-2}{3},\frac{1}{2}\left(1-\sqrt{\frac{3}{\beta+1}}\right)\right\}$. *Then the following statements hold:*

(1) For all $k>0$ *the total load of nodes* $v\in V_k$ *with property* $\langle\mathrm{D4}\rangle$ *is at most*

$$\mathcal{O}\left(\omega_k^{\frac{2-\beta}{3(1-\varepsilon)}}+\omega_k^{\frac{(2\beta^2-11\beta+14)(\beta-2)}{27(1-\varepsilon)}}+n^{\frac{3+1-\varepsilon+2(\beta-2)(1-\varepsilon)^2}{6}-1}\right.$$
$$\left.+\exp\left(-\omega_k^{\frac{1-\varepsilon}{3}}/4\right)\omega_k^{\frac{1}{1-\varepsilon}+(\beta-2)}+\frac{\mathrm{polylog}(n)}{\sqrt{n}}\right)m\qquad \textit{w. h. p.}$$

(2) For $k=0$ *there are no* $\langle\mathrm{D4}\rangle$*-nodes w. h. p.*

Proof. First we want to show the second statement, i.e. that there cannot be any $\langle\mathrm{D4}\rangle$-nodes in the core. To become a $\langle\mathrm{D4}\rangle$-node, a node from the core has to lose at least $\left(\omega_0\frac{W_1}{W}\right)^{2/3}=\Theta\left(n^{\frac{1+(1-\varepsilon)(\beta-2)}{3}}\right)$ edges. As we already observed that there are no $\langle\mathrm{D2}\rangle$-nodes in V_1 w.h.p., these edges can only be lost by $\langle\mathrm{D3}\rangle$-nodes. Now we only need to show that there are not enough $\langle\mathrm{D3}\rangle$-nodes in V_1. The number of $\langle\mathrm{D3}\rangle$-nodes in V_1 is at most

$$6\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\cdot n\cdot\frac{\omega_0^{1-\beta}}{\omega_0^{1/3}}+3\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\cdot n\cdot\omega_1^{1-\beta}\cdot\omega_1^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}+2$$

due to Lemma 6.10 and Lemma 6.11. This term is dominated by $n\cdot\omega_1^{1-\beta}\cdot\omega_1^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}=n\cdot\omega_0^{(1-\beta)(1-\varepsilon)+\varepsilon-\frac{1}{3}}$. Recall $\omega_0=\Theta(n^{1/2})$. Now we only need to show $\frac{(1-\beta)(1-\varepsilon)+\varepsilon-1/3}{2}+1<\frac{1+(1-\varepsilon)(\beta-2)}{3}$. Then the number of $\langle\mathrm{D3}\rangle$ nodes on V_1 will be asymptotically smaller than the number of edges a core node needs to lose to have property $\langle\mathrm{D4}\rangle$. The inequality we need to show is equivalent to the following

$$5(2-\beta)+\varepsilon(\beta-4)<0,$$

which holds since $2<\beta<3$ and $0<\varepsilon<1$. This shows the second statement of the Lemma.

To get the number of $\langle\mathrm{D4}\rangle$-nodes in V_k we simply have to divide the number of edges between V_k and $\langle\mathrm{D2}\rangle$- or $\langle\mathrm{D3}\rangle$-nodes in $V_{k+1}$by the number of edges a node has to lose to become a $\langle\mathrm{D4}\rangle$-node, which is $\left(\omega_k\frac{W_{k+1}}{W}\right)^{2/3}$. As each $\langle\mathrm{D4}\rangle$-node $v\in V_k$ gets a load of at most $(1+\varepsilon_k)\frac{w_v}{W_k}m\leqslant(1+\varepsilon_k)\frac{\omega_{k-1}}{W_k}m$ due to

Lemma 6.2, we have to multiply the resulting number of ⟨D4⟩-nodes by this value. For the sake of readability we want to calculate the resulting dead-end load for each term independently.

We begin with the load lost by ⟨D3⟩-nodes. Recall the bound from Lemma 6.11. The requirement for this bound was $\varepsilon < 1/3$ which is ensured by $\varepsilon < \min\left\{\frac{\beta-2}{3}, \frac{1}{2}\left(1-\sqrt{\frac{3}{\beta+1}}\right)\right\}$. The share of load lost due to missing nodes in the border around ω_k is at most

$$\left(\frac{12}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\frac{\omega_k^{2-\beta}}{\omega_k^{1/3}}+\frac{4\cdot\omega_k}{d\cdot n}\right)W_k\cdot(1+\varepsilon_k)\frac{\omega_{k-1}}{\left(\omega_k\frac{W_{k+1}}{W}\right)^{2/3}W_k}m.$$

If we apply the lower bound $W_{k+1}\geqslant\frac{\gamma}{2}n\omega_{k+1}^{2-\beta}$ and use $\omega_{k+1}=\omega_k^{1-\varepsilon}$ and $\omega_{k-1}=\omega_k^{\frac{1}{1-\varepsilon}}$, we get an expression as follows

$$\begin{aligned}&\left(\frac{2\cdot d}{\gamma}\right)^{2/3}(1+\varepsilon_k)\left(\frac{12}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\omega_k^{1-\beta+\frac{1}{1-\varepsilon}+\frac{2}{3}(1-\varepsilon)(\beta-2)}\right.\\&+\left.\frac{4\cdot\omega_k^{\frac{1}{3}+\frac{1}{1-\varepsilon}+\frac{2}{3}(1-\varepsilon)(\beta-2)}}{d\cdot n}\right)m.\end{aligned}\tag{6.7}$$

We can see that

$$\begin{aligned}1-\beta+\tfrac{1}{1-\varepsilon}+\tfrac{2}{3}(1-\varepsilon)(\beta-2)&=\frac{1}{1-\varepsilon}\left((\beta-2)\left(\tfrac{2}{3}(1-\varepsilon)^2-1\right)+\varepsilon(\beta-1)\right)\\&=\frac{2}{3}\frac{\beta-2}{\beta-1}\varepsilon^2+\left(1-\frac{4}{3}\frac{\beta-2}{\beta-1}\right)\varepsilon-\frac{1}{3}\frac{\beta-2}{\beta-1}.\end{aligned}$$

This is a quadratic expression in ε, which is at most $\frac{(2\beta^2-11\beta+14)(\beta-2)}{27(1-\varepsilon)}$ if $0<\varepsilon<\frac{\beta-2}{3}$. It now holds that $\frac{(2\beta^2-11\beta+14)(\beta-2)}{27(1-\varepsilon)}<0$ for all $2<\beta<3$.

Now we look at the share of load lost due to missing nodes in the border around ω_{k+1}. This value is at most

$$\begin{aligned}&\left(\frac{6}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\omega_{k+1}^{2-\beta}\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)\right.\\&\left.+2\frac{\omega_{k+1}\left(1+\omega_{k+1}^{\frac{3\varepsilon-1}{3(1-\varepsilon)}}\right)}{d\cdot n}\right)W_k\cdot(1+\varepsilon_k)\frac{\omega_{k-1}}{\left(\omega_k\frac{W_{k+1}}{W}\right)^{2/3}W_k}m.\end{aligned}$$

If we apply the same substitutions and lower bounds as before, we get the expression

$$\left(\frac{2\cdot d}{\gamma}\right)^{2/3}(1+\varepsilon_k)\left(\frac{12}{d}\left(d\frac{\beta-2}{\beta-1}\right)^{\beta-1}\omega_k^{\frac{(1-\varepsilon)(2-\beta)}{3}+\frac{1}{1-\varepsilon}-\frac{2}{3}+\frac{3\varepsilon-1}{3}}\right.$$
$$\left.+\quad\frac{4\cdot\omega_k^{(1-\varepsilon)-\frac{2}{3}+\frac{1}{1-\varepsilon}+\frac{2}{3}(1-\varepsilon)(\beta-2)}}{d\cdot n}\right)m. \tag{6.8}$$

We can see that

$$\frac{(1-\varepsilon)(2-\beta)}{3}+\tfrac{1}{1-\varepsilon}-\frac{2}{3}+\frac{3\varepsilon-1}{3}=\frac{3-(\beta+1)(1-\varepsilon)^2}{3(1-\varepsilon)}.$$

This is a quadratic expression in ε which is at most $\frac{2-\beta}{12(1-\varepsilon)}<0$ if $0<\varepsilon<\frac{1}{2}\left(1-\sqrt{\frac{3}{\beta+1}}\right)$.

The former bounds only account for the first terms of equation (6.7) and equation (6.8). Now we want to show that the second terms are asymptotically small in poly(n), too. We can easily see, that the term

$$\left(\frac{2\cdot d}{\gamma}\right)^{2/3}(1+\varepsilon_k)\,\frac{4\cdot\omega_k^{\frac{1}{3}+\frac{1}{1-\varepsilon}+\frac{2}{3}(1-\varepsilon)(\beta-2)}}{d\cdot n}$$

from equation (6.7) dominates the term from equation (6.8). For $\omega_k\leqslant n^{\frac{1-\varepsilon}{2}}$ it holds that

$$\frac{\omega_k^{\frac{1}{3}+\frac{1}{1-\varepsilon}+\frac{2}{3}(1-\varepsilon)(\beta-2)}}{n}\leqslant n^{\frac{1-\varepsilon}{6}+\frac{1}{2}+\frac{1}{3}(\beta-2)(1-\varepsilon)^2-1}.$$

Now we only need to show $\frac{1-\varepsilon}{6}+\frac{1}{2}+\frac{1}{3}(\beta-2)(1-\varepsilon)^2<1$, which clearly holds, since $\beta-2<1$ and $1-\varepsilon<1$. This means that the second terms from equations (6.8) and (6.7) only account to a load of $\Theta\left(n^{\frac{3+1-\varepsilon+2(\beta-2)(1-\varepsilon)^2}{6}-1}m\right)$ for all layers but the core.

Now we need to take care of the contribution of load lost due to $\langle$D2$\rangle$-nodes from V_{k+1}. From Lemma 6.12 we know that the number of edges to these nodes is at most

$$6\exp\left(-\omega_{k+1}^{1/3}/4\right)W_{k+1}+\left(4\omega_k\frac{W_k}{W}+6(c'+3)\ln n\right)\frac{d\sqrt{c'}}{\sqrt{2}}\sqrt{n\ln n}+n^{-1}$$

with probability at least $1-3\cdot n^{-c'}$ for some constant $c'>3$. As before we multiply this value with $(1+\varepsilon_k)\frac{\omega_{k-1}}{\left(\omega_k\frac{W_{k+1}}{W}\right)^{2/3}W_k}m$ to get an upper bound on the load lost due to these missing edges. The first term of this expression amounts to

$$\Theta\left(\exp\left(\omega_k^{\frac{1-\varepsilon}{3}}/4\right)\omega_k^{(1-\varepsilon)\frac{2-\beta}{3}+\frac{1}{1-\varepsilon}+\beta-2-\frac{2}{3}}m\right) = \mathcal{O}\left(\exp\left(\omega_k^{\frac{1-\varepsilon}{3}}/4\right)\omega_k^{\frac{1}{1-\varepsilon}+(\beta-2)}m\right)$$

using the bounds $W_k = \Theta(n\cdot\omega_k^{2-\beta})$, the substitutions $\omega_{k-1} = \omega_k^{\frac{1}{1-\varepsilon}}$ and $\omega_{k+1} = \omega_k^{1-\varepsilon}$ and $W = d\cdot n$. This is only a small fraction of m for sufficiently large constant ω_k. The second term of the expression amounts to

$$\Theta\left(\omega_k^{\frac{1}{3}+\frac{1}{1-\varepsilon}+\frac{2}{3}(\beta-2)(1-\varepsilon)}\frac{\sqrt{n\ln n}}{n} + \omega_k^{\beta-2-\frac{2}{3}+\frac{1}{1-\varepsilon}+\frac{2}{3}(1-\varepsilon)(\beta-2)}\frac{\sqrt{n}(\ln n)^{3/2}}{n}\right)m$$

using the same bounds and substitutions as before. We may now assume $\omega_k = \mathcal{O}(\mathrm{polylog}(n))$, as only for those layers ⟨D2⟩-nodes may appear. This means that the former expression is bounded by $\mathcal{O}\left(\frac{\mathrm{polylog}(n)}{\sqrt{n}}m\right)$. The last term n^{-1} is clearly dominated by the term before and can therefore be ignored.

Putting all load contributions together we get the first statement as desired. □

Finally, we bound the total load left on dead-ends during the top-down phase.

Lemma 6.14. *For every constant c, there exists a constant c′ such that if we run the top-down phase on layers with* $\omega_i \geqslant c'$*, then with probability at least* $1-1/n^{-c}$ *we obtain a total load of at most* $m/2$ *on all dead-ends on these layers.*

Proof. The upper bound on the dead-end load from Lemma 6.7 gives us a fraction

$$\begin{aligned}
&\mathcal{O}\left(\frac{\omega_{k-1}}{W_k}\ln n\cdot m + \frac{\omega_{k-1}}{W_k}n_k\exp\left(-\frac{1}{4}\left(\omega_k\frac{W_{k-1}}{W}\right)^{1/3}\right)\right)\\
= \;&\mathcal{O}\left(\frac{\omega_{k-1}\omega_k^{\beta-2}\ln n}{n} + \omega_k^{\frac{\varepsilon}{1-\varepsilon}}\exp\left(-\frac{1}{4}\left(\frac{1}{2\overline{d}}\right)^{1/3}\omega_k^{(3-\beta)/6}\right)\right)
\end{aligned}$$

of m, where we used the bounds for the W_k and n_k we established at the beginning of section 6. As we know that ⟨D1⟩-nodes only appear for $\omega_k = \mathcal{O}(\mathrm{polylog}(n))$, the first term is a small constant fraction for large enough n if we sum up over all $\mathcal{O}(\log\log n)$ layers. For a large enough constant ω_k the first term sums up to a small constant fraction over all layers, as it shrinks exponentially fast with growing ω_k.

Lemma 6.8 states that the contribution of ⟨D2⟩-nodes from V_k to all layers is at most a fraction

$$\mathcal{O}\left(\omega_k^{\frac{4-\beta+\varepsilon(\beta-1)}{1-\varepsilon}}\cdot\exp\left(-\omega_k^{1/3}/4\right) + \frac{\mathrm{polylog}(n)}{\sqrt{n}}\right)$$

of m. Again, the first term sums up to a small constant fraction over all layers for $\omega_k \geqslant \widehat{w}$, where $\widehat{w}$ is a large enough constant. The second term is a small constant fraction for large enough n if we sum up over all $\mathcal{O}(\log\log n)$ layers. As we already argued in section 6, we still need to take care of nodes with load at most $\widehat{w}$. To deal with them we stop analysis at a constant $\widehat{w}_2 > \widehat{w}$. For all nodes with weight at least $\widehat{w}$ we demand property $\langle$D2$\rangle$, but for nodes v with $w_v < \widehat{w}$ we now demand a degree of at least $\widehat{w}_2$, because otherwise they would end up in one of the lower layers which we ignore. As the weight of these nodes is at most $\widehat{w}$ this results in a degree deviation of at least $w_v \frac{\widehat{w}_2 - \widehat{w}}{\widehat{w}}$. Then the fraction of load these nodes contribute is at most

$$\mathcal{O}\left(\widehat{w}^{\frac{4-\beta+\varepsilon(\beta-1)}{1-\varepsilon}} \cdot \exp\left(-\frac{\widehat{w}_2 - \widehat{w}}{4\widehat{w}}\right) + \frac{\operatorname{polylog}(n)}{\sqrt{n}}\right)$$

for all nodes with $w_v \geqslant 1$. If we choose a big enough constant $\widehat{w}_2 > \widehat{w}$ this becomes a small constant fraction

Lemma 6.9 and Lemma 6.10 give us a fraction of at most

$$\mathcal{O}\left(\omega_k^{-\varepsilon(\beta-2)} + \frac{\omega_k^{\beta-2}\omega_{k+1}^{\beta-2}\ln n}{n}\right)$$

of the total load m per layer. If we choose a large enough constant ω_k the first terms sum up to at most a small constant fraction of over all layers. As $\omega_{k+1} < \omega_k < n^{1/2}$, the second term is at most $n^{\beta-3}$, which is a small constant fraction for large enough n if we sum up over all $\mathcal{O}(\log\log n)$ layers.

Lemma 6.13 shows that $\langle$D4$\rangle$-nodes contribute at most a fraction of

$$\mathcal{O}\left(\omega_k^{\frac{2-\beta}{3(1-\varepsilon)}} + \omega_k^{\frac{(2\beta^2-11\beta+14)(\beta-2)}{27(1-\varepsilon)}} + \exp\left(\omega_k^{\frac{1-\varepsilon}{3}}/4\right)\omega_k^{\frac{1}{1-\varepsilon}+(\beta-2)} \right.$$
$$\left. +\ n^{\frac{3+1-\varepsilon+2(\beta-2)(1-\varepsilon)^2}{6}-1} + \frac{\operatorname{polylog}(n)}{\sqrt{n}}\right)$$

of m to the total dead-end load. As the exponents of ω_k in the first two terms are negative constants and the third term shrinks exponentially with growing ω_k, these terms sum up to a small constant fraction over all layers for ω_k at least a large enough constant. Note that the exponent of the fourth term is negative. Therefore, the last two terms sum up to at most a constant fraction over all $\mathcal{O}(\log\log n)$ layers for large enough n.

As all four node types give arbitrarily small constant fractions for ω_k at least a large enough constant and n sufficiently large, the total load left on dead-ends during the downward distribution is at most a small constant fraction of m. The fact that all these results hold with probability at least $1-n^{-c}$ for an arbitrarily large constant c gives us the result as desired. □

The last lemma implies that with high probability, for a suitably chosen *key layer* at most half of the load is left on dead-ends during the top-down phase on this and the above layers. In particular, our upper bound on the load of non-dead-end nodes in Lemma 6.2 implies that on this layer, every such node gets at most a load of $(1+\varepsilon_k) \cdot m \cdot w_v / W_k$. On the other hand, a load of $m/2$ passes through this layer w.h.p. In the worst case all non-dead-ends get the maximum load of $(1+\varepsilon_k) \cdot m \cdot w_v / W_k$. This results in at least $n \frac{\gamma}{4(1+\varepsilon_k)} {\omega_k}^{-\frac{\beta-1}{(1-\varepsilon)}}$ nodes which absorb m/n load each, causing a decrease of unassigned load by a constant fraction of at least $\frac{\gamma}{4(1+\varepsilon_k)} {\omega_k}^{-\frac{\beta-1}{(1-\varepsilon)}}$. Here, $\omega_k \geqslant c'$ where c' is as chosen in Lemma 6.14.

7 Analysis of Iterative Absorption

Algorithm 1 sends all unassigned load back to the top, balances it within the top layer, and restarts the top-down distribution step. Observe that all the arguments made for the analysis of the downward propagation can be applied for any value of m. The absorption of load during these iterations is adjusted according to the following scheme. We let each of the nodes absorb at most a load of $m/(n \cdot t^2)$ in round t. This scheme is executed for $t = \log\log n$ rounds and then repeated in chunks of $\log\log n$ rounds until all load is assigned. We will show that with high probability after a constant number of repetitions, all load is assigned. In addition, as $\sum_{t=1}^{\infty} 1/t^2 = \Pi^2/6$, each node receives a load of $(1+\mathcal{O}(1)) \cdot m/n$.

In particular, our aim is to show that using this scheme we need only $\mathcal{O}(\log\log n)$ top-down distribution steps to reduce the total unassigned load in the system to $m' = m/\log^c n$, for any constant c. This is shown in the lemma below. Given this result, we run the protocol long enough such that c becomes a sufficiently large constant. We want to show that, if this is the case, each node on a layer with polylogarithmic degree gets a load of at most m/n, resulting in all remaining load being absorbed. As each non-dead-end on this layer gets a share of at most $w_v \frac{W_k}{W} m' = \frac{\mathrm{polylog}(n)}{n} m' = m/n$ they fulfill the requirement. The same bound holds for $\langle$D4$\rangle$-nodes by definition. As $\langle$D1$\rangle$- and $\langle$D2$\rangle$-nodes do not appear on layers of at least polylogarithmic degree, we can ignore them as well. All we need to care about now are $\langle$D3$\rangle$-nodes. We can derive upper bounds on their load similar to the ones for non-dead-ends using results on the expected number of edges these nodes have into both of the possible next-highest layers as in the proof of Lemma 4.2. It now remains to show the following lemma.

Lemma 7.1. *Using the repeated absorption scheme of Algorithm 1, for any constant c, only $\mathcal{O}(\log\log n)$ rounds suffice to reduce the unassigned load in the network to $m/\log^c n$.*

Proof. In the analysis of the top-down distribution step above, we saw that for arbitrary m we absorb a constant fraction of the load within a key layer. We denote by m^t the total unassigned load in round t. If the nodes would know m^t, they could absorb a load of m^t/n in each round, which would decrease the load to $m/\log^c n$ in $\mathcal{O}(\log\log n)$ rounds w.h.p. However, we assume that nodes do not know the current load. Still, we can use this imaginary process as a benchmark process for our protocol.

Instead, in our scheme we absorb a load of at most $m/(nt^2)$ in round t. We denote by $\varepsilon > 0$ a lower bound on the constant fraction of load that is absorbed in any round of the benchmark process. Note that for large t, our absorption of $m/(nt^2)$ per node is actually more than the $m(1-\varepsilon)^t/n$ that would at best result from the benchmark process. Thus, let t^* be the value such that

$$(t^*)^2(1-\varepsilon)^{t^*} = 1 \ .$$

Obviously, after round t^*, the absorption in our process dominates the one in the benchmark process, and we obtain at least the same load reduction. For the first t^* rounds, we simply restrict our attention to the first round of the first t^* repetitions of our scheme. In each of these rounds, a load of m/n is absorbed, which dominates the absorption of the first t^* rounds of the benchmark process. Then, starting in the next repetition, we consider round t^* and the following (including also the next repetitions of our scheme). In these rounds the absorption dominates the absorption of the remaining rounds of the benchmark process. Thus, within $\mathcal{O}(\log\log n)$ rounds of our repeated scheme, for each round of the benchmark process there is a round in our process, which dominates in terms of load absorption. Hence, after this many rounds, we obtain the same result as the benchmark process. □

8 Discussion

To the best of our knowledge, we have presented the first double-logarithmic load balancing protocol for a realistic network model. Our algorithm reaches a balanced state in time *less than the diameter* of the graph, which is a common lower bound for other protocols (e.g. [10]). Note that our Theorem 4.1 can be interpreted outside of the intended domain: It reproves (without using the fact) that the giant component is of size $\Theta(n)$ (known from [4]) and that rumor spreading to most vertices can be done in $\mathcal{O}(\log\log n)$ (known from [11]).

Our algorithm works fully distributed, and nodes decide how many tokens should be sent or received based only on their current load (and those of its neighbors). We expect our wave algorithm to perform very robust against node and edge failures as it does *not* require global information on distances [10] or the computation of a balancing flow [8].

Our Theorem 4.3 allows initial load on nodes with degree $\Omega(\operatorname{polylog} n)$. Future work includes a further relaxation of this assumption, for instance, by employing results about greedy local-search based algorithms to find high degree nodes [3, 6]. Another interesting direction is to translate our load balancing protocol into an algorithm which samples a random node using the analogy between load and probability distributions. Such sampling algorithms are crucial for crawling large-scale networks such as online social networks like Facebook, where direct sampling is not supported [12].

Bibliography

[1] D. Angluin and L. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. *J. Comput. Syst. Sci.*, 18:155–193, 1979.

[2] J. E. Boillat. Load balancing and poisson equation in a graph. *Concurrency: Pract. Exper.*, 2:289–313, 1990.

[3] C. Borgs, M. Brautbar, J. T. Chayes, S. Khanna, and B. Lucier. The power of local information in social networks. In *8th Intl. Workshop Internet & Network Economics (WINE)*, pp. 406–419, 2012.

[4] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99: 15879–15882, 2002.

[5] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Phys. Rev. Lett.*, 90:058701, 2003.

[6] C. Cooper, T. Radzik, and Y. Siantos. A fast algorithm to find all high degree vertices in graphs with a power law degree sequence. In *9th Intl. Workshop Algorithms and Models for the Web Graph (WAW)*, pp. 165–178, 2012.

[7] G. Cybenko. Load balancing for distributed memory multiprocessors. *J. Parallel and Distributed Comput.*, 7:279–301, 1989.

[8] R. Diekmann, A. Frommer, and B. Monien. Efficient schemes for nearest neighbor load balancing. *Parallel Computing*, 25:789–812, 1999.

[9] D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009.

[10] R. Elsässer and T. Sauerwald. Discrete load balancing is (almost) as easy as continuous load balancing. In *29th Symp. Principles of Distributed Computing (PODC)*, pp. 346–354, 2010.

[11] N. Fountoulakis, K. Panagiotou, and T. Sauerwald. Ultra-fast rumor spreading in social networks. In *23rd Symp. Discrete Algorithms (SODA)*, pp. 1642–1660, 2012.

[12] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou. Walking in Facebook: A case study of unbiased sampling of OSNs. In *29th IEEE Conf. Computer Communications (INFOCOM)*, pp. 2498–2506, 2010.

[13] F. C. Graham and M. Radcliffe. On the spectra of general random graphs. *Electr. J. Comb.*, 18, 2011.

[14] K. H. Huebner, D. L. Dewhirst, D. E. Smith, and T. G. Byrom. *The Finite Element Methods for Engineers.* Wiley, 2001.

[15] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times.* AMS, 2008.

[16] S. Muthukrishnan, B. Ghosh, and M. H. Schultz. First- and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory Comput. Syst.*, 31:331–354, 1998.

[17] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:pp. 167–256, 2003.

[18] R. Subramanian and I. D. Scherson. An analysis of diffusive load-balancing. In *6th Symp. Parallelism in Algorithms and Architectures (SPAA)*, pp. 220–225, 1994.

[19] S. Surana, B. Godfrey, K. Lakshminarayanan, R. Karp, and I. Stoica. Load balancing in dynamic structured peer-to-peer systems. *Performance Evaluation*, 63:217–240, 2006.

[20] R. van der Hofstad. Random graphs and complex networks. Available at `www.win.tue.nl/~rhofstad/NotesRGCN.pdf`, 2011.

[21] D. Zhanga, C. Jianga, and S. Li. A fast adaptive load balancing method for parallel particle-based simulations. *Simulation Modelling Practice and Theory*, 17:1032–1042, 2009.