Randomized Load Balancing on Networks with Stochastic Inputs *†

Leran Cai[1] and Thomas Sauerwald[1]

1 **University of Cambridge, UK**
(lc647|tms41)@cl.cam.ac.uk

Abstract

Iterative load balancing algorithms for indivisible tokens have been studied intensively in the past, e.g., [21, 18, 24]. Complementing previous worst-case analyses, we study an average-case scenario where the load inputs are drawn from a fixed probability distribution. For cycles, tori, hypercubes and expanders, we obtain almost matching upper and lower bounds on the discrepancy, the difference between the maximum and the minimum load. Our bounds hold for a variety of probability distributions including the uniform and binomial distribution but also distributions with unbounded range such as the Poisson and geometric distribution. For graphs with slow convergence like cycles and tori, our results demonstrate a substantial difference between the convergence in the worst- and average-case. An important ingredient in our analysis is a new upper bound on the t-step transition probability of a general Markov chain, which is derived by invoking the evolving set process.

1998 ACM Subject Classification G.3 Probability and Statistics

Keywords and phrases random walks, randomized algorithms, parallel computing

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.

1 Introduction

In the last decade, large parallel networks became widely available for industrial and academic users. An important prerequisite for their efficient usage is to balance their work efficiently. Load balancing is known to have applications to scheduling [27], routing [9], numerical computation such as solving partial differential equations [29, 28, 26], and finite element computations [13]. 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. In this work we will study a decentralized and iterative load balancing protocol where a processor knows only its current load and that of the neighboring processors. Based on this, decides how much load should be sent (or received).

Load Balancing Models. A widely used approach is diffusion, e.g., the first-order-diffusion scheme [9, 18], where the amount of load sent along each edge in each round is proportional to the load difference between the incident nodes. In this work, we consider the alternative, the so-called matching model, where in each round only the edges of a matching are used to average the load locally. In comparison to diffusion, the matching model reduces the

* The second author was supported by the ERC Starting Grant "Dynamic March".
† See [8] for a full version of this work.

communication in the network and moreover tends to behave in a more "monotone" fashion than diffusion, since it avoids concurrent load exchanges which may increase the maximum load or decrease the minimum load in certain cases.

We measure the smoothness of the load distribution by the so-called *discrepancy*, which is the difference between the maximum and minimum load among all nodes. In view of more complex scenarios where jobs are eventually removed or new jobs are generated, the discrepancy seems to be a more appropriate measure than the *makespan*, which only considers the maximum load.

Many studies in load balancing assume that load is arbitrarily divisible. In this so-called *continuous case*, load balancing corresponds to a Markov chain on the graph and one can resort to a wide range of established techniques to analyze the convergence speed [6, 11, 18]. In particular, the *spectral gap* captures the time to reach a small discrepancy fairly accurately, e.g., see [25, 21] for the diffusion and see [7, 17] for the matching model.

However, in many applications a processor's load may consist of tasks which are not further divisible. That is why the continuous case has been also referred to as "idealized case" [21]. A natural way to model indivisible tasks is the *unit-size token model* where one assumes a smallest load entity, the unit-size token, and load is always represented by a multiple of this smallest entity. In the following, we will refer to the unit-size token model as the *discrete case*.

Initiated by the work of [21], there has been a number of studies on load balancing in the discrete case. Unlike the deterministic rounding in [21], [24] analyzed a randomized rounding based strategy, meaning that an excess token will be distributed uniformly at random among the two communicating nodes. The authors of [24] proved that with this strategy the time to reach constant discrepancy in the discrete case is essentially the same as the corresponding time in the continuous case. Their results hold both for the *random matching model*, where in each round a new random matching is generated by a simple distributed protocol, and the *balancing circuit model* (a.k.a. dimension exchange), where a fixed sequence of matching is applied periodically. In this work, we will focus on the *balancing circuit model*, which is particularly well suited for highly structured graphs such as cycles, tori or hypercubes.

Worst-Case vs. Average-Case Inputs. Previous work has almost always adopted the usual worst-case framework for deriving bounds on the load discrepancy [21]. That means any upper bound on the discrepancy holds for an arbitrary input, i.e., an arbitrary initial load vector. While it is of course very natural and desirable to have such general bounds, the downside is that for graphs with poor expansion like cycles or 2D-tori, the convergence is rather slow, i.e., quadratic or linear in the number of nodes n. This serves as a *motivation* to explore an average-case input. Specifically, we assume that the number of load items at each node is sampled independently from a fixed distribution. Our main results demonstrate that the convergence of the load vector is considerably quicker (measured by the load discrepancy), especially on networks with slow convergence in the worst-case such as cycles and 2D-tori.

We point out that many related problems including scheduling on parallel machines or load balancing in a dynamic setting (meaning that jobs are continuously added and processed) have been studied under random inputs, e.g., [3, 12, 2]. To the best of our knowledge, only very few works have studied this question in iterative load balancing. One exception is [22], which investigated the performance of continuous load balancing on tori in the diffusion model. In contrast to this work, however, only upper bounds are given and they hold for the multiplicative ratio between maximum and minimum load, rather than the discrepancy. Another related work is [4], which presents a distributed algorithm for community detection that is based on averaging a random $\{-1, 1\}$ initial load vector.

1.1 Notation and Background

We assume that $G = (V, E)$ is an undirected, connected graph with n nodes labeled in $[0, n-1]$. Unless stated otherwise, all logarithms are to the base e. The notations $\mathbb{P}[\mathcal{E}]$ and $\mathbb{E}[X]$ denote the probability of an event $\mathcal{E}$ and the expectation of a random variable X, respectively. For any n-dimensional vector x, $\operatorname{disc}(x) = \max_i x_i - \min_i x_i$ denotes the *discrepancy*. By $\frac{\mathbf{1}}{\mathbf{n}}$ we denote the vector with all values being $\frac{1}{n}$.

Matching Model. In the *matching model* (sometimes also called *dimension exchange model*), every two matched nodes in round t balance their load as evenly as possible. This can be expressed by a symmetric n by n matching matrix $\mathbf{M}^{(t)}$, where with slight abuse of notation we use the same symbol for the matching and the corresponding matching matrix. Formally, matrix $\mathbf{M}^{(t)}$ is defined by $\mathbf{M}^{(t)}_{u,u} = 1/2$, $\mathbf{M}^{(t)}_{v,v} = 1/2$ and $\mathbf{M}^{(t)}_{u,v} = \mathbf{M}^{(t)}_{v,u} = 1/2$ if $\{u, v\} \in \mathbf{M}^{(t)} \subseteq E$, and $\mathbf{M}^{(t)}_{u,u} = \mathbf{M}^{(t)}_{v,v} = 1$, $\mathbf{M}^{(t)}_{u,v} = 0$ $(u \neq v)$ if u, v are not matched.

Balancing Circuit. In the *balancing circuit model*, a specific sequence of matchings is applied periodically. More precisely, let $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(d)}$ be a sequence of d matching matrices, also called *period* [1]. Then in step $t \geq 1$, we apply the matching matrix $\mathbf{M}^{(t)} := \mathbf{M}^{(((t-1) \mod d)+1)}$. We define the *round matrix* by $\mathbf{M} := \prod_{s=1}^{d} \mathbf{M}^{(s)}$. If $\mathbf{M}$ is symmetric, we define $\lambda(\mathbf{M})$ to be its second largest eigenvalue (in absolute value). Following [21], if $\mathbf{M}$ is not symmetric (which is usually the case), we define $\lambda(\mathbf{M})$ as the second largest eigenvalue of the symmetric matrix $\mathbf{M} \cdot \mathbf{M}^T$, where $\mathbf{M}^T$ is the transpose of $\mathbf{M}$. We always assume that $\lambda(\mathbf{M}) < 1$, which is guaranteed to hold if the matrix $\mathbf{M}$ is irreducible. Since $\mathbf{M}$ is doubly stochastic, all powers of $\mathbf{M}$ are doubly stochastic. A natural choice for the d matching matrices is given by an edge coloring of G. There are various efficient distributed edge coloring algorithms, e.g. [20, 19].

Balancing Circuit on Specific Topologies. For *cycles*, we will consider the natural "Odd-Even" scheme meaning that for $\mathbf{M}^{(1)}$, the matching consists of all edges $\{j, (j+1) \pmod n\}$ for any odd j, while for $\mathbf{M}^{(2)}$, the matching consists of all edges $\{j, (j+1) \pmod n\}$ for any even j. More generally, for r-dimensional tori with vertex set $[0, n^{1/r} - 1]^r$, we will have $2 \cdot r$ matchings in total, meaning that for every dimension $1 \leq i \leq r$ we have two matchings along dimension i, similar to the definition of matchings for the cycle. For *hypercubes*, the canonical choice is dimension exchange consisting of $d = \log_2 n$ matching matrices $\mathbf{M}^{(i)}$ by $\mathbf{M}^{(i)}_{u,v} = 1/2$ if and only if the bit representation of u and v differ only in bit i.

Continuous Case vs. Discrete Case. In the continuous case, load is arbitrarily divisible. Let $\xi^{(0)} \in \mathbb{R}^n$ be the initial load represented as a row vector, and in every round two matched nodes average their load perfectly. We consider the load vector $\xi^{(t)}$ after t rounds in the balancing circuit model (that means, after the executions of $t \cdot d$ matchings in total). This process corresponds to a linear system $\xi^{(t)} = \xi^{(t-1)} \mathbf{M}$, which results in $\xi^{(t)} = \xi^{(0)} \mathbf{M}^t$.

Let us now turn to the discrete case with indivisible, unit-size tokens. Let $x^{(0)} \in \mathbb{N}^n$ be the initial load vector with average load $\overline{x} := \sum_{w \in V} x^{(0)}_w / n$, and $x^{(t)}$ be the load vector at the end of round t. In case the sum of tokens of the two paired nodes is odd, we employ the *random orientation* (or *randomized rounding*) [21, 24]. More precisely, if there are two nodes

[1] Note that d may be different from the maximal degree (or degree) of the underlying graph.

u and v with load a and b being paired by matching $\mathbf{M}^{(t)}$, then node u gets either $\lceil \frac{a+b}{2} \rceil$ or $\lfloor \frac{a+b}{2} \rfloor$ tokens, with probability 1/2 each. The remaining tokens are assigned to node v.

The Average-Case Setting. We consider a setting where each entry of the initial load vector $x^{(0)}$ is chosen from an exponentially concentrated probability distribution D with expectation μ and variance σ^2 (see Definition 1.1). Our main results in this paper hold for all distributions satisfying the following definition.

▶ **Definition 1.1.** A distribution D over $\mathbb{N} \cup \{0\}$ with expectation μ and variance σ^2 is *exponentially concentrated* if there is a constant $\kappa > 0$ so that for any $X \sim D$, $\delta > 0$,

$$\mathbb{P}\left[|X - \mu| \geq \delta \cdot \sigma\right] \leq \exp\left(-\kappa\delta\right).$$

In the following, we refer to **average-case** when the initial number of load items on each vertex is drawn independently from a fixed exponentially concentrated distribution.

▶ **Lemma 1.2.** *The uniform distribution, binomial distribution, geometric distribution and Poisson distribution are all exponentially concentrated.*

▶ **Lemma 1.3.** *Let D be an exponentially concentrated distribution and let $X \sim D$. Then,*

$$\mathbb{P}\left[|X - \mu| \leq 8/\kappa \cdot \sigma \log n\right] \geq 1 - n^{-3}.$$

In particular, the initial discrepancy satisfies $\operatorname{disc}(x^{(0)}) = O(\sigma \cdot \log n)$ *with probability* $1 - n^{-2}$.

The advantage of Lemma 1.3 is that we can use a simple conditioning trick to work with distributions that have a finite range and can therefore be analyzed by Hoeffding's inequality. Therefore in the analysis we may simply work with a bounded-range distribution $\widetilde{D}$, which is D under the condition that only values in the interval $[\mu - 8/\kappa \cdot \sigma \log n, \mu + 8/\kappa \cdot \sigma \log n]$ occur.

1.2 Our results

Our first contribution is a general formula that allows us to express the load difference between an arbitrary pair of nodes in round t.

▶ **Theorem 1.4.** *Consider the balancing circuit model with an arbitrary round matrix $\mathbf{M}$ in the average case. Then for any pair of nodes u, v and round t, it holds for any $\delta > 0$ that*

$$\mathbb{P}\left[\left|x_u^{(t)} - x_v^{(t)}\right| \geq \delta \cdot 16\sqrt{2}/\kappa \cdot \sigma \cdot \log n \cdot \left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2 + \sqrt{64 \log n}\right] \leq 2 \cdot e^{-\delta^2} + 2n^{-3}.$$

Furthermore, for any pair of vertices u, v, we have the following lower bound:

$$\mathbb{P}\left[\left|x_u^{(t)} - x_v^{(t)}\right| \geq \sigma/(2\sqrt{2\log_2 \sigma}) \cdot \left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2 - \sqrt{64 \log n}\right] \geq \frac{1}{16}.$$

▶ Remark. The lower bound above is useless if σ is small, say, at most a constant. However for sufficiently large σ, the lower bound gives a useful result (see also Section 4 & 5).

The proof of the upper bound Theorem 1.4 is the easier direction, and it relies on a previous result relating continuous and discrete load balancing from [24]. The lower bound is technically more challenging and applies a generalized version of the central limit theorem.

Together, the upper and lower bound in the above result establish that the load deviation between any two nodes u and v is essentially captured by $\left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2$. However, in some instances it might be desirable to have a more tangible estimate at the expense of generality. To this end, we first observe that $\left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2^2 \leq 4 \cdot \max_{k \in V} \left\|\mathbf{M}^t_{\cdot,k} - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2^2$ (see Lemma 3.1). Hence we are left with the problem of bounding the t-step probability vector $\mathbf{M}^t_{\cdot,k}$.

For reversible Markov chains, the last expression has been analyzed in several works. For example, [15, Lemma 3.6] implies that for random walks on graphs, $\mathbf{P}^t_{u,v} = O(\deg(v)/\sqrt{t})$. However, the Markov chain associated to $\mathbf{M}$ is not reversible in general. For irreversible Markov chains, [14] used the so-called evolving set process to derive a similar bound. Specifically, they proved in [14, Theorem 17.17] that if $\mathbf{P}$ denotes the transition matrix of a lazy random walk (i.e., a random walk with loop probability at least 1/2) on a graph with maximal degree Δ, π the stationary distribution of $\mathbf{P}$, then for any vertex $x \in V$:

$$\left|\mathbf{P}^t_{x,x} - \pi_x\right| \leq \frac{\sqrt{2}\Delta^{5/2}}{\sqrt{t}}.$$

Such estimates have been used in applications besides load balancing, including distributed random walks and spanning tree enumeration [23, 15]. Here we generalize this result to Markov chains with an arbitrary loop probability and to arbitrary t-step transition probabilities:

▶ **Theorem 1.5.** *Let $\mathbf{P}$ be the transition matrix of an irreducible Markov chain and π its stationary distribution. Then we have for all states x, y and step t,*

$$\left|\mathbf{P}^t_{x,y} - \pi_y\right| \leq \frac{\pi_{\max}^{3/2}}{\pi_{\min}^{3/2}} \cdot \frac{2}{\beta^{1/2}\alpha}\sqrt{\frac{1-\beta+\alpha}{\alpha t}},$$

where $\alpha := \min_{u \neq v} \mathbf{P}_{u,v} > 0$ and $\beta := \min_u \mathbf{P}_{u,u} > 0$.

Applying this bound to a round matrix $\mathbf{M}$ formed of $d = O(1)$ matchings we obtain $\left|\mathbf{M}^t_{u,v} - 1/n\right| = O(t^{-1/2})$. It should be noted that [24, Lemma 2.5] proved a weaker version where the upper bound is only $O(t^{-1/8})$ instead of $O(t^{-1/2})$. As proven in Lemma 4.2, the bound $O(t^{-1/2})$ is asymptotically tight if we consider the balancing circuit model on *cycles*.

Combining the bound in Theorem 1.5 with the upper bound in Theorem 1.4 yields:

▶ **Theorem 1.6.** *Consider the balancing circuit model with an arbitrary round matrix $\mathbf{M}$ consisting of $d = O(1)$ matchings in the average case. The discrepancy after t rounds is $O(t^{-1/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ with probability $1 - O(n^{-1})$.*

Since the initial discrepancy in the average case is $O(\sigma \cdot \log n)$ (see Lemma 1.3), Theorem 1.6 implies that in the average case, there is a significant decrease (roughly of order $t^{-1/4}$) in the discrepancy, regardless of the underlying topology. For round matrices $\mathbf{M}$ with small second largest eigenvalue, the next result provides a significant improvement:

▶ **Theorem 1.7.** *Consider the balancing circuit model with an arbitrary round matrix $\mathbf{M}$ consisting of d matchings in the average case. We can derive that the discrepancy after t rounds is $O(\lambda(\mathbf{M})^{t/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ with probability $1 - O(n^{-1})$.*

In Section 4, we derive bounds on the discrepancy for concrete topologies (see Figure 1).

Finally, we discuss our results and compare them with the convergence of the discrepancy in the worst-case in Section 5. On a high level, these results demonstrate that on all the considered topologies, we have much faster convergence in the average-case than in the worst-case. However, if we are only interested in the time to achieve a very small, say, constant or poly-logarithmic discrepancy, then we reveal an interesting dichotomy: we have a quicker convergence than in the worst-case iff the standard deviation σ is smaller than some threshold depending on the topology. We observe the same phenomena in our experiments.

Graph	$\mathrm{disc}(x^{(t)})$
Cycle	$t^{-1/4} \cdot \sigma$
r-dim. Torus	$t^{-r/4} \cdot \sigma$
Expander	$\lambda^{t/4} \cdot \sigma$
Hypercube	$2^{-t/2} \cdot \sigma$

Figure 1 Discrepancy bounds (without logarithmic factors) for different topologies.

2 Proof of the General Bound (Theorem 1.4)

2.1 Proof of Theorem 1.4 (Upper Bound)

We will use the following result from [24] that bounds the deviation between the continuous and discrete load, assuming that we have $\xi^{(0)} = x^{(0)}$.

▶ **Theorem 2.1** ([24, Theorem 3.6(i)]). *Consider the balancing circuit model with an arbitrary round matrix* $\mathbf{M}$. *Then for any round* $t \geq 1$ *it holds that*

$$\mathbb{P}\left[\max_{w\in V}\left|x_w^{(t)} - \xi_w^{(t)}\right| \leq \sqrt{16 \cdot \log n}\right] \geq 1 - 2n^{-3}.$$

The basic proof idea is as follows. Since $\xi_u^{(t)} - \xi_v^{(t)} = \sum_{w\in V} \xi_w^{(0)} \cdot (\mathbf{M}_{w,u}^t - \mathbf{M}_{w,v}^t)$, it is a weighted sum of n i.i.d. random variables and its expectation is 0. We then can apply Hoeffding's inequality to obtain the theorem.

2.2 Proof of Theorem 1.4 (Lower Bound)

The proof of the lower bound will use the following quantitative version of a central limit type theorem for independent but non-identical random variables.

▶ **Theorem 2.2** (Berry-Esseen [5, 10] for non-identical r.v.). *Let* $X_1, X_2, ..., X_n$ *be independently distributed with* $\mathbb{E}[X_i] = 0$, $\mathbb{E}\left[X_i^2\right] = \mathrm{Var}(X_i) = \sigma_i^2$, *and* $\mathbb{E}\left[|X_i|^3\right] = \rho_i < \infty$. *If* $F_n(x)$ *is the distribution of* $\frac{X_1+\ldots+X_n}{\sqrt{\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2}}$ *and* $\Phi(x)$ *is the standard normal distribution, then*

$$|F_n(x) - \Phi(x)| \leq C_0\psi_0,$$

where $\psi_0 = \left(\sum_{i=1}^n \sigma_i^2\right)^{-3/2} \cdot \sum_{i=1}^n \rho_i$ *and* $C_0 > 0$ *is a constant.*

With this concentration tool at hand, we are able to prove the lower bound in Theorem 1.4. Unfortunately, it appears quite difficult to apply Theorem 2.2 directly to $\xi_u^{(t)} - \xi_v^{(t)}$, since we need a good bound on the error term ψ_0. To this end, we will first partition the vertex set V into buckets with equal contribution to $\xi_u^{(t)} - \xi_v^{(t)}$. Then we will apply Theorem 2.2 to the bucket with the largest variance.

Proof of Theorem 1.4 (Lower Bound). We first consider $\xi_u^{(t)} - \xi_v^{(t)}$:

$$dev := \xi_u^{(t)} - \xi_v^{(t)} = \sum_{w\in V} \xi_w^{(0)} \cdot \left(\mathbf{M}_{w,u}^t - \mathbf{M}_{w,v}^t\right),$$

which is a weighted sum of i.i.d. random variables with expectation μ and variance σ^2. As mentioned earlier, we have $\mathbb{E}[dev] = \sum_{w\in V} \mathbb{E}\left[\xi_w^{(0)}\right] \cdot \left(\mathbf{M}_{w,u}^t - \mathbf{M}_{w,v}^t\right) = 0$ since $\mathbf{M}$ is a doubly stochastic matrix. Of course, we could apply Theorem 2.2 directly to dev, but it appears difficult to control the error term ψ_0. Therefore we will first partition the above sum into buckets where the weights of the random variables are roughly the same.

More precisely, we will partition V into $2\log_2\sigma$ buckets, where for each i we have $V_i := \left\{w \in V\colon |\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}| \in (2^{-i-1}, 2^{-i}]\right\}$ for $1 \le i \le 2\log_2\sigma - 1$, and $V_{2\log_2\sigma} := \left\{w \in V\colon \left|\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right| \le \frac{1}{\sigma^2}\right\}$.

Further, let us consider the variance of dev, since $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$ and the inputs are independent random variables:

$$\sigma^2_{dev} = \sum_{w\in V} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2 \sigma^2.$$

Then by the pigeonhole principle there exists an index $1 \le i \le 2\log_2\sigma$ such that

$$\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2 \sigma^2 \ge \frac{1}{2\log_2\sigma} \cdot \sigma^2_{dev}.$$

This is equivalent to

$$\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2 \ge \frac{1}{2\log_2\sigma} \cdot \sum_{w\in V} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2.$$

Firstly, if such index i is just $2\log_2\sigma$, we can prove the lower bound easily. Now we can derive that, for all w in V_i,

$$\left\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\right\|_2^2 \le \left\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\right\|_\infty \cdot \left\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\right\|_1 \le \frac{1}{\sigma^2} \cdot 2 = O(\sigma^{-2}).$$

Then in Theorem 1.4 $\left\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\right\|_2$ is $O(\sigma^{-1})$ and the lower bound holds trivially. Therefore, we will assume in the remainder of the proof that $i < 2\log_2\sigma$.

We now decompose dev into $dev = S + S^c$, where

$$S := \sum_{w\in V_i} \xi^{(0)}_w \cdot \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right) \text{ and } S^c := \sum_{w\notin V_i} \xi^{(0)}_w \cdot \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right).$$

Let us first analyze S. We need to apply Theorem 2.2 to S. Before applying Theorem 2.2, we scale the original distribution to $\xi'^{(0)}_w = \xi^{(0)}_w - \mu$ to make the expectation be 0. In preparation for this, let us first upper bound ψ_0. Using the definition of exponentially concentrated distributions, it follows that for any constant k, the first k moments of $\xi'^{(0)}_w$ are all bounded from above by $O(\sigma^k)$. Hence,

$$\psi_0 = \frac{\sum_{w\in V_i} \mathbb{E}\left[\left|\xi'^{(0)}_w \cdot \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)\right|^3\right]}{\left(\sum_{w\in V_i} \mathbb{E}\left[\left(\xi'^{(0)}_w \cdot \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)\right)^2\right]\right)^{3/2}} \le \frac{O(\sigma^3)\cdot \sum_{w\in V_i}\left|\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right|^3}{\sigma^3\left(\sum_{w\in V_i}\left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2\right)^{3/2}}.$$

Recalling that for any $w \in V_i$, $\left|\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right| \in (2^{-i-1}, 2^{-i}]$, we can simplify the above expression as follows:

$$\psi_0 = O\left(\frac{|V_i| \cdot 2^{-3i}}{|V_i|^{3/2} \cdot 2^{-3i}}\right) = O(|V_i|^{-1/2}).$$

In the following, we will assume that $|V_i| \geq C_1$, where $C_1 > 0$ is a sufficiently large constant to be specified later. Since $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$, we have

$$F_n(x) = \mathbb{P}\left[\frac{\sum_{w\in V_i} \xi_w^{'(0)} \cdot (\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v})}{\sigma\sqrt{\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2}} \leq x\right]$$
$$= \mathbb{P}\left[S - \mathbb{E}\left[S\right] \leq x\sigma\sqrt{\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2}\right].$$

Since $|V_i| \geq C_1$, there is a constant $C_2 = C_2(C_1, C_0) > 0$ such that $C_0 \cdot \psi_0 \leq C_2$ and

$$\mathbb{P}\left[S - \mathbb{E}\left[S\right] \geq x\sigma\sqrt{\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2}\right] \geq \Phi(-x) - C_0\psi_0 \geq \Phi(-x) - C_2.$$

Now let $\Phi^c(x)$ denote the complement of the standard normal distribution. By using [1, Formula 7.1.13] and substitution we get:

$$\frac{1}{\sqrt{\pi}(x + \sqrt{x^2+2})e^{x^2}} < \Phi^c(x) \leqslant \frac{1}{\sqrt{\pi}(x + \sqrt{x^2+4/\pi})e^{x^2}}.$$

Hence by $\Phi(-x) = \Phi^c(x)$, choosing $x = 1$ and C_1 sufficiently large,

$$\mathbb{P}\left[S - \mathbb{E}\left[S\right] \geq \sigma\sqrt{\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2}\right] \geq \frac{1}{16}.$$

Similarly, we can derive that

$$\mathbb{P}\left[\mathbb{E}\left[S\right] - S \geq \sigma\sqrt{\sum_{w\in V_i} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2}\right] \geq \frac{1}{16}.$$

Hence, independent of what the value S^c is, there is still a probability of at least $1/16$ so that $|S + S^c| \geq \sigma/2 \cdot \sqrt{1/(2\log_2 \sigma)} \cdot \sqrt{\sum_{w\in V} \left(\mathbf{M}^t_{w,u} - \mathbf{M}^t_{w,v}\right)^2}$, which completes the proof for the case $|V_i| \geq C_1$. The case $|V_i| < C_1$ is similar and omitted here. The basic idea is not to apply Berry-Esseen but simply use the fact that any exponentially distributed random variable deviates from the expectation by $\Omega(\sigma)$ with constant probability. ◀

3 Proof of the Universal Bounds (Theorem 1.6, Theorem 1.7)

In the previous section we proved that the deviation between the loads of two nodes u and v is essentially captured by $\left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2$. However, in some cases it might be hard to compute or estimate this quantity for arbitrary vertices u and v. Therefore we will establish Theorem 1.6 which gives a more concrete estimate.

3.1 Proof of Theorem 1.6

The proof of Theorem 1.6 is fairly involved and we sketch the high level ideas. We first show that $\left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2^2$ can be upper bounded in terms of the ℓ_2-distance to the stationary distribution.

▶ **Lemma 3.1.** *Consider the balancing circuit model with an arbitrary round matrix* $\mathbf{M}$. *Then for all* $u, v \in V$, *we have* $\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\|_2^2 \leq 4 \cdot \max_{k \in V} \|\mathbf{M}^t_{.,k} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2$. *Further, for any* $u \in V$ *we have* $\max_{v \in V} \|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\|_2^2 \geq \|\mathbf{M}^t_{.,u} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2$.

The next step and main ingredient of the proof of Theorem 1.6 is to establish that $\|\mathbf{M}^t_{.,k} - \frac{\mathbf{1}}{\mathbf{n}}\|_\infty = O(1/\sqrt{t})$. This result will be a direct application of a general bound on the t-step probabilities of an arbitrary, possibly non-reversible Markov chain, as given in Theorem 1.5 from page 5:

In this subsection we prove Theorem 1.6, assuming the correctness of Theorem 1.5 whose proof is deferred to Section 3.2.

Proof of Theorem 1.6. By Theorem 1.4 and Lemma 3.1, we obtain

$$\mathbb{P}\left[\left|x_u^{(t)} - x_v^{(t)}\right| \geq \delta \cdot 16\sqrt{2}/\kappa \cdot \sigma \cdot \log n \cdot \max_{k \in V} \left\|\mathbf{M}^t_{.,k} - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2 + \sqrt{64 \log n}\right] \leq 2e^{-\delta^2} + 2n^{-3}.$$

Hence we can find a $\delta = \sqrt{3 \log n}$ so that the latter probability gets smaller than $4n^{-3}$. Further, by applying Theorem 1.5 with $\alpha = \beta = 2^{-d}$ to $\mathbf{P} = \mathbf{M}$ we conclude that $\|\mathbf{M}^t_{.,k} - \frac{\mathbf{1}}{\mathbf{n}}\|_\infty = O(t^{-1/2})$, since $d = O(1)$. Using $\|.\|_2^2 \leq \|.\|_\infty \cdot \|.\|_1$, $\|\mathbf{M}^t_{.,k} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = O(t^{-1/2})$ and the union bound, $\mathrm{disc}(x^{(t)}) = O(t^{-1/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ with probability at least $1 - 4n^{-1}$. ◀

3.2 Proof of Theorem 1.5

This section is devoted to the proof of Theorem 1.5. Our proof is based on the evolving-set process, which is a Markov chain based on any given irreducible, not necessarily reversible Markov chain on Ω. For the definition of the evolving set process, we closely follow the exposition in [14, Chapter 17].

Let $\mathbf{P}$ denote the transition matrix of an irreducible Markov chain and π its stationary distribution. $\mathbf{P}^t$ is the t-step transition probability matrix. The *edge measure* Q is defined by $Q_{x,y} := \pi_x \mathbf{P}_{x,y}$ and $Q(A, B) = \sum_{x \in A, y \in B} Q_{x,y}$.

▶ **Definition 3.2.** Given a transition matrix $\mathbf{P}$, the **evolving-set process** is a Markov chain on subsets of Ω defined as follows. Suppose the current state is $S \subset \Omega$. Let U be a random variable which is uniform on $[0, 1]$. The next state of the chain is the set

$$\tilde{S} = \left\{ y \in \Omega : \frac{Q(S, y)}{\pi_y} \geq U \right\}.$$

This chain is *not irreducible* because $\varnothing$ and Ω are absorbing states. It follows that

$$\mathbb{P}\left[y \in S_{t+1} \mid S_t\right] = \frac{Q(S_t, y)}{\pi_y}$$

since the probability that $y \in S_{t+1}$ is equal to the probability of the event that the chosen value of U is less than $\frac{Q(S_t, y)}{\pi_y}$.

▶ **Proposition 3.3** ([14, Proposition 17.19])**.** *Let* (M_t) *be a non-negative martingale with respect to* (Y_t), *and define* $T_h := \min\{t \geq 0 : M_t = 0 \text{ or } M_t \geq h\}$. *Assume that for any* $h \geq 0$

- *For* $t < T_h$, $\mathrm{Var}(M_{t+1} \mid Y_0, \ldots, Y_t) \geq \sigma^2$, *and*
- $M_{T_h} \leq Dh$.

Let $T := T_1$. *If* M_0 *is a constant, then* $\mathbb{P}[T > t] \leq \frac{2M_0}{\sigma}\sqrt{\frac{D}{t}}$.

We now generalize [14, Lemma 17.14] to cover arbitrarily small loop probabilities.

▶ **Lemma 3.4.** *Let (U_t) be a sequence of independent random variables, each uniform on $[0,1]$, such that S_{t+1} is generated from S_t using U_{t+1}. Then with $\beta := \min_u \mathbf{P}_{u,u} > 0$,*

$$\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \le \beta, S_t = S\right] \ge \pi(S) + Q(S, S^c),$$
$$\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} > \beta, S_t = S\right] \le \pi(S) - \frac{\beta Q(S, S^c)}{1-\beta}.$$

The derivation of the next lemma closely follows the analysis in [14, Chapter 17].

▶ **Lemma 3.5.** *For any two states x, y, $\left|\mathbf{P}^t_{x,y} - \pi_y\right| \le \frac{\pi_y}{\pi_x} \cdot \mathbb{P}_{\{x\}}\left[\tau > t\right]$.*

Now we want to use Proposition 3.3 to bound $\mathbb{P}_{\{x\}}\left[\tau > t\right]$. To apply it, we substitute the following parameters: M_0 is be $\pi(\{x\})$, Y_t is S_t, and $T = T_1 := \min\{t \ge 0 : \pi(S_t) = 0 \textit{ or } \pi(S_t) \ge 1\}$. Hence in our case, τ is the same as T (or T_1) in the proposition. The following two lemmas elaborate on the two preconditions (i) and (ii) of Proposition 3.3.

▶ **Lemma 3.6.** *For any time t and $S_0 = \{x\}$, $\mathrm{Var}_{S_t}(\pi(S_{t+1})) \ge \beta \pi_{\min}^2 \alpha^2$.*

Finally, we derive an upper bound on the amount by which S_t can increase in one iteration.

▶ **Lemma 3.7.** *For any time t and $S_0 = \{x\}$, $\pi(S_{t+1}) \le \left(\frac{1-\beta}{\alpha} + 1\right) \frac{\pi_{\max}}{\pi_{\min}} \cdot \pi(S_t)$.*

The proof of Theorem 1.5 follows then by combining Proposition 3.3, Lemma 3.4, Lemma 3.5, Lemma 3.6 and Lemma 3.7.

3.3 Proof of Theorem 1.7

We now prove the following discrepancy bound that depends on the $\lambda(\mathbf{M})$, as defined in Section 1.1.

Proof of Theorem 1.7. By [24, Lemma 2.4], for any pair of vertices $u, v \in V$, $\left|\mathbf{M}^t_{u,v} - \frac{1}{n}\right| \le \lambda(\mathbf{M})^{t/2}$. Hence by Lemma 3.1, $\left\|\mathbf{M}^t_{\cdot,u} - \mathbf{M}^t_{\cdot,v}\right\|_2 = O(\lambda(\mathbf{M})^{t/4})$. The bound on the discrepancy follows from Theorem 1.4 and the union bound over all vertices. ◀

4 Applications to Different Graph Topologies

Cycles. Recall that for the cycle, $V = \{0, \ldots, n-1\}$ is the set of vertices, and the distance between two vertices is $\mathrm{dist}(x,y) = \min\{y - x, x + n - y\}$ for any pair of vertices $x < y$.

The upper bound on the discrepancy follows directly from Theorem 1.6, and it only remains to prove the lower bound. To this end, we will apply the lower bound in Theorem 1.4 and need to derive a lower bound on $\|\mathbf{M}^t_{\cdot,u} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2$. Intuitively, if we had a simple random walk, we could immediately infer that this quantity is $\Omega(1/\sqrt{t})$. Since after t steps, the random walk is with probability $\approx 1/\sqrt{t}$ at any vertex with distance at most $O(\sqrt{t})$. To prove that this also holds for the load balancing process, we first derive a concentration inequality that upper bounds the probability for the random walk to reach a distant state:

▶ **Lemma 4.1.** *Consider the standard balancing circuit model on the cycle with round matrix $\mathbf{M}$. Then for any $u \in V$ and $\delta \in (0, n/2 - 1)$, we have*

$$\sum_{v \in V:\ \mathrm{dist}(u,v) \ge \delta} \mathbf{M}^t_{u,v} \le 2 \cdot \exp\left(-\frac{(\delta-2)^2}{8t}\right).$$

With the help of Lemma 4.1, we can indeed verify our intuition:

▶ **Lemma 4.2.** *Consider the standard balancing circuit model on the cycle with round matrix* $\mathbf{M}$*. Then for any vertex* $u \in V$*,* $\|\mathbf{M}^t_{.,u} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = \Omega(1/\sqrt{t})$*.*

Lemma 4.2 also proves that the factor $\sqrt{1/t}$ in the upper bound in Theorem 1.5 is the best possible. The lower bound on the discrepancy now follows by combining Lemma 4.2 with Theorem 1.4 and Lemma 3.1 stating that for any vertex $u \in V$, there exists another vertex $v \in V$ such that $\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\|_2^2 \geq \|\mathbf{M}^t_{.,u} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = \Omega(1/\sqrt{t})$.

Tori. In this section we consider r-dimensional tori, where $r \geq 1$ is any constant. For the upper bound, note that the computation of $\mathbf{M}^t_{.,.}$ can be decomposed to independent computations in the r dimensions, and each dimension has the same distribution as the cycle on $n^{1/r}$ vertices. Specifically, if we denote by $\widetilde{\mathbf{M}}$ the round matrix of the standard balancing circuit scheme on the cycle with $n^{1/r}$ vertices and $\mathbf{M}$ is the round matrix of the r-dimensional torus with n vertices, then for any pair of vertices $x = (x_1, \ldots, x_r), v = (y_1, \ldots, y_r)$ on the torus we have $\mathbf{M}^t_{x,y} = \prod_{i=1}^r \widetilde{\mathbf{M}}^t_{x_i,y_i}$. From Theorem 1.5, $|\widetilde{\mathbf{M}}^t_{x_i,y_i} - \frac{1}{n^{1/r}}| = O(t^{-1/2})$, and therefore, since r is a constant,

$$\mathbf{M}^t_{x,y} \leq \prod_{i=1}^r \left(\frac{1}{n^{1/r}} + O(t^{-1/2})\right) = O(t^{-r/2} + n^{-1}),$$

and thus $\left\|\mathbf{M}^t_{x,y} - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2^2 = O(t^{-r/2})$ for any pair of vertices x, y. Hence by Lemma 3.1, $\left\|\mathbf{M}^t_{.,u} - \mathbf{M}^t_{.,v}\right\|_2^2 = O(t^{-r/2})$. Plugging this bound into Theorem 1.4 yields that the load difference between any pair of the nodes u and v at round t is at most $O(t^{-r/4} \cdot \sigma \cdot \log^{3/2} n + \sqrt{\log n})$ with probability at least $1 - 4n^{-1}$. The bound on the discrepancy now simply follows by the union bound.

We now turn to the lower bound on the discrepancy. With the same derivation as in Lemma 4.2 we obtain the following result:

▶ **Lemma 4.3.** *Consider the standard balancing circuit model on the* r*-dimensional torus with round matrix* $\mathbf{M}$*. Then for any vertex* $u \in V$*,* $\|\mathbf{M}^t_{.,u} - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = \Omega(t^{-r/2})$*.*

The lower bound on the torus follows by combining Lemma 4.3 and Theorem 1.4.

Expanders. The upper bound $O(\lambda(\mathbf{M})^{t/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ for expanders follows immediately from Theorem 1.7. For the lower bound, since the round matrix consists of d matchings, it is easy to verify that whenever $\mathbf{M}^t_{u,v} > 0$, we have $\mathbf{M}^t_{u,v} \geq 2^{-d \cdot t}$. Consequently, for any vertex $u \in V$, $\left\|\mathbf{M}^t_{.,u} - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2^2 = \Omega(2^{-d \cdot t})$. Plugging this into Theorem 1.4 yields a lower bound on the discrepancy which is $\Omega(2^{-d \cdot t/2} \cdot \sigma / \sqrt{\log \sigma})$.

Hypercubes. For the hypercube, there is a worst-case bound of $\log_2 \log_2 n + O(1)$ [16, Theorem 5.1 & 5.3] for any input after $\log_2 n$ iterations of the dimension-exchange, i.e., after one execution of the round matrix. Hence, we will only analyze the discrepancy after s matchings, where $1 \leq s < \log_2 n$. By applying the same analysis as in Theorem 1.7, but now with $|\prod_{s=1}^t \mathbf{M}^{(s)}_{u,v} - \frac{1}{n}| \leq 2^{-t}$, we obtain that the discrepancy is $O(2^{-t/2} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$. Applying Theorem 1.4, we obtain the lower bound $\Omega(2^{-t/2} \cdot \sigma / \sqrt{\log \sigma})$.

5 Discussion and Empirical Results

We will now compare our average-case to a worst-case scenario on cycles, 2D-tori and hypercubes. For the sake of concreteness, we always assume that the input is drawn from a uniform distribution $\mathsf{Uni}[0, 2K]$. Our choice for the worst-case load vector will have (roughly) the same number of tokens and initial discrepancy. However, the exact definition of the vector will depend on the underlying topology.

Cycles and 2D-Tori For the worst-case setting on cycles, fix an arbitrary node $u \in V$ and let all nodes with distance at most $n/4$ initially have a load of $2K$ while all the other nodes have load 0. This gives rise to a load vector with nK tokens and initial discrepancy $2K$. For 2D-tori, fix an arbitrary node $u \in V$ and assign a load of $2K$ to the $n/2$ nearest neighbors and load 0 otherwise. This defines a load vector with nK tokens and initial discrepancy $2K$.

The next result provides a lower bound on the discrepancy for cycles and 2D-tori in the aforementioned worst-case setting. It essentially shows that for worst-case inputs, $\Omega(n^2)$ rounds and $\Omega(n)$ rounds are necessary for the cycle, 2D-tori, respectively, in order to reduce the discrepancy by more than a constant factor. This stands in sharp contrast to Theorem 1.6, proving a decay of the discrepancy by $\approx t^{-1/4}$, starting from the first round.

▶ **Proposition 5.1.** *For the aforementioned worst-case setting on the cycle, it holds for any round $t > 0$ that $\operatorname{disc}(x^{(t)}) \geq \frac{1}{8} \cdot K \cdot \left(1 - \exp\left(-\frac{n^2}{2048t}\right)\right) - \sqrt{64 \log n}$, with probability at least $1 - n^{-1}$. Further, for 2D-tori, it holds for any round $t > 0$ that $\operatorname{disc}(x^{(t)}) \geq \frac{1}{8} \cdot K \cdot \left(1 - \exp\left(-\frac{n}{2048t}\right)\right) - \sqrt{64 \log n}$, with probability at least $1 - n^{-1}$.*

Hypercube. We will consider only $\log_2 n$ rounds, since the discrepancy is $\log_2 \log_2 n + O(1)$ after $\log_2 n$ rounds and $O(1)$ after $2 \log_2 n$ rounds [16]. A natural corresponding worst-case distribution is to have load $2K$ on all nodes whose $\log_2 n$-th bit is equal to 1 and 0 otherwise. This way the discrepancy is only reduced in the final round $\log_2 n$.

Experiments. For each of the cycle, 2D-torus and hypercube, we consider two comparative experiments with an average-case initial load vector and a worst-case initial load vector.

The first experiment considers a “lightly loaded case”, where the theoretical results suggest that a small (i.e., constant or logarithmic) discrepancy is reached before the expected worst-case load balancing time, which are $\approx n^2$ for cycles and $\approx n$ for 2D-tori. The second experiment considers a “heavily loaded case”, where the theoretical results suggest that a small discrepancy is not reached faster than in the worst-case.

For cycles and 2D-tori we choose for the lightly loaded case $K = \sqrt{n}$ and for the heavily loaded case $K = n^2$. The experiments confirm the theoretical results in the sense that for both choices of K, we have a much quicker convergence of the discrepancy than in the worst case. However, the experiments also demonstrate that only in the lightly loaded case we reach a small discrepancy quickly, whereas in the heavily loaded case there is no big difference between worst-case and average-case if it comes to the time to reach a small discrepancy.

On the hypercube, our bounds on the discrepancy suggest a smaller K in comparison to the experiments on cycles and 2D-tori. That is why we choose $K = n^{1/4}$ in the lightly loaded case and $K = n$ in the heavily loaded case. With these adjustments of K in both cases, the experimental results of the hypercube are inline with the ones for the cycle and 2D-tori. More details including the plots can be found in the full version [8].

Acknowledgements.

We are grateful to the reviewers for their valuable comments. We are also thankful to Robert Elsässer and Yuchen Yang for helpful discussions.

References

1 Milton Abramowitz, Irene A Stegun, et al. Handbook of mathematical functions. *Applied mathematics series*, 55:62, 1966.

2 Dan Alistarh, Keren Censor-Hillel, and Nir Shavit. Are lock-free concurrent algorithms practically wait-free? *J. ACM*, 63(4):31:1–31:20, 2016.

3 Aris Anagnostopoulos, Adam Kirsch, and Eli Upfal. Load balancing in arbitrary network topologies with stochastic adversarial input. *SIAM J. Comput.*, 34(3):616–639, 2005.

4 Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Find your place: Simple distributed algorithms for community detection. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'17)*, pages 940–959, 2017.

5 Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

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

7 S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized Gossip Algorithms. *IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking*, 52:2508–2530, 2006.

8 Leran Cai and Thomas Sauerwald. Randomized load balancing on networks with stochastic inputs, 2017. `arXiv:1703.08702`.

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

10 Carl-Gustaf Esseen. *On the Liapounoff limit of error in the theory of probability*. Almqvist & Wiksell, 1942.

11 Bhaskar Ghosh, Frank Thomson Leighton, Bruce M. Maggs, S. Muthukrishnan, C. Greg Plaxton, Rajmohan Rajaraman, Andréa W. Richa, Robert Endre Tarjan, and David Zuckerman. Tight analyses of two local load balancing algorithms. *SIAM Journal on Computing*, 29(1):29–64, 1999.

12 Ashish Goel and Piotr Indyk. Stochastic load balancing and related problems. In *40th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 579–586, 1999.

13 Kenneth H. Huebner, Donald L. Dewhirst, Douglas E. Smith, and Ted G. Byrom. *The Finite Element Methods for Engineers*. Wiley, 2001.

14 David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

15 Russell Lyons. Asymptotic enumeration of spanning trees. *Combinatorics, Probability & Computing*, 14(4):491–522, 2005.

16 Marios Mavronicolas and Thomas Sauerwald. The impact of randomization in smoothing networks. *Distributed Computing*, 22(5-6):381–411, 2010.

17 S. Muthukrishnan and Bhaskar Ghosh. Dynamic load balancing by random matchings. *J. Comput. Syst. Sci.*, 53:357–370, 1996.

18 S. Muthukrishnan, Bhaskar Ghosh, and Martin H. Schultz. First- and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory Comput. Syst.*, 31:331–354, 1998.

19 A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM Journal on Computing*, 26(2):350–368, 1997.

20 Alessandro Panconesi and Aravind Srinivasan. Improved distributed algorithms for coloring and network decomposition problems. In *Proc. 24th Symp. Theory of Computing (STOC)*, pages 581–592, 1992.

21 Yuval Rabani, Alistair Sinclair, and Rolf Wanka. Local divergence of Markov chains and the analysis of iterative load balancing schemes. In *Proc. 39th Symp. Foundations of Computer Science (FOCS)*, pages 694–705, 1998.

22 Peter Sanders. Analysis of nearest neighbor load balancing algorithms for random loads. *Parallel Computing*, 25(8):1013–1033, 1999.

23 Atish Das Sarma, Danupon Nanongkai, Gopal Pandurangan, and Prasad Tetali. Distributed random walks. *J. ACM*, 60(1):2:1–2:31, 2013. URL: `http://doi.acm.org/10.1145/2432622.2432624`, `doi:10.1145/2432622.2432624`.

24 Thomas Sauerwald and He Sun. Tight bounds for randomized load balancing on arbitrary network topologies. In *Proc. 53rd Symp. Foundations of Computer Science (FOCS)*, pages 341–350, 2012.

25 A.J. Sinclair and M.R. Jerrum. Approximate counting, uninform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

26 Raghu Subramanian and Isaac D. Scherson. An analysis of diffusive load-balancing. In *Proc. 6th Symp. Parallelism in Algorithms and Architectures (SPAA)*, pages 220–225, 1994.

27 Sonesh Surana, Brighten Godfrey, Karthik Lakshminarayanan, Richard Karp, and Ion Stoica. Load balancing in dynamic structured peer-to-peer systems. *Performance Evaluation*, 63(3):217–240, 2006.

28 Roy D. Williams. Performance of dynamic load balancing algorithms for unstructured mesh calculations. *Concurrency: Practice and Experience*, 3(5):457–481, 1991.

29 Dongliang Zhanga, Changjun Jianga, and Shu Li. A fast adaptive load balancing method for parallel particle-based simulations. *Simulation Modelling Practice and Theory*, 17(6):1032–1042, 2009.