Journal of Computer and System Sciences 81 (2015) 159–185

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Randomized diffusion for indivisible loads

Petra Berenbrink [a], Colin Cooper [b], Tom Friedetzky [c], Tobias Friedrich [d,*], Thomas Sauerwald [e]

[a] Simon Fraser University, Burnaby, Canada
[b] King's College London, UK
[c] Durham University, UK
[d] Friedrich-Schiller-Universität Jena, Germany
[e] University of Cambridge, UK

ARTICLE INFO

Article history:
Received 3 December 2010
Received in revised form 31 March 2014
Accepted 17 April 2014
Available online 5 June 2014

Keywords:
Distributed computing
Load balancing
Diffusion
Randomized algorithms
Random walk

ABSTRACT

We present a new randomized diffusion-based algorithm for balancing indivisible tasks (tokens) on a network. Our aim is to minimize the discrepancy between the maximum and minimum load. The algorithm works as follows. Every vertex distributes its tokens as evenly as possible among its neighbors and itself. If this is not possible without splitting some tokens, the vertex redistributes its excess tokens among all its neighbors randomly (without replacement). In this paper we prove several upper bounds on the load discrepancy for general networks. These bounds depend on some expansion properties of the network, that is, the second largest eigenvalue, and a novel measure which we refer to as refined local divergence. We then apply these general bounds to obtain results for some specific networks. For constant-degree expanders and torus graphs, these yield exponential improvements on the discrepancy bounds. For hypercubes we obtain a polynomial improvement.

1. Introduction

During the last years, 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 also known to have applications to scheduling, routing, numerical computation, and finite element computations.

In this paper we analyze a very simple neighborhood-based load balancing algorithm. We assume that the processors are connected by an arbitrary d-regular network. In the beginning, every vertex has a certain number of tokens (load) and the goal is to distribute the tokens as evenly as possible. More precisely, we aim at minimizing the difference between the minimum and maximum load, which we call *discrepancy*.

Neighborhood-based load balancing algorithms normally operate in parallel steps. In each step, every processor is allowed to probe the load of all of its neighbors (*diffusion load balancing*), or to probe the load of one neighbor (*dimension exchange*). Then each processor has to decide how much load it will forward to the neighbors in question. Here we consider a very natural diffusion-based approach where every processor tries to balance the load locally. This means that along each edge, a load of *load-difference*$/(d+1)$ is sent to the vertex with less load. This is exactly the approach in the *continuous* diffusion model where tokens can be split arbitrarily. This method balances the load perfectly.

* Corresponding author.

http://dx.doi.org/10.1016/j.jcss.2014.04.027
0022-0000/

In contrast to continuous diffusion, we consider the (arguably more realistic [25]) case of *discrete* diffusion where tokens are indivisible. Quantifying by how much the integrality assumption decreases the efficiency of load balancing is an interesting question and has been posed by many authors (e.g., [12,16,20,23–25]).

In the common *edge-oriented* view of e.g. [13,15,24], for each edge one has to decide between transferring either $\lceil \textit{load-difference}/(d+1) \rceil$ or $\lfloor \textit{load-difference}/(d+1) \rfloor$ tokens (referred to as rounding up or rounding down). Rounding up results in a load balancing algorithm that keeps sending tokens back and forth between processors with a small load difference. Another disadvantage is that the approach can generate "negative loads" for vertices with only a few tokens. On the other hand, always rounding down cannot balance better than $d \cdot \mathrm{diam}(G)$, where $\mathrm{diam}(G)$ denotes the diameter of the underlying graph G. To overcome these problems we adopt a *vertex-oriented* view in this paper. We present a randomized diffusion load balancing algorithm where the vertices (not edges) decide randomly how much they are sending.

1.1. Related work

Due to the vast amount of literature on load balancing, we consider only previous work dealing with diffusion load balancing, or randomized algorithms for neighborhood-based load balancing. We do not consider the dimension exchange model in general, or token distribution model where only one token can be sent to a neighbor per step.

Continuous diffusion. The diffusion model was first studied by Cybenko [5] and, independently, Boillat [3]. Cybenko [5] (see also [23,25]) shows 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 $\mathbf{P}$ with $\mathbf{P}_{ij} = 1/(d+1)$ if $\{i, j\} \in E$. Subramanian and Scherson [25] observe similar relations between convergence time and certain properties of the underlying network like electrical and fluid conductance.

Muthukrishnan et al. [23] refer to the above diffusion model as the *first order scheme* and generalize it to the so called *second order scheme*. Here the load transferred over an edge (i, j) in step t does not only depend on the load difference of i and j, but also on the amount of load transferred over the edge in step $t-1$. Diekmann et al. [7] extend the idea of [23] and propose a general framework to analyze the convergence behavior of a wide range of diffusion type methods.

Discrete diffusion. In order to approximate the idealized process by a discrete process with indivisible load, Rabani et al. [24] consider a diffusion algorithm (called RSW algorithm in the following) which always rounds down the indivisible load on each edge. To quantify the deviation of the discrete load from the idealized process, they propose a natural measure, the *local divergence* Ψ_1. The local divergence measures the sum of load differences across all edges in the network, aggregated over time. They give a general bound on the divergence in terms of $\lambda_{\max}$, which denotes the absolute value of the second largest eigenvalue of the diffusion matrix $\mathbf{P}$. By a more careful analysis, they also get an improved upper bound on Ψ_1 for tori, resulting in a tight bound on the discrepancy achieved by their algorithm.

Discrete load balancing via random walks. Elsässer et al. [10–12] proposed an algorithm based on random walks. They show that after $\mathcal{O}(\log(Kn)/(1-\lambda_{\max}))$ steps, the maximum load is at most the average load plus a constant [11]. In comparison to our algorithm, their algorithm is more complicated and different from the usual diffusion framework. For example, vertices require an estimate of n and have to compute the average load during the balancing procedure. Moreover, the final stage uses concurrent random walks (representing tokens) to reduce the maximum load. In this stage, the load transfer along an edge may be much smaller (or higher) than $\textit{load-difference}/(d+1)$.

Discrete neighborhood load balancing with randomization. In [13] the last two authors analyze a randomized version of the dimension-exchange algorithm using randomly generated or deterministic matchings. In their algorithm, the decision to round up or down is randomized. For detailed results see Table 1. Note that in their case every node exchanges load with at most one neighbor. This is typically much easier to analyze than diffusion algorithms.

Friedrich et al. [15] analyze a deterministic modification of the standard diffusion algorithm for hypercubes and constant-dimensional tori. The idea is that each edge keeps tracks of its own rounding errors. In each step an edge's decision to round up or down is done such that the sum of its rounding errors is minimized. Again, the detailed results can be found in Table 1. Friedrich et al. [15] also consider a randomized version of the diffusion algorithm. Their approach is edge-based. Edges decide independently at random whether to round up or down. The probabilities are chosen such that, in expectation, the behavior of the continuous diffusion algorithm is mimicked. They present a general upper bound for their approach in terms of $\lambda_{\max}$. Note that both algorithms in [15] may generate negative load due to the edge-based rounding.

Source of inspiration. We wish to point out that our work was inspired by recent combinatorial results regarding so-called *rotor-router walks* [4,8]. Unlike in a random walk, in a rotor-router walk each vertex serves its neighbors in a fixed order. The resulting (completely deterministic) walk nevertheless closely resembles a random walk in several respects. Similarly, one can say that in each round of our load-balancing algorithm a vertex chooses a random order of its neighbors (and itself) and sends around all its tokens in this order in a round-robin fashion.

1.2. Our contribution

Algorithm. We consider a vertex-based randomized diffusion algorithm for the discrete model with indivisible tokens. Let d be the degree of the (regular) network and let X_i be the load of vertex i. Our algorithm works as follows. First, vertex i sends $\lfloor X_i/(d+1) \rfloor$ tokens to each neighbor and keeps the same amount of tokens for itself. Then the remaining

Table 1
Discrepancy of neighborhood load balancing after $\tau(G,K)=\Theta(\log(Kn)/1-\lambda_{\max})$ rounds.

Graph class	FS [13]	RSW [24]	FGS [15] det.	FGS [14] rand.	Our algorithm
d-reg. graph	$\mathcal{O}(\Psi_2(G)\sqrt{\log n})$	$\mathcal{O}(\Psi_1(G))$	–	–	$\mathcal{O}(\Upsilon_2(G)\sqrt{d\log n})$
	$\mathcal{O}(\frac{d\log\log n}{1-\lambda_{\max}})$	–	–	$\mathcal{O}(\frac{d\log\log n}{1-\lambda_{\max}})$	$\mathcal{O}(\frac{d\log\log n}{1-\lambda_{\max}})$
	$\mathcal{O}(\sqrt{\frac{d\log n}{1-\lambda_{\max}}})$	$\mathcal{O}(\frac{d\log n}{1-\lambda_{\max}})$	–	–	$\mathcal{O}(d\sqrt{\log n}+\sqrt{\frac{d\log n\log d}{1-\lambda_{\max}}})$
d-reg. expander	$\mathcal{O}(d\log\log n)$	$\mathcal{O}(d\log n)$	–	$\mathcal{O}(d\log\log n)$	$\mathcal{O}(d\log\log n)$
Hypercube	$\mathcal{O}(\log^2 n)$	$\Theta(\log^2 n)$	$\mathcal{O}(\log^{3/2} n)$	$\mathcal{O}(\log^2 n\log\log n)$	$\mathcal{O}(\log n)$
r-dim. torus	$\mathcal{O}(n^{1/(2r)}\sqrt{\log n})$	$\Theta(n^{1/r})$	$\mathcal{O}(1)$	$\mathcal{O}(n^{1/r}\log\log n)$	$\mathcal{O}(\sqrt{\log n})$
Properties	**FS [13]**	**RSW [24]**	**FGS [15] det.**	**FGS [14] rand.**	**Our algorithm**
Diffusion	✗	✓	✓	✓	✓
No neg. load	✓	✓	✗	✗	✓

$X_i-(d+1)\lfloor X_i/(d+1)\rfloor$ tokens (called *excess tokens*) are randomly distributed (without replacement) among vertex i and its d neighbors.

Results. To state our results formally, we let $\tau(G,K)=\mathcal{O}(\log(Kn)/(1-\lambda_{\max}))$ be the number of steps after which the continuous process achieves a constant discrepancy for any initial load distribution with discrepancy K (cf. Fact 2.5, [24]). All our bounds on the discrepancy are independent of the initial load vector, and hold with high probability (w.h.p.), i.e., with probability at least $1-n^{-\Omega(1)}$.

Theorem 1.1. *Let G be an arbitrary d-regular graph and let K be the initial discrepancy. Then the discrepancy after $\tau(G,K)=\mathcal{O}(\log(Kn)/(1-\lambda_{\max}))$ rounds is w.h.p. at most*

(1) $\mathcal{O}(\Upsilon_2(G)\sqrt{d\log n})$,
(2) $\mathcal{O}(d+\sqrt{d\log(n)((\Upsilon_2(G))^2-d)})$,
(3) $\mathcal{O}(d\frac{\log\log n}{1-\lambda_{\max}})$.

The role of $\Upsilon_2(G)$ is similar to the local divergence $\Psi_1(G)$ used in [24] (cf. Definitions 2.7 and 2.8). $\Upsilon_2(G)$ accounts for the more balanced reallocation of the excess tokens due to our randomized approach and is much smaller than $\Psi_1(G)$, i.e., $\Upsilon_2(G)\le\sqrt{\Psi_1(G)}$ for any graph G.

The next theorem provides more specific bounds on the discrepancy. It is derived by first bounding $\Upsilon_2(G)$ and then applying Theorem 1.1.

Theorem 1.2. *The following upper bounds on the discrepancy after $\tau(G,K)=\mathcal{O}(\log(Kn)/(1-\lambda_{\max}))$ rounds hold w.h.p.*

(1) $\mathcal{O}(d\sqrt{\log n}+\sqrt{\frac{d\log n\log d}{1-\lambda_{\max}}})$,
(2) *d-regular expander:* $\mathcal{O}(d\log\log n)$,
(3) *r-dim. torus, $r=\mathcal{O}(1)$:* $\mathcal{O}(\sqrt{\log n})$,
(4) *Hypercube:* $\mathcal{O}(\log n)$.

Let us compare our results to the RSW algorithm [24] as it is also very natural, considers diffusion and avoids negative loads. More comparisons can be found in Table 1. For d-regular expanders, [24] proves a discrepancy bound of $\mathcal{O}(d\log n)$ after $\tau(G,K)$ rounds. This is almost tight, as $d\cdot\operatorname{diam}(G)$ is a simple lower bound for the RSW algorithm. Hence for small d, we obtain an exponential improvement in terms of the discrepancy.

For the r-dimensional torus graph, [24, Theorem 8] proved a bound of $\mathcal{O}(n^{1/r})$ on the discrepancy after $\tau(G,K)$ rounds. This is tight due to the lower bound of $\operatorname{diam}(G)$. Again, our new result represents an exponential improvement.

For the hypercube with n vertices, [24, Theorem 4] implies a discrepancy bound of $\mathcal{O}(\log^3 n)$ after $\tau(G,K)$ rounds. The techniques used to analyze our new algorithm can be also used to prove a tight bound of $\Theta(\log^2 n)$ on the discrepancy for the RSW algorithm. For our new algorithm, we obtain a smaller bound of $\mathcal{O}(\log n)$ on the discrepancy.

Techniques. The key ingredient of the analyses in [13,15,24] is "*an appropriate edge-oriented view of the rounding errors in each balancing step, which allows them to be handled independently*" (as stated by Rabani et al. [24]). The problem with vertex-oriented algorithms are the dependencies between the rounding results for edges incident to the same vertex. To deal with these dependencies we use a different analysis compared to [13,15], based on martingale tail estimates. The other main technical contribution is the use of the new parameter $\Upsilon_2(G)$ (Definition 2.8) as opposed to the local divergence $\Psi_1(G)$ as used in [24].

2. Algorithms and notation

We use standard graph-theoretical notation. We only consider graphs $G=(V,E)$ that are connected, undirected, d-regular and simple. The n vertices of G are given by $[n]:=\{1,2,\dots,n\}$. The neighborhood of a vertex i is denoted by $N(i)$. For a pair of vertices $i,j\in V(G)$, let $\operatorname{dist}(i,j)$ be the length of a shortest path between i and j, and $\operatorname{diam}(G)$ be the diameter of G. $[i:j]$ refers to an edge $\{i,j\}\in E$ with $i<j$. This notation will be useful to have a unique representative for each edge $\{i,j\}\in E$. Every vertex in the graph has a certain amount of load items (tokens). We assume that the load is indivisible and each token is of unit-size.

We denote by $\mathbf{P}$ the *transition matrix*, i.e., $\mathbf{P}_{i,j}=\frac{1}{d+1}$ if $\{i,j\}\in E$ or $i=j$, and $\mathbf{P}_{i,j}=0$ otherwise. We will often use $\mathbf{P}^t$ which means that we raise the matrix $\mathbf{P}$ to the power of t. Note that $\mathbf{P}^t_{i,j}$ can be also seen as the probability for a random walk being located at vertex j at step t, when having started from vertex i.

For the estimation of the convergence of our processes, the absolute value of the second largest eigenvalues of $\mathbf{P}$ plays a crucial role. Let us denote the eigenvalues of $\mathbf{P}$ by $1=\lambda_1\ge\lambda_2\ge\lambda_3\ge\dots\ge\lambda_n>-1$ and define $\lambda_{\max}:=\max\{\lambda_2,|\lambda_n|\}$.

Lemma 2.1. *For any d-regular graph G, $1/(1-\lambda_{\max}(G))\le n^3$.*

Proof. Let $d=\mu_1\ge\mu_2\ge\dots\ge\mu_n$ be the n eigenvalues of the adjacency matrix of the d-regular graph G. As shown in [19, Problem 11.29], $d-\mu_2\ge\frac{1}{\operatorname{diam}(G)n}$. Hence, $\lambda_2=\mu_2/(d+1)\le\frac{d}{d+1}-\frac{1}{\operatorname{diam}(G)(d+1)n}\le 1-\frac{1}{n^3}$. Moreover, it is proved in [1, Theorem 1.1] that $\mu_n\ge -d+\frac{1}{(\operatorname{diam}(G)+1)n}$ and therefore $\lambda_n\ge-\frac{d}{d+1}+\frac{1}{(\operatorname{diam}(G)+1)(d+1)n}\ge -1+\frac{1}{n^3}$. □

We also need the following elementary inequalities.

Lemma 2.2. *The following two inequalities hold.*

(1) *For any integer m and any non-negative numbers $x_1,x_2,\dots,x_n\in\mathbb{R}^+$, $(x_1+x_2+\dots+x_n)^m\le n^{m-1}\cdot(x_1^m+x_2^m+\dots+x_n^m)$.*
(2) *Moreover, for any three numbers $x,y,z\in\mathbb{R}$, $(x-y)^2\le 2((x-z)^2+(y-z)^2)$.*

Proof. For the first statement, recall that $x\mapsto x^m$ is a convex function for non-negative x. Hence,

$$\left(\frac{x_1+x_2+\dots+x_n}{n}\right)^m\le\frac{x_1^m+x_2^m+\dots+x_n^m}{n}.$$

Rearranging yields the first claim. For the second statement, we apply the first statement to obtain $(x-y)^2=(x-z+z-y)^2\le 2((x-z)^2+(y-z)^2)$. □

2.1. Our discrete process

Our balancing procedure proceeds in rounds $1,2,\dots$. Fix a vertex i at some step and let X_i be the current load of this vertex. Then, i sends $\lfloor X_i/(d+1)\rfloor$ tokens to each of its neighbors and keeps $\lfloor X_i/(d+1)\rfloor$ tokens for itself. The remaining $X_i-(d+1)\lfloor X_i/(d+1)\rfloor\in[0,d]$ *excess-tokens* are distributed randomly (without replacement) among i and its d neighbors.

To describe our processes more formally, we first present our notation that is based on [24]. For any round t, let $X^{(t)}$ be the n-dimensional load-vector at (the end of) step t (load vectors are always regarded as column-vectors here). The *discrepancy* of the load vector $X^{(t)}$ at step t is defined as $\max_{i,j\in[n]}|X_i^{(t)}-X_j^{(t)}|$.

For each edge $\{i,j\}\in E$ we define a random variable $Z_{i,j}^{(t)}$ which is one if i sends an excess token to j at step t, and $Z_{i,j}^{(t)}$ is zero otherwise. Similarly, let $Z_{i,i}^{(t)}$ be one if i keeps an excess token for itself, and zero otherwise. Note that each $Z_{i,j}^{(t)}$ with $j\in N(i)\cup\{i\}$ is a Bernoulli random variable with

$$\mathbf{Pr}\left[Z_{i,j}^t=1\right]=\frac{X_i^{(t-1)}}{d+1}-\left\lfloor\frac{X_i^{(t-1)}}{d+1}\right\rfloor.$$

Additionally, the number of excess tokens sent out by i satisfies

$$Z_{i,i}^{(t)}+\sum_{j:\{i,j\}\in E}Z_{i,j}^{(t)}=X_i^{(t-1)}-(d+1)\left\lfloor\frac{X_i^{(t-1)}}{d+1}\right\rfloor. \tag{2.1}$$

Note that $Z_{i,j}$ and $Z_{j,i}$ are independent for $i\ne j$. Now we can describe the discrete process as follows,

$$X_i^{(t)}=\left\lfloor\frac{X_i^{(t-1)}}{d+1}\right\rfloor+Z_{i,i}^{(t)}+\sum_{j:\{i,j\}\in E}\left(\left\lfloor\frac{X_j^{(t-1)}}{d+1}\right\rfloor+Z_{j,i}^{(t)}\right). \tag{2.2}$$

2.2. The continuous process

We also look at the corresponding *continuous process*, where the load is arbitrarily divisible. The load vector of this process is denoted by $\xi^{(t)}$ in round t. To analyze $X^{(t)}$, we shall bound its deviation from $\xi^{(t)}$ and use the fact that the evolution of $\xi^{(t)}$ in t is well-understood by Markov chain theory. The reason is that $\xi^{(t)}$ is given by the recurrence $\xi^{(t)} = \xi^{(t-1)}\mathbf{P}$, which results in $\xi^{(t)} = \xi^{(0)}\mathbf{P}^t$. Alternatively, we can write this as

$$\xi_i^{(t)} = \xi_i^{(t-1)} + \sum_{j:\{i,j\}\in E} \frac{\xi_j^{(t-1)} - \xi_i^{(t-1)}}{d+1}. \tag{2.3}$$

We define the average load as $\bar{\xi} := \sum_{i=1}^{n} \xi_i^{(0)}/n$. We will use the following result that bounds the load difference of the vertices to the average load in step t for the continuous process.

Lemma 2.3. *(See [23, Lemma 1].) Let $G = (V, E)$ be an arbitrary connected graph. Then for any initial vector $\xi^{(0)}$ and time step $t \geq 0$, $\sum_{i=1}^{n}(\xi_i^{(t)} - \bar{\xi})^2 \leq \lambda_{\max}^{2t} \sum_{i=1}^{n}(\xi_i^{(0)} - \bar{\xi})^2$.*

We will use the following immediate consequence of this lemma.

Corollary 2.4. *Let $G = (V, E)$ be any graph. Then for any time step $t \geq 0$ and any vertex $k \in V$, $\sum_{i=1}^{n}(\mathbf{P}_{i,k}^t - \frac{1}{n})^2 \leq \lambda_{\max}^{2t}$.*

Proof. Let $\xi^{(0)}$ be the unit-vector with 1 at entry k and 0 otherwise. Observe that $\bar{\xi} = 1/n$ and $\xi_i^{(t)} = \sum_{j=1}^{n} \xi_j^{(0)} \mathbf{P}_{j,i}^t = \mathbf{P}_{k,i}^t = \mathbf{P}_{i,k}^t$. Hence applying Lemma 2.3 leads to

$$\sum_{i=1}^{n}\left(\mathbf{P}_{i,k}^t - \frac{1}{n}\right)^2 \leq \lambda_{\max}^{2t} \sum_{i=1}^{n}\left(\xi_i^{(0)} - \frac{1}{n}\right)^2 = \lambda_{\max}^{2t}\left(1 - \frac{1}{n}\right) \leq \lambda_{\max}^{2t}. \qquad \square$$

The following well-known result bounds the discrepancy of ξ.

Fact 2.5. *(See [24, Theorem 1].) Let G be any graph with n vertices. For the continuous process, the discrepancy is reduced to $\epsilon > 0$ after $\frac{2}{1-\lambda_{\max}} \ln(\frac{Kn^2}{\epsilon})$ steps, where K is the discrepancy of the initial load vector.*

By $\tau(G, K)$ we denote the number of steps required for the continuous process to achieve a discrepancy of 1 for any initial load vector with discrepancy K. Fact 2.5 implies that $\tau(G, K) = \mathcal{O}(\frac{\log(Kn)}{1-\lambda_{\max}})$.

2.3. Difference between continuous process and discrete process

To obtain results for the discrete process, we upper bound the deviation between the discrete and continuous process at a step t when initialized with the same load vector. The step t is chosen just large enough to ensure that continuous process has achieved a discrepancy of at most 1 for every load vector with initial discrepancy K (cf. Fact 2.5). Hence, the discrepancy of the discrete process is upper bounded by the deviation between the discrete and continuous process (plus 1).

Similar to [13,15,24], we first express the difference between the discrete and idealized process by a sum of weighted rounding errors (Eq. (3.1)). In this sum, the rounding errors are weighted by powers of the transition probabilities. In contrast to [13,15,24], the rounding errors (of the same time step) are not independent for all edges. This is due to our vertex-based approach and complicates the analysis.

To find a recursion for the discrete process, similar to Eq. (2.3) for the continuous process, plug Eq. (2.1) into Eq. (2.2) to obtain

$$\begin{aligned} X_i^{(t)} &= \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor - \left(\sum_{j:\{i,j\}\in E} Z_{i,j}^{(t)}\right) + X_i^{(t-1)} - (d+1)\left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor + \sum_{j:\{i,j\}\in E}\left(\left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor + Z_{j,i}^{(t)}\right) \\ &= X_i^{(t-1)} + \sum_{j:\{i,j\}\in E}\left(\left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor - \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor + Z_{j,i}^{(t)} - Z_{i,j}^{(t)}\right). \end{aligned} \tag{2.4}$$

Comparing Eq. (2.4) to (2.3) motivates the definition of a random variable $\Delta_{i,j}^{(t)}$ for the rounding error made by the vertex i on the edge from i to j at step t:

$$\Delta_{i,j}^{(t)} := -\frac{X_j^{(t-1)}}{d+1} + \frac{X_i^{(t-1)}}{d+1} + \left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor - \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor + Z_{j,i}^{(t)} - Z_{i,j}^{(t)}. \tag{2.5}$$

This allows us to write

$$X_i^{(t)} = X_i^{(t-1)} + \sum_{j:\{i,j\}\in E} \frac{X_j^{(t-1)} - X_i^{(t-1)}}{d+1} + \Delta_{i,j}^{(t)}. \tag{2.6}$$

Now we state some basic properties of the rounding errors.

Lemma 2.6. *Let $G = (V, E)$ be an arbitrary connected graph.*

(1) *For every $\{i, j\} \in E$ and time step t, $\Delta_{i,j}^{(t)} = -\Delta_{j,i}^{(t)}$, $|\Delta_{i,j}^{(t)}| \le 2$ and $\mathbf{E}[\Delta_{i,j}^{(t)}] = 0$.*
(2) *Consider two vertex-disjoint edges $\{i, j\} \in E$ and $\{k, \ell\} \in E$ and assume that $X^{(t-1)}$ is fixed. Then $\Delta_{i,j}^{(t)}$ and $\Delta_{k,\ell}^{(t)}$ are independent.*

Proof. (1): The antisymmetry of Δ follows directly by the definition. The absolute value of Δ can be bounded as follows:

$$\left|\Delta_{i,j}^{(t)}\right| \le \left| \underbrace{\frac{X_i^{(t-1)}}{d+1} - \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor}_{\in[0,1)} - \left(\underbrace{\frac{X_j^{(t-1)}}{d+1} - \left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor}_{\in[0,1)} \right) \right| + \big| \underbrace{Z_{j,i}^{(t)} - Z_{i,j}^{(t)}}_{\in[-1,1]} \big| \le 1 + 1 = 2.$$

Finally, linearity of expectations and the definition of $Z_{i,j}^{(t)}$ and $Z_{j,i}^{(t)}$ gives

$$\mathbf{E}\big[\Delta_{i,j}^{(t)}\big] = -\frac{X_j^{(t-1)}}{d+1} + \frac{X_i^{(t-1)}}{d+1} + \left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor - \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor + \frac{X_j^{(t-1)}}{d+1} - \left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor - \frac{X_i^{(t-1)}}{d+1} + \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor = 0.$$

(2): Recall that by assumption, the load vector $X^{(t-1)}$ is fixed. By the definition of $\Delta_{i,j}^{(t)}$ in Eq. (2.5), $\Delta_{i,j}^{(t)}$ depends only on the random variables $Z_{i,j}^{(t)}$ and $Z_{j,i}^{(t)}$. In addition, Eq. (2.1) describes a relation between $Z_{i,j}^{(t)}$ and all other $Z_{i,\ell}^{(t)}$ with $\{i, \ell\} \in E$ (and the same holds for $Z_{j,i}^{(t)}$ as well). Hence $\Delta_{i,j}^{(t)}$ depends on

$$Z_{i,j}^{(t)}, \qquad Z_{j,i}^{(t)}, \qquad \text{all } Z_{i,s}^{(t)} \quad \text{with } \{i,s\} \in E, \quad \text{and} \quad \text{all } Z_{j,s}^{(t)} \quad \text{with } \{j,s\} \in E.$$

Similarly, $\Delta_{k,\ell}^{(t)}$ depends on

$$Z_{k,\ell}^{(t)}, \qquad Z_{\ell,k}^{(t)}, \qquad \text{all } Z_{k,s}^{(t)} \quad \text{with } \{k,s\} \in E, \quad \text{and} \quad \text{all } Z_{\ell,s}^{(t)} \quad \text{with } \{\ell,s\} \in E.$$

Hence if the edges $\{i, j\} \in E$ and $\{k, \ell\} \in E$ are vertex-disjoint, the set of random variables in the two sets above are disjoint. Hence, $\Delta_{i,j}^{(t)}$ and $\Delta_{k,\ell}^{(t)}$ are independent. □

We now continue by returning to Eq. (2.6). For any vertex $i \in V$ and step t, let us define an error vector $\Delta^{(t)}$ with $\Delta_i^{(t)} := \sum_{j:\{i,j\}\in E} \Delta_{i,j}^{(t)}$. With this notation we have, $X^{(t)} = X^{(t-1)}\mathbf{P} + \Delta^{(t)}$.

Solving this recursion (see [24]) and setting $\xi^{(0)} = X^{(0)}$ results in

$$X^{(t)} = X^{(0)}\mathbf{P}^t + \sum_{s=0}^{t-1} \Delta^{(t-s)}\mathbf{P}^s = \xi^{(t)} + \sum_{s=0}^{t-1} \Delta^{(t-s)}\mathbf{P}^s,$$

where $\mathbf{P}^0$ is the $n \times n$-identity matrix. Hence, for any vertex $k \in V$

$$\begin{aligned} X_k^{(t)} - \xi_k^{(t)} &= \sum_{s=0}^{t-1}\sum_{i=1}^{n} \Delta_i^{(t-s)}\mathbf{P}_{i,k}^s = \sum_{s=0}^{t-1}\sum_{i=1}^{n}\sum_{j:\{i,j\}\in E} \Delta_{i,j}^{(t-s)}\mathbf{P}_{i,k}^s \\ &= \sum_{s=0}^{t-1}\sum_{[i:j]\in E} \Delta_{i,j}^{(t-s)}\big(\mathbf{P}_{i,k}^s - \mathbf{P}_{j,k}^s\big), \end{aligned} \tag{2.7}$$

where the last equality uses $\Delta_{i,j}^{(t-s)} = -\Delta_{j,i}^{(t-s)}$ shown in Lemma 2.6(1).

2.4. Definition of local divergence and refined local divergence

Eq. (2.7) and $|\Delta_{i,j}^{(t-s)}| \le 2$ suggests to consider $\sum_{s=0}^{t-1}\sum_{[i:j]\in E}|\mathbf{P}_{i,k}^{s}-\mathbf{P}_{j,k}^{s}|$, which is a parameter that only depends on the graph G, but not on the behavior of the load balancing algorithm. We first adjust a definition from [13] that generalizes the original definition of local divergence from [24] for $p=1$.

Definition 2.7. (See [13,24].) For any $p \in \mathbb{N}_{>0}$, the local p-divergence of a graph $G=(V,E)$ is

$$\Psi_p(G) := \max_{k\in V}\left(\sum_{t=0}^{\infty}\sum_{[i:j]\in E}\left|\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right|^{p}\right)^{1/p}.$$

Note that $\Psi_2(G)^2 \le \Psi_1(G)$, since $|\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}| \le 1$ for all t,i,k. As pointed out in [24], "$\Psi_1(G)$ *is a natural quantity that measures the sum of load differences across all edges in the network, aggregated over time (and suitably normalized) which may be of independent interest*". Here, we will mainly consider a natural extension of $\Psi_1(G)$ to the ℓ_2-norm, $\Psi_2(G)$ and its sibling $\Upsilon_2(G)$ which is defined below.

Definition 2.8. For any $p \in \mathbb{N}_{>0}$, the refined local p-divergence of a graph $G=(V,E)$ is

$$\Upsilon_p(G) := \max_{k\in V}\left(\frac{1}{2}\sum_{t=0}^{\infty}\sum_{i=1}^{n}\max_{j\in N(i)}\left|\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right|^{p}\right)^{1/p}.$$

Note that $\Upsilon_p(G) \le \Psi_p(G)$, since for each $\{i,j\}\in E(G)$ the term $|\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}|^p$ appears once in $\Psi_p(G)$ and at most twice in $\Upsilon_p(G)$ (this also explains why we include the factor of $1/2$ in the definition of $\Upsilon_p(G)$).

The analysis of our algorithm will be based on $\Upsilon_2(G)$. The following lemma shows that only early time steps can have a significant contribution to $\Upsilon_2(G)$.

Lemma 2.9. *Let $G=(V,E)$ be any graph and define $\kappa := (4\ln n)/(1-\lambda_{\max})$. Then for an arbitrary vertex $k\in V$,*

$$\sum_{t=\kappa}^{\infty}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2 = \mathcal{O}(1).$$

Proof. Using Lemma 2.2, we get

$$\begin{aligned}\sum_{t=\kappa}^{\infty}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2 &\le 2\cdot\sum_{t=\kappa}^{\infty}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\frac{1}{n}\right)^2+\left(\mathbf{P}_{j,k}^{t}-\frac{1}{n}\right)^2\\ &= 2d\cdot\sum_{t=\kappa}^{\infty}\sum_{i\in V}\left(\mathbf{P}_{i,k}^{t}-\frac{1}{n}\right)^2\\ &\le 2d\cdot\sum_{t=\kappa}^{\infty}\lambda_{\max}^{2t},\end{aligned}$$

where the last inequality is due to Corollary 2.4. Using the fact that $x^{1/(1-x)} \le 1/e$ for $x\in[0,1)$ and the inequality $1-\lambda_{\max} \le n^{-3}$ (Lemma 2.1), we can bound this term as follows,

$$2d\cdot\sum_{t=\kappa}^{\infty}\lambda_{\max}^{2t} = 2d\cdot\frac{\lambda_{\max}^{2\kappa}}{1-(\lambda_{\max})^2} \le 2d\cdot\frac{\lambda_{\max}^{(8\ln n)/(1-\lambda_{\max})}}{1-\lambda_{\max}} \le 2d\cdot\frac{e^{-8\ln n}}{n^{-3}} = \mathcal{O}(1). \qquad \square$$

3. Proof of Theorem 1.1

We now bound the discrepancy of our discrete process in terms of the local divergence $\Upsilon_2(G)$. We do this by upper bounding the deviation between the discrete and the continuous process. A similar approach was used in Rabani et al. [24] who bounded this deviation in terms of $\Psi_1(G)$. They showed that reducing the initial discrepancy from K to $\mathcal{O}(\Psi_1(G))$ can be achieved within $\mathcal{O}(\log(Kn)/(1-\lambda_{\max}))$ steps for any initial load vector. However, it turns out that our randomized process can be bounded in terms of $\Upsilon_2(G)$. Note that $\Upsilon_2(G)$ is in general much smaller than $\Upsilon_1(G)$ (or $\Psi_1(G)$) (cf. the remarks after Definition 2.7). We will use the following concentration inequality for martingales, which is commonly known as the "method of average bounded differences".

Theorem 3.1. *(See [9, p. 83].) Let $Y_1, \dots, Y_n$ be an arbitrary set of random variables and let f be a function of these random variables satisfying the property that for each $\ell \in [n]$, there is a non-negative c_ℓ such that*

$$\left|\mathbf{E}[f \mid Y_\ell, Y_{\ell-1}, \dots, Y_1] - \mathbf{E}[f \mid Y_{\ell-1}, \dots, Y_1]\right| \le c_\ell.$$

Then for any $\delta > 0$,

$$\mathbf{Pr}\left[\left|f - \mathbf{E}[f]\right| > \delta\right] \le 2\exp\left(-\frac{\delta^2}{2c}\right),$$

where $c := \sum_{\ell=1}^n c_\ell^2$.

Proof of Theorem 1.1. *Proof of the first statement.* Let us now fix a vertex $k \in V$ and a time step t. Recall from Eq. (2.7) that

$$X_k^{(t)} - \xi_k^{(t)} = \sum_{s=0}^{t-1} \sum_{[i:j]\in E} \Delta_{i,j}^{(t-s)} \left(\mathbf{P}_{i,k}^s - \mathbf{P}_{j,k}^s\right) = \sum_{s=1}^{t} \sum_{[i:j]\in E} \Delta_{i,j}^{(s)} \left(\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right). \tag{3.1}$$

Consider the random variable $X_k^{(t)} - \xi_k^{(t)}$. By Lemma 2.6, $\mathbf{E}[X_k^{(t)} - \xi_k^{(t)}] = 0$. Our goal is to apply the martingale tail estimate from Theorem 3.1 to $f_k := X_k^{(t)} - \xi_k^{(t)}$. We first rewrite f_k,

$$\begin{aligned} f_k &= \sum_{s=1}^{t} \sum_{[i:j]\in E} \Delta_{i,j}^{(s)} \left(\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right) \\ &= \sum_{s=1}^{t} \sum_{[i:j]\in E} \left(-\frac{X_j^{(t-1)}}{d+1} + \frac{X_i^{(t-1)}}{d+1} + \left\lfloor \frac{X_j^{(t-1)}}{d+1} \right\rfloor - \left\lfloor \frac{X_i^{(t-1)}}{d+1} \right\rfloor + Z_{j,i}^{(t)} - Z_{i,j}^{(t)}\right) \cdot \left(\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right), \end{aligned}$$

where the last equality follows by the definition of $\Delta_{i,j}^{(s)}$.

We observe that for a fixed load vector $X^{(0)}$ the function f_k depends only on the randomly chosen destinations of the excess tokens. There are t time steps, n nodes, and at most d excess tokens per node per time step. We describe these random choices by a sequence of $t \cdot n \cdot d$ random variables, $Y_1, Y_2, \dots, Y_{tnd}$. For any ℓ with $1 \le \ell \le tnd$, let $(s, i, r) \in [t] \times [n] \times [d]$ be such that $\ell = (s-1)nd + (i-1)d + r$ (note that (s, i, r) is the ℓ-th largest element in an increasing lexicographic ordering of $[t] \times [n] \times [d]$). Then Y_ℓ refers to the destination of the r-th excess token of vertex i at step s (if there is one). More precisely,

$$Y_\ell := \begin{cases} j & \text{if } r \le X_i^{(s-1)} - (d+1)\lfloor \frac{X_i^{(s-1)}}{d+1} \rfloor \text{ and the } r\text{-th excess token of vertex } i \text{ at step } s \text{ is sent to } j, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $Y_\ell \in N(i) \cup \{i\}$. In order to apply Theorem 3.1, we have to upper bound

$$\left|\mathbf{E}[f_k \mid Y_\ell, Y_{\ell-1}, \dots, Y_1] - \mathbf{E}[f_k \mid Y_{\ell-1}, \dots, Y_1]\right|. \tag{3.2}$$

Consider a fixed ℓ that corresponds to (s_1, i_1, r_1) in the lexicographic ordering.

To bound Eq. (3.2), we use Eq. (3.1) to get

$$\begin{aligned} &\left|\mathbf{E}[f_k \mid Y_\ell, Y_{\ell-1}, \dots, Y_1] - \mathbf{E}[f_k \mid Y_{\ell-1}, \dots, Y_1]\right| \\ &\quad \le \sum_{s=1}^{t} \sum_{[i:j]\in E} \left|\mathbf{E}\left[\Delta_{i,j}^{(s)} \,\middle|\, Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[\Delta_{i,j}^{(s)} \,\middle|\, Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|. \end{aligned}$$

In the remainder of the proof we now split the sum over s into the three parts $1 \le s < s_1$, $s = s_1$, and $s_1 < s \le t$. We prove that the parts $s < s_1$ and $s > s_1$ both equal zero while the part $s = s_1$ is upper bounded by $2 \cdot \max_{j \in N(i_1)} |\mathbf{P}_{i_1,k}^{t-s_1} - \mathbf{P}_{j,k}^{t-s_1}|$.

$s < s_1$: For every $\{i, j\} \in E$, $\Delta_{i,j}^{(s)}$ is already determined by $Y_{\ell-1}, \dots, Y_1$. Hence,

$$\sum_{s=1}^{s_1-1} \sum_{[i:j]\in E} \left|\mathbf{E}\left[\Delta_{i,j}^{(s)} \,\middle|\, Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[\Delta_{i,j}^{(s)} \,\middle|\, Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right| = 0. \tag{3.3}$$

$s = s_1$: This is the most involved case due to the dependencies among $\{\Delta_{i,j}^{(s)}: \{i,j\} \in E\}$.

$$\sum_{[i:j]\in E} \left|\mathbf{E}\left[\Delta_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[\Delta_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|$$

$$\begin{aligned}
&\le \sum_{[i:j]\in E} \left|\mathbf{E}\left[-\frac{X_j^{(s-1)}}{d+1} + \frac{X_i^{(s-1)}}{d+1} + \left\lfloor \frac{X_j^{(s-1)}}{d+1} \right\rfloor - \left\lfloor \frac{X_i^{(s-1)}}{d+1} \right\rfloor + Z_{j,i}^{(s)} - Z_{i,j}^{(s)} \,\middle|\, Y_\ell, Y_{\ell-1}, \dots, Y_1\right]\right. \\
&\quad \left. - \mathbf{E}\left[-\frac{X_j^{(s-1)}}{d+1} + \frac{X_i^{(t-1)}}{d+1} + \left\lfloor \frac{X_j^{(s-1)}}{d+1} \right\rfloor - \left\lfloor \frac{X_i^{(s-1)}}{d+1} \right\rfloor + Z_{j,i}^{(s)} - Z_{i,j}^{(s)} \,\middle|\, Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right| \\
&= \sum_{[i:j]\in E} \left|\mathbf{E}\left[Z_{j,i}^{(s)} - Z_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[Z_{j,i}^{(s)} - Z_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right| \qquad (3.4) \\
&\le \sum_{[i:j]\in E} \left(\left|\mathbf{E}\left[Z_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[Z_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|\right. \\
&\quad \left. + \left|\mathbf{E}\left[Z_{j,i}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[Z_{j,i}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\right]\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|\right) \\
&= \sum_{i\in V} \sum_{j\in N(i)} \left|\Lambda_{i,j}^{(s)}\right| \cdot \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right| \le \sum_{i\in V} \left(\max_{j\in N(i)} \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|\right) \sum_{j\in N(i)} \left|\Lambda_{i,j}^{(s)}\right|, \qquad (3.5)
\end{aligned}$$

where we used $\Lambda_{i,j}^{(s)} := \mathbf{E}[Z_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1] - \mathbf{E}[Z_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1]$ to simplify the notation. Eq. (3.4) follows as $Y_{\ell-1}, \dots, Y_1$ determine the load vector $X^{(s-1)}$. To bound Eq. (3.5) we consider $\sum_{j\in N(i)} |\Lambda_{i,j}^{(s)}|$ for $i = i_1$ and $i \ne i_1$ separately.

Case 1: Let $i = i_1$. Assume first $Y_\ell = 0$. This means that node i_1 has less than r_1 extra tokens at step t_1. Hence $|\Lambda_{i,j}^s| = 0$. Now we assume that $Y_\ell \ne 0$. This means that node i_1 has at least r_1 extra tokens at step t_1. Let $b \ge r_1$ be the number of extra tokens of i_1 at step s_1. Clearly, b and the destinations of the extra tokens considered in the previous rounds, $Y_{\ell-r_1+1}, \dots, Y_{\ell-1}$, are already determined by $Y_{\ell-1}, \dots, Y_1$ (note that if $r_1 = 1$ then this set is empty). The remaining $Y_{\ell+1}, \dots, Y_{\ell+b-r_1}$ are chosen uniformly at random among $(N(i_1) \cup \{i_1\}) \setminus \{Y_{\ell-r_1+1}, \dots, Y_{\ell-1}\} =: \widetilde{N}(i_1)$ without replacement. Let $w \in \widetilde{N}(i_1)$ be the destination of the r_1-th excess token of i_1 at step s_1, that is, $Y_\ell = w$ and consequently, $Z_{i_1,w}^{(s_1)} = 1$. Clearly, $0 < \Lambda_{i_1,w}^{(s_1)} \le 1$, and for all $j \in \widetilde{N}(i_1) \setminus \{w\}$, $\Lambda_{i_1,j}^{(s_1)} < 0$. For the vertices $j \in \{Y_{\ell-r_1+1}, \dots, Y_{\ell-1}\}$, $\Lambda_{i_1,j}^{(s_1)} = 0$, as $Y_{\ell-1}, \dots, Y_1$ already determined that $Z_{i_1,j}^{(s_1)} = 1$. Linearity of expectations yields

$$\sum_{j\in N(i_1)\cup\{i_1\}} \Lambda_{i_1,j}^{(s_1)} = \mathbf{E}\left[\sum_{j\in N(i_1)\cup\{i_1\}} Z_{i_1,j}^{(s_1)} \,\middle|\, Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[\sum_{j\in N(i_1)\cup\{i_1\}} Z_{i_1,j}^{(s_1)} \,\middle|\, Y_{\ell-1}, \dots, Y_1\right] = 0.$$

The last equality holds since $\sum_{j\in N(i_1)\cup\{i_1\}} Z_{i_1,j}^{(s_1)} = b$ and b is determined by $Y_{\ell-1}, \dots, Y_1$. Hence,

$$\begin{aligned}
\sum_{j\in N(i_1)\cup\{i_1\}} \left|\Lambda_{i_1,j}^{(s_1)}\right| &= \sum_{\substack{j\in N(i_1)\cup\{i_1\}:\\ \Lambda_{i_1,j}^{(s_1)}>0}} \Lambda_{i_1,j}^{(s_1)} - \sum_{\substack{j\in N(i_1)\cup\{i_1\}:\\ \Lambda_{i_1,j}^{(s_1)}\le 0}} \Lambda_{i_1,j}^{(s_1)} \\
&= 2 \cdot \sum_{\substack{j\in N(i_1)\cup\{i_1\}:\\ \Lambda_{i_1,j}^{(s_1)}>0}} \Lambda_{i_1,j}^{(s_1)} = 2\left|\Lambda_{i_1,w}^{(s_1)}\right| \le 2. \qquad (3.6)
\end{aligned}$$

Case 2: $i \ne i_1$. As ℓ corresponds to (s_1, i_1, r_1), the random variable $Z_{i,j}^{(s_1)}$ is independent of Y_ℓ, which is the choice of the r_1-th excess token of vertex i_1 at step s_1. Hence

$$\sum_{j\in N(i)} \left|\Lambda_{i,j}^{(s_1)}\right| = \sum_{j\in N(i)} \left|\mathbf{E}\left[Z_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\right] - \mathbf{E}\left[Z_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\right]\right| = 0.$$

Combining Case 1 and Case 2 we obtain

$$\begin{aligned}
(3.5) &= \left(\max_{j\in N(i_1)} \left|\mathbf{P}_{i_1,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|\right) \sum_{j\in N(i_1)} \left|\Lambda_{i_1,j}^{(s)}\right| + \sum_{i\in V, i\ne i_1} \left(\max_{j\in N(i)} \left|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right|\right) \sum_{j\in N(i)} \left|\Lambda_{i,j}^{(s)}\right| \\
&\le \max_{j\in N(i_1)} \left|\mathbf{P}_{i_1,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\right| \cdot 2 + 0. \qquad (3.7)
\end{aligned}$$

$s > s_1$: Let $\widetilde{\ell}$ be the largest integer such that $Y_{\widetilde{\ell}}$ corresponds to time step $s-1$. Since $s > s_1$, we have $s-1 \geq s_1$ and therefore $\widetilde{\ell} \geq \ell$. By the choice of $\widetilde{\ell}$, $Y_{\widetilde{\ell}}, \dots, Y_1$ determine the load vector at the end of step s_1, $X^{(s_1)}$. By Lemma 2.6, we obtain $\mathbf{E}[\Delta_{i,j}^{(s)} \mid Y_{\widetilde{\ell}}, \dots, Y_1] = 0$, and by the chain rule of expectations,

$$\begin{aligned}\mathbf{E}\big[\Delta_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\big] &= \mathbf{E}\big[\mathbf{E}\big[\Delta_{i,j}^{(s)} \mid Y_{\widetilde{\ell}}, \dots, Y_1\big] \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\big] \\ &= \mathbf{E}[0 \mid Y_\ell, Y_{\ell-1}, \dots, Y_1] = 0.\end{aligned}$$

With the same arguments, $\mathbf{E}[\Delta_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1] = 0$, and therefore

$$\sum_{s=s_1+1}^{t} \sum_{[i:j]\in E} \big|\mathbf{E}\big[\Delta_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\big] - \mathbf{E}\big[\Delta_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\big]\big| \cdot \big|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\big| = 0. \tag{3.8}$$

This finishes the case distinction. Combining Eqs. (3.3), (3.7) and (3.8) for the three cases $s < s_1$, $s = s_1$, and $s > s_1$, we obtain that for every fixed $1 \leq \ell \leq tnd$,

$$\begin{aligned}&\big|\mathbf{E}[f_k \mid Y_\ell, Y_{\ell-1}, \dots, Y_1] - \mathbf{E}[f_k \mid Y_{\ell-1}, \dots, Y_1]\big| \\ &\quad \leq \sum_{s=1}^{t} \sum_{[i:j]\in E} \big|\mathbf{E}\big[\Delta_{i,j}^{(s)} \mid Y_\ell, Y_{\ell-1}, \dots, Y_1\big] - \mathbf{E}\big[\Delta_{i,j}^{(s)} \mid Y_{\ell-1}, \dots, Y_1\big]\big| \cdot \big|\mathbf{P}_{i,k}^{t-s_1} - \mathbf{P}_{j,k}^{t-s_1}\big| \\ &\quad = 0 + \max_{j\in N(i_1)} \big|\mathbf{P}_{i_1,k}^{t-s_1} - \mathbf{P}_{j,k}^{t-s_1}\big| \cdot 2 + 0 = 2 \cdot \max_{j\in N(i_1)} \big|\mathbf{P}_{i_1,k}^{t-s_1} - \mathbf{P}_{j,k}^{t-s_1}\big| =: c_\ell.\end{aligned}$$

To apply Theorem 3.1, we consider $\sum_{\ell=1}^{tnd}(c_\ell)^2$.

$$\begin{aligned}\sum_{\ell=1}^{tnd}(c_\ell)^2 &= \sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{r=1}^{d}\Big(2\max_{j\in N(i)}\big|\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\big|\Big)^2 = 4d\sum_{s=0}^{t-1}\sum_{i=1}^{n}\max_{j\in N(i)}\big(\mathbf{P}_{i,k}^{s} - \mathbf{P}_{j,k}^{s}\big)^2 \\ &\leq 4d\max_{k\in V}\Bigg(\sum_{s=0}^{\infty}\sum_{i=1}^{n}\max_{j\in N(i)}\big(\mathbf{P}_{i,k}^{s} - \mathbf{P}_{j,k}^{s}\big)^2\Bigg) = 8d\big(\Upsilon_2(G)\big)^2.\end{aligned} \tag{3.9}$$

By Theorem 3.1, we have for any $\delta \geq 0$, $\mathbf{Pr}[|f_k| > \delta] \leq 2\exp(-\delta^2/(2\sum_{\ell=1}^{tnd}(c_\ell)^2))$. Hence by choosing $\delta := \Upsilon_2(G)\sqrt{32d\ln n}$, the probability above gets smaller than $2n^{-2}$. Applying the union bound we obtain $\mathbf{Pr}[\forall k \in V: |f_k| > \delta] \leq n2n^{-2} = 2n^{-1}$. By Eq. (3.1), $\max_{k\in[n]} X_k^{(t)} \leq |\xi_k^{(t)}| + |f_k|$. For $t := \tau(G, K)$, we obtain $|\xi_k^{(t)} - \bar{\xi}| \leq 1$ for every vertex k. Hence $\max_{i,j\in[n]} |X_i^{(t)} - X_j^{(t)}| \leq 2|f_k| + 2$. This implies $\mathbf{Pr}[\max_{i,j\in[n]} |X_i^{(t)} - X_j^{(t)}| \leq 2\delta + 2] \geq 1 - 2n^{-1}$, as needed.

Proof of the second statement. Fix a vertex $k \in V$. Recall from Eq. (3.1) that

$$X_k^{(t)} - \xi_k^{(t)} = \sum_{s=1}^{t}\sum_{[i:j]\in E}\Delta_{i,j}^{(s)}\big(\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\big). \tag{3.10}$$

We split the right hand side of Eq. (3.10) at step $t-1$ to obtain

$$\underbrace{\sum_{s=1}^{t-1}\sum_{[i:j]\in E}\Delta_{i,j}^{(s)}\big(\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\big)}_{=:f_k} + \underbrace{\sum_{[i:j]\in E}\Delta_{i,j}^{(s)}\big(\mathbf{P}_{i,k}^{0} - \mathbf{P}_{j,k}^{0}\big)}_{=:h_k}.$$

We can bound h_k using the triangle inequality as follows,

$$|h_k| \leq \sum_{[i:j]\in E}\big|\Delta_{i,j}^{(t)}\big| \cdot \big|\mathbf{P}_{i,k}^{0} - \mathbf{P}_{j,k}^{0}\big| \leq 2 \cdot \sum_{[i:j]\in E}\big|\mathbf{P}_{i,k}^{0} - \mathbf{P}_{j,k}^{0}\big| = 2d,$$

since $|\Delta_{i,j}^{(t)}| \leq 2$ and $\sum_{i=1}^{n}\mathbf{P}_{i,k}^{0} = 1$. To bound f_k, we use the same approach as in the proof of the first statement. Also here, we use the same definition of variables Y_ℓ with $1 \leq \ell \leq (t-1)nd$. In order to apply Theorem 3.1, we have to estimate the differences c_ℓ, $1 \leq \ell \leq (t-1)nd$. As in Eq. (3.9) we obtain

$$\sum_{\ell=1}^{(t-1)nd}(c_\ell)^2 \leq 4d\sum_{s=1}^{t-1}\sum_{i=1}^{n}\max_{j\in N(i)}\big(\mathbf{P}_{i,k}^{t-s} - \mathbf{P}_{j,k}^{t-s}\big)^2.$$

Since $\sum_{i=1}^{n} \max_{j\in N(i)} (\mathbf{P}^0_{i,k} - \mathbf{P}^0_{j,k})^2 = 2d$, we obtain that

$$4d \sum_{s=1}^{t-1} \sum_{i=1}^{n} \max_{j\in N(i)} \left(\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right)^2 \le 8d \cdot \left(\left(\Upsilon_2(G)\right)^2 - d\right).$$

By Theorem 3.1, we obtain that

$$\mathbf{Pr}\left[|f_k| > \delta\right] \le 2\exp\left(-\delta^2 / \left(16d \cdot \left(\left(\Upsilon_2(G)\right)^2 - d\right)\right)\right).$$

Hence by choosing $\delta := \sqrt{32\log(n)\, d((\Upsilon_2(G))^2 - d)}$ we get $\mathbf{Pr}[|f_k| > \delta] \le 2n^{-2}$. Hence,

$$\mathbf{Pr}\left[\left|X^{(t)}_k - \xi^{(t)}_k\right| > 2d + \delta\right] \le \mathbf{Pr}\left[|h_k| > 2d\right] + \mathbf{Pr}\left[|f_k| > \sqrt{32\log(n)\, d\left(\left(\Upsilon_2(G)\right)^2 - d\right)}\right] \le 2n^{-2}.$$

Taking the union bound over all vertices k yields

$$\mathbf{Pr}\left[\forall k \in V: \left|X^{(t)}_k - \xi^{(t)}_k\right| \le 2d + \sqrt{32\log(n)\, d\left(\left(\Upsilon_2(G)\right)^2 - d\right)}\right] \le n2n^{-2} = 2n^{-1}. \tag{3.11}$$

Now choosing $t := \tau(G, K)$, we obtain the second statement by using exactly the same arguments as in the proof of the first statement.

Proof of the third statement. The third statement is shown by a similar approach. Again, fix a vertex $k \in V$ and a time step t. Now we split the right hand side of Eq. (3.10) at step $t - \vartheta$, where $\vartheta := (4\ln\ln n)/(1 - \lambda_{\max})$.

$$\sum_{s=1}^{t} \sum_{[i:j]\in E} \Delta^{(s)}_{i,j}\left(\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right) = \underbrace{\sum_{s=1}^{t-\vartheta} \sum_{[i:j]\in E} \Delta^{(s)}_{i,j}\left(\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right)}_{=:f_k} + \underbrace{\sum_{s=t-\vartheta+1}^{t} \sum_{[i:j]\in E} \Delta^{(s)}_{i,j}\left(\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right)}_{=:h_k}.$$

We first bound the last part directly by applying the triangle inequality as follows.

$$\begin{aligned} |h_k| &\le \sum_{s=t-\vartheta+1}^{t} \sum_{[i:j]\in E} \left|\Delta^{(s)}_{i,j}\right|\left|\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right| \\ &\le 2\vartheta \sum_{[i:j]\in E} \left(\mathbf{P}^{t-s}_{i,k} + \mathbf{P}^{t-s}_{j,k}\right) \le 2\vartheta d, \end{aligned}$$

where the first inequality holds since $|\Delta^{(s)}_{i,j}| \le 2$ and where the last inequality holds since $\sum_{i=1}^{n} \mathbf{P}^{t-s}_{i,k} = 1$ for every k.

To bound f_k, we use the same approach as in the proof of the first (and second) statement. Also here, we use the same definition of random variables Y_ℓ with $1 \le \ell \le (t-\vartheta)nd$. In order to apply Theorem 3.1, we have to estimate the differences c_ℓ, $1 \le \ell \le (t-\vartheta)nd$. As in Eq. (3.9) we obtain

$$\sum_{\ell=1}^{(t-\vartheta)nd} (c_\ell)^2 \le 8d \sum_{s=1}^{t-\vartheta} \sum_{i=1}^{n} \max_{j\in N(i)} \left(\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right)^2.$$

By Theorem 3.1, we obtain that

$$\mathbf{Pr}\left[|f_k| > \delta\right] \le 2\exp\left(-\delta^2 \Big/ \left(16d \sum_{s=1}^{t-\vartheta} \sum_{i=1}^{n} \max_{j\in N(i)} \left(\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k}\right)^2\right)\right).$$

Hence by choosing $\delta := \sqrt{32\log(n)\, d \sum_{s=1}^{t-\vartheta} \sum_{i=1}^{n} \max_{j\in N(i)} (\mathbf{P}^{t-s}_{i,k} - \mathbf{P}^{t-s}_{j,k})^2}$ we get $\mathbf{Pr}[|f_k| > \delta] \le 2n^{-2}$. Hence,

$$\mathbf{Pr}\left[\left|X^{(t)}_k - \xi^{(t)}_k\right| > 2\vartheta d + \delta\right] \le \mathbf{Pr}\left[|h_k| > 2\vartheta d\right] + \mathbf{Pr}\left[|f_k| > \delta\right] \le 0 + 2n^{-2} = 2n^{-2}.$$

Taking the union bound over all vertices k yields

$$\mathbf{Pr}\left[\forall k \in V: \left|X^{(t)}_k - \xi^{(t)}_k\right| \le 2\vartheta d + \delta\right] \le n2n^{-2} = 2n^{-1}. \tag{3.12}$$

In order to complete the proof, it remains to prove that $2\vartheta d + \delta = \mathcal{O}((d\log\log n)/(1 - \lambda_{\max}))$. To upper bound δ, we first consider

$$\begin{aligned}
\sum_{s=1}^{t-\vartheta}\sum_{i=1}^{n}\max_{j\in N(i)}\left(\mathbf{P}^{t-s}_{i,k}-\mathbf{P}^{t-s}_{j,k}\right)^2 &= \sum_{s=\vartheta}^{t-1}\sum_{i=1}^{n}\max_{j\in N(i)}\left(\mathbf{P}^{s}_{i,k}-\mathbf{P}^{s}_{j,k}\right)^2\\
&\le 2\sum_{s=\vartheta}^{t}\sum_{i=1}^{n}\max_{j\in N(i)}\left(\left(\mathbf{P}^{s}_{i,k}-\frac{1}{n}\right)^2+\left(\mathbf{P}^{s}_{j,k}-\frac{1}{n}\right)^2\right)\\
&= 2\sum_{s=\vartheta}^{t-1}\sum_{i=1}^{n}\left(\mathbf{P}^{s}_{i,k}-\frac{1}{n}\right)^2+2\sum_{s=\vartheta}^{t-1}\sum_{i=1}^{n}\max_{j\in N(i)}\left(\mathbf{P}^{s}_{j,k}-\frac{1}{n}\right)^2\\
&\le 2\sum_{s=\vartheta}^{t-1}\sum_{i=1}^{n}\left(\mathbf{P}^{s}_{i,k}-\frac{1}{n}\right)^2+2\sum_{s=\vartheta}^{t-1}\sum_{j=1}^{n}d\left(\mathbf{P}^{s}_{j,k}-\frac{1}{n}\right)^2\\
&\le (2d+2)\sum_{s=\vartheta}^{t-1}\lambda_{\max}^{2s},
\end{aligned}$$

where the first inequality uses Lemma 2.2 and the last inequality follows from Corollary 2.4. The last term can be now bounded as follows,

$$(2d+2)\sum_{s=\vartheta}^{\infty}\lambda_{\max}^{2s}\le(2d+2)\frac{\lambda_{\max}^{2(\frac{4\ln\ln n}{1-\lambda_{\max}})}}{1-(\lambda_{\max})^2}\le(2d+2)\frac{e^{-8\ln\ln n}}{1-\lambda_{\max}}=(2d+2)\frac{(\log n)^{-8}}{1-\lambda_{\max}},$$

where the second last inequality uses the fact that $x^{1/(1-x)}\le 1/e$ for $x\in[0,1)$. We can now use this bound to get a more explicit expression for the bound in Eq. (3.12),

$$\mathbf{Pr}\left[\forall k\in V: \left|X_k^{(t)}-\xi_k^{(t)}\right|\le\frac{4d\ln\ln n}{1-\lambda_{\max}}+\sqrt{32\log(n)\,d\cdot(d+2)\frac{(\log n)^{-8}}{1-\lambda_{\max}}}\right]\le 2n^{-1}.$$

We choose $t=\tau(G,K)$ to get $|\xi_k^{(t)}-\bar{\xi}|\le 1$ for every vertex k. As in the proof of the first statement, this yields

$$\mathbf{Pr}\left[\max_{i,j\in[n]}\left|X_i^{(t)}-X_j^{(t)}\right|\le 2(\vartheta d+\delta)+2\right]\ge 1-4n^{-1}.$$

This completes the proof. □

4. Proof of Theorem 1.2

This section contains 3 subsections in which we derive three upper bounds on the local divergence. The first bound holds for general graphs, the second for tori and the third for hypercubes. In detail, we show the following.

(1) For any graph G, $\Upsilon_2(G)=\mathcal{O}(\sqrt{d+\frac{\log d}{1-\lambda_{\max}}})$ (Theorem 4.1).
(2) For the r-dimensional torus graph G with $r=\mathcal{O}(1)$, $\Upsilon_2(G)\le\Psi_2(G)=\mathcal{O}(1)$ (Theorem 4.2).
(3) For the hypercube G with n vertices (Theorem 4.14)

$$\Psi_1(G)=\frac{\log_2(n)+1}{n}\sum_{p=0}^{\log_2(n)-1}\sum_{\ell=p+1}^{\log_2 n}\binom{\log_2 n}{\ell}=\Theta\left(\log^2 n\right).$$

Theorem 1.2 follows from these results. Theorem 1.2(1) follows from Theorem 1.1(1) and Theorem 4.1. Theorem 1.2(2) follows from Theorem 1.1(3). Theorem 1.2(3) follows from Theorem 1.1(1) and Theorem 4.2. Theorem 1.2(4) follows from Theorem 1.1(2) and Theorem 4.11.

4.1. General graphs

Theorem 4.1. *For any graph G, $\Upsilon_2(G)=\mathcal{O}(\sqrt{d+\frac{\log d}{1-\lambda_{\max}}})$.*

Proof. For simplicity, we consider $(\Psi_2(G))^2$. Let $k\in V$ be an arbitrary but fixed vertex. For some integer value τ to be specified later, we split the time into three parts, $t=0$, $1\le t\le\tau-1$ and $t\ge\tau$:

$$\sum_{t=0}^{\infty}\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2=\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{0}-\mathbf{P}_{j,k}^{0}\right)^2+\sum_{t=1}^{\tau-1}\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2+\sum_{t=\tau}^{\infty}\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2.$$

We start with the first term. Since $(\mathbf{P}_{i,k}^{0}-\mathbf{P}_{j,k}^{0})^2\le 1$, we conclude that

$$\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{0}-\mathbf{P}_{j,k}^{0}\right)^2\le\sum_{i=1}^{n}\sum_{[i:j]\in E}\left|\mathbf{P}_{i,k}^{0}-\mathbf{P}_{j,k}^{0}\right|\le d,$$

since each row sum of $\mathbf{P}$ is 1. Let us now consider the second term. We observe that for any two vertices r, s and any time step $t\ge 1$, $\mathbf{P}_{r,s}^{t}\le 1/(d+1)$. This allows us to bound the second term as follows,

$$\begin{aligned}\sum_{t=1}^{\tau-1}\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2&\le\sum_{t=1}^{\tau-1}\sum_{i=1}^{n}\sum_{[i:j]\in E}\left(\left(\mathbf{P}_{i,k}^{t}\right)^2+\left(\mathbf{P}_{j,k}^{t}\right)^2\right)\\&=d\sum_{t=1}^{\tau-1}\sum_{i=1}^{n}\left(\mathbf{P}_{i,k}^{t}\right)^2\\&\le d\sum_{t=1}^{\tau-1}\left((d+1)\left(\frac{1}{d+1}\right)^2+(n-d-1)\cdot 0\right)\\&\le\tau-1.\end{aligned}$$

Let us now consider the third term. Using Lemma 2.2 we obtain

$$\begin{aligned}\sum_{t=\tau}^{\infty}\sum_{[i:j]\in E}\left(\mathbf{P}_{i,k}^{t}-\mathbf{P}_{j,k}^{t}\right)^2&\le\sum_{t=\tau}^{\infty}\sum_{[i:j]\in E}2\left(\left(\mathbf{P}_{i,k}^{t}-\frac{1}{n}\right)^2+\left(\mathbf{P}_{j,k}^{t}-\frac{1}{n}\right)^2\right)\\&=d\sum_{t=\tau}^{\infty}\sum_{i=1}^{n}\left(\mathbf{P}_{i,k}^{t}-\frac{1}{n}\right)^2\\&\le d\sum_{t=\tau}^{\infty}\lambda_{\max}^{2t}\le d\frac{(\lambda_{\max})^{2\tau}}{1-(\lambda_{\max})^2}\le\frac{(\lambda_{\max})^{\tau}}{1-\lambda_{\max}},\end{aligned}$$

where we have used Corollary 2.4 in the second last inequality. Choosing $\tau:=\frac{\ln d}{1-\lambda_{\max}}$ and recalling that $x^{1/(1-x)}\le 1/e$ for any $x\in[0,1)$ yields the following bound:

$$\begin{aligned}\left(\Upsilon_2(G)\right)^2&\le d+\frac{\ln d}{1-\lambda_{\max}}+d\frac{(\lambda_{\max})^{\frac{\ln d}{1-\lambda_{\max}}}}{1-\lambda_{\max}}\\&\le d+\frac{\ln d}{1-\lambda_{\max}}+\frac{1}{1-\lambda_{\max}}=\mathcal{O}\left(d+\frac{\log d}{1-\lambda_{\max}}\right).\end{aligned}$$

Taking the square root yields $\Upsilon_2(G)=\mathcal{O}(\sqrt{d+\frac{\log d}{1-\lambda_{\max}}})$. □

4.2. Torus

Since for r-dimensional tori $1/(1-\lambda_{\max})=\Theta(n^{2/r})$ and for hypercubes $1/(1-\lambda_{\max})=\Theta(\log n)$, the following theorems represent improvements over the bound in Theorem 4.1 for these specific networks.

Theorem 4.2. *For the r-dimensional torus graph with $r=\mathcal{O}(1)$, $\Upsilon_2(G)\le\Psi_2(G)=\mathcal{O}(1)$.*

The proof of this result is rather long and technical. Hence, we further divide this subsection. In Section 4.2.1 we record some elementary inequalities. In Section 4.2.2 we relate the random walk on the (finite) r-dimensional torus graph to a random walk on the set $\mathbb{Z}^r$. For the latter, we can apply a local central limit theorem [18] which approximates transition probabilities by a multivariate normal distribution. In Section 4.2.3 we present the proof of Theorem 4.2.

4.2.1. Technical inequalities

The r-dimensional torus graph is defined as follows, where we assume for simplicity that $\sqrt[r]{n}-1$ is an even integer. The set of vertices is $V=\{-(\sqrt[r]{n}-1)/2,\dots,0,\dots,(\sqrt[r]{n}-1)/2\}^r$ and the set of edges are between vertices that differ exactly in one coordinate by one (hereby, we identify $-(\sqrt[r]{n}-1)/2-1$ with $(\sqrt[r]{n}-1)/2$ and $(\sqrt[r]{n}-1)/2+1$ with $-(\sqrt[r]{n}-1)/2$). As r is a fixed constant, we will assume that n is large enough such that $n^{1/r}/2\ge 2$.

Before we prove Theorem 4.2, we present some technical tools.

First we recall a higher dimensional version of the well-known bound $\sum_{k=1}^{\infty} k^{-(1+\epsilon)}=\mathcal{O}(1)$ for $\varepsilon>0$ that is also mentioned in [15].

Lemma 4.3. *For any constant $r\in\mathbb{N}$ and any constant $\epsilon>0$,*

$$\sum_{k\in\mathbb{Z}^r_{\neq 0}} \|k\|_2^{-(r+\epsilon)}=\mathcal{O}(1),$$

where $\mathbb{Z}^r_{\neq 0}=\mathbb{Z}^r\setminus\{0^r\}$.

Proof. By Lemma 2.2, $k_1^2+\dots+k_r^2\ge \frac{1}{r}(|k_1|+\dots+|k_r|)^2$ which implies

$$\begin{aligned}\sum_{k\in\mathbb{Z}^r_{\neq 0}} \|k\|_2^{-(r+\epsilon)} &= \sum_{k\in\mathbb{Z}^r_{\neq 0}} \left(k_1^2+\dots+k_r^2\right)^{-(r+\epsilon)/2} \le r\sum_{x=1}^{\infty}\sum_{\substack{k\in\mathbb{Z}^r\\ \|k\|_1=x}} x^{-(r+\epsilon)}\\ &= r\sum_{x=1}^{\infty}(2x+1)^{r-1}x^{-(r+\epsilon)} \le r4^{r-1}\sum_{x=1}^{\infty}x^{-(1+\epsilon)}=\mathcal{O}(1). \qquad \square\end{aligned}$$

The following inequalities are simple consequences of the triangle inequality for norms.

Lemma 4.4. *Let $i\in\mathbb{Z}^r$ and v be a vector with ± 1 at one position and zeros elsewhere. Then*

(1) $\|i\|_2^2-\|i+v\|_2^2\le 2\|i\|_2+1$.
(2) $\|i+v\|_2^2-\|i\|_2^2\le 2\|i\|_2+1$.
(3) *For any $p\in\mathbb{Z}^r_{\neq 0}$ and $i\in\mathbb{Z}^r$ with $\|i\|_1\le r\cdot n^{1/r}/2$, $\|i+p\cdot n^{1/r}\|_2\le (r/2+1)\cdot\|p\cdot n^{1/r}\|_2$.*
(4) *For any $p\in\mathbb{Z}^r_{\neq 0}$ and $i\in\mathbb{Z}^r$ with $\|i\|_\infty\le n^{1/r}/2$, $\|i+pn^{1/r}\|_2\ge \frac{1}{2r}\|pn^{1/r}\|_2$.*

Proof. The first statement is obvious if $\|i\|_2\le\|i+v\|_2$. Hence we assume that $\|i\|_2\ge\|i+v\|_2$. Using this and the triangle inequality of the ℓ_2-norm, we get

$$\|i\|_2^2-\|i+v\|_2^2\le\left(\|i+v\|_2+\|v\|_2\right)^2-\|i+v\|_2^2=2\|i+v\|_2+1\le 2\|i\|_2+1.$$

The second statement can be shown similarly. Using the triangle inequality, we obtain that

$$\|i+v\|_2^2-\|i\|_2^2\le\left(\|i\|_2+\|v\|_2\right)^2-\|i\|_2^2=2\|i\|_2+1.$$

To see the third statement, note that

$$\begin{aligned}\left\|i+p\cdot n^{1/r}\right\|_2 &\le \|i\|_2+\left\|p\cdot n^{1/r}\right\|_2\le\|i\|_1+\left\|p\cdot n^{1/r}\right\|_2\\ &\le r\cdot n^{1/r}/2+\left\|p\cdot n^{1/r}\right\|_2\le r\cdot n^{1/r}/2\cdot\|p\|_2+\left\|p\cdot n^{1/r}\right\|_2=(r/2+1)\cdot\left\|p\cdot n^{1/r}\right\|_2.\end{aligned}$$

Finally, for the fourth statement, we have

$$\left\|i+pn^{1/r}\right\|_2\ge\left\|i+pn^{1/r}\right\|_\infty\ge\left\|\frac{pn^{1/r}}{2}\right\|_\infty\ge\frac{1}{2r}\left\|pn^{1/r}\right\|_2,$$

where the inequality in the middle holds as all coordinates of i are bounded in absolute value by $n^{1/r}/2$, while $\|p\cdot n^{1/r}\|_\infty\ge n^{1/r}$, as $p\neq 0$. $\square$

Lemma 4.5. *For any constant $r\in\mathbb{N}$ and any $\ell>0$,*

$$\sum_{x=1}^{\infty}\exp\left(-\frac{x^2}{\ell^2}\right)\cdot x^r\le C\cdot\ell^{r+1},$$

where $C>0$ is a constant that can depend on r but not on ℓ.

Proof. Consider first the case $\ell \geq 1$. Define $\alpha := \min\{x \in \mathbb{N}_0 : \forall y \geq x : \exp(-y^2/2) \leq (y+1)^{-r}\}$. As α is a constant only depending on r, we have

$$\begin{aligned}
\sum_{x=1}^{\infty} \exp\left(-\frac{x^2}{\ell^2}\right) \cdot x^r &= \sum_{x=0}^{\infty} \sum_{p=1}^{\ell} \exp\left(-\frac{(x\ell+p)^2}{\ell^2}\right)(x\ell+p)^r \\
&\leq \sum_{x=0}^{\infty} \exp\left(-\frac{(x\ell)^2}{\ell^2}\right) \sum_{p=1}^{\ell} \big((x+1)\ell\big)^r \\
&= \sum_{x=0}^{\infty} \exp\big(-x^2\big)\ell(x+1)^r \ell^r \\
&= \sum_{x=0}^{\alpha-1} \exp\big(-x^2\big) \cdot (x+1)^r \cdot \ell^{r+1} + \sum_{x=\alpha}^{\infty} \exp\big(-x^2\big)(x+1)^r \cdot \ell^{r+1} \\
&\leq \alpha \cdot 1 \cdot (\alpha)^r \ell^{r+1} + \sum_{x=\alpha}^{\infty} \exp\big(-x^2/2\big)\ell^{r+1} \\
&= \mathcal{O}\big(\ell^{r+1}\big).
\end{aligned}$$

The second case is $0 < \ell < 1$. Here, we define $\beta := \min\{x \in \mathbb{N} : \forall y \geq x : \exp(-y^2/2) \leq y^{-r}\}$, which is again a constant only depending on r. Note that $\beta \geq 1$. Then,

$$\begin{aligned}
\sum_{x=1}^{\infty} \exp\left(-\frac{x^2}{\ell^2}\right) x^r &= \sum_{x=1}^{\beta-1} \exp\left(-\frac{x^2}{\ell^2}\right) \cdot x^r + \sum_{x=\beta}^{\infty} \exp\left(-\frac{x^2}{\ell^2}\right) \cdot x^r \\
&\leq (\beta-1)^{r+1} \exp\left(-\frac{1}{\ell^2}\right) + \sum_{x=\beta}^{\infty} \exp\left(-\frac{x^2}{2\ell^2}\right) \exp\left(-\frac{x^2}{2}\right) x^r \\
&= \mathcal{O}(1) \exp\left(-\frac{1}{\ell^2}\right) + \sum_{x=\beta}^{\infty} \exp\left(-\frac{x^2}{2\ell^2}\right) \\
&\leq \mathcal{O}(1) \exp\left(-\frac{1}{\ell^2}\right) + \sum_{x=\beta}^{\infty} \exp\left(-\frac{x^2}{4\ell^2}\right) \cdot \exp\left(-\frac{x^2}{4\ell^2}\right) \\
&\leq \mathcal{O}(1) \exp\left(-\frac{1}{\ell^2}\right) + \exp\left(-\frac{1}{4\ell^2}\right) \sum_{x=\beta}^{\infty} \exp\left(-\frac{x^2}{4}\right) \\
&= \mathcal{O}(1) \exp\left(-\frac{1}{4\ell^2}\right).
\end{aligned}$$

It remains to upper bound $\exp(-1/(4\ell^2))$ by $\mathcal{O}(\ell^{r+1})$. For this, let $\gamma := \max\{0 < x < 1 : \forall 0 < y \leq x : -1/(4y^2) \leq (r+1) \cdot \ln(y)\}$. Observe that γ is a constant only depending on r. This implies for $\ell \geq \gamma$ that $\exp(-1/(4\ell^2)) = \mathcal{O}(1)$. On the other hand, for $\ell \leq \gamma$ we get $\exp(-1/(4\ell^2)) \leq \exp(-(r+1)\ln(\ell)) = \ell^{r+1}$ by the definition of γ, which completes the proof. □

We continue with another simple analytic lemma.

Lemma 4.6. *Let $k \in \mathbb{Z}^r$ and v be a vector with ± 1 at one position and zeros elsewhere. Then if $\|k\|_2 \leq \|k+v\|_2$ or $\|k\|_\infty \geq n^{1/r}/2$,*

$$\left|\exp\left(-\frac{r\|k\|_2^2}{t}\right) - \exp\left(-\frac{r\|k+v\|_2^2}{t}\right)\right| \leq \exp\left(-\frac{\|k\|_2^2}{4t}\right) \cdot \frac{r(2\|k\|_2+1)}{t}.$$

Proof. We first consider the case $\|k\|_2 \leq \|k+v\|_2$. There,

$$\left|\exp\left(-\frac{r\|k\|_2^2}{t}\right) - \exp\left(-\frac{r\|k+v\|_2^2}{t}\right)\right| = \left|\exp\left(-\frac{r\|k\|_2^2}{t}\right) \cdot \left(1 - \exp\left(-\frac{r\|k+v\|_2^2}{t} + \frac{r\|k\|_2^2}{t}\right)\right)\right|.$$

Let us consider the second factor (which is positive by assumption). Using the second statement of Lemma 4.4, we obtain that

$$1-\exp\left(-\frac{r\|k+v\|_2^2}{t}+\frac{r\|k\|_2^2}{t}\right)\le 1-\exp\left(-\frac{r(2\|k\|_2+1)}{t}\right)\le\frac{r(2\|k\|_2+1)}{t},$$

where the last inequality follows from $\exp(-x)\ge 1-x$.

The second case to consider is $\|k\|_2>\|k+v\|_2$ and $\|k\|_\infty\ge n^{1/r}/2$. Then

$$\left|\exp\left(-\frac{r\|k\|_2^2}{t}\right)-\exp\left(-\frac{r\|k+v\|_2^2}{t}\right)\right|=\left|\exp\left(-\frac{r\|k+v\|_2^2}{t}\right)-\exp\left(-\frac{r\|k\|_2^2}{t}\right)\right|$$
$$=\left|\exp\left(-\frac{r\|k+v\|_2^2}{t}\right)\cdot\left(1-\exp\left(-\frac{r\|k\|_2^2}{t}+\frac{r\|k+v\|_2^2}{t}\right)\right)\right|.$$

Using the first statement of Lemma 4.4, we obtain for the second factor that

$$1-\exp\left(-\frac{r\|k\|_2^2}{t}+\frac{r\|k+v\|_2^2}{t}\right)\le 1-\exp\left(-\frac{r(2\|k\|_2+1)}{t}\right)\le\frac{r(2\|k\|_2+1)}{t}.$$

Moreover,

$$\|k+v\|_2^2\ge\|k+v\|_\infty^2\ge\|k/2\|_\infty^2\ge\frac{1}{4r}\cdot\|k\|_2^2,$$

where the second inequality holds because of $\|k+v\|_\infty^2\ge(\|k\|_\infty-1)^2\ge(\|k\|_\infty/2)^2$, since $\|k\|_\infty\ge n^{1/r}/2\ge 2$ (recall that we assumed that n is large enough such that $n^{1/r}/2\ge 2$). Hence

$$\left|\exp\left(-\frac{r\|k\|_2^2}{t}\right)-\exp\left(-\frac{r\|k+v\|_2^2}{t}\right)\right|\le\exp\left(-\frac{\|k\|_2^2}{4t}\right)\cdot\frac{r(2\|k\|_2+1)}{t}.\qquad\square$$

4.2.2. From the torus graph to $\mathbb{Z}^d$

We now follow an idea from [15] that relates a random walk on the r-dimensional torus graph $G=(V,E)$ with n vertices to a random walk on the infinite grid $\mathbb{Z}^r$. The (infinite) transition matrix of a random walk on the infinite grid $\mathbb{Z}^r$ is given by

$$\bar{\mathbf{P}}_{i,j}:=\begin{cases}\frac{1}{2r+1} & \text{if } \|i-j\|_1=1,\\ \frac{1}{2r+1} & \text{if } i=j,\\ 0 & \text{otherwise.}\end{cases}$$

Note that there is a natural relation between a random walk on the infinite grid $\mathbb{Z}^r$ and a random walk on V by projecting the random walk on $\mathbb{Z}^r$ to the finite set V. To make this more precise, we define for any vertex $(i_1,i_2,\dots,i_r)\in V$,

$$H(i):=\left(i_1+\mathbb{Z}\cdot\sqrt[r]{n},i_2+\mathbb{Z}\cdot\sqrt[r]{n},\dots,i_r+\mathbb{Z}\cdot\sqrt[r]{n}\right)\subseteq\mathbb{Z}^r.$$

With $\mathbf{P}_i^t:=\mathbf{P}_{0,i}^t$ and $\bar{\mathbf{P}}_i^t:=\bar{\mathbf{P}}_{0,i}^t$, we obtain that

$$\mathbf{P}_i^t=\sum_{k\in H(i)}\bar{\mathbf{P}}_k^t. \tag{4.1}$$

We also record a simple observation that follows from the definition and the fact that all coordinates of vertices in V are between $-(\sqrt[r]{n}-1)/2$ and $+(\sqrt[r]{n}-1)/2$.

Observation 4.7. For any $i\in V$ and any $k\in H(i)$, $\|i\|_2\le\|k\|_2$.

The reason why the relation given in Eq. (4.1) is useful is that $\bar{\mathbf{P}}_k^t$ can be approximated in terms of a multivariate normal distribution by a local central limit theorem [18]. That is, we will use an appropriate local central limit theorem to approximate the transition probabilities $\bar{\mathbf{P}}_k^t$ of $\mathbb{Z}^r$ with a multivariate normal distribution. To derive the limiting distribution $\widetilde{\mathbf{P}}_k^t$ of our random walk $\bar{\mathbf{P}}_{i,j}$, we follow Lawler and Limic [18] and let $X=(X_1,\dots,X_r)$ be a $\mathbb{Z}^r$-valued random variable with $\mathbf{Pr}[X=z]=1/(2r+1)$ for every vector z with one ±1 and zeros elsewhere, and $\mathbf{Pr}[X=0^r]=1/(2r+1)$. Observe that $\mathbf{E}[X_jX_k]=0$ for $j\neq k$ since not both of them can be non-zero simultaneously. Moreover, $\mathbf{E}[X_jX_j]=\frac{1}{(2r+1)}(-1)^2+\frac{1}{(2r+1)}(+1)^2=\frac{2}{2r+1}$ for all $1\le j\le r$. Hence the covariance matrix is

$$\varGamma:=\left[\mathbf{E}[X_jX_k]\right]_{1\le j,k\le r}=(r+1/2)^{-1}\cdot\mathbf{I},$$

where $\mathbf{I}$ is the $r\times r$-identity matrix.

Applying a local central limit from [18] to our setting yields the following:

Lemma 4.8. *(Cf. [18, Theorem 2.3.6] and [18, Eq. (2.2)].) For all $k, j \in \mathbb{Z}^r$ and all $t \in \mathbb{N}$,*

$$\left|\left(\bar{\mathbf{P}}^t_{k+j} - \bar{\mathbf{P}}^t_k\right) - \left(\widetilde{\mathbf{P}}^t_{k+j} - \widetilde{\mathbf{P}}^t_k\right)\right| = \mathcal{O}\left(\frac{\|j\|_1}{t^{(r+3)/2}}\right),$$

where

$$\widetilde{\mathbf{P}}^t_k := \frac{1}{(2\pi)^r t^{r/2}} \int_{\mathbb{R}^r} \exp\left(\mathrm{i}\frac{x \cdot k}{\sqrt{t}}\right) \exp\left(-\frac{x \cdot \Gamma x}{2}\right) d^r x, \tag{4.2}$$

where $\mathrm{i} = \sqrt{-1}$ denotes the imaginary unit.

The next lemma computes similarly to [15] the integral in Eq. (4.2).

Lemma 4.9. *With the notation of Lemma 4.8,*

$$\widetilde{\mathbf{P}}^t_k = \left(\frac{2r+1}{4\pi t}\right)^{r/2} \exp\left(\frac{-r\|k\|_2^2}{t}\right).$$

Proof. We calculate

$$\begin{aligned}
\widetilde{\mathbf{P}}^t_k &= \frac{1}{(2\pi)^r t^{r/2}} \int_{\mathbb{R}^r} \exp\left(\mathrm{i}\frac{x \cdot k}{\sqrt{t}} - \frac{x \cdot \Gamma x}{2}\right) d^r x \\
&= \frac{1}{(2\pi)^r t^{r/2}} \int_{\mathbb{R}^r} \exp\left(\mathrm{i}\frac{x \cdot k}{\sqrt{t}} - \frac{\|x\|_2^2}{2r+1}\right) d^r x \\
&= \frac{1}{(2\pi)^r t^{r/2}} \int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left(\|x\|_2^2 - 2\mathrm{i}\frac{r+1/2}{\sqrt{t}} x \cdot k\right)\right) d^r x.
\end{aligned} \tag{4.3}$$

To evaluate the integral we complete the square, which yields

$$\begin{aligned}
&\int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left(\|x\|_2^2 - 2\mathrm{i}\frac{r+1/2}{\sqrt{t}} x \cdot k\right)\right) d^r x \\
&\quad = \int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left(\|x\|_2^2 - 2\mathrm{i}\frac{r+1/2}{\sqrt{t}} x \cdot k - \frac{(2r+1)^2}{t}\|k\|_2^2 + \frac{(2r+1)^2}{t}\|k\|_2^2\right)\right) d^r x \\
&\quad = \exp\left(-\frac{r}{t}\|k\|_2^2\right) \int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left\|x - \mathrm{i}\frac{r+1/2}{\sqrt{t}} k\right\|_2^2\right) d^r x.
\end{aligned} \tag{4.4}$$

By substituting $z = x - \mathrm{i}\frac{r+1/2}{\sqrt{s}} k$ we get

$$\begin{aligned}
&\int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left\|x - \mathrm{i}\frac{r+1/2}{\sqrt{t}} k\right\|_2^2\right) d^r x \\
&\quad = \int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left(\|z\|_2^2\right)\right) d^r z \\
&\quad = \int \cdots \int_{\mathbb{R}^r} \exp\left(-\frac{1}{2r+1}\left(\sum_{i=1}^{r} z_i^2\right)\right) dz_r \cdots dz_1 \\
&\quad = \int \cdots \int_{\mathbb{R}^{r-1}} \exp\left(-\frac{1}{2r+1}\left(\sum_{i=1}^{r-1} z_i^2\right)\right) \left(\int_{\mathbb{R}} \exp\left(-\frac{1}{2r+1} z_r^2\right) dz_r\right) dz_{r-1} \cdots dz_1 \\
&\quad = \left(\sqrt{\pi(2r+1)}\right) \cdot \int \cdots \int_{\mathbb{R}^{r-1}} \exp\left(-\frac{1}{2r+1}\left(\sum_{i=1}^{r} z_i^2\right)\right) dz_{r-1} \cdots dz_1 \\
&\quad = \left(\sqrt{\pi(2r+1)}\right)^r.
\end{aligned} \tag{4.5}$$

Combining Eqs. (4.3), (4.4) and (4.5), we get

$$\widetilde{\mathbf{P}}_k^t = \frac{1}{(2\pi)^r t^{r/2}} \exp\left(-\frac{r}{t}\|k\|_2^2\right)\left(\sqrt{\pi(2r+1)}\right)^r = \left(\frac{2r+1}{4\pi t}\right)^{r/2} \exp\left(\frac{-r\|k\|_2^2}{t}\right),$$

as stated in the lemma. □

4.2.3. Proof of Theorem 4.2

We are now in a position to prove Theorem 4.2.

Proof of Theorem 4.2. Since $\Upsilon_2(G) \le \Psi_2(G)$ by definition, it is sufficient to prove that $(\Psi_2(G))^2$ is upper bounded by a constant. Since the torus graph is vertex-transitive, it suffices to consider the case $k = 0$. Hence it suffices to upper bound

$$\sum_{t=0}^{\infty} \sum_{[i:j]\in E} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2.$$

We first split this sum into three parts.

$$\left(\Psi_2(G)\right)^2 = \underbrace{\sum_{[i:j]\in E} \left(\mathbf{P}_i^0 - \mathbf{P}_j^0\right)^2}_{=:A} + \underbrace{\sum_{t=1}^{\kappa} \sum_{[i:j]\in E} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2}_{=:B} + \underbrace{\sum_{t=\kappa+1}^{\infty} \sum_{[i:j]\in E} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2}_{=:C},$$

where $\kappa := (4\ln n)/(1-\lambda_{\max}) = \mathcal{O}(n^{2/r}\log n)$.

Note that $A = d = 2r = \mathcal{O}(1)$. To bound C, we use Lemma 2.9 to get $C = \mathcal{O}(1)$. Hence it only remains to consider B. We first rewrite B as follows,

$$B = \sum_{[i:j]\in E} \sum_{t=1}^{\kappa} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2 = \sum_{(i,j)\in\vec{E}} \sum_{t=1}^{\kappa} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2,$$

where $\vec{E} \subseteq V \times V$ is an orientation of the edges E such that for all edges $\{u, v\} \in E$, either $(u, v) \in \vec{E}$ or $(v, u) \in \vec{E}$. In the following, we choose an orientation $\vec{E}$ such that for all $(u, v) \in \vec{E}$, $\|u\|_2 \le \|v\|_2$. Additionally, it will also be handy to use the following notation,

$$\vec{E}_{\neq 0} := \left\{(i, j) \in \vec{E}\colon \|i\|_2 > 0\right\},$$
$$\vec{E}_0 := \left\{(i, j) \in \vec{E}\colon \|i\|_2 = 0\right\}.$$

For each $(i, j) \in \vec{E}$, we split the inner sum of B at time

$$\sigma(i) := \begin{cases} 0 & \text{if } i = 0,\\ \frac{c^2\cdot\|i\|_2^2}{\log^2(2\|i\|_2^2)} & \text{otherwise,}\end{cases}$$

where c is a sufficiently small constant that satisfies $0 < c \le 1/r$. This gives

$$B = \underbrace{\sum_{(i,j)\in\vec{E}} \sum_{t=1}^{\sigma(i)} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2}_{=:B_1} + \underbrace{\sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\mathbf{P}_i^t - \mathbf{P}_j^t\right)^2}_{=:B_2}. \tag{4.6}$$

Let us first consider B_1. For $t \le \sigma(i)$ we use Lawler [17, Lemma 1.5.1(a)] saying that for random walks on infinite grids,

$$\sum_{\|k\|_2\ge\lambda\sqrt{t}} \bar{\mathbf{P}}_k^t = \mathcal{O}\left(e^{-\lambda}\right), \tag{4.7}$$

for all $t > 0$ and $\lambda > 0$. In particular, this gives

$$\mathbf{P}_i^t = \sum_{k\in H(i)} \bar{\mathbf{P}}_k^t \le \sum_{k\in\mathbb{Z}^r : \|k\|_2\ge\|i\|_2} \bar{\mathbf{P}}_k^t = \mathcal{O}\left(e^{-\|i\|_2/\sqrt{t}}\right),$$

where we have used Observation 4.7 saying that for any $k \in H(i)$, $\|i\|_2 \le \|k\|_2$. For any $(i, j) \in \vec{E}$ we have $\|i\|_2 \le \|j\|_2$ and therefore

$$\mathbf{P}_j^t = \sum_{k \in H(j)} \bar{\mathbf{P}}_k^t \le \sum_{k \in \mathbb{Z}^r : \|k\|_2 \ge \|j\|_2} \bar{\mathbf{P}}_k^t \le \sum_{k \in \mathbb{Z}^r : \|k\|_2 \ge \|i\|_2} \bar{\mathbf{P}}_k^t = \mathcal{O}\big(e^{-\|i\|_2/\sqrt{t}}\big).$$

Hence,

$$\begin{aligned}
B_1 &= \sum_{(i,j) \in \vec{E}} \sum_{t=1}^{\sigma(i)} \big(\mathbf{P}_i^t - \mathbf{P}_j^t\big)^2 \\
&\le \sum_{(i,j) \in \vec{E}_{\neq 0}} \sum_{t=1}^{\sigma(i)} \big(\max\{\mathbf{P}_i^t, \mathbf{P}_j^t\}\big)^2 + \sum_{(i,j) \in \vec{E}_0} \sum_{t=1}^{\sigma(0)} \big(\max\{\mathbf{P}_i^t, \mathbf{P}_j^t\}\big)^2 \\
&\le \sum_{(i,j) \in \vec{E}_{\neq 0}} \sum_{t=1}^{\sigma(i)} \mathcal{O}\big(\exp\big(-2\|i\|_2/\sqrt{t}\big)\big) \quad \big(\text{since } \sigma(0) = 0\big) \\
&\le \sum_{i \in V_{\neq 0}} \sum_{t=1}^{\sigma(i)} \mathcal{O}\big(\exp\big(-2\|i\|_2/\sqrt{t}\big)\big) \\
&\le \sum_{i \in V_{\neq 0}} \sigma(i) \cdot \mathcal{O}\big(\exp\big(-2\log\big(2\|i\|_2^2\big)/c\big)\big) \\
&\le \sum_{i \in V_{\neq 0}} \frac{\|i\|_2^2}{\log^2(2\|i\|_2^2)} \cdot \mathcal{O}\big(\|i\|_2^{-4r}\big) \quad (\text{since } c \le 1/r) \\
&= \mathcal{O}\bigg(\sum_{i \in V_{\neq 0}} \|i\|_2^{-4r+5/2}\bigg).
\end{aligned}$$

Applying Lemma 4.3 on the last term finally gives $B_1 = \mathcal{O}(1)$.

Rewriting the second part B_2 of Eq. (4.6) yields

$$\begin{aligned}
B_2 = \sum_{(i,j) \in \vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \big(\mathbf{P}_i^t - \mathbf{P}_j^t\big)^2 &= \sum_{(i,j) \in \vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \bigg(\sum_{k \in H(i)} \bar{\mathbf{P}}_k^t - \sum_{\ell \in H(j)} \bar{\mathbf{P}}_\ell^t\bigg)^2 \\
&= \sum_{(i,j) \in \vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \bigg(\sum_{k \in H(i)} \big(\bar{\mathbf{P}}_k^t - \bar{\mathbf{P}}_{k+(j-i)}^t\big)\bigg)^2.
\end{aligned}$$

We now define $H(i,t)$ as a subset of $H(i)$ by

$$H(i,t) := \big\{k \in H(i) \colon \|k\|_\infty \le 3\log n \cdot \sqrt{t}\big\} \subseteq H(i)$$

and split B_2 as follows,

$$\begin{aligned}
B_2 &= \sum_{(i,j) \in \vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \bigg(\sum_{k \in H(i) \setminus H(i,t)} \big(\bar{\mathbf{P}}_k^t - \bar{\mathbf{P}}_{k+(j-i)}^t\big) + \sum_{k \in H(i,t)} \big(\bar{\mathbf{P}}_k^t - \bar{\mathbf{P}}_{k+(j-i)}^t\big)\bigg)^2 \\
&\le 2 \underbrace{\sum_{(i,j) \in \vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \bigg(\sum_{k \in H(i) \setminus H(i,t)} \big(\bar{\mathbf{P}}_k^t - \bar{\mathbf{P}}_{k+(j-i)}^t\big)\bigg)^2}_{=:B_{2,1}} \\
&\quad + 2 \underbrace{\sum_{(i,j) \in \vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \bigg(\sum_{k \in H(i,t)} \big(\bar{\mathbf{P}}_k^t - \bar{\mathbf{P}}_{k+(j-i)}^t\big)\bigg)^2}_{=:B_{2,2}},
\end{aligned}$$

where the last line follows from Lemma 2.2. Recall Eq. (4.7) which states that $\sum_{\|k\|_2 \ge \lambda\sqrt{t}} \bar{\mathbf{P}}_k^t = \mathcal{O}(e^{-\lambda})$ for all $t > 0$ and $\lambda > 0$. This gives

$$\sum_{k \in H(i) \setminus H(i,t)} \bar{\mathbf{P}}_k^t \le \sum_{k \in \mathbb{Z}^r : \|k\|_\infty \ge 3\log n \cdot \sqrt{t}} \bar{\mathbf{P}}_k^t \le \sum_{k \in \mathbb{Z}^r : \|k\|_2 \ge 3\log n \cdot \sqrt{t}} \bar{\mathbf{P}}_k^t = \mathcal{O}\big(e^{-3\log n}\big) \le n^{-2}.$$

If $\|k\|_\infty \geq 3\log n \cdot \sqrt{t}$, then for any $\{i,j\} \in E$, $\|k + (j-i)\|_\infty \geq 3\log n \cdot \sqrt{t} - 1$. Hence,

$$\sum_{k \in H(i) \setminus H(i,t)} \bar{\mathbf{P}}^t_{k+(j-i)} = \mathcal{O}\big(e^{-3\log n + 1/\sqrt{t}}\big) \leq n^{-2}.$$

This allows us to upper bound $B_{2,1}$ as follows,

$$\begin{aligned} B_{2,1} &= \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i)\setminus H(i,t)} \big(\bar{\mathbf{P}}^t_k - \bar{\mathbf{P}}^t_{k+(j-i)}\big)\right)^2 \\ &\leq \sum_{(i,j)\in\vec{E}} \sum_{t=1}^{\kappa} \left(\sum_{k\in H(i)\setminus H(i,t)} \max\big\{\bar{\mathbf{P}}^t_k, \bar{\mathbf{P}}^t_{k+(j-i)}\big\}\right)^2 \\ &\leq \sum_{(i,j)\in\vec{E}} \sum_{t=1}^{\kappa} \big(n^{-2}\big)^2 \leq n^{-4} \sum_{(i,j)\in\vec{E}} \kappa = n^{-4}\mathcal{O}\big(n \cdot n^{2/r}\log n\big) = o(1). \end{aligned}$$

To bound $B_{2,2}$, we relate $\bar{\mathbf{P}}$ to the multivariate normal distribution given by $\widetilde{\mathbf{P}}$ which was defined in Lemma 4.8. Using the triangle inequality, we obtain that

$$\begin{aligned} B_{2,2} &= \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i,t)} \big(\bar{\mathbf{P}}^t_k - \bar{\mathbf{P}}^t_{k+(j-i)}\big)\right)^2 \\ &\leq \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i,t)} \big|\big(\bar{\mathbf{P}}^t_k - \bar{\mathbf{P}}^t_{k+(j-i)}\big) - \big(\widetilde{\mathbf{P}}^t_k - \widetilde{\mathbf{P}}^t_{k+(j-i)}\big)\big| + \big|\widetilde{\mathbf{P}}^t_k - \widetilde{\mathbf{P}}^t_{k+(j-i)}\big|\right)^2 \\ &\leq \underbrace{2\sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i,t)} \big|\big(\bar{\mathbf{P}}^t_k - \bar{\mathbf{P}}^t_{k+(j-i)}\big) - \big(\widetilde{\mathbf{P}}^t_k - \widetilde{\mathbf{P}}^t_{k+(j-i)}\big)\big|\right)^2}_{=:B_{2,2,1}} \\ &\quad + \underbrace{2\sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i,t)} \big|\widetilde{\mathbf{P}}^t_k - \widetilde{\mathbf{P}}^t_{k+(j-i)}\big|\right)^2}_{=:B_{2,2,2}}. \end{aligned}$$

Again we bound each part of the sum above separately and start with $B_{2,2,1}$. By Lemma 4.8,

$$\begin{aligned} B_{2,2,1} &= \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i,t)} \big|\big(\bar{\mathbf{P}}^t_k - \bar{\mathbf{P}}^t_{k+(j-i)}\big) - \big(\widetilde{\mathbf{P}}^t_k - \widetilde{\mathbf{P}}^t_{k+(j-i)}\big)\big|\right)^2 \\ &\leq \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i,t)} \mathcal{O}\big(t^{-(r+3)/2}\big)\right)^2. \end{aligned}$$

Note that the number of vertices in $H(i,t)$ can be bounded by

$$\big|H(i,t)\big| \leq \left(\left\lceil \frac{3\log n \cdot \sqrt{t}}{n^{1/r}} \right\rceil\right)^r \leq \left(\frac{3\log n \cdot \sqrt{t}}{n^{1/r}} + 1\right)^r,$$

since all coordinates of a vertex $k \in H(i,t)$ are bounded by $3\log n \cdot \sqrt{t}$ and additionally, the difference between a coordinate of k and the respective coordinate of i must be a multiple of $n^{1/r}$. Using Lemma 2.2, we can further estimate this by

$$\big|H(i,t)\big| \leq 2^{r-1}\left(\frac{(\log n \cdot \sqrt{t})^r}{n} + 1^r\right) = \mathcal{O}\left(\frac{(\log n)^r t^{r/2}}{n} + 1\right).$$

Therefore,

$$\begin{aligned} B_{2,2,1} &\leq \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \big(\big|H(i,t)\big| \cdot \mathcal{O}\big(t^{-(r+3)/2}\big)\big)^2 \\ &= \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \mathcal{O}\left(\frac{(\log n)^{2r} t^{-3}}{n^2}\right) + \sum_{(i,j)\in\vec{E}} \sum_{t=\sigma(i)+1}^{\kappa} \mathcal{O}\big(t^{-(r+3)}\big) \end{aligned}$$

$$
\begin{aligned}
&= \frac{(\log n)^{2r}}{n^2} \sum_{(i,j)\in\overrightarrow{E}} \sum_{t=\sigma(i)+1}^{\kappa} \mathcal{O}\big(t^{-3}\big) + \sum_{(i,j)\in\overrightarrow{E}} \sum_{t=\sigma(i)+1}^{\kappa} \mathcal{O}\big(t^{-(r+3)}\big) \\
&= \frac{(\log n)^{2r}}{n^2} \left(\sum_{(i,j)\in\overrightarrow{E}_{\neq 0}} \sum_{t=\sigma(i)+1}^{\infty} \mathcal{O}\big(t^{-3}\big) + \sum_{(i,j)\in\overrightarrow{E}_{0}} \sum_{t=\sigma(i)+1}^{\kappa} \mathcal{O}\big(t^{-3}\big) \right) + \sum_{(i,j)\in\overrightarrow{E}_{\neq 0}} \sum_{t=\sigma(i)+1}^{\kappa} \mathcal{O}\big(t^{-(r+3)}\big) \\
&= \frac{(\log n)^{2r}}{n^2} \left(\sum_{(i,j)\in\overrightarrow{E}_{\neq 0}} \mathcal{O}\left(\frac{\log^4(2\|i\|_2^2)}{\|i\|_2^4} \right) + \mathcal{O}(1) \right) + \sum_{(i,j)\in\overrightarrow{E}_{\neq 0}} \mathcal{O}\left(\frac{(\log(2\|i\|_2^2))^{(2r+4)}}{\|i\|_2^{(2r+4)}} \right) \\
&= \frac{(\log n)^{2r}}{n^2} \sum_{(i,j)\in\overrightarrow{E}_{\neq 0}} \mathcal{O}(1) + \sum_{i\in V} \mathcal{O}\big(\|i\|_2^{-2r-3}\big) \\
&= \mathcal{O}(1),
\end{aligned}
$$

where the last line follows from Lemma 4.3.

We continue to bound the remaining part $B_{2,2,2}$. We first use $H(i,t) \subseteq H(i)$, then apply Lemma 4.9 twice, and obtain

$$
\begin{aligned}
B_{2,2,2} &\le \sum_{(i,j)\in\overrightarrow{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\sum_{k\in H(i)} \big| \widetilde{\mathbf{P}}_k^t - \widetilde{\mathbf{P}}_{k+(j-i)}^t \big| \right)^2 \\
&= \sum_{(i,j)\in\overrightarrow{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\frac{2r+1}{4\pi t} \right)^r \left(\sum_{k\in H(i)} \left| \exp\left(-\frac{r\|k\|_2^2}{t} \right) - \exp\left(-\frac{r\|k+(j-i)\|_2^2}{t} \right) \right| \right)^2 \\
&\le \sum_{(i,j)\in\overrightarrow{E}} \sum_{t=\sigma(i)+1}^{\kappa} \left(\frac{2r+1}{4\pi t} \right)^r \left(\sum_{k\in H(i)} \exp\left(-\frac{\|k\|_2^2}{4t} \right) \frac{r(2\|k\|_2+1)}{t} \right)^2,
\end{aligned}
$$

where we used Lemma 4.6 in the last line (note that in the case $k = i$ we have $\|i\|_2^2 \le \|j\|_2^2$ by the definition of $\overrightarrow{E}$ and when $k \in H(i) \setminus \{i\}$, we have $\|k\|_\infty \ge n^{1/r}/2$).

Recalling that $H(i) = \{i + n^{1/r} p \colon p \in \mathbb{Z}^d\}$ and $r^{-1}\|p\|_1 \le \|p\|_2 \le \|p\|_1$, we obtain

$$
\begin{aligned}
&\sum_{k\in H(i)} \exp\left(-\frac{\|k\|_2^2}{4t} \right) \frac{r(2\|k\|_2+1)}{t} \\
&= \sum_{p\in\mathbb{Z}^r} \exp\left(-\frac{\|i+pn^{1/r}\|_2^2}{4t} \right) \frac{r(2\|i+pn^{1/r}\|_2+1)}{t} \\
&\le \frac{2r\|i\|_2+1}{t} + \sum_{p\in\mathbb{Z}^r_{\neq 0}} \exp\left(-\frac{\|pn^{1/r}\|_2^2}{16tr^2} \right) \frac{r((r+2)\|pn^{1/r}\|_2+1)}{t}
\end{aligned}
$$

where the last inequality follows by using the fourth and third inequality of Lemma 4.4. We continue to upper bound the last term:

$$
\begin{aligned}
&= \frac{2r\|i\|_2+1}{t} + \sum_{p\in\mathbb{Z}^r_{\neq 0}} \exp\left(-\frac{\|pn^{1/r}\|_2^2}{16tr^2} \right) \frac{(r^2+2r)\|pn^{1/r}\|_2+r}{t} \\
&\le \frac{2r\|i\|_2+1}{t} + \sum_{p\in\mathbb{Z}^r_{\neq 0}} \exp\left(-\frac{\|pn^{1/r}\|_2^2}{16tr^2} \right) \frac{(r^2+2r)\|pn^{1/r}\|_2+r\|pn^{1/r}\|_2}{t} \\
&\le \frac{2r\|i\|_2+1}{t} + \sum_{p\in\mathbb{Z}^r_{\neq 0}} \exp\left(-\frac{\|pn^{1/r}\|_1^2}{16tr^3} \right) \frac{(r^2+3r)\|pn^{1/r}\|_1}{t} \\
&\le \frac{2r\|i\|_2+1}{t} + \sum_{\beta=1}^{\infty} \sum_{p\in\mathbb{Z}^r:\|p\|_1=\beta} \exp\left(-\frac{\|p\|_1^2 n^{2/r}}{16tr^3} \right) \frac{4r^2\|p\|_1 n^{1/r}}{t} \\
&= \frac{2r\|i\|_2+1}{t} + \mathcal{O}\left(\frac{n^{1/r}}{t} \sum_{\beta=1}^{\infty} (2\beta+1)^{r-1} \exp\left(-\beta^2 \Big/ \frac{16tr^3}{n^{2/r}} \right) \beta \right)
\end{aligned}
$$

$$= \frac{2r\|i\|_2+1}{t} + \mathcal{O}\left(\frac{n^{1/r}}{t}\sum_{\beta=1}^{\infty}\exp\left(-\beta^2 \Big/ \frac{16tr^3}{n^{2/r}}\right)\beta^r\right).$$

By applying Lemma 4.5, the second summand can be upper bounded by

$$\mathcal{O}\left(\frac{n^{1/r}}{t}\left(\sqrt{\frac{16tr^3}{n^{2/r}}}\right)^{r+1}\right) = \mathcal{O}\left(t^{(r-1)/2}n^{1/r-1/r(r+1)}\right) = \mathcal{O}\left(t^{(r-1)/2}n^{-1}\right).$$

Plugging this into our upper bound on $B_{2,2,2}$, we obtain

$$\begin{aligned}
B_{2,2,2} &\le \left(\frac{2r+1}{4\pi}\right)^r \sum_{(i,j)\in\vec{E}}\sum_{t=\sigma(i)+1}^{\kappa}\left(\frac{2r+1}{4\pi t}\right)^r\left(\mathcal{O}\left(t^{(r-1)/2}n^{-1}\right)+\frac{2r\|i\|_2+1}{t}\right)^2 \\
&= \mathcal{O}\left(\sum_{(i,j)\in\vec{E}}\sum_{t=1}^{\kappa}t^{-1}n^{-2}\right) + \mathcal{O}\left(\sum_{(i,j)\in\vec{E}}\sum_{t=\sigma(i)+1}^{\kappa}\frac{\|i\|_2^2+1}{t^{r+2}}\right) \\
&= \mathcal{O}\left(n^{-2}\sum_{(i,j)\in\vec{E}}\log(\kappa)\right) + \mathcal{O}\left(\sum_{(i,j)\in\vec{E}_{\neq 0}}\sum_{t=\sigma(i)+1}^{\infty}\frac{\|i\|_2^2+1}{t^{r+2}}\right) + \mathcal{O}\left(\sum_{(i,j)\in\vec{E}_{=0}}\sum_{t=1}^{\kappa}\frac{\|i\|_2^2+1}{t^{r+2}}\right) \\
&= \mathcal{O}(1) + \mathcal{O}\left(\sum_{(i,j)\in\vec{E}_{\neq 0}}\|i\|_2^2\left(\sigma(i)\right)^{-r-1}\right) + \mathcal{O}\left(\sum_{t=1}^{\kappa}t^{-r-2}\right) \\
&= \mathcal{O}(1) + \mathcal{O}\left(\sum_{(i,j)\in\vec{E}_{\neq 0}}\|i\|_2^2\left(\frac{\|i\|_2^2}{\log^2(2\|i\|_2^2)}\right)^{-r-1}\right) + \mathcal{O}(1) \\
&= \mathcal{O}(1) + \mathcal{O}\left(\sum_{i\in\mathbb{Z}^r_{\neq 0}}\|i\|_2^{-2r+1/2}\right) + \mathcal{O}(1) = \mathcal{O}(1),
\end{aligned}$$

where the last inequality is due to Lemma 4.3. This completes the proof. □

4.3. Hypercube

Before we can calculate $\Upsilon_2(G)$ and $\Psi_1(G)$ on the hypercube we state the following result which can be easily derived from [22].

Lemma 4.10. *For the d-dimensional hypercube with $n = 2^d$ vertices the following statements hold.*

(1) *For any edge $\{i, j\} \in E$, any vertex $k \in V$ and any time step $\vartheta \in \mathbb{N}$,*

$$\sum_{t=\vartheta}^{\infty}\left|\mathbf{P}^t_{i,k} - \mathbf{P}^t_{j,k}\right| = \left|\sum_{t=\vartheta}^{\infty}\left(\mathbf{P}^t_{i,k} - \mathbf{P}^t_{j,k}\right)\right|.$$

(2) *Let 0 also denote the vertex $0^{\log_2 n} \in V$. For any two vertices i, j with $\{i, j\} \in E$, $\|i\|_1 = p$ and $\|j\|_1 = p+1$ we have*

$$\sum_{t=0}^{\infty}\left(\mathbf{P}^t_{i,0} - \mathbf{P}^t_{j,0}\right) = \frac{1}{n}\cdot\frac{\log_2(n)+1}{\binom{\log_2 n}{p}(\log_2(n)-p)}\cdot\sum_{\ell=p+1}^{\log_2 n}\binom{\log_2 n}{\ell}.$$

Proof. First we show (1). As shown in [6, Lemma 6], it holds for all triples of vertices i, j, k with $\text{dist}(i,k) \le \text{dist}(j,k)$ that for all time steps $t \in \mathbb{N}$, $\mathbf{P}^t_{i,k} \ge \mathbf{P}^t_{j,k}$. This immediately implies that

$$\sum_{t=\vartheta}^{\infty}\left|\mathbf{P}^t_{i,k} - \mathbf{P}^t_{j,k}\right| = \left|\sum_{t=\vartheta}^{\infty}\left(\mathbf{P}^t_{i,\ell} - \mathbf{P}^t_{j,k}\right)\right|.$$

The second claim is a slight reformulation of [22, Theorem 5] (see also [21, Corollary 3.35]). Note that in the notation of [21,22], we have $\delta = 1$ and $w_u - w_v = \alpha n\delta(\sum_{t=0}^{\infty}\mathbf{M}^t_{0,i} - \mathbf{M}^t_{0,j})$ with $\alpha = \frac{1}{\log_2(n)+1}$ ($\mathbf{M}$ is the same matrix as $\mathbf{P}$). □

Theorem 4.11. *For the d-dimensional hypercube G with $n = 2^d$ vertices, $\Upsilon_2(G) = \sqrt{d + \mathcal{O}(1)}$.*

Note that since $\Upsilon_2(G) \geq \sqrt{d}$ for any d-regular network, this bound is almost tight.

Proof of Theorem 4.11. First note that since the d-dimensional hypercube is vertex-transitive, it suffices to consider the case $k = 0 = 0^{\log_2 n}$ in the definition of $\Upsilon_2(G)$. Therefore,

$$\big(\Upsilon_2(G)\big)^2 = \frac{1}{2} \sum_{i \in \{0,1\}^d} \sum_{t=0}^{\infty} \max_{j \in N(i)} \big(\mathbf{P}^t_{i,0} - \mathbf{P}^t_{j,0}\big)^2.$$

Our aim is to prove that $(\Upsilon_2(G))^2 = d + \mathcal{O}(1)$. Using Lemma 2.9, we obtain for $\kappa := (4 \ln n)/(1 - \lambda_{\max}) = \mathcal{O}(\log^2 n)$ that

$$\sum_{i \in \{0,1\}^d} \sum_{t=\kappa+1}^{\infty} \max_{j \in N(i)} \big(\mathbf{P}^t_{i,0} - \mathbf{P}^t_{j,0}\big)^2 = \mathcal{O}(1).$$

Furthermore for $t = 0$, we have

$$\sum_{i \in \{0,1\}^d} \max_{j \in N(i)} \big(\mathbf{P}^0_{i,0} - \mathbf{P}^0_{j,0}\big)^2 = 2d.$$

Hence,

$$\big(\Upsilon_2(G)\big)^2 = \frac{1}{2} \sum_{i \in \{0,1\}^d} \sum_{t=1}^{\kappa} \max_{j \in N(i)} \big(\mathbf{P}^t_{i,0} - \mathbf{P}^t_{j,0}\big)^2 + d + \mathcal{O}(1)$$

and it remains to consider only the time steps between 1 and κ in the following.

We now move on to exploit further symmetries of the hypercube. As the hypercube is distance-transitive [2], we have for any two vertices i, j with $\|i\|_1 = \|j\|_1$ and for any $t \in \mathbb{N}_0$,

$$\mathbf{P}^t_{i,0} = \mathbf{P}^t_{j,0}. \tag{4.8}$$

A simple consequence of this fact is that for any vertex i,

$$\mathbf{P}^t_{i,0} \leq 1 \Big/ \binom{\log_2 n}{i}. \tag{4.9}$$

To simplify the notation, we also define for $p \in \mathbb{N}$ with $0 \leq p \leq \log_2 n$,

$$\mathbf{P}^t_{p,0} := \mathbf{P}^t_{0^p 1^{\log_2 n - p},0}.$$

We will use the following result from [6].

Lemma 4.12. *(See [6, Lemma 6].) For any fixed t, $\mathbf{P}^t_{p,0}$ is decreasing in p ($0 \leq p \leq \log_2 n$).*

Combining Eq. (4.9) and Lemma 4.12, we obtain the following lemma.

Lemma 4.13. *For any $t \in \mathbb{N}_{>0}$, $\mathbf{P}^t_{0,0} \leq \frac{1}{\log_2 n + 1}$. Moreover, for any $1 \leq p \leq \log_2(n)/2$ and $t \in \mathbb{N}$,*

$$\mathbf{P}^t_{p,0} \leq \frac{1}{\binom{\log_2 n}{p}},$$

and for any $\log_2(n)/2 \leq p \leq \log_2 n$,

$$\mathbf{P}^t_{p,0} \leq \frac{1}{\binom{\log_2 n}{\log_2(n)/2}}.$$

With the notation identifying all vectors $i \in \{0,1\}^d$ with $\|i\|_1 = p$

$$\begin{aligned}
\sum_{i \in \{0,1\}^d} \sum_{t=1}^{\kappa} \max_{j \in N(i)} \big(\mathbf{P}^t_{i,0} - \mathbf{P}^t_{j,0}\big)^2 &= \sum_{p=0}^{\log_2 n} \binom{\log_2 n}{p} \sum_{t=1}^{\kappa} \max_{j \in \{p-1, p+1\}} \big(\mathbf{P}^t_{p,0} - \mathbf{P}^t_{j,0}\big)^2 \\
&\leq \sum_{p=1}^{\log_2(n)-1} \binom{\log_2 n}{p} \left(\sum_{t=1}^{\kappa} \big(\mathbf{P}^t_{p,0} - \mathbf{P}^t_{p-1,0}\big)^2 + \sum_{t=1}^{\kappa} \big(\mathbf{P}^t_{p,0} - \mathbf{P}^t_{p+1,0}\big)^2 \right)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{1,0}-\mathbf{P}^t_{0,0}\right)^2+\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{\log_2 n,0}-\mathbf{P}^t_{\log_2(n)-1,0}\right)^2\\
&\le \sum_{p=0}^{\log_2(n)-1}\left(\binom{\log_2 n}{p}+\binom{\log_2 n}{p+1}\right)\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{p,0}-\mathbf{P}^t_{p+1,0}\right)^2\\
&= \sum_{p=0}^{\log_2(n)-1}\binom{\log_2 n+1}{p+1}\left(\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{p,0}-\mathbf{P}^t_{p+1,0}\right)\right)^2.
\end{aligned}
$$

We split the outer sum in three parts:

$$
\begin{aligned}
\sum_{p=0}^{\log_2(n)}\binom{\log_2 n+1}{p+1}\left(\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{p,0}-\mathbf{P}^t_{p+1,0}\right)\right)^2 &= \underbrace{\sum_{p=0}^{5}\binom{\log_2 n+1}{p+1}\left(\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{p,0}-\mathbf{P}^t_{p+1,0}\right)\right)^2}_{=:B_1}\\
&\quad+\underbrace{\sum_{p=6}^{\log_2(n)-6}\binom{\log_2 n+1}{p+1}\left(\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{p,0}-\mathbf{P}^t_{p+1,0}\right)\right)^2}_{=:B_2}\\
&\quad+\underbrace{\sum_{p=\log_2(n)-5}^{\log_2(n)}\binom{\log_2 n+1}{p+1}\left(\sum_{t=1}^{\kappa}\left(\mathbf{P}^t_{p,0}-\mathbf{P}^t_{p+1,0}\right)\right)^2}_{=:B_3}.
\end{aligned}
$$

We first consider B_2. By using the second statement of Lemma 4.10 and recalling that $\mathbf{P}^t_{p,0}-\mathbf{P}^{t,0}_{p+1}\ge 0$, we can upper bound B_2 by

$$
\begin{aligned}
B_2 &\le \sum_{p=6}^{\log_2 n-6}\frac{\binom{\log_2 n+1}{p+1}}{\binom{\log_2 n}{p}^2}\left(\frac{1}{n}\cdot\frac{\log_2(n)+1}{\log_2(n)-p}\sum_{\ell=p+1}^{\log_2 n}\binom{\log_2 n}{\ell}\right)^2\\
&= \sum_{p=6}^{\log_2 n-6}\frac{\log_2 n+1}{p+1}\frac{1}{\binom{\log_2 n}{p}}\left(\frac{1}{n}\cdot\frac{\log_2(n)+1}{\log_2(n)-p}\sum_{\ell=p+1}^{\log_2 n}\binom{\log_2 n}{\ell}\right)^2.
\end{aligned}
$$

Using that $\sum_{\ell=p+1}^{\log_2 n}\binom{\log_2 n}{\ell}\le n$ and plugging in the bound $6\le p\le \log_2 n-6$ we can continue with

$$
\begin{aligned}
&\le \sum_{p=6}^{\log_2 n-6}\frac{(\log_2 n+1)^3}{7}\frac{1}{\binom{\log_2 n}{p}}\\
&= \frac{(\log_2 n+1)^3}{7}\left(\sum_{p=6}^{\log_2 n/2}\frac{1}{\binom{\log_2 n}{p}}+\sum_{p=\log_2 n/2+1}^{\log_2 n-6}\frac{1}{\binom{\log_2 n}{p}}\right)\\
&= \mathcal{O}\left(\log^3 n\right)\sum_{p=6}^{\log_2(n)/2}\frac{1}{\binom{\log_2 n}{p}}\\
&\le \mathcal{O}\left(\log^3 n\right)\left(\sum_{p=6}^{\sqrt{\log n}}\left(\frac{p}{\log_2 n}\right)^p+\sum_{p=\sqrt{\log n}}^{\log_2(n)/2}\left(\frac{p}{\log_2 n}\right)^p\right)\\
&\le \mathcal{O}\left(\log^3 n\right)\left(\sum_{p=6}^{\sqrt{\log n}}(\log_2 n)^{-p/2}+\sum_{p=\sqrt{\log n}}^{\log_2(n)/2}2^{-p}\right)\\
&= \mathcal{O}\left(\log^3 n\right)\cdot\mathcal{O}\left(\frac{1}{\log^3(n)}+2^{-\sqrt{\log n}}\right)\\
&= \mathcal{O}(1).
\end{aligned}
$$

To upper bound B_3, we use again the second statement of Lemma 4.10 to obtain

$$B_3 \le \sum_{p=\log_2 n-5}^{\log_2 n} \frac{\log_2 n+1}{p+1} \frac{1}{\binom{\log_2 n}{p}} \left(\frac{1}{n} \cdot \frac{\log_2(n)+1}{\log_2(n)-p} \sum_{\ell=p+1}^{\log_2 n} \binom{\log_2 n}{\ell} \right)^2 = o(1),$$

as $\frac{1}{n}$ dominates all other factors which are at most polylogarithmic in n.

It remains to upper bound B_1. Using Lemma 4.12, we obtain

$$B_1 = \sum_{p=0}^{5} \binom{\log_2 n+1}{p+1} \sum_{t=1}^{\kappa} \left(\mathbf{P}^t_{p,0} - \mathbf{P}^t_{p+1,0}\right)^2 \le \sum_{p=0}^{5} 2\binom{\log_2 n+1}{p+1} \cdot \sum_{t=1}^{\kappa} \left(\mathbf{P}^t_{p,0}\right)^2.$$

We now split the inner sum into two parts: $1 \le t \le 14$, and $15 \le t \le \kappa$. We first consider the sum with $1 \le t \le 14$. For bounding the term $\mathbf{P}^t_{p,0}$, we now use Lemma 4.13 to obtain that

$$\begin{aligned} \sum_{p=0}^{5} 2\binom{\log_2 n+1}{p+1} \cdot \sum_{t=1}^{14} \left(\mathbf{P}^t_{p,0}\right)^2 &\le 2\binom{\log_2 n+1}{1} \cdot \frac{14}{(\log_2 n+1)^2} + \sum_{p=1}^{5} 2\binom{\log_2 n+1}{p+1} \cdot \frac{14}{\left(\binom{\log_2 n}{p}\right)^2} \\ &= \mathcal{O}(1). \end{aligned} \tag{4.10}$$

For larger time steps $t \ge 15$, we examine $\mathbf{P}^t_{p,0}$ for $0 \le p \le 5$ more closely. Observe that a random walk that starts from p increases the distance to $0^{\log_2 n}$ in step t with probability at least $1 - \frac{t+p}{\log_2 n+1}$. Moreover, in order to arrive at a vertex q with $0 \le q \le 5$ at step 15, the random walk can at most 10 times increase the distance to $0^{\log_2 n}$ during 15 steps. This implies that for all $0 \le p \le 5$,

$$\sum_{q=0}^{5} \mathbf{P}^{15}_{p,q} \le \binom{15}{5} \cdot \left(\frac{20}{\log_2 n+1} \right)^5 = \mathcal{O}\left((\log n)^{-5}\right). \tag{4.11}$$

Combining Eq. (4.11) and Lemma 4.13, we obtain that

$$\begin{aligned} \sum_{t=15}^{\kappa} \left(\mathbf{P}^t_{p,0}\right)^2 &\le \left(\sum_{t=15}^{\kappa} \mathbf{P}^t_{p,0} \right)^2 \\ &\le \left(\sum_{t=15}^{\kappa} \left(\sum_{q=0}^{5} \mathbf{P}^{15}_{p,q} \cdot \mathbf{P}^{t-15}_{q,0} + \sum_{q=6}^{\log n} \mathbf{P}^{15}_{p,q} \cdot \mathbf{P}^{t-15}_{q,0} \right) \right)^2 \\ &\le \left(\sum_{t=15}^{\kappa} \left(\sum_{q=0}^{5} \mathbf{P}^{15}_{p,q} \cdot 1 + \sum_{q=6}^{\log(n)/2} 1 \cdot \frac{1}{\binom{\log_2 n}{q}} + \sum_{q=\log(n)/2+1}^{\log n} 1 \cdot \frac{1}{\binom{\log_2 n}{\log_2(n)/2}} \right) \right)^2 \\ &\le \left(\kappa \cdot \left(\mathcal{O}\left((\log n)^{-5}\right) + \mathcal{O}\left((\log n)^{-5}\right) \right) \right)^2 \\ &= \mathcal{O}\left((\log n)^{-6}\right). \end{aligned}$$

Hence we obtain

$$\sum_{p=0}^{5} 2\binom{\log_2 n+1}{p} \cdot \sum_{t=15}^{\kappa} \left(\mathbf{P}^t_{p,0}\right)^2 \le \sum_{p=0}^{5} 2\binom{\log_2 n+1}{p} \cdot \mathcal{O}\left((\log n)^{-6}\right) = \mathcal{O}(1). \tag{4.12}$$

Combining Eq. (4.10) and Eq. (4.12), we find that also $B_3 = \mathcal{O}(1)$, which finishes the proof. □

4.4. Hypercube

In the following we give an *exact* bound on $\Psi_1(G)$.

Theorem 4.14. *Let G be a hypercube with n vertices. Then,*

$$\Psi_1(G) = \frac{\log_2(n)+1}{n} \sum_{p=0}^{\log_2(n)-1} \sum_{\ell=p+1}^{\log_2 n} \binom{\log_2 n}{\ell} = \Theta\left(\log^2 n\right).$$

Proof. By symmetry, it suffices to consider $k = 0 = 0^{\log_2 n}$ for $\Psi_1(G)$. By using Lemma 4.10 twice, we get

$$\begin{aligned}
\Psi_1(G) &= \sum_{t=0}^{\infty} \sum_{[i:j]\in E} \left|\mathbf{P}_{i,0}^t - \mathbf{P}_{j,0}^t\right| \\
&= \sum_{p=0}^{\log_2(n)-1} \sum_{\substack{\{i,j\}\in E: \\ \|i\|_1=p, \|j\|_1=p+1}} \left| \sum_{t=0}^{\infty} \left(\mathbf{P}_{i,0}^t - \mathbf{P}_{j,0}^t\right) \right| \\
&= \sum_{p=0}^{\log_2(n)-1} \cdot \sum_{\substack{\{i,j\}\in E: \\ \|i\|_1=p, \|j\|_1=p+1}} \frac{1}{n} \cdot \frac{\log_2(n)+1}{\binom{\log_2 n}{p}(\log_2(n)-p)} \cdot \sum_{\ell=p+1}^{\log_2 n} \binom{\log_2 n}{\ell} \\
&= \frac{\log_2(n)+1}{n} \sum_{p=0}^{\log_2(n)-1} \sum_{\ell=p+1}^{\log_2 n} \binom{\log_2 n}{\ell},
\end{aligned} \tag{4.13}$$

where in the last equality we have used the fact that for any $0 \le p \le \log_2(n) - 1$, there are $\binom{\log_2 n}{p}(\log_2 n - p)$ edges $\{i, j\} \in E$ with $\|i\|_1 = p$ and $\|j\|_1 = p + 1$. We can upper bound this term by

$$\Psi_1(G) \le \frac{\log_2(n)+1}{n} \log_2(n) 2^{\log_2 n} = \mathcal{O}\big(\log^2 n\big).$$

For the lower bound on Eq. (4.13),

$$\Psi_1(G) \ge \frac{\log_2(n)+1}{n} \sum_{p=0}^{\log_2(n)/2} \sum_{\ell=\log_2(n)/2}^{\log_2 n} \binom{\log_2 n}{\ell} = \Omega\big(\log^2 n\big). \qquad \square$$

As the discrepancy of the RSW algorithm is at most $\Psi_1(G)$ after $\tau(G, K)$ rounds [24, Corollary 3], we obtain:

Corollary 4.15. *The discrepancy of the RSW algorithm [24] is at most $\mathcal{O}(\log^2 n)$ after $\tau(G, K) = \mathcal{O}(\log(Kn) \cdot \log^2 n)$ time steps.*

Note that the best possible result from [24, Theorem 4] yields only a weaker bound of $\mathcal{O}(\log^3 n)$. Our result is tight since $d \cdot \mathrm{diam}(G) = (\log_2 n)^2$ is a simple lower bound.

5. Discussion

We presented a new diffusion-based load-balancing scheme which is very simple and avoids negative load. We show bounds on the discrepancy for general graphs depending on the local (or refined local) divergence and the eigenvalue gap of the graph. For (constant-degree) expander graphs we prove a discrepancy of $\mathcal{O}(\log\log n)$, for hypercubes of $\mathcal{O}(\log n)$, and for r-dimensional torus graphs of $\mathcal{O}(\sqrt{\log n})$.

We also note that our proof techniques are not restricted to the algorithm presented in this paper. For example an adversarial algorithm where the adversary is allowed to specify the destinations of the excess tokens could also be analyzed. Adapting the proof of Theorem 1.1 to this algorithm, one can show that the deviation is at most $\mathcal{O}(d\Upsilon_1(G))$.

References

[1] N. Alon, B. Sudakov, Bipartite subgraphs and the smallest eigenvalue, Comb. Probab. Comput. 9 (1) (2000).
[2] N. Biggs, Algebraic Graph Theory, Cambridge University Press, 1993.
[3] J.E. Boillat, Load balancing and Poisson equation in a graph, Concurr. Comput. 2 (1990) 289–313.
[4] J. Cooper, J. Spencer, Simulating a random walk with constant error, Comb. Probab. Comput. 15 (2006) 815–822.
[5] G. Cybenko, Load balancing for distributed memory multiprocessors, J. Parallel Distrib. Comput. 7 (1989) 279–301.
[6] P. Diaconis, R. Graham, J. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, Random Struct. Algorithms 1 (1990) 51–72.
[7] R. Diekmann, A. Frommer, B. Monien, Efficient schemes for nearest neighbor load balancing, Parallel Comput. 25 (1999) 789–812.
[8] B. Doerr, T. Friedrich, Deterministic random walks on the two-dimensional grid, Comb. Probab. Comput. 18 (2009) 123–144.
[9] D. Dubhashi, A. Panconesi, Concentration of Measure for the Analysis of Randomized Algorithms, Cambridge University Press, 2009.
[10] R. Elsässer, B. Monien, Load balancing of unit size tokens and expansion properties of graphs, in: 15th ACM Symposium on Parallel Algorithms and Architectures, SPAA'03, 2003, pp. 266–273.
[11] R. Elsässer, T. Sauerwald, Discrete load balancing is (almost) as easy as continuous load balancing, in: 29th Annual ACM Principles of Distributed Computing, PODC'10, 2010, pp. 346–354.
[12] R. Elsässer, B. Monien, S. Schamberger, Distributing unit size workload packages in heterogeneous networks, J. Graph Algorithms Appl. 10 (2006) 51–68.
[13] T. Friedrich, T. Sauerwald, Near-perfect load balancing by randomized rounding, in: 41st Annual ACM Symposium on Theory of Computing, STOC'09, 2009, pp. 121–130.

[14] T. Friedrich, M. Gairing, T. Sauerwald, Quasirandom load balancing, in: 21st Annual ACM–SIAM Symposium on Discrete Algorithms, SODA'10, 2010, pp. 1620–1629.
[15] T. Friedrich, M. Gairing, T. Sauerwald, Quasirandom load balancing, SIAM J. Comput. 41 (4) (2012) 747–771.
[16] B. Ghosh, F.T. Leighton, B.M. Maggs, S. Muthukrishnan, C.G. Plaxton, R. Rajaraman, A.W. Richa, R.E. Tarjan, D. Zuckerman, Tight analyses of two local load balancing algorithms, SIAM J. Comput. 29 (1999) 29–64.
[17] G. Lawler, Intersections of Random Walks, Probab. Appl., Birkhäuser, 1991.
[18] G.F. Lawler, V. Limic, Random Walk: A Modern Introduction, Cambridge Stud. Adv. Math., Cambridge University Press, 2010.
[19] L. Lovász, Combinatorial Problems and Exercises, Elsevier North-Holland, 1993.
[20] L. Lovász, P. Winkler, Mixing of random walks and other diffusions on a graph, Surv. Comb. (1995) 119–154.
[21] H. Meyerhenke, Disturbed diffusive schemes for solving partitioning problems on graphs, PhD thesis, University of Paderborn, Germany, 2008, http://www2.cs.uni-paderborn.de/cs/ag-monien/PERSONAL/HENNINGM/pubs.html.
[22] H. Meyerhenke, T. Sauerwald, Analyzing disturbed diffusion on networks, in: 17th International Symposium on Algorithms and Computation, ISAAC'06, 2006, pp. 429–438.
[23] S. Muthukrishnan, B. Ghosh, M.H. Schultz, First and second-order diffusive methods for rapid, coarse, distributed load balancing, Theory Comput. Syst. 31 (1998) 331–354.
[24] Y. Rabani, A. Sinclair, R. Wanka, Local divergence of Markov chains and the analysis of iterative load balancing schemes, in: 39th Annual IEEE Symposium on Foundations of Computer Science, FOCS'98, 1998, pp. 694–705.
[25] R. Subramanian, I.D. Scherson, An analysis of diffusive load-balancing, in: 6th ACM Symposium on Parallel Algorithms and Architectures, SPAA'94, 1994, pp. 220–225.