IEEE Control Systems Letters paper presented at
2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Low complexity convergence rate bounds for the synchronous gossip subclass of push-sum algorithms

Balázs Gerencsér   Miklós Kornyik

***Abstract*— We develop easily accessible quantities for bounding the almost sure exponential convergence rate of push-sum algorithms. We analyze the scenario of i.i.d. synchronous gossip, every agent communicating independently towards at most a single target at every step. Multiple bounding expressions are developed depending on the generality of the setup, all functions of the network's spectrum. Numerical experiments demonstrate the quality of the bounds obtained together with the computational speedup for acquiring them.**

I. INTRODUCTION

Average consensus algorithms have been around for some time [1], [2], with the fundamental goal of computing the average of input values on a network in a distributed manner with only local communication and simple operations. Often some symmetry is imposed on the communication, in terms of the matrix describing the linear update of the vector of values to be either doubly stochastic, or even symmetric. This condition is quite well understood [3], see the survey [4] also for applications, further discussion and references.

However, the interest for distributed averaging algorithms capable of handling asynchronous directed communication emerged, naturally driving away the representing update matrix from being doubly stochastic, still with the intent to compute the exact average. As a result, the successful scheme of *push-sum* was proposed [5], later referred to as *ratio consensus* [6] and joined by variants such as *weighted gossip* [7]. The goal of these algorithms is the same, but now using only local, directed communication and without requiring message passing to happen synchronously or consistently across the network. Given the simple objective of the algorithm, it also serves as a building block for more complex tasks, e.g., the spectral analysis of the network [8] or distributed optimization algorithms [9]. For the latter, the distributed tracking of the gradient implicitly raises the challenge for understanding the two-timescale variant of the baseline algorithm.

With other real-life communication challenges taken into account, the concept has been extended in multiple ways to handle such aspects, e.g., including packet loss [10] [11] or delay [6].

Related work. An essential question in the analysis for usability and efficiency is understanding the asymptotics of the processes, their convergence and the rate at which it happens. Already in [5], the exponential convergence of the original push-sum scheme has been proven, however without elaborating on the exact rate yet. An important step ahead was made in [12] providing an impressive upper bound along an unspecified, infinite subset of the timeline for the almost sure (a.s.) rate of convergence. More recently, the *tight bound for the rate of a.s. convergence* has been identified [13] as the spectral gap of the Lyapunov exponents of the random update matrix series with generous applicability. While being a clean representation, the concern is present that this Lyapunov spectral gap is known to be *incomputable* in general [14]. As a follow-up, it was possible to combine the inspiration of [12] and the tool-set of [13] to obtain an actual upper bound on the a.s. rate for the i.i.d. case [15], now formulated by manipulating a single (random) update matrix, thus leading to a *computable quantity*. The bounds are solid, however for a graph on N vertices, matrices of size $N^2 \times N^2$ have to be analyzed, quickly *increasing in complexity*.

Contributions. Our goal is to provide even simpler convergence rate estimates. For this purpose, we focus our attention to the natural setup, where a weighted network determines the communication scheme driving the consensus process. In particular, we assume synchronized gossip message passing, i.e. every node sending a single packet to a single (random) recipient at each time slot, or possibly skipping a slot if it would message itself.

The bounds provided can be computed directly once the standard spectral description of the network is available. We are to formulate *multiple variants*, both to provide general, but more conservative estimates, and also sharper ones for a more restricted setting with stronger symmetries.

Layout. The rest of the paper is structured as follows. In the next section we formally define the averaging process and state our results. Section III builds a framework for the proof of the theorems, while Section IV completes the proofs. Detailed numerical performance analysis and concluding remarks are provided in Section V and Section VI, respectively.

II. MAIN RESULTS

Let us recall the push-sum algorithm [5] along with some notations to be used. Given is a finite graph $G = (V, E)$ with the vertex set $V = [N] := \{1, 2, \dots, N\}$, having degree sequence $d_1, \dots, d_N$. There is an initial vector of values $x(0) \in \mathbb{R}^N$ at the vertices to be averaged. The process is also using an auxiliary vector initialized at $w(0) = \mathbf{1} \in \mathbb{R}^N$. At each time step, representing local communication, a linear

The research was supported by NRDI (National Research, Development and Innovation Office) grant KKP 137490. B. Gerencsér and M. Kornyik are with the HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary (email: gerencser.balazs@renyi.hu and kornyik.miklos@renyi.hu). B. Gerencsér is also with the Eötvös Loránd University, Department of Probability and Statistics, Budapest, Hungary.

column-stochastic update $A(t)$ is performed on both vectors:

$$x(t) = A(t)x(t-1), \qquad w(t) = A(t)w(t-1).$$

The average $\bar{x} := \frac{1}{N}\sum_i x_i(0)$ is then locally estimated by $x_i(t)/w_i(t)$. There is a wide generality of how $(A(t))_{t\geq 0}$ can be chosen [7], [13]. In the current paper, we focus on the scenario of i.i.d. matrix series, and when at each step, every vertex i sends a single message to a random recipient β_i (or possibly none) with a constant fixed proportion $q \in [0,1]$ and all these choices are independent from one another. Formally, $A(t) \stackrel{d}{=} A = (1-q)I + q\sum_i e_{\beta_i}e_i^T$. This is a significant generalization of the setup used in [12], where the underlying topology was given by the complete graph with $q = 1/2$.

By setting $p_{ij} := \mathbb{P}(\beta_j = i)$, we obtain an overall transition probability matrix P which by construction has to be compatible with the adjacency matrix of G. For convenience, we introduce the notation $P_q = (1-q)I + qP$. It is easy to check that $\mathbb{E}A = P_q$. In case P has only real eigenvalues let λ_i denote its i^{th} largest eigenvalue and let $\lambda_{q,i} = (1-q) + q\lambda_i$ denote that of P_q.

Following the system description let us state our main results about the exponential convergence rate to consensus.

Theorem 1: Let us consider a push-sum algorithm with symmetric message probability matrix P. Then

$$\limsup \frac{1}{t}\max_i \log\left|\frac{x_i(t)}{w_i(t)} - \bar{x}\right| \leq \frac{1}{2}\log\left\{(1-q)^2 + 2q(1-q)\lambda_2 + q^2\right\} =: \frac{\gamma}{2} \qquad (1)$$

Remark 1: In case each vertex chooses a recipient uniformly among its neighbors, the bounding quantity in Theorem 1 will depend only on the graph structure. Indeed, it is easy to check that in this case $\mathbb{E}A = P_q$ with $P = MD^{-1}$, where D denotes the diagonal matrix consisting of the degrees of the underlying graph's vertices, while M denotes the graph's adjacency matrix.

A better bound can be obtained for cases with stronger symmetries: a message probability matrix is said to be transitive if for any pair (i,j) there exists a permutation matrix Π with $\Pi_{ij} = 1$ such that $\Pi P \Pi^{-1} = P$.

Theorem 2: Suppose that the message probability matrix P is symmetric and transitive. Then

$$\limsup \frac{1}{t}\max_i \log\left|\frac{x_i(t)}{w_i(t)} - \bar{x}\right| \leq \frac{1}{2}\log \xi_1 =: \frac{\alpha}{2}, \qquad (2)$$

with ξ_1 being the largest root of the polynomial

$$f(\xi) = \prod_{i>1}(\xi - \lambda_{q,i}^2) - \frac{q^2}{N}\sum_{i>1}(1-\lambda_i^2)\prod_{j:i\neq j>1}(\xi - \lambda_{q,i}^2).$$

Remark 2: If G is a transitive graph and each vertex chooses a recipient uniformly among its neighbors, then the corresponding message probability matrix satisfies the assumption of Theorem 2.

Let us emphasize the gain in computation complexity, stemming from the smaller matrices analyzed despite using their full spectrum: for the stated bounds it is $O(N^3)$ in the worst case [16], whereas for the bounds of [12], [15] it is $O(N^4)$ [17] meaning that the speed-up is at least of order N. Later in Section V we will show that it can be even better in practice.

III. Mathematical Setup

Let us build a framework using an appropriate matrix transformation and corresponding tools in a general setting.

First we remark that the elements $w_i(t)$ are all positive because the diagonal elements of the nonnegative $A(t)$ are strictly positive. Using the notations $H(t) = A(t)A(t-1)\cdots A(1)$, $J = \mathbf{1}\mathbf{1}^\top/N$ easy calculation yields

$$\begin{aligned} x(t) - \bar{x}w(t) &= H(t)x_0 - \bar{x}w(t) \\ &= \big(H(t)(I-J) + H(t)J\big)x_0 - \bar{x}w(t) \\ &= H(t)(I-J)x_0, \end{aligned}$$

meaning that for some constant $C > 0$ we have

$$\begin{aligned} \max_i\left|\frac{x_i(t)}{w_i(t)} - \bar{x}\right| &\leq C\frac{||x_0||_2 \cdot ||H(t)(I-J)||_2}{\min_i w_i(t)} \\ &\leq C\frac{||x_0||_2 \cdot ||H(t)(I-J)||_F}{\min_i w_i(t)}. \end{aligned} \qquad (3)$$

We are interested in the almost sure convergence rate of the quantity on the left of (3). As we will see the dominant term will be $||H(t)(I-J)||_F$. To get a handle on this factor let us analyze the expectation of $||H(t)(I-J)||_F^2$:

$$\begin{aligned} \mathbb{E}||H(t)(I-J)||_F^2 &= \mathbb{E}\ \mathrm{Tr}\{(I-J)H(t)^\top H(t)(I-J)\} \\ &= \mathrm{Tr}\{(I-J)\mathbb{E}[H(t)^\top H(t)](I-J)\}. \end{aligned}$$

According to the definition of $H(t)$ we can write

$$\begin{aligned} \mathbb{E}[H(t)^\top H(t)] &= \mathbb{E}\big[\mathbb{E}[A(1)^\top \widetilde{H}(t)^\top \widetilde{H}(t)A(1) \mid A(1)]\big] \\ &= \mathbb{E}[A(1)^\top \mathbb{E}[\widetilde{H}(t)^\top \widetilde{H}(t)]A(1)], \end{aligned} \qquad (4)$$

where $\widetilde{H}(t) = A(t)A(t-1)\cdots A(2)$, and by the i.i.d. nature of the updates $\widetilde{H}(t) \stackrel{d}{=} H(t-1)$. This motivates the introduction of the linear operator $\Phi : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$:

$$\Phi(X) := \mathbb{E}[A^\top X A],$$

which we will need to understand for further developing (4).

Remark 3: Since $\mathrm{Tr}\{(I-J)H(t)H(t)^\top(I-J)\} = \mathrm{Tr}\{H(t)^\top(I-J)H(t)\}$ and the adjoint of the linear map $f : X \mapsto A^\top X A$ is the map $f^* : X \mapsto AXA^\top$, it follows that $\Phi^*(X) = \mathbb{E}[AXA^\top]$ for any symmetric matrix X.

For satisfactory notation, before we progress let us introduce the linear operator $\Psi : \mathbb{R}^{N\times N} \to \mathbb{R}^N$ together with its pseudo-inverse $\Psi^- : \mathbb{R}^N \to \mathbb{R}^{N\times N}$:

$$(\Psi(X))_i = x_{ii},$$

$$(\Psi^-(v))_{ij} = \begin{cases} v_i & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 1: For any matrix X, we have

$$\Phi(X) = P_q^\top X P_q + q^2\big\{\Psi^-[P^\top\ \Psi(X)] - \Psi^-\Psi(P^\top X P)\big\}$$

reminding that $P_q = (1-q)I + qP$.

The proof is postponed to the Appendix. In order to obtain a bound on $\mathrm{Tr}\,\{(I-J)\mathbb{E}[H(t)^\top H(t)](I-J)\}$ it is enough to understand Φ, since

$$\mathrm{Tr}\,\{(I-J)\mathbb{E}[H(t)^\top H(t)](I-J)\} = \mathrm{Tr}\,\{(I-J)\Phi^t(I)(I-J)\},$$

where $\Phi^t(\cdot)$ denotes the t-fold application of Φ.

Proposition 1: The map Φ has the following properties:

(P1) Φ is linear,
(P2) $\Phi(X^\top) = \Phi(X)^\top$,
(P3) for any skew-symmetric matrix X, $\Phi(X) = P_q^\top X P_q$,
(P4) if $X \geq 0$ then $\Phi(X) \geq 0$, i.e. Φ keeps the positive semi-definite property,
(P5) if $x_{kl} \geq 0\ \forall(k,l)$, then $\Phi(X)_{kl} \geq 0\ \forall(k,l)$,
(P6) J is an eigenmatrix of Φ with eigenvalue 1: $\Phi(J) = J$,
(P7) for $X \geq 0$, and $P = P^\top$ we have $\mathrm{Tr}\,\Phi(X) \leq \mathrm{Tr}\,X$.

For the adjoint map Φ^* the following properties hold:

(P*1) $\Phi^*(Y) = P_q Y P_q^\top + q^2\{\Psi^-[P\Psi(Y)] - P(\Psi^-\Psi Y)P^\top\}$,
(P*2) if $X \geq 0$ then $\Phi^*(X) \geq 0$, i.e. Φ^* also keeps the positive semi-definite property,
(P*3) if $x_{kl} \geq 0\ \forall(k,l)$, then $\Phi^*(X) \geq 0\ \forall(k,l)$,
(P*4) if $X\mathbf{1} = 0$ then $\Phi^*(X)\mathbf{1} = 0$.

The proof can be found in the Appendix.

Remark 4: According to the definition of A, we have $JA = J$, and so

$$\begin{aligned}(I-J)A(I-J) &= A(I-J),\\ \mathbb{E}[(I-J)A^\top X A(I-J)]&\\ = \mathbb{E}[(I-J)A^\top(I-J)X(I-J)A(I-J)].&\end{aligned}$$

Corollary 1: Let us define the operator $\widehat{\Phi}$ as

$$\begin{aligned}\widehat{\Phi} : X \mapsto (I-J)\Phi(X)(I-J)&\\ \in \mathrm{End}(\{Y \in \mathbb{R}^{N\times N} : Y = Y^\top, YJ = 0\}),&\end{aligned}$$

then according to the properties of Φ combined with Remark 4, for any $X \in \{Y \in \mathbb{R}^{N\times N} : Y = Y^\top, YJ = 0\}$ we have

$$\begin{aligned}\widehat{\Phi}^t(X) &= (I-J)\Phi^t(X)(I-J),\\ \Phi((I-J)X(I-J)) &= \Phi(X) + JXP_q + P_q^\top XJ - JXJ.\end{aligned}$$

IV. Proofs

The next proposition is the backbone of Theorem 1.

Proposition 2: Let P be a symmetric, doubly stochastic matrix. Then

$$\mathrm{Tr}\,(\Phi^*)^t(I-J) \leq N(\lambda_{q,2}^2 + q^2(1-\lambda_2^2))^t.$$

Proof: Due to the properties of Φ^* the subspace $\mathcal{X}_0 := \{X \in \mathbb{R}^{N\times N} : X = X^\top, XJ = 0\}$ is invariant under its action. For the target expression we see

$$\begin{aligned}\mathrm{Tr}\,(\Phi^*)^t(I-J) &\leq N\|(\Phi^*)^t(I-J)\|_2\\ &\leq N\|(\Phi^*)^t\|_{\mathcal{X}_0\to\mathcal{X}_0}\|(I-J)\|_2 \leq N\|\Phi^*\|^t_{\mathcal{X}_0\to\mathcal{X}_0}\end{aligned}$$

where $\|\Phi^*\|_{\mathcal{X}_0\to\mathcal{X}_0} = \max\{\|\Phi^*(X)\|_2 : \|X\|_2 \leq 1, X \in \mathcal{X}_0\}$. It is easy to show that $\arg\max\|\Phi^*(X)\|_2 \geq 0$: let $\widetilde{X} = \arg\max\|\Phi^*(X)\|_2$ and let us consider its decomposition $\widetilde{X} = X^+ - X^-$ with $X^+, X^- \geq 0$. Due to Φ^* keeping the semi-definite property we have

$$\begin{aligned}v^\top\Phi^*(\widetilde{X})v &= v^\top\Phi^*(X^+)v - v^\top\Phi^*(X^-)v\\ &\leq v^\top\Phi^*(X^+)v + v^\top\Phi^*(X^-)v,\end{aligned}$$

which would lead to a contradiction if X^- was not 0. Thus

$$\begin{aligned}\|\Phi^*\|_{\mathcal{X}_0\to\mathcal{X}_0} = \max\{v^\top\Phi^*(X)v : X \in \mathcal{X}_0, X \geq 0,&\\ \|X\|_2 \leq 1, \|v\| \leq 1, v \perp \mathbf{1}\}.&\end{aligned}$$

meaning that it is enough to bound $v^\top\Phi^*(X)v$ from above. Let $v \perp \mathbf{1}$ and $X \in \mathcal{X}_0$ with $\|X\|_2 \leq 1$, then

$$\begin{aligned}v^\top\Phi^*(X)v = (P_q v)^\top X P_q v&\\ + q^2\Big(\sum_{i,j} p_{ji}x_{ii}v_j^2 - \sum_i (Pv)_i^2 x_{ii}\Big).&\end{aligned}$$

Due to the conditions imposed on X, we have $x_{ii} \in [0,1]$, furthermore $\sum_j p_{ji} - (Pv)_i^2 \geq 0$ for any i, thus

$$\begin{aligned}v^\top\Phi^*(X)v &\leq \|P_q v\|^2 + q^2\big(\sum_{i,j} p_{ji}v_j^2 - \sum_i (Pv)_i^2\big)\\ &= v^\top\big(P_q^2 + q^2(I-P^2)\big)v\end{aligned}$$

where it is easy to check that the second eigenvector is the maximizer, leading to the upper bound of the statement. ■

With all the tools at our hands we can prove Theorem 1.

Proof: [of Theorem 1] According to Lemma 10 in [15] whose assumptions are clearly satisfied we have

$$\limsup_t \frac{1}{t}\log\frac{1}{\min_i w_i(t)} \leq 0,$$

thus considering the quantity in (3) we can infer that

$$\begin{aligned}\limsup_t \frac{1}{t}\log\left(\frac{1}{\min_i w_i(t)}||x_0||\cdot||H(t)(I-J)||_F\right)&\\ \leq \limsup_t \frac{1}{t}\log||H(t)(I-J)||_F.&\end{aligned}$$

The matrix $H(t)(I-J)$ can we written as a product of the i.i.d. random matrices $A(t)(I-J)$, therefore due to the Fürstenberg-Kesten theorem it follows that

$$\begin{aligned}&\limsup_t \frac{1}{t}\log||H(t)(I-J)||_F = \lim_t \frac{1}{t}\mathbb{E}\log||H(t)(I-J)||_F\\ &\leq \lim_t \frac{1}{2t}\log\mathbb{E}\|H(t)(I-J)\|_F^2,\end{aligned}$$

where we applied Jensen's inequality implicitly for the squares of the norm inside the logarithm. By expressing the expectation we obtain

$$\mathbb{E}||H(t)(I-J)||_F^2 = \mathrm{Tr}\,(\Phi^*)^t(I-J).$$

The rate of decay of the r.h.s. can be controlled using Proposition 2:

$$\begin{aligned}&\limsup_t \frac{1}{2t}\log\mathrm{Tr}\,(\Phi^*)^t(I-J)\\ &\leq \frac{1}{2}\big((1-q)^2 + 2q(1-q)\lambda_2 + q^2\big).\end{aligned}$$

Linking the series of calculations above we conclude. ■

Now we will turn to the case when the underlying system is transitive. In this scenario it is possible to give stronger bounds, but before doing so we need to reformulate the problem. Let us define $X_t = (\Phi^*)^t(I - J)$. The following two statements will help us to understand this process.

Lemma 2: Assume that P is the kernel of a symmetric transitive Markov chain, implying that any diagonal element $p_{ii}^{(t)}$ of P^t depends solely on t and not on i. Then we have

1) $X_t \in \mathcal{P}_t := \mathrm{Span}\{P^k, J \;;\; 0 \le k \le 2t\}$, hence the diagonal elements of X_t depend solely on t,
2) $X_{t+1} = P_q X_t P_q + q^2 r_t (I - P^2)$, where r_t denotes the common diagonal element of X_t.

Proof: We prove by induction. For $t = 0$, $X_0 = I - J$, trivially a polynomial of P and J. For the induction step $t \to t + 1$ assume that $X_t \in \mathcal{P}_t$, then X_{t+1} is equal to

$$\Phi^*(X_t) = P_q X_t P_q + q^2\{(\Psi^- P \Psi X_t) - P(\Psi^- \Psi X_t)P\}$$

and, since X_t is a polynomial of P and J, it is also transitive. Noting $\Psi X_t = r_t \mathbf{1}$ and $\Psi^- \Psi X_t = r_t I$, where r_t denotes the common diagonal element of X_t, we can derive the recursion

$$X_{t+1} = P_q X_t P_q + q^2 r_t (I - P^2), \tag{5}$$

which shows that $X_{t+1} \in \mathcal{P}_{t+1}$. ■

As a consequence, if P is symmetric and transitive then X_t and P possess the same eigenvectors.

Proposition 3: If P is symmetric and transitive and v is an eigenvector of P and X_t corresponding to the eigenvalue λ and μ_t respectively, then the following recursion holds:

$$\mu_{t+1} = \lambda_q^2 \mu_t + q^2 r_t (1 - \lambda^2), \tag{6}$$

where r_t, same as before, denotes the common diagonal element of X_t. Furthermore the largest eigenvalue of X_t is asymptotically bounded:

$$\limsup_t \frac{1}{t} \log \max_i \mu_{t,i} \le \log \zeta_2,$$

where ζ_2 is the second largest root of the polynomial

$$p(x) = \prod_i (x - \lambda_{q,i}^2)\left(1 + \frac{q^2}{N}\sum_{j>1} \frac{1 - \lambda_j^2}{x - \lambda_{q,j}^2}\right).$$

Proof: Using $r_t = \frac{1}{N}\mathrm{Tr}\, X_t = \frac{1}{N}\sum_i \mu_{t,i}$ we can write the recursion described in (6) for all eigenvalues jointly as

$$\boldsymbol{y}_{t+1} = \left(D + \frac{q^2}{N}\mathbf{b}\mathbf{1}^\top\right)\boldsymbol{y}_t \tag{7}$$

with the vectors $(\boldsymbol{y}_t)_i = \mu_{t,i}$, $\mathbf{b}_i = 1 - \lambda_i^2$, and $D = \mathrm{diag}\,(\lambda_{q,1}^2, \ldots, \lambda_{q,N}^2)$, where $\lambda_{q,i}$ denotes the i^{th} largest eigenvalue of P_q and $\mu_{t,i}$ denotes the eigenvalue of X_t corresponding to the same eigenvector. In this case $\mu_{t,1} = 0$, since $\lambda_{q,1} = 1$ and $X_t \mathbf{1} = 0$ for any t.

Let (λ, v) denote an eigen-pair of P. According to the previous lemma X_t is a polynomial of P and J, hence v is an eigenvector of X_t, furthermore, due to the symmetry of P, we have $JP = PJ = J$. Recursion (5) then yields

$$\begin{aligned} X_{t+1} v &= P_q X_t P_q v + q^2 r_t (I - P^2) v \\ &= \lambda_q^2 \mu_t v + q^2 r_t (1 - \lambda^2) v \end{aligned}$$

and we have

$$\mu_{t+1} = \lambda_q^2 \mu_t + q^2 r_t (1 - \lambda^2),$$

proving the first part.

Before continuing with the proof of the second part, let us note that e_1 is a left eigenvector of the matrix $D + q^2 \mathbf{b}\mathbf{1}^\top/N$ corresponding to the eigenvalue 1, meaning that the right eigenvectors corresponding to different eigenvalues are orthogonal to e_1. Let us choose the following new basis: $f_1 = e_1$, $f_i = e_i - e_1, i > 1$, then for $i > 1$ we have $f_i \perp \mathbf{1}$. The matrix $D + q^2\mathbf{b}\mathbf{1}^\top/N$ of the recursion in the basis $\{f_i\}$ can be described by the block matrix

$$\begin{bmatrix} q^2/N \sum_j b_j + \lambda_{q,1}^2 & \ell^\top \\ q^2/N (b_i)_{i\ge 2} & D_{N-1} \end{bmatrix} \tag{8}$$

where $\ell = (\lambda_{q,i}^2 - \lambda_{q,1}^2)_{i \ge 2}$ and D_{N-1} is the lower right $(N-1) \times (N-1)$ submatrix of D.

The characteristic polynomial of the matrix in (8) can be computed via expanding along the first column, leading to

$$\begin{aligned} p(x) &= \left(x - \frac{q^2}{N}\sum_i b_i - \lambda_{q,1}^2\right)\prod_{i>1}(x - \lambda_{q,i}^2) \\ &\quad + \sum_{i>1} \frac{q^2}{N} b_i (\lambda_{q,1}^2 - \lambda_{q,i}^2) \prod_{j:1<j\ne i} (x - \lambda_{q,j}^2) \\ &= \prod_{i>1}(x - \lambda_{q,i}^2)\Bigg\{x - \lambda_{q,1}^2 - \frac{q^2}{N}\sum_i (1 - \lambda_i^2) \\ &\quad + \frac{q^2}{N}\sum_{j>1}(1 - \lambda_j^2)\frac{(1 - \lambda_{q,j}^2)}{x - \lambda_{q,j}^2}\Bigg\}. \end{aligned}$$

Exploiting the fact $\lambda_1 = \lambda_{q,1} = 1$ we obtain for $p(x)$

$$\prod_{i>1}(x - \lambda_{q,i}^2)\left\{x - 1 + \frac{q^2}{N}\sum_{j>1}\left[(1 - \lambda_j^2)\left(\frac{1 - \lambda_{q,j}^2}{x - \lambda_{q,j}^2} - 1\right)\right]\right\}$$

$$= \prod_i (x - \lambda_{q,i}^2)\left(1 - \frac{q^2}{N}\sum_{j>1}\frac{1 - \lambda_j^2}{x - \lambda_{q,j}^2}\right).$$

We note here that due the initialization $\boldsymbol{y}_0 = (0, 1, \ldots, 1)^\top$ and the fact that e_1 is a left eigenvector of $D + q^2\mathbf{b}\mathbf{1}^\top/N$, we have for each subsequent vector $\boldsymbol{y}_t \perp e_1$. This means that we are only interested in the second largest root ζ_2 of the characteristic polynomial to obtain the asymptotic growth rate of μ_t, i.e. $||\mu_t||_\infty \le C\zeta_2^t\, ||\mu_0||_\infty$ with some $C > 0$. ■

Proof: [of Theorem 2] The first part of the proof, up until the point where we have to bound the quantity $\mathbb{E}||(I - J)H(t)||_F^2$ from above, is analogous to the proof of Theorem 1, hence it will be omitted here. To relate the process investigated, observe

$$\begin{aligned} \mathbb{E}||H(t)(I - J)||_F^2 &= \mathrm{Tr}\,\{(\Phi^*)^t(I - J)\} \\ &= \sum_j \mu_{t,j} \le N \max_j \mu_{t,j}, \end{aligned}$$

with $\mu_{t,j}$ as in Proposition 3. Due to $X_0 = I - J$ and $\lambda_1 = 1$ with $v_1 = c\mathbf{1}$, we have $\mu_{0,1} = 0$ and according to recursion (6), $\mu_{t,1} = 0$ for any $t > 0$. In Proposition 3 it was shown that

$||\mu_t||_\infty \leq C\zeta_2^t||\mu_0|| = C\zeta_2^t$, where ζ_2 denoted the second largest root of the polynomial

$$p(x) = \prod_i (x - \lambda_{q,i}^2) - \frac{q^2}{N}\sum_{i>1}(1-\lambda_i^2)\prod_{j:j\neq i}(x-\lambda_{q,j}^2).$$

Exploiting the fact that 1 is a root of p we have

$$\frac{p(x)}{x-1} = \prod_{i>1}(x - \lambda_{q,i}^2) - \frac{q^2}{N}\sum_{i>1}(1-\lambda_i^2)\prod_{j:1<j\neq i}(x-\lambda_{q,j}^2),$$

thus by denoting the largest root of $p(x)/(x-1)$ with ξ_1 we can write $||\mu_t||_\infty \leq C\xi_1^t$. Altogether

$$\limsup_t \frac{1}{2t}\log \mathbb{E}||H(t)(I-J)||_F^2 \leq \frac{1}{2}\log \xi_1,$$

and this concludes the proof. ■

V. Numerical experiments

In order to evaluate the performance and computational cost of the bounds obtained on the convergence rate, we performed a numerical comparison of the available quantities. For the simulated process we set $t = 1500$ and, for complexity and memory usage reduction, we used the modified rate approximation

$$\zeta := \frac{1}{t}\log\left\|\frac{1}{\sqrt{M}}[\Psi^-(w(t))]^{-1}H(t)X^M(0)\right\|_F,$$

where $X^M(0)$ is an $N \times M$ matrix of uniform random independent columns in $\mathbf{1}^\perp$ with unit norm, representing various initializations, reaching the principal rate with high probability. We chose $M = \lfloor\sqrt{N}\rfloor$. This was compared with

- $\eta_2/2 := \log\rho\big(\mathbb{E}(A(1)^{\otimes 2})(I-J)^{\otimes 2}\big)/2$ from [15],
- $\gamma/2$ of Theorem 1 and $\alpha/2$ of Theorem 2.

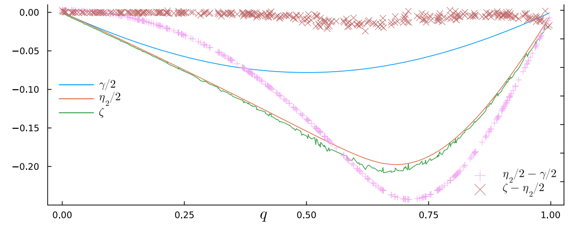
The underlying topologies were random transitive graphs generated as Cayley-graphs of the symmetric groups up to S_5 with 3 random generators, and random regular graphs of matching sizes and degrees. In all cases the recipients were chosen uniformly at random.

In case of the transitive graph we can see a tight fit by the bound of Theorem 2 through $q \in (0,1)$ as shown in Figure 1b. In fact, we recover $\eta_2/2$ which is expected from the exact analysis carried out in the proof, and the errors from the simulated rates ζ are an order less than the rates themselves. In the case of symmetric P, Theorem 1 provides a still usable bound as shown in Figure 1a. The resulting curve captures the linear trend when q is near 0, it is a factor ≈ 2 off from the numerical value when $q \leq 0.5$ and then deteriorates.

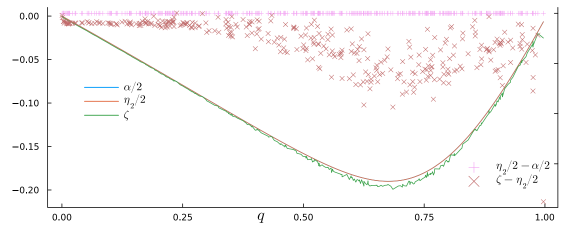
The most distinctive difference lies in the runtime (and memory usage) of the methods, shown in Table I, including smaller instances. The simulations were carried out on AMD EPYC 7643 CPUs using the Julia computing language [18].

	our bounds $\gamma/2, \alpha/2$, resp.			$\eta_2/2$		
N	24	60	120	24	60	120
reg. g.	0.0012	0.0029	0.0071	0.155	1.71	15.33
trans. g.	0.0097	0.0119	0.0358	0.132	1.50	24.85

TABLE I: Runtimes[sec] of computing the different bounds



(a) Regular case, $N = 120$, $d = 6$, with $q \in (0,1)$.



(b) Transitive case, $N = 120, d = 6$, with $q \in (0,1)$.

Fig. 1: Bounds and simulated rates

VI. Summary and future plans

In this work we have presented multiple bounds with accuracy depending on the level of symmetry of the underlying topology. For $N = 120$ vertices, the computational cost of these bounds are ≈ 3 orders of magnitude less than that of computing $\eta_2/2$ from [15] or [12] confirming their usefulness in assessing the efficiency of various push-sum algorithms.

Along the way proving our main results we have developed a framework as described in Section III relying on matrix operators Φ, Φ^* that we hope can be useful when analyzing similar dynamics.

Concerning our future plans, we observed in some numerical experiments that the expression in (2) was also a valid bound for regular graphs. This lead us to state the following:

Conjecture 1: The bound given in 2 provides a useful approximation in the simple symmetric case, e.g. for arbitrary regular graphs. Possibly not an exact upper bound but a quantity within small controllable error of the true rate.

Appendix

Proof: [of Lemma 1] Let $L := \sum_i e_{\beta_i}e_i^\top$ then

$$\begin{aligned}\mathbb{E}[A^\top XA] &= (1-q)^2X + q(1-q)\mathbb{E}[XL]\\ &\quad + q(1-q)\mathbb{E}[L^\top X] + q^2\mathbb{E}[L^\top XL]\\ &= P_q^\top XP_q - q^2P^\top XP + q^2\mathbb{E}[L^\top XL].\end{aligned}$$

Next we will compute the term $\mathbb{E}[L^\top XL]$ as

$$\begin{aligned}\mathbb{E}[L^\top XL] &= \sum_{i,i'}\mathbb{E}[e_ie_{\beta_i}^\top Xe_{\beta_{i'}}e_{i'}^\top]\\ &= \sum_{i\neq i'}\mathbb{E}x_{\beta_i,\beta_{i'}}e_ie_{i'}^\top + \sum_i \mathbb{E}x_{\beta_i,\beta_i}e_ie_i^\top\\ &= \sum_{\substack{i\neq i'\\ j,j'}}p_{ji}p_{j'i'}x_{jj'}e_ie_{i'}^\top + \sum_{i,j}p_{ji}x_{jj}e_ie_i^\top\\ &= P^\top XP - \sum_{i,j,j'}p_{ji}p_{j'i}x_{jj'}e_ie_i^\top + \sum_{i,j}p_{ji}x_{jj}e_ie_i^\top\\ &= P^\top XP - \Psi^-\Psi(P^\top XP) + \Psi^-(P^\top\ \Psi(X)).\end{aligned}$$

Thus putting together the two parts reads

$$\Phi(X) = P_q^\top X P_q + q^2\{\Psi^-(P^\top \Psi(X)) - \Psi^-\Psi(P^\top X P)\}$$

and this concludes the proof. ■

Proof: [of Proposition 1] The first three properties follow directly from the definition of Φ, their proofs are left to the respected reader.

Property (P4) can be proved as follows. Let $X \geq 0$ and let w be an arbitrary vector, then

$$w^\top \Phi(X) w = w^\top \mathbb{E}[A^\top X A] w = \mathbb{E}[(Aw)^\top X A w] \geq 0.$$

(P5) is analogous to (P4), i.e. $\Phi(X)_{kl} = (\mathbb{E}[A^\top X A])_{kl} \geq 0$.

Property (P6) is the result of a short series of calculations.

$$\begin{aligned}\Phi(J) &= P^\top J P + q^2(\Psi^- P^\top \Psi J - \Psi^-\Psi(P^\top J P)) \\ &= J + q^2(1/N \cdot I - 1/N \cdot I) = J,\end{aligned}$$

due to the facts $JP = P^\top J = J$ and $\Psi J = \mathbf{1}/N$.

Before proving (P7) let us note that due to $X \geq 0$ and the linearity of Φ it is enough to prove this property for $X = xx^\top$. Using the definition of $A = (1-q)I + q^2 \sum_i e_{\beta_i} e_i^\top$ we have

$$\begin{aligned}\mathrm{Tr}\,\Phi(xx^\top) &= \mathrm{Tr}\,\mathbb{E}\{(1-q)^2 xx^\top \\ &\quad + q(1-q)(L^\top xx^\top + xx^\top L) + q^2 L^\top xx^\top L\} \\ &= (1-q)^2 ||x||_2^2 \\ &\quad + 2q(1-q) x^\top P^\top x + q^2 \sum_{i,j} p_{ji} x_j^2 \leq ||x||_2^2\end{aligned}$$

where in the last step we used the facts $P = P^\top$, $P\mathbf{1} = \mathbf{1}$ and $x^\top P x \leq \lambda_1(P)||x||_2^2 = ||x||_2^2$. Now we proceed with proving the properties of the adjoint.

The proof of (P*1) is based on the following series of calculations. Due to the equivalences

$$\begin{aligned}&\mathrm{Tr}\,\{\Psi^-(P^\top\ \Psi X) Y^\top\} = \sum_{i,k} p_{ki} x_{kk} y_{ii} \\ &= \sum_k x_{kk} \sum_i p_{ki} y_{ii} = \mathrm{Tr}\,\{X \Psi^-(P \Psi Y^\top)\}, \\ &\mathrm{Tr}\,\{\Psi^-\Psi(P^\top X P) Y^\top\} = \sum_{i,k,l} p_{ki} x_{kl} p_{li} y_{ii} \\ &= \sum_{k,l} x_{kl} \sum_i p_{ki} y_{ii} p_{li} = \mathrm{Tr}\,\{X P \Psi(Y^\top) P^\top\}\end{aligned}$$

we have

$$\begin{aligned}&\langle \Phi(X), Y\rangle = \mathrm{Tr}\,\{\Phi(X) Y^\top\} = \mathrm{Tr}\,\{P_q^\top X P_q Y^\top\} \\ &\quad + q^2 \mathrm{Tr}\left\{[\Psi^-(P^\top \Psi(X))] Y^\top - [\Psi^-\Psi(P^\top X P)] Y^\top\right\} \\ &= \mathrm{Tr}\,\{X P_q Y^\top P_q^\top\} + q^2 \mathrm{Tr}\,\{X \Psi^-(P \Psi Y^\top) \\ &\quad - X P \Psi(Y^\top) P^\top\} = \langle X, \Phi^*(Y)\rangle.\end{aligned}$$

Properties (P*2) and (P*3) can be confirmed analogously to (P4), (P5).

(P*4) is a result of the short derivation

$$\begin{aligned}\Phi^*(X)\mathbf{1} &= P_q X P_q^\top \mathbf{1} + q^2(\Psi^- P \Psi X \mathbf{1} - P(\Psi^-\Psi X) P^\top \mathbf{1}) \\ &= 0 + q^2(\Psi^-(P\Psi(X))\mathbf{1} - P(\Psi^-\Psi X)\mathbf{1})\end{aligned}$$

since $P_q^\top \mathbf{1} = P^\top \mathbf{1} = \mathbf{1}$ and we assumed $X\mathbf{1} = 0$. For the second term we have

$$\begin{aligned}(\Psi^-(P\Psi(X))\mathbf{1})_i &= \sum_j p_{ij} x_{jj} \\ (P(\Psi^-\Psi X)\mathbf{1})_i &= \sum_j p_{ij} x_{jj},\end{aligned}$$

so $\Phi^*(X)\mathbf{1} = 0$. This concludes the proof. ■

References

[1] V. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *44th IEEE Conference on Decision and Control*, pp. 2996–3000, 2005.

[2] J. N. Tsitsiklis, *Problems in decentralized decision making and computation*. PhD thesis, Massachusetts Institute of Technology, 1984.

[3] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 225–230, 2009.

[4] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[5] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proceedings of 44th Annual IEEE Symposium on Foundations of Computer Science*, pp. 482–491, 2003.

[6] C. N. Hadjicostis and T. Charalambous, "Average consensus in the presence of delays in directed graph topologies," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 763–768, 2014.

[7] F. Bénézit, V. Blondel, P. Thiran, J. N. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proceedings of 2010 IEEE International Symposium on Information Theory (ISIT)*, pp. 1753–1757, 2010.

[8] D. Kempe and F. McSherry, "A decentralized algorithm for spectral analysis," *Journal of Computer and System Sciences*, vol. 74, no. 1, pp. 70–83, 2008.

[9] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2014.

[10] C. N. Hadjicostis, N. H. Vaidya, and A. D. Domínguez-García, "Robust distributed average consensus via exchange of running sums," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1492–1507, 2015.

[11] A. Olshevsky, I. C. Paschalidis, and A. Spiridonoff, "Fully asynchronous push-sum with growing intercommunication intervals," in *2018 Annual American Control Conference (ACC)*, pp. 591–596, IEEE, 2018.

[12] F. Iutzeler, P. Ciblat, and W. Hachem, "Analysis of sum-weight-like algorithms for averaging in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2802–2814, 2013.

[13] B. Gerencsér and L. Gerencsér, "Tight bounds on the convergence rate of generalized ratio consensus algorithms," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1669–1684, 2022.

[14] J. N. Tsitsiklis and V. Blondel, "The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate," *Mathematics of Control, Signals and Systems*, vol. 10, no. 1, pp. 31–40, 1997.

[15] B. Gerencsér, "Computable convergence rate bound for ratio consensus algorithms," *IEEE Control Systems Letters*, vol. 6, pp. 3307–3312, 2022.

[16] X. Dai, J. Xu, and A. Zhou, "Convergence and optimal complexity of adaptive finite element eigenvalue computations," *Numerische Mathematik*, vol. 110, no. 3, pp. 313–355, 2008.

[17] J. Lee, V. Balakrishnan, C.-K. Koh, and D. Jiao, "From o(k 2 n) to o(n): A fast complex-valued eigenvalue solver for large-scale on-chip interconnect analysis," in *2009 IEEE MTT-S International Microwave Symposium Digest*, pp. 181–184, IEEE, 2009.

[18] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.