# Accelerated Gossip in Networks of Given Dimension Using Jacobi Polynomial Iterations[*]

### Raphaël Berthier[†], Francis Bach[†], and Pierre Gaillard[†]

**Abstract.** Consider a network of agents connected by communication links, where each agent holds a real value. The gossip problem consists in estimating the average of the values diffused in the network in a distributed manner. We develop a method for solving the gossip problem that depends only on the spectral dimension of the network, that is, in the communication network set-up, the dimension of the space in which the agents live. This contrasts with previous work that required the spectral gap of the network as a parameter, or suffered from slow mixing. Our method shows an important improvement over existing algorithms in the nonasymptotic regime, i.e., when the values are far from being fully mixed in the network. Our approach stems from a polynomial-based point of view on gossip algorithms, as well as an approximation of the spectral measure of the graphs with a Jacobi measure. We show the power of the approach with simulations on various graphs, and with performance guarantees on graphs of known spectral dimension, such as grids and random percolation bonds. An extension of this work to distributed Laplacian solvers is discussed. As a side result, we also use the polynomial-based point of view to show the convergence of the message passing algorithm for gossip of Moallemi and Van Roy on regular graphs. The explicit computation of the rate of the convergence shows that message passing has a slow rate of convergence on graphs with small spectral gap.

**Key words.** gossip, consensus, averaging, multiagent, distributed, polynomial iterations, Jacobi orthogonal polynomials, Laplacian solvers

**AMS subject classifications.** 90C35, 68M14, 68W15

**DOI.** 10.1137/19M1244822

**1. Introduction.** The averaging problem, or gossip problem, is a fundamental primitive of distributed algorithms. Given a network composed of agents and undirected communication links between them, we assign to each agent $v$ a real value $\xi_v$, called an observation. The goal is to design an iterative communication procedure allowing each agent to know the average of the initial observations in the network as quickly as possible.

The landmark paper [6] suggests the natural following protocol to solve the averaging problem: at each iteration, each agent replaces his current observation by some average of the observations of its neighbors in the network. We will refer to this method in the following by the term *simple gossip*. More precisely, we are given a weight matrix $W = (W_{v,w})_{v,w \in V}$, called the gossip matrix, indexed by the vertices $v \in V$ of the network graph, satisfying the

[†]Inria and École Normale Supérieure, PSL Research University, 75006 Paris, France (raphael.berthier@inria.fr, francis.bach@inria.fr, pierre.gaillard@inria.fr).

property that $W_{v,w}$ is nonzero only if $v \sim w$, that is $v$ and $w$ are connected in the graph. Then the simple gossip iteration writes

$$(1.1) \qquad x_v^0 = \xi_v \,, \qquad\qquad x_v^{t+1} = \sum_{w:w \sim v} W_{v,w} x_w^t \,, \qquad t \geqslant 0 \,.$$

The paper [6] proves the linear convergence of the observations to their average.

However, the rate of the linear convergence was shown to worsen significantly in many networks of interest as the size of the network increases. More precisely, define the diameter $D$ of the network as the largest number of communication links needed to connect any two agents. While, obviously, $D$ steps of averaging are needed for any gossip method to spread information in the network, the simple gossip method may require up to $\Theta(D^2)$ communication steps to estimate the average, as for instance on the line graph, the two-dimensional grid, or the random geometric graph (see [8] or [6, section IV.A]). To reach the $O(D)$ bound, a diverse set of ideas were proposed, including second-order recursions [7, 25], message passing algorithms [18], lifted Markov chain techniques [28], methods using Chebychev polynomial iterations [2, 27], inspiration arising from advection-diffusion processes [26], and geographic gossip [10], a method that uses the knowledge of the location of the agents on a field. To the best of our knowledge, all of these accelerated methods assume that the agents hold additional information about the network graph, such as its spectral gap. For instance, second-order methods typically take the form (see [7])

$$(1.2) \qquad \begin{aligned} & x_v^0 = \xi_v \,, \qquad x_v^1 = \sum_{w:w \sim v} W_{v,w} x_w^0 \,, \\ & x_v^{t+1} = \omega \sum_{w:w \sim v} W_{v,w} x_w^t + (1-\omega) x_v^{t-1} \,, \qquad t \geqslant 1 \,, \end{aligned}$$

where $\omega$ is some simple function of the spectral gap $\gamma$, that is the distance between the largest and the second largest eigenvalues of $W$. This iteration obtains optimal asymptotic convergence on many graphs, with a relaxation time of the linear convergence on the order of $1/\sqrt{\gamma}$ as $\gamma \to 0$.

In this paper, we develop a gossip method based not on the spectral gap $\gamma$ but on the density of eigenvalues of $W$ near the upper edge of the spectrum. Looking at the upper part of the spectrum at a broader scale allows us to improve the local averaging of the gossip algorithm in the regime $t < 1/\sqrt{\gamma}$. This improvement is worthwhile as the spectral gap $\gamma$ can get arbitrarily small in large graphs. For instance, in the case of the line graph or the two-dimensional grid, the relaxation time $1/\sqrt{\gamma}$ is of the order of the diameter $D$ of the graph. Thus the regime $t < 1/\sqrt{\gamma}$ can be relevant for applications.

Remarkably, the spectral density of $W$ near the upper edge can be described by a very natural parameter: the spectral dimension $d$. The network is of spectral dimension $d$ if the number of eigenvalues of $W$ in $[1 - \Lambda, 1]$ is of the order of $\Lambda^{d/2}$ for small $\Lambda$ ($\gamma \ll \Lambda \ll 1$); see section 5.3 for rigorous definitions. We will see with examples that this definition coincides with our intuition of the dimension of the graph, which is the dimension of the manifold on which the agents live. For instance, the grid with nodes $\mathbb{Z}^d$ where the nodes at distance 1 are connected is a graph of dimension $d$. Thus the parameter $d$ is much easier to know than the spectral gap $\gamma$.

In real-world situations, the practitioner reasonably knows if the network on which she implements the gossip method is of finite dimension, and if so, she also knows the dimension $d$. In this paper, we argue that she should run a second-order iteration with time-dependent weights

$$x_v^0 = \xi_v, \qquad x_v^1 = a_0 \sum_{w:w \sim v} W_{v,w} x_w^0 + b_0 x_v^0,$$

(1.3)

$$x_v^{t+1} = a_t \sum_{w:w \sim v} W_{v,w} x_w^t + b_t x_v^t - c_t x_v^{t-1}, \quad t \geqslant 1,$$

where the recurrence weights $a_t, b_t, c_t$ are given by the formulas

(1.4)

$$a_0 = \frac{d+4}{2(2+d)}, \qquad\qquad b_0 = \frac{d}{2(2+d)},$$

$$a_t = \frac{(2t+d/2+1)(2t+d/2+2)}{2(t+1+d/2)^2}, \qquad b_t = \frac{d^2(2t+d/2+1)}{8(t+1+d/2)^2(2t+d/2)},$$

$$c_t = \frac{t^2(2t+d/2+2)}{(t+1+d/2)^2(2t+d/2)}, \qquad t \geqslant 1.$$

The motivation for these choices of weights $a_t, b_t, c_t$ is not obvious at first sight. It follows from a *polynomial-based* point of view on gossip algorithms: it consists in seeing the iterations (1.1), (1.2), and (1.3) as sequences $P_0, P_1, P_2, \ldots$ of polynomials in the gossip matrix $W$. The correspondence is given by the relation $x^t = P_t(W)\xi$ where $x^t = (x_v^t)_{v \in V}$ and $\xi = (\xi_v)_{v \in V}$. This approach is inspired by similar work done in the resolution of linear systems [12] and on the load balancing problem [9]. The choice of an iteration is reframed as the choice of a sequence of polynomials, and the performance of the resulting gossip method depends on the spectrum of $W$. As the dimension of the graph gives the rate of decrease of the spectral density near the edge of the spectrum, it suggests the sequence of polynomials one should take: we choose a parametrized sequence of polynomials called *Jacobi polynomials* that is well known in the literature on orthogonal polynomials (see Definition SM2.2 of the Jacobi polynomials). This actually leads to the iteration (1.3), which we call the *Jacobi polynomial iteration*.

The Jacobi polynomial iteration (1.3) improves the convergence of the gossip method in the transitive phase $t < 1/\sqrt{\gamma}$ but loses the optimal rate of convergence of second-order gossip because it does not use the spectral gap $\gamma$. We argue that in most applications of gossip methods, the asymptotic rate of convergence is not of practical importance, especially if the transient time is long. However, we also build a gossip iteration that uses both parameters $d$ and $\gamma$ and achieves both the efficiency in the nontransitive regime and the fast rate of convergence.

This resolution of the gossip problem with inner-product free polynomial-based iterations is new, and could lead to other interesting algorithms on other types of graphs. Here, the phrase "inner-product free" comes from the literature on polynomial-based iterations for linear systems [12] and refers to the fact that recurrence coefficients $a_t, b_t, c_t$ are computed without using the gossip matrix $W$ (but parametrized using the knowledge of $d$). Indeed, as the

knowledge of the gossip matrix $W$ is distributed across the graph, it would be a challenging distributed problem to compute the recurrence coefficients if they depended on $W$.

Although our work is inspired by iterative methods for linear systems, the Jacobi iteration that we developed for gossip can be transposed into a new idea in this literature which can be useful for the distributed resolution of Laplacian systems over multiagent networks.

**Outline of the paper.** Section 2 sets some notation used in the remainder of the paper. In section 3, we give simulations in different types of networks of dimensions 2 and 3. We show that the recursion (1.3) brings important benefits over existing methods in the nonasymptotic regime, i.e., when the observations are far from being fully mixed in the graph.

In sections 4–5, we develop the derivation of the Jacobi polynomial iteration. Section 4 describes an optimal way to design polynomial-based gossip algorithms, following the lines of [12, 9], and discusses its feasibility. Section 5 uses the notion of spectral dimension of a graph to inspire the practical Jacobi polynomial iteration (1.3).

In section 6, we present the adaptation of the Jacobi polynomial iteration to the case where the spectral gap $\gamma$ of $W$ is given to improve the asymptotic rate of convergence.

In section 7, we describe the parallel between gossip methods and iterative methods for linear systems and discuss the contributions that our work can bring to the distributed resolution of Laplacian systems over networks.

**Code.** The code that generated the simulations is available online [4].

**2. Problem setting.** A network of agents is modeled by an undirected finite graph $G = (V, E)$, where $V$ is the set of vertices of the graph, or agents, and $E$ the set of edges, or communication links. We assume each agent $v$ holds a real value $\xi_v$. Our goal is to design an iterative algorithm that quickly gives each agent the average $\bar{\xi} = (1/n) \sum_{v \in V} \xi_v$, where $n = |V|$ is the number of agents. A fundamental operation to estimate the average $\bar{\xi}$ consists in averaging the observations of neighbors in the network. We formalize this notion using a gossip matrix.

Definition 2.1. *A gossip matrix $W = (W_{v,w})_{v,w \in V}$ on the graph $G$ is a matrix with entries indexed by the vertices of the graph satisfying the following properties:*
  - *$W$ is nonnegative: for all $v, w \in V$, $W_{v,w} \geqslant 0$.*
  - *$W$ is supported by the graph $G$: for all distinct vertices $v, w$ such that $W_{v,w} > 0$, $\{v, w\}$ must be an edge of $G$.*
  - *$W$ is normalized: for all $v \in V$, $\sum_{w \in V} W_{v,w} = 1$.*
  - *$W$ is symmetric: for all $v, w \in V$, $W_{v,w} = W_{w,v}$.*

If $W$ is a gossip matrix and $x = (x_v)_{v \in V}$ is a set of values stored by the agents $v$, the product $Wx$ is interpreted as the computation by each agent $v$ of a weighted average of the values $x_w$ of its neighbors $w$ in the graph (and of its own value $x_v$). This average is computed simultaneously for all agents $v$; indeed, in this paper we deal only with *synchronous* gossip. Note that we do not need the symmetry assumption on $W$ to interpret $W$ as an averaging operation. This assumption is usual in gossip frameworks as it allows one to use the spectral theory for $W$, on which our analysis relies heavily. It appears, for instance, in the works [6, 7, 25].

In a $d$-regular graph $G$ ($\forall v, \deg v = d$), a typical gossip matrix is $W = A(G)/d =$

$(\mathbf{1}_{\{\{v,w\}\in E\}}/d)_{v,w\in V}$, where $A(G)$ is the adjacency matrix of the graph. More generally, if the graph has all vertices of degree bounded by some quantity $d_{\max}$, then a natural gossip matrix is

$$(2.1) \qquad\qquad W = I + \frac{1}{d_{\max}}(A - D),$$

where $D$ is the degree matrix, which is the diagonal matrix such that $D_{v,v} = \deg v$.

Note that any gossip matrix $W$ is an operator on $\mathbb{R}^V$ bounded by 1. Indeed, if $x \in \mathbb{R}^V$, by Jensen's inequality,

$$\|Wx\|_2^2 = \sum_{v\in V} x_v^2 = \sum_{v\in V}\left(\sum_{w\in V} W_{v,w}x_w\right)^2 \leqslant \sum_{v\in V}\sum_{w\in V} W_{v,w}x_w^2 = \sum_{w\in V} x_w^2 = \|x\|_2^2\,.$$

**Definition 2.2** (spectral gap). *Denote by $1 \geqslant \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant -1$ the real eigenvalues of the symmetric matrix $W$. As $W$ is normalized, $W\mathbf{1} = \mathbf{1}$; we can take $\lambda_1 = 1$, which corresponds to the eigenvector $\mathbf{1} = (1,\ldots,1)$. We define*
   1. *the spectral gap $\gamma = 1 - \lambda_2$ as the distance between the two largest eigenvalues of $W$;*
   2. *the absolute spectral gap $\tilde\gamma = \min(1 - \lambda_2, \lambda_n + 1)$ as the difference between the moduli of the two largest eigenvalues of $W$ in magnitude.*

We now discuss different iterations for the gossip problem.

**Simple gossip.** Simple gossip is a natural algorithm solving the gossip problem that consists in repeatedly averaging values in the graph [6]. More precisely, we choose a gossip matrix $W$ on the graph $G$, initialize $x^0 = \xi = (\xi_v)_{v\in V}$, and, at each communication round $t$, compute

$$(2.2) \qquad\qquad x^{t+1} = Wx^t\,.$$

Note that the latter equation is simply a compact rewriting of (1.1). We can rewrite this iteration as $x^t = W^t\xi$. Note that in this last equation, we used the notation $.^t$ to denote both the index of $x$ and the power of the square matrix $W$. We will frequently make use of the indexation $.^t$ when vectors indexed by the vertices (or the edges) also depend on time.

We describe the speed of convergence of this method using ideas from [6].

**Proposition 2.3.** *Let $\xi$ be an arbitrary family of initial observations and $x^t$ the iterates of simple gossip defined by (2.2). Denote by $\tilde\gamma$ the absolute spectral gap of $W$. Then*

$$\limsup_{t\to\infty} \left\|x^t - \bar\xi\mathbf{1}\right\|_2^{1/t} \leqslant 1 - \tilde\gamma\,.$$

*Moreover, the upper bound is reached if and only if there exists an eigenvector $u$ of $W$, corresponding to an eigenvalue of magnitude $1 - \tilde\gamma$, such that $\langle\xi, u\rangle \neq 0$.*

*Proof.* Let $u^1 = \mathbf{1}/\sqrt{n}, u^2, \ldots, u^n$ be the eigenvectors of $W$ associated to the eigenvalues

$\lambda_1, \ldots, \lambda_n$, normalized such that $\|u^i\|_2 = 1$. Then

$$\|x^t - \bar{\xi}\mathbf{1}\|_2^2 = \|W^t\xi - \bar{\xi}\mathbf{1}\|_2^2 = \left\|\sum_{i=1}^n \lambda_i^t \langle \xi, u^i \rangle u^i - \langle \xi, u^1 \rangle u^1\right\|_2^2 = \left\|\sum_{i=2}^n \lambda_i^t \langle \xi, u^i \rangle u^i\right\|_2^2$$

$$\leqslant \sum_{i=2}^n \left\|\lambda_i^t \langle \xi, u^i \rangle u^i\right\|_2^2 = \sum_{i=2}^n \lambda_i^{2t} \langle \xi, u^i \rangle^2 \leqslant (1-\tilde{\gamma})^{2t} \sum_{i=2}^n \langle \xi, u^i \rangle . \qquad \blacksquare$$

**Shift-register gossip.** Several acceleration schemes of gossip [7, 25] store some past iterates to compute higher-order recursions (that thus depend on powers of $W$). For instance, the shift-register iteration of [7] is of the form

$$(2.3) \qquad x^0 = \xi, \qquad\qquad x^1 = W\xi, \qquad\qquad x^{t+1} = \omega W x^t + (1-\omega)x^{t-1},$$

where $\omega$ is a parameter that needs to be tuned.

Proposition 2.4 (from [16, Theorem 2]). *Let $\xi$ be an arbitrary family of initial observations and $x^t$ the iterates of shift-register gossip defined in* (2.3) *with parameter*

$$\omega = 2\frac{1 - \sqrt{\tilde{\gamma}(1 - \tilde{\gamma}/4)}}{(1 - \tilde{\gamma}/2)^2},$$

*where $\tilde{\gamma}$ is the absolute spectral gap of the gossip matrix $W$. Then*

$$\limsup_{t\to\infty} \|x^t - \bar{\xi}\mathbf{1}\|_2^{1/t} \leqslant 1 - 2\frac{\sqrt{\tilde{\gamma}(1 - \tilde{\gamma}/4)} - \tilde{\gamma}/2}{1 - \tilde{\gamma}}.$$

*Moreover, the upper bound is reached if and only if there exists an eigenvector $u$ of $W$, corresponding to an eigenvalue of magnitude $1 - \tilde{\gamma}$, such that $\langle \xi, u \rangle \neq 0$.*

The important consequence of this result is that the rate of convergence of the shift-register method behaves like $1 - 2\sqrt{\tilde{\gamma}} + o(\sqrt{\tilde{\gamma}})$ as $\tilde{\gamma} \to 0$. This differs from simple gossip where the rate of convergence behaves like $1 - \tilde{\gamma}$. This means that in graphs with a small spectral gap, shift-register enjoys a much better rate of convergence than simple gossip: this is why we say that shift-register enjoys an *accelerated* rate of convergence as opposed to simple gossip which has a *diffusive* or *unaccelerated* rate. This effect on the asymptotic rate of convergence can be seen in Figures 2 and 3.

**Polynomial gossip.** More abstractly, we define a polynomial gossip method as any method combining the past iterates of the simple gossip method:

$$(2.4) \qquad\qquad\qquad x^t = P_t(W)\xi,$$

where $P_t$ is a polynomial of degree smaller than or equal to $t$ satisfying $P_t(1) = 1$. The constraint $P_t(1) = 1$ ensures that $x^t = \bar{\xi}\mathbf{1}$ if all initial observations are the same, i.e., $\xi = \bar{\xi}\mathbf{1}$. The constraint $\deg P_t \leqslant t$ ensures that the iterate $x^t$ can be computed in $t$ time steps. Simple gossip corresponds to the particular case of the polynomial $P_t(\lambda) = \lambda^t$. Shift-register gossip is a polynomial gossip method whose corresponding polynomials can be expressed using the
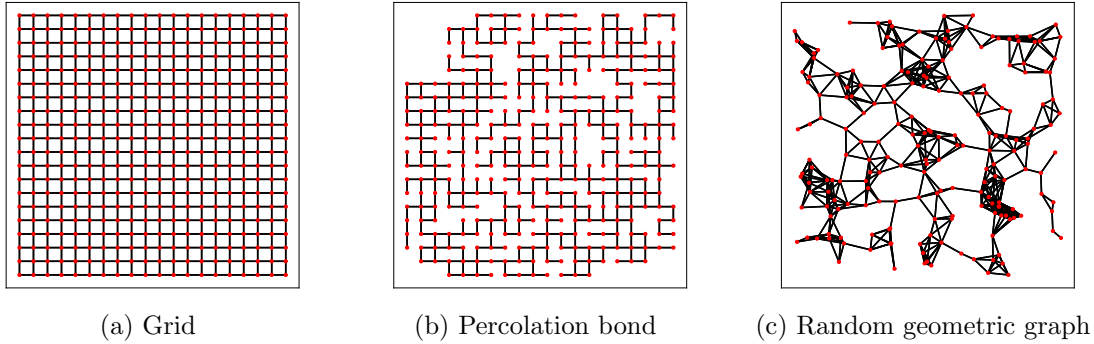
(a) Grid                (b) Percolation bond        (c) Random geometric graph

**Figure 1.** *The three types of* 2D *graphs considered in simulations.*

Chebyshev polynomials (see Proposition SM8.4). The method (1.3) will be derived as the polynomial iteration corresponding to some Jacobi polynomials.

In this paper, we design polynomial gossip methods whose polynomials $P_t$, $t \geqslant 0$, satisfy a second-order recursion. This key property ensures that the resulting iterates $x^t = P_t(W)\xi$ can be computed recursively.

**3. Simulations: Comparison of simple gossip, shift-register gossip, and the Jacobi polynomial iteration.** In this section, we run our methods on grids, percolation bonds, and random geometric graphs; the latter is a widely used model for real-world networks [23, section 1.1]. In each case, we consider both the two-dimensional (2D) structure and its three-dimensional (3D) counterpart. We refer the reader to Figure 1 for visualizations of the 2D structures and to section SM1 for details about the parameters used.

We compare our Jacobi polynomial iteration (1.3) with the simple gossip method (1.1) and the shift-register algorithm (1.2). We found experimentally that the behavior of the shift-register algorithm was typical of methods based on the spectral gap such as the splitting algorithm of [25] and the Chebyshev polynomial acceleration scheme [2, 27]; to avoid redundancy we do not present the similar behavior of these methods. We also compare with local averaging, which is given by the formula

$$x_v^t = \frac{1}{|B_t(v)|} \sum_{w \in B_t(v)} \xi_w \,,$$

where $|B_t(v)|$ denotes the ball in $G$, centered in $v$, of radius $t$, for the shortest path distance. Note that local averaging does not correspond in general to any computationally cheap iteration, as opposed to the algorithms we present here. Thus it should not be considered as a gossip method, but rather as a lower bound on the performance achievable by any gossip method. (This is made fully rigorous in the statistical gossip framework of section SM7.)

In our simulations, we change the graph $G$ that we run our algorithms on, but we always sample $\xi_v \sim_{i.i.d.} \mathcal{N}(0,1), v \in V$, and measure the performance of gossip methods through the quantity $\|x^t - \bar{\xi}\mathbf{1}\|_2/\sqrt{n}$. Thus the performance of the algorithms is random because the initial values $\xi_v$ are random, and also because percolation bonds and random geometric graphs are

random. The results we present here are averaged over 10 realizations of the graph and the initial values, which is sufficient to give stable results.

**Tuning.** The optimal tuning of the shift-register gossip method as a function of the spectral gap was determined in [16, Theorem 2]; it is given by the formula (2.4); this is the tuning that we use in our simulations. The Jacobi polynomial iteration is tuned by choosing $d = 2$ in the 2D grid, 2D percolation bonds, and 2D random geometric graphs, and $d = 3$ for their 3D analogues.

**Interpretation of the results.** The results of the simulations are exposed in Figure 2. The qualitative picture remains the same across different graphs. Simple gossip performs better than shift-register gossip in a first phase, but in a large $t$ asymptotic, simple gossip converges slowly whereas shift-register gossip converges quickly. The Jacobi polynomial iteration enjoys the quick diffusion of simple gossip in the first phase and reaches the full mixing before shift-register gossip. As a consequence, the Jacobi polynomial iteration gets considerably closer to the local averaging optimal bound, especially in very regular structures like grids.

These results should be mitigated with the large $t$ asymptotic: in Figure 3, we show the comparison of gossip methods on a longer time scale, in linear and log-scale $y$-axis. We only present the results on the 2D grid as they are typical of the behavior on other structures. We observe that shift-register gossip enjoys a much better asymptotic rate of convergence than simple gossip and the Jacobi polynomial iteration.

Methods that use the spectral gap are designed to achieve the best possible asymptotic (see [7], [25]); thus the above observation is not surprising. These methods, however, fail in the nonasymptotic regime, where they are outperformed by the Jacobi polynomial iteration and simple gossip. We believe that in applications where a high precision on the average is not needed, the Jacobi polynomial iteration brings important improvements over existing methods, let alone the fact that it is considerably easier to tune. However, in section 6, we present a Jacobi polynomial iteration that uses the spectral gap of the gossip matrix to obtain the accelerated convergence rate.

**4. Design of best polynomial gossip iterations.** We now turn to the design of efficient polynomial iterations of the form $x^t = P_t(W)\xi$. An important result of this section is that the best iterates of this form can be computed in an online fashion as they result from a second-order recurrence relation.

The approach presented in this section is similar to [9, section 3.3], although therein it is applied to the slightly different problem of load balancing. We repeat here the derivations as we take a slightly different approach: here we derive the best polynomial $P_t$ with fixed $W$ and $\xi$, while in [9] the matrix $W$ is fixed, but a polynomial $P_t$ efficient uniformly over $\xi$ is sought. We then discuss why the resulting recursion may be impractical. The next section introduces some approximation of the impractical scheme that leads to the practical iteration (1.3).

Our measure of performance of a polynomial gossip iteration is the sum of squared errors over the agents of the network:

$$\mathcal{E}(P_t) = \sum_{v \in V} (x_v^t - \bar{\xi})^2 = \|x^t - \bar{\xi}\mathbf{1}\|_2^2 = \|P_t(W)\xi - \bar{\xi}\mathbf{1}\|_2^2.$$

Denote by $\lambda_1, \lambda_2, \ldots, \lambda_n$ the real eigenvalues of the symmetric matrix $W$ and by $u^1, u^2, \ldots, u^n$
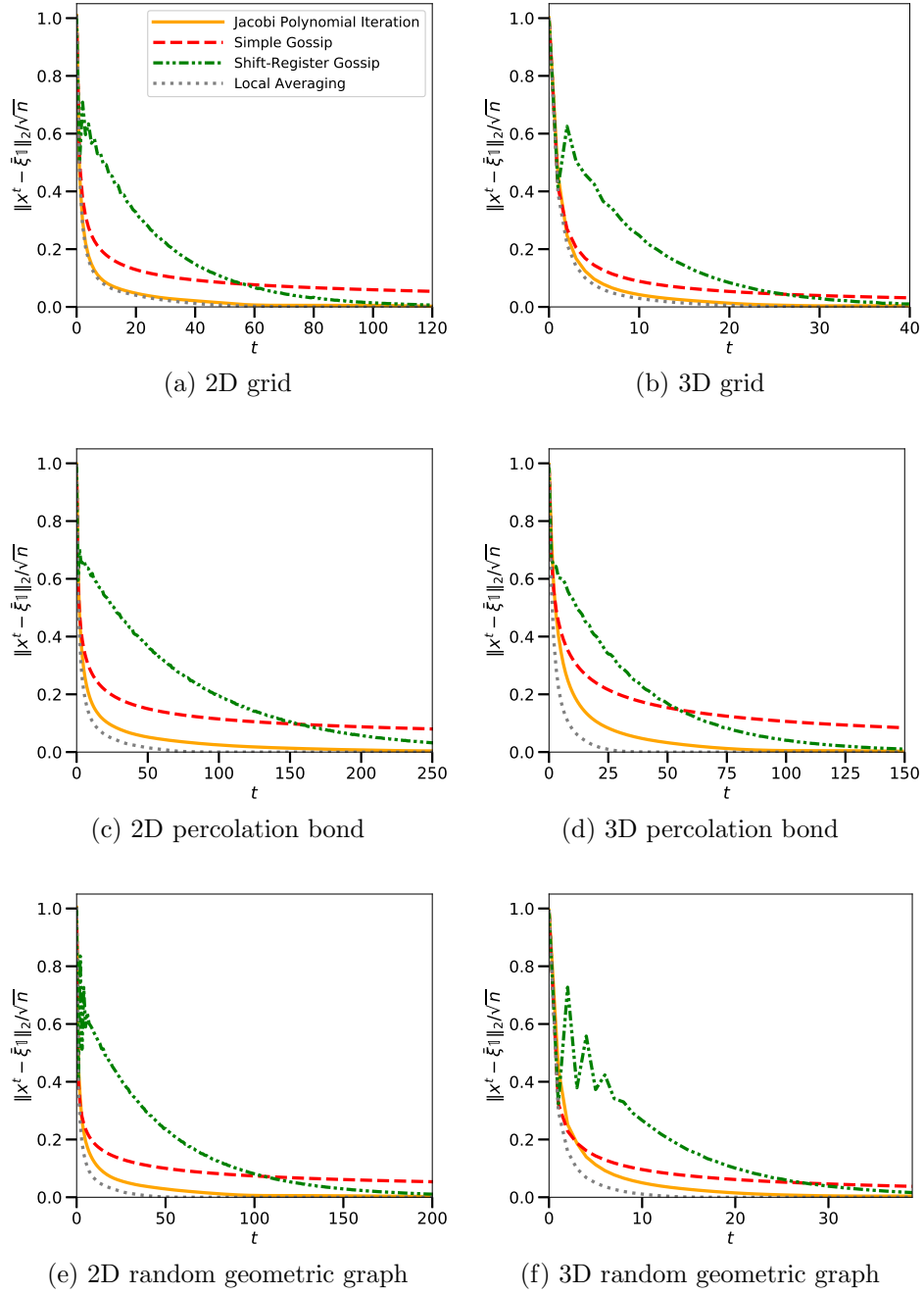
(a) 2D grid

(b) 3D grid

(c) 2D percolation bond

(d) 3D percolation bond

(e) 2D random geometric graph

(f) 3D random geometric graph

**Figure 2.** *Performance of different gossip algorithms running on graphs with an underlying low-dimensional geometry, as measured by $\|x^t - \bar{\xi}\mathbf{1}\|_2/\sqrt{n}$.*

the associated eigenvectors, normalized such that $\|u^i\|_2 = 1$. The diagonalization of $W$ gives

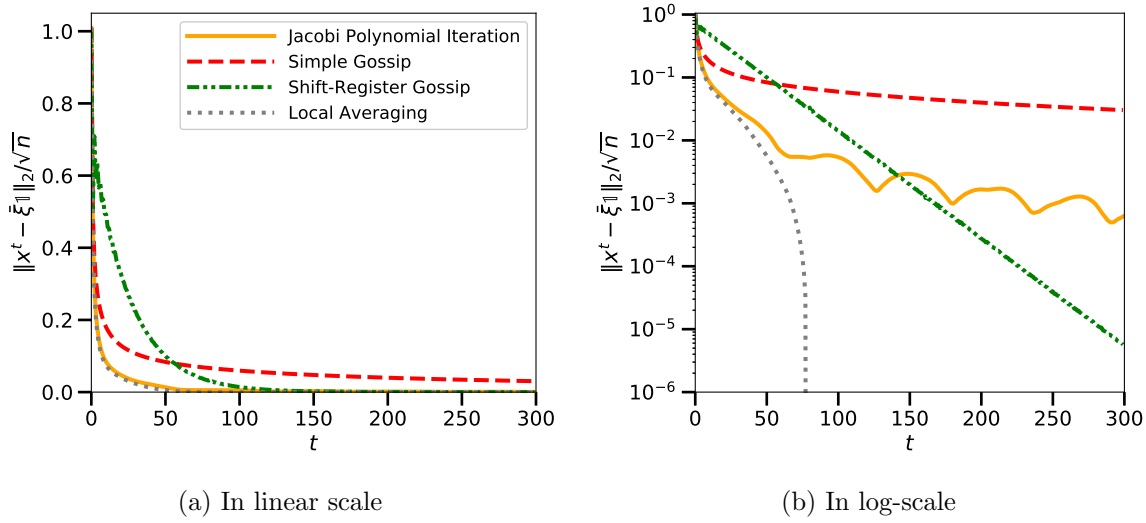(a) In linear scale                              (b) In log-scale

**Figure 3.** *Performance of different gossip algorithms running on the 2D grid.*

the new expression of the error

$$(4.1) \qquad \mathcal{E}(P_t) = \sum_{i=2}^{n} \langle \xi, u^i \rangle^2 P_t(\lambda_i)^2 = \int_{-1}^{1} P_t(\lambda)^2 \mathrm{d}\sigma(\lambda), \qquad \mathrm{d}\sigma(\lambda) = \sum_{i=2}^{n} \langle \xi, u^i \rangle^2 \delta_{\lambda_i},$$

where $\langle ., . \rangle$ denotes the canonical scalar product on $\mathbb{R}^n$ and $\delta_\lambda$ is the Dirac mass at $\lambda$.

The polynomial $\pi_t$ minimizing the error $\mathcal{E}(P_t)$ must be chosen as

$$(4.2) \qquad \pi_t \in \operatorname*{argmin}_{P(1)=1,\, \deg P \leqslant t} \int_{-1}^{1} P(\lambda)^2 \mathrm{d}\sigma(\lambda).$$

We now show that the sequence of best polynomials $\pi_0, \pi_1, \pi_2, \ldots$ can be computed as the result of a second-order recursion, which leads to a second-order gossip method, whose coefficients depend on $\sigma$. As noted in [7], having iterates $x^t$ that satisfy a low-order recurrence relation is valuable as it ensures that they can be computed online with limited memory cost. In order to prove this property for our iterates, we use that these polynomials are orthogonal with respect to (w.r.t.) some measure $\tau$.

**Definition 4.1** (orthogonal polynomials w.r.t. $\tau$). *Let $\tau$ be a measure on $\mathbb{R}$ with finite moments. Endow the set of polynomials $\mathbb{R}[X]$ with the scalar product*

$$\langle P, Q \rangle_\tau = \int_{\mathbb{R}} P(\lambda) Q(\lambda) \mathrm{d}\tau(\lambda).$$

*Denote by $T \in \mathbb{N} \cup \{\infty\}$ the cardinal of the support of $\tau$. Then there exists a family $\pi_0, \pi_1, \ldots, \pi_{T-1}$ of polynomials, such that for all $t < T$, $\pi_0, \pi_1, \ldots, \pi_t$ form an orthogonal basis of $(\mathbb{R}_t[X], \langle ., . \rangle_\tau)$, where $\mathbb{R}_t[X]$ denotes the set of polynomials of degree smaller than or equal to $t$. In other words, for all $s, t < T$,*

$$\deg \pi_t = t, \qquad\qquad \langle \pi_s, \pi_t \rangle_\tau = 0 \qquad \text{if } s \neq t.$$

$\pi_0, \pi_1, \ldots, \pi_{T-1}$ *is called a sequence of* orthogonal polynomials w.r.t. $\tau$. *Moreover, the family of orthogonal polynomials* $\pi_0, \pi_1, \ldots, \pi_{T-1}$ *is unique up to a rescaling of each of the polynomials.*

An extensive reference on orthogonal polynomials is the book [29]. An introduction from the point of view of applied mathematics can be found in [13]. In section SM2, we recall the results from the theory of orthogonal polynomials that we use in this paper. The next proposition states that the optimal polynomials sought in (4.2) are orthogonal polynomials.

**Proposition 4.2.** *Let* $\sigma$ *be some finite measure on* $[-1,1]$, *and let* $T \in \mathbb{N} \cup \{\infty\}$ *be the cardinal of* $\mathrm{Supp}\,\sigma - \{1\}$. *For* $0 \leqslant t \leqslant T - 1$, *the minimizer* $\pi_t$ *of*

$$\min_{P(1)=1,\, \deg P \leqslant t} \int_{-1}^{1} P(\lambda)^2 \mathrm{d}\sigma(\lambda)$$

*is unique. Moreover,* $\pi_0, \ldots, \pi_{T-1}$ *is the unique sequence of orthogonal polynomials w.r.t.* $\mathrm{d}\tau(\lambda) = (1 - \lambda)\mathrm{d}\sigma(\lambda)$ *normalized such that* $\pi_t(1) = 1$.

This result is well known and usually stated without proof [21, sections 3, 4.1], [22, section 2]; we give the short proof in section SM4. In the following, the phrase "the orthogonal polynomials w.r.t. $\tau$" will refer to the unique family of orthogonal polynomials w.r.t. $\tau$ *and normalized such that* $\pi_t(1) = 1$.

*Remark* 4.3. When $T$ is finite and $t \geqslant T$, finding a minimizer of $\int_{-1}^{1} P(\lambda)^2 \mathrm{d}\sigma(\lambda)$ over the set of polynomials such that $P(1) = 1$, $\deg P \leqslant t$ is trivial. Indeed, one can consider the polynomial

$$\pi_T(\lambda) = \frac{\prod_{\lambda' \in \mathrm{Supp}\,\sigma - \{1\}}(\lambda - \lambda')}{\prod_{\lambda' \in \mathrm{Supp}\,\sigma - \{1\}}(1 - \lambda')}$$

which is of degree $T$, satisfies $\pi_T(1) = 1$, and $\int_{-1}^{1} \pi_T(\lambda)^2 \mathrm{d}\sigma(\lambda) = \sigma(\{1\})$. This is the best value that a polynomial $P$ of any degree, such that $P(1) = 1$, can get.

A fundamental result on orthogonal polynomials states that they follow a second-order recursion.

**Proposition 4.4** (three-term recurrence relation, from [29, Theorem 3.2.1]). *Let* $\pi_0, \ldots, \pi_{T-1}$ *be a sequence of orthogonal polynomials w.r.t. some measure* $\tau$. *There exist three sequences of coefficients* $(a_t)_{1 \leqslant t \leqslant T-2}$, $(b_t)_{1 \leqslant t \leqslant T-2}$, *and* $(c_t)_{1 \leqslant t \leqslant T-2}$ *such that, for* $1 \leqslant t \leqslant T - 2$,

$$\pi_{t+1}(\lambda) = (a_t \lambda + b_t)\pi_t(\lambda) - c_t \pi_{t-1}(\lambda).$$

The classical proof of this proposition is given in section SM2.1. Taking $\sigma$ to be the spectral measure of (4.1) in Proposition 4.2, we get that the best polynomial gossip algorithm is a second-order method whose coefficients are determined by the graph $G$, the gossip matrix $W$, and the vertex $v$. Indeed, as $\pi_0, \ldots, \pi_{T-1}$ is a family of orthogonal polynomials, there exist coefficients $a_t, b_t, c_t$ such that

$$\pi_{t+1}(\lambda) = (a_t \lambda + b_t)\pi_t(\lambda) - c_t \pi_{t-1}(\lambda),$$

and thus

$$\pi_{t+1}(W) = a_t W \pi_t(W) + b_t \pi_t(W) - c_t \pi_{t-1}(W) \,.$$

Decomposing $\pi_1(\lambda) = a_0 \lambda + b_0$ and applying the previous relation in $\xi$ gives the second-order recursion for the best polynomial estimators $x^t = \pi_t(W)\xi$:

(4.3)         $x^0 = \xi \,,$         $x^1 = a_0 W \xi + b_0 \xi \,,$         $x^{t+1} = a_t W x^t + b_t x^t - c_t x^{t-1} \,.$

Note that the dependence of the gossip method on the graph $G$, the gossip matrix $W$, and the vertex $v$ is entirely hidden in the coefficients $a_t, b_t, c_t$. Thus the choice of the coefficients is central. In [9], it is argued that the coefficients can be computed in a "preprocessing step." Indeed, the coefficients can be computed in a centralized or decentralized manner, at the cost of extra communication steps. The gossip method that consists in computing the optimal coefficients $a_t, b_t, c_t$ and running (4.3) will be referred to as *parameter-free polynomial iteration*, as it does not require any tuning of parameters, by analogy with the terminology used in polynomial methods for the resolution of linear systems (see [12, section 6]). It corresponds to the optimal polynomial iteration. For a detailed exposition on the parameter-free polynomial iteration and a discussion of its practicability, see section SM5.

However, in dynamic networks that are constantly changing, it is not a valid option to keep repeating the preprocessing step to update the coefficients $a_t, b_t, c_t$. Our approach consists in observing that there are sequences of coefficients like (1.4) that, despite not being optimal, work reasonably well on a large set of graphs. This implies that even if the details of the graph are not known to the algorithmic designer, she can make a choice of coefficients that have a fair performance.

More formally, we approximate the true spectral measure $\sigma$ of the graph with a simpler measure $\tilde{\sigma}$, whose associated polynomials have known recursion coefficients $a_t, b_t, c_t$. We show that in some cases, substituting the orthogonal polynomials w.r.t. $\sigma$ with the ones orthogonal to $\tilde{\sigma}$ does not worsen the efficiency of the gossip method much. In the next sections, we argue for two choices of the approximating measure $\tilde{\sigma}$. The first uses only the spectral dimension $d$ of the network and gives the Jacobi polynomial iteration (1.3). The second uses both the spectral dimension $d$ and the spectral gap $\gamma$ of $W$ and gives the Jacobi polynomial iteration with spectral gap.

Figure 4 reproduces Figure 3 and adds the performance of the parameter-free polynomial iteration and the Jacobi polynomial iteration with spectral gap. It shows that in linear scale, the performance of the parameter-free polynomial iteration is indistinguishable from the performance of the Jacobi polynomial iterations with or without spectral gap, which are obtained through approximations of the spectral measure $\sigma$. However, the figure in log-scale shows that the asymptotic convergence of the methods depends on the coarseness of the approximation. The relevance of this asymptotic convergence to the practice depends on the application.

*Remark* 4.5. The shift-register iteration $x^t = P_t(W)\xi$ defined in (2.3) can be seen as a best polynomial gossip iteration with some approximating measure. Indeed, the polynomials $P_t$, $t \geqslant 0$, are the orthogonal polynomials w.r.t. some measure whose support is strictly included in $[-1, 1]$ (see Proposition SM8.5).

(a) In linear scale
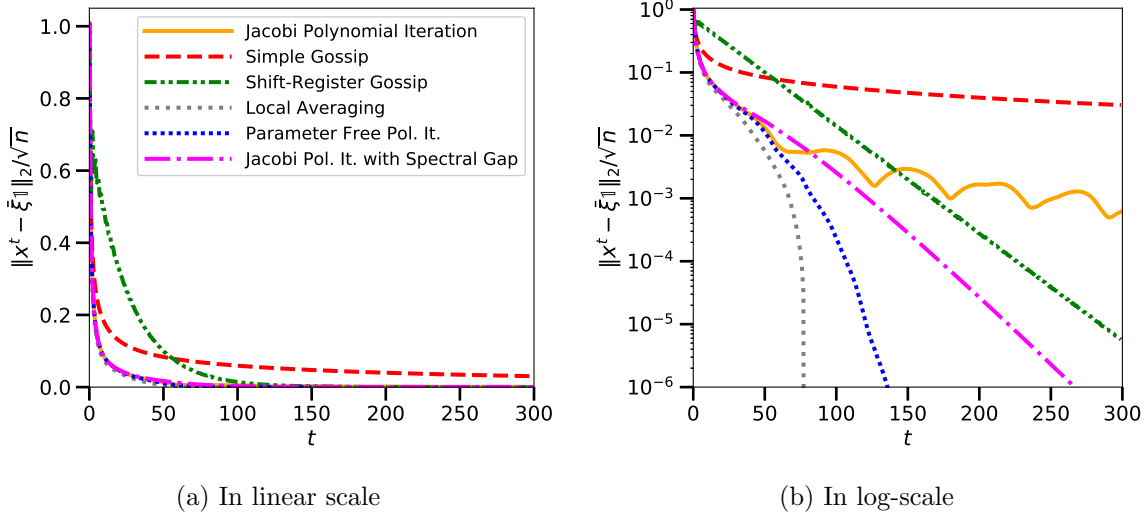
(b) In log-scale

**Figure 4.** *Performance of different gossip algorithms running on the 2D grid.*

## 5. Design of polynomial gossip algorithms for graphs of given spectral dimension.

### 5.1. The dimension $d$ and the rate of decrease of the spectral measure near 1. We now assume that we are given a graph $G$ on which we would like to run the optimal polynomial gossip algorithm (4.3). However, we do not know the spectral measure $\sigma$ or the coefficients $a_t, b_t, c_t$. In this section, we give a heuristic motivating an approximation $\tilde{\sigma}$ of the spectral measure $\sigma$ using only the dimension $d$ of the graph. The heuristic is supported by the simulations of section 3 and some rigorous theoretical support in section SM7.

Our approximation is given by the following nonrigorous intuition:

$$(5.1) \qquad \text{the graph } G \text{ is of dimension } d \quad \Leftrightarrow \quad \sigma([1-\Lambda, 1]) \approx C\Lambda^{d/2} \quad \text{as } \Lambda \ll 1\,,$$

for some constant $C$. Of course, we have neither defined the dimension of a graph nor given a rigorous signification of the symbols $\approx$ and $\ll$. We come back to these questions in section 5.3, but for now we assume that the reader has an intuitive understanding of these notions and finish drawing the heuristic picture.

Intuition (5.1) describes the repartition of the mass of $\sigma$ near 1. This mass near 1 challenges the design of polynomial methods as the gossip polynomials $P$ are constrained to satisfy $P(1) = 1$ while minimizing $\int P^2 \mathrm{d}\sigma$. Moreover, eigenvalues of a graph close to 1 are known to describe the large-scale structure of the graph and thus must be central in the design of gossip methods. The traditional design of gossip algorithms considered the spectral gap $\gamma$ between 1 and the second largest eigenvalue, a quantity that typically gets very small in large graphs. Intuition (5.1) also describes the behavior of the spectrum near 1, but on a larger scale than the spectral gap. It describes how the set of the largest eigenvalues is distributed around 1.

### 5.2. The Jacobi iteration for graphs of given dimension. When a spectral measure satisfies the edge estimate (5.1), we approximate it with a measure satisfying the same estimate,
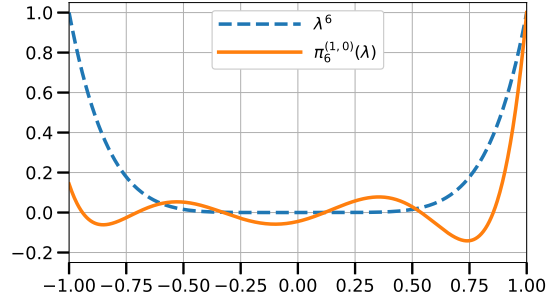
**Figure 5.** *Comparison of the Jacobi polynomial $\pi_6^{(1,0)}(\lambda)$ with the polynomial of simple gossip $\lambda^6$.*

namely

$$d\tilde{\sigma}(\lambda) = (1 - \lambda)^{d/2-1}\mathbf{1}_{\{\lambda\in(-1,1)\}}d\lambda\,.$$

Note that we do not elaborate on the normalization of the approximate measure $d\tilde{\sigma}$ as it is only used to define an orthogonality relation between polynomials, in which the normalization does not matter. The orthogonal polynomials w.r.t. the modified spectral measure $(1 - \lambda)d\tilde{\sigma}(\lambda) = (1 - \lambda)^{d/2}\mathbf{1}_{\{\lambda\in(-1,1)\}}d\lambda$ and their recursion coefficients are known as they correspond to the well-studied Jacobi polynomials [29, Chapter IV]:

$$\begin{aligned}
a_0^{(d)} &= \frac{d + 4}{2(2 + d)}\,, & b_0^{(d)} &= \frac{d}{2(2 + d)}\,,\\
a_t^{(d)} &= \frac{(2t + d/2 + 1)(2t + d/2 + 2)}{2(t + 1 + d/2)^2}\,, & b_t^{(d)} &= \frac{d^2(2t + d/2 + 1)}{8(t + 1 + d/2)^2(2t + d/2)}\,,\\
c_t^{(d)} &= \frac{t^2(2t + d/2 + 2)}{(t + 1 + d/2)^2(2t + d/2)}\,.
\end{aligned}$$

(5.2)

These coefficients are derived in section SM6.2. This approximation of the spectral measure gives the practical recursion

$$(5.3)\qquad x^0 = \xi\,,\quad x^1 = a_0^{(d)}W\xi + b_0^{(d)}\xi\,,\quad x^{t+1} = a_t^{(d)}Wx^t + b_t^{(d)}x^t - c_t^{(d)}x^{t-1}\,,$$

which only depends on $d$. It is just a rewriting of the Jacobi polynomial iteration (1.3) given in the introduction of this paper. The Jacobi polynomial $\pi_t^{(d/2,0)}(\lambda)$ such that $x^t = \pi_t^{(d/2,0)}(W)\xi$ is plotted in Figure 5 with $d = 2$ and $t = 6$, along with the polynomial $\lambda^6$ associated with simple gossip. The Jacobi polynomial is smaller in magnitude near 1.

**5.3. Spectral dimension of a graph.** In this section, we discuss the meaning of intuition (5.1). There are several definitions of the dimension of a graph.

When referring to the dimension of a graph, many authors actually refer to some quantity $d$ that has been used in the construction of the graph. An example is the $d$-dimension grid $\{1,\ldots,n\}^d$. Another example consists in removing edges in $\mathbb{Z}^d$ with probability $1 - p$, independently of one another. The resulting graph $G$ is called a percolation bond [14]. It is natural to consider that this graph is of dimension $d$. A more complicated example is the

random geometric graph: choose $d \geqslant 1$, sample $n$ points uniformly in the $d$-dimensional cube $[0, 1]^d$, and connect with an edge all pairs of points closer than some chosen distance $r > 0$. It is natural to say that this random geometric graph is $d$-dimensional as it is the dimension of the surface it is built on.

Mathematicians have developed more intrinsic definitions of the dimension of a graph [11]; here we use the notion of *spectral dimension*. This definition is of interest only for infinite graphs $G = (V, E)$. Here, we consider only locally finite graphs, meaning that each node has only a finite number of neighbors. As with Definition 2.1, one can define a gossip matrix $W$ with entries indexed by $V \times V$. If $G$ is infinite, $W$ is a doubly infinite array, but with only a finite number of nonzero elements in each line and column as the graph is locally finite.

The spectral dimension of a graph $G$ is defined using a random walk on the graph— typically the simple random walk on $G$—but here we consider the lazy random walk with transition matrix $\tilde{W} = (I + W)/2$. (We take the *lazy* random walk to avoid periodicity issues.)

**Definition 5.1 (spectral dimension).** *Denote by $p_t$ the probability that the lazy random walk, when started from $v$, returns at $v$ at time $t$. The spectral dimension of the graph, if it exists and is finite, is the limit*

$$d_s = d_s(G, W, v) = -2 \lim_{t \to \infty} \frac{\ln p_t}{\ln t}.$$

If the graph is connected and $W$ is the transition matrix of the simple random walk, this definition does not depend on the choice of the vertex $v$. Motivations for this definition are as follows.

**Proposition 5.2.** *The spectral dimension of $(\mathbb{Z}^d, W)$ with $W = A(\mathbb{Z}^d)/d$ is $d$.*

*Proof.* The return probability $p_t$ of the lazy random walk on $\mathbb{Z}^d$ is equivalent to $C/t^{d/2}$ for some constant $C$. It is, for instance, a consequence of the local central limit theorem for random walks on $\mathbb{Z}^d$ [15, Theorem 2.1.1]. Thus the spectral dimension of $\mathbb{Z}^d$ is $d$. ∎

The spectral dimension of a graph is related to the decay of the spectrum of $W$ near 1.

**Definition 5.3 (spectral measure of a possibly infinite graph).** *Let $G$ be a graph and $W$ its gossip matrix. Fix $v \in V$. As $W$ is an auto-adjoint operator, bounded by 1, acting on $\ell^2(V)$, there exists a unique positive measure $\sigma = \sigma(G, W, v)$ on $[-1, 1]$, called the spectral measure, such that for all polynomial $P$,*

$$\langle e_v, P(W)e_v \rangle_{\ell^2(V)} = \int_{-1}^{1} P(\lambda)\mathrm{d}\sigma(\lambda).$$

For a deeper presentation of spectral graph theory, see [19] and references therein. Note that when the graph $G$ is finite, it is easy to check that the spectral measure is the discrete measure $\sigma(G, W, v) = \sum_{i=1}^{n} (u_v^i)^2 \delta_{\lambda_i}$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $W$ and $u^1, \ldots, u^n$ are the associated normalized eigenvectors. However, when the graph $G$ is infinite, the spectrum may exhibit a continuous part w.r.t. the Lebesgue measure.

**Proposition 5.4 (the spectral dimension is the spectral decay).** *Let $G$ be a graph, $W$ a gossip matrix on $G$, and $v$ a vertex. We denote by $d_s = d_s(G, W, v)$ the spectral dimension and by*

$\sigma = \sigma(G, W, v)$ *the spectral measure. Then the limit* $\lim_{\Lambda \to 0} \ln \sigma([1 - \Lambda, 1])/\ln \Lambda$ *exists and is finite if and only if* $d_s$ *exists and is finite. In that case,*

$$\lim_{\Lambda \to 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} = \frac{d_s}{2}.$$

This proposition gives a rigorous equivalent to intuition (5.1). It uses the spectral dimension of the graph, which is an intrinsic property of the graph and turns out to coincide with our intuition of the dimension of a graph in examples of interest. Note that in section 4, the spectral measure $\sigma$ is defined as $\mathrm{d}\sigma(\lambda) = \sum \langle \xi, u^i \rangle^2 \delta_{\lambda_i}$, whereas in this section it is defined for finite graphs as $\mathrm{d}\sigma(\lambda) = \sum (u_v^i)^2 \delta_{\lambda_i}$. Roughly speaking, intuition (5.1) is valid for the former if $\xi$ projects evenly on all eigenvectors $u^i$. This is the case if $\xi$ has random i.i.d. components, for instance; this is used in section SM7.

*Proof of Proposition* 5.4. We first assume that $l = \lim_{\Lambda \to 0} \ln \sigma([1 - \Lambda, 1])/\ln \Lambda$ exists and is finite. We show that $d_s$ exists and that $l = d_s/2$. To this end, we define

$$\underline{d_s} = -2 \limsup_{t \to \infty} \frac{\ln p_t}{\ln t}, \qquad \qquad \bar{d_s} = -2 \liminf_{t \to \infty} \frac{\ln p_t}{\ln t},$$

where $p_t$ is defined as in Definition 5.1. Note that

(5.4)
$$p_t = \left\langle e_v, \left( \frac{I + W}{2} \right)^t e_v \right\rangle \overset{\text{(Definition 5.3)}}{=} \int \left( \frac{1 + \lambda}{2} \right)^t \mathrm{d}\sigma(\lambda).$$

*Proof that* $\bar{d_s}/2 \leqslant l$. Consider $l_+ > l$. Then there exists constants $c_1, c_2 > 0$ such that for all $\Lambda \in [0, 2]$,

$$\sigma([1 - \Lambda, 1]) \geqslant c_1 \Lambda^{l_+} = c_2 \sigma^{(l_+ - 1, 0)}([1 - \Lambda, 1]),$$

where $\sigma^{(l_+ - 1, 0)}(\mathrm{d}\lambda) = (1 - \lambda)^{l_+ - 1} \mathrm{d}\lambda$. Then

$$p_t \overset{(5.4)}{=} \int_{[-1,1]} \left( \frac{1 + \lambda}{2} \right)^t \mathrm{d}\sigma(\lambda) \overset{\text{(Lemma SM3.1)}}{\geqslant} c_2 \int_{-1}^{1} \left( \frac{1 + \lambda}{2} \right)^{t-1} (1 - \lambda)^{l_+ - 1} \mathrm{d}\lambda$$

$$\overset{(u = (1 + \lambda)/2)}{=} c_3 \int_0^1 u^t (1 - u)^{l_+ - 1} \mathrm{d}u = c_3 B(t + 1, l_+) \underset{t \to \infty}{\sim} \frac{c_4}{t^{l_+}}$$

for some constant $c_3, c_4 > 0$. Thus $\liminf_{t \to \infty} \frac{\ln p_t}{\ln t} \geqslant -l_+$, i.e., $\bar{d_s}/2 \leqslant l_+$. This being true for all $l_+ > l$, this proves $\bar{d_s}/2 \leqslant l$.

*Proof that* $\underline{d_s}/2 \geqslant l$. Consider $l_- < l$. Then there exist constants $C_1, C_2$ such that for all $\Lambda \in [0, 2]$,

$$\sigma([1 - \Lambda, 1]) \leqslant C_1 \Lambda^{l_-} = C_2 \sigma^{(l_- - 1, 0)}([1 - \Lambda, 1]),$$

where $\sigma^{(l_- - 1, 0)}(\mathrm{d}\lambda) = (1 - \lambda)^{l_- - 1} \mathrm{d}\lambda$. Then

$$p_t \overset{(5.4)}{=} \int_{[-1,1]} \left( \frac{1 + \lambda}{2} \right)^t \mathrm{d}\sigma(\lambda) \overset{\text{(Lemma SM3.1)}}{\leqslant} C_2 \int_{-1}^{1} \left( \frac{1 + \lambda}{2} \right)^t (1 - \lambda)^{l_- - 1} \mathrm{d}\lambda$$

$$\overset{(u = (1 + \lambda)/2)}{=} C_3 \int_0^1 u^t (1 - u)^{l_- - 1} \mathrm{d}u = C_3 B(t + 1, l_-) \underset{t \to \infty}{\sim} \frac{C_4}{t^{l_-}}$$

for some constants $C_3, C_4$. Thus $\limsup_{t\to\infty} \frac{\ln p_t}{\ln t} \leqslant -l_-$, which implies $\underline{d}_s/2 \geqslant l_-$. This being true for all $l_- < l$, this proves $\underline{d}_s/2 \geqslant l$.

Finally, we have proven $l \leqslant \underline{d}_s/2 \leqslant \bar{d}_s/2 \leqslant l$. Thus the limit $d_s = -2\lim_{t\to\infty} \ln p_t / \ln t$ exists and is equal to $2l$.

Conversely, we assume now that $d_s$ exists and is finite. We show that $l = \lim_{\Lambda\to 0} \ln \sigma([1 - \Lambda, 1])/\ln \Lambda$ exists and that $l = d_s/2$. To this end, we define

$$\underline{l} = \liminf_{\Lambda\to 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda}, \qquad \bar{l} = \limsup_{\Lambda\to 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda}.$$

*Proof that $\underline{l} \geqslant d_s/2$.* For any $t \in \mathbb{N}$, we have

$$\mathbf{1}_{\{\lambda\geqslant 1-\Lambda\}} \leqslant \left(1 - \frac{\Lambda}{2}\right)^{-t} \left(\frac{1+\lambda}{2}\right)^t;$$

thus, by integrating against $d\sigma(\lambda)$,

$$\sigma([1 - \Lambda, 1]) \leqslant \left(1 - \frac{\Lambda}{2}\right)^{-t} \int \left(\frac{1+\lambda}{2}\right)^t \sigma(d\lambda),$$

$$\frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} \geqslant \frac{\ln \int \left(\frac{1+\lambda}{2}\right)^t \sigma(d\lambda)}{\ln t} \frac{\ln t}{\ln \Lambda} - \frac{t \ln\left(1 - \frac{\Lambda}{2}\right)}{\ln \Lambda}.$$

We choose $t(\Lambda) = \lfloor \Lambda^{-1}\rfloor$. Then we get

$$\underline{l} = \liminf_{\Lambda\to 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} \geqslant -\frac{d_s}{2}(-1) - 0 = \frac{d_s}{2}.$$

*Proof that $\bar{l} \leqslant d_s/2$.* For any $t \in \mathbb{N}$, we have $((1+\lambda)/2)^t - (1 - \Lambda/2)^t \leqslant \mathbf{1}_{\{\lambda\geqslant 1-\Lambda\}}$; thus, by integrating against $d\sigma(\lambda)$,

$$\int \left(\frac{1+\lambda}{2}\right)^t d\sigma(\lambda) - \left(1 - \frac{\Lambda}{2}\right)^t \leqslant \sigma([1 - \Lambda, 1]).$$

Let $d > d_s$. There exists a constant $c > 0$ such that $\int ((1+\lambda)/2)^t d\sigma(\lambda) \geqslant c/t^{d/2}$. Then

$$\ln\left(\frac{c}{t^{d/2}} - \left(1 - \frac{\Lambda}{2}\right)^t\right) \leqslant \ln \sigma([1 - \Lambda, 1]).$$

Let $\alpha > 1$. We choose $t(\Lambda) = \lceil \Lambda^{-\alpha}\rceil$. Then

$$\left(1 - \frac{\Lambda}{2}\right)^{t(\Lambda)} = \exp\left(t(\Lambda)\ln\left(1 - \frac{\Lambda}{2}\right)\right) \leqslant \exp\left(-\frac{t(\Lambda)\Lambda}{2}\right) \leqslant \exp\left(-\frac{1}{2}\Lambda^{1-\alpha}\right)$$

decreases superpolynomially fast as $\Lambda \to 0$. Since $ct(\Lambda)^{-d/2} \underset{\Lambda\to 0}{\sim} c\Lambda^{\alpha d/2}$, it yields

$$\bar{l} = \limsup_{\Lambda\to 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} \leqslant \frac{\alpha d}{2}.$$

As this is true for all $\alpha > 1, d > d_s$, we have $\bar{l} \leqslant d_s/2$.

Finally, we have proven that $d_s/2 \leqslant \underline{l} \leqslant \bar{l} \leqslant d_s/2$. Then the limit $l = \lim_{\Lambda\to 0} \ln \sigma([1 - \Lambda, 1])/\ln \Lambda$ exists and $l = d_s/2$. ∎

Proposition 5.4 allows us to prove the following generalization of Proposition 5.2.

**Proposition 5.5** (spectral dimension of the supercritical percolation cluster). *Let $G_0$ be a supercritical percolation bond in $\mathbb{Z}^d$ with edge probability $p \in (p_c, 1]$; i.e., a.s., there is an infinite connected component $G$ in $G_0$. Endow $G$ with the gossip matrix $W = I + (A - D)/(2d)$, where $A$ and $D$ are, respectively, the adjacency and the degree matrices of $G$. Fix $v \in \mathbb{Z}^d$. Then a.s. on the event $\{v \in G\}$, $d_s(G, W, v) = d$.*

This proposition suggests that the spectral dimension is unrelated to the small-scale structure of the graph.

*Proof of Proposition 5.5.* The return probabilities of the random walk on the supercritical percolation cluster have rather been studied in continuous time. The continuous-time random walk is defined as follows: the random walk at $w$ waits at an exponential time of parameter 1 before picking a site $w'$ out of the $2d$ neighboring sites uniformly randomly. If there is an edge in the percolation configuration between $w$ and $w'$, the random walk jumps to $w'$; otherwise it stays in $w$ and starts again. Denote by $X_t$ the continuous-time random walk and by $\mathbb{P}_w$ the probability w.r.t. this random walk when it is started from some vertex $w$.

**Lemma 5.6.** *There exist two constants $c = c(d, p), C = c(d, p) > 0$ such that, a.s. on the set $\{v \in G\}$, there exists a random time $t_0$ such that for $t \geqslant t_0$,*

$$\frac{c}{t^{d/2}} \leqslant \mathbb{P}_v(X_t = v) \leqslant \frac{C}{t^{d/2}} .$$

*Proof.* The upper bound is proved in [17, Theorem 1.2]. As noted in [5, Lemma 5.1], the lower bound can be proved using a central limit theorem on $X_t$; we repeat the argument here as our random walk differs slightly from theirs. As $X_t$ is reversible w.r.t. the uniform measure on $G$,

$$\mathbb{P}_v(X_{2t} = v) = \sum_{w \in G} \mathbb{P}_v(X_t = w)\mathbb{P}_w(X_t = v) = \sum_{w \in G} \mathbb{P}_v(X_t = w)^2 .$$

By the Cauchy–Schwarz inequality,

$$\mathbb{P}_v(\|X_t - v\|_2 \leqslant \sqrt{t})^2 = \left( \sum_{x \in G} \mathbf{1}_{\{\|x - v\|_2 \leqslant \sqrt{t}\}} \mathbb{P}_v(X_t = x) \right)^2$$

$$\leqslant \left| \{\|x \in G : x - v\|_2 \leqslant \sqrt{t}\} \right| \left( \sum_{w \in G} \mathbb{P}_v(X_t = w)^2 \right)$$

$$\leqslant C_1 t^{d/2} \mathbb{P}_v(X_{2t} = v)$$

for some constant $C_1$. Now, using [1, Theorem 1.1(a)], there exists a deterministic variance $\sigma^2$ such that the law of $(X_t - v)/\sqrt{t}$ converges a.s. on the event $\{v \in G\}$ to a centered Gaussian with variance $\sigma^2$. Thus there exist a deterministic constant $c_1 > 0$ and a random time $t_1$ such that for $t \geqslant t_1$, $\mathbb{P}_t(\|X_t - v\|_2 \leqslant \sqrt{t})^2 \geqslant c_1$. This finishes the proof of the lower bound. ∎

We now finish the proof of the proposition using Lemma 5.6. If $\mu^t$ denotes the law of $X_t$, we have $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\mu^t\right] = (W - I)\mu^t$. This yields $\mu^t = e^{t(W-I)}\mu^0$, which implies

$$\mathbb{P}_v(X_t = v) = \langle \delta_v, \mu^t \rangle = \langle \delta_v, e^{t(W-I)}\delta_v \rangle \overset{\text{(Definition 5.3)}}{=} \int e^{t(\lambda-1)}\mathrm{d}\sigma(\lambda) .$$

As a consequence, Lemma 5.6 translates into bounds on the Laplace transform of $\sigma$: a.s. on $\{v \in G\}$, for $t$ large enough,

$$\frac{c}{t^{d/2}} \leqslant \int e^{t(\lambda-1)} \mathrm{d}\sigma(\lambda) \leqslant \frac{C}{t^{d/2}} \,.$$

Some bounds on the spectral density of $\sigma$ near 1 easily follow (see [20, Lemma 4.5]): there exist constants $c', C' > 0$ such that a.s. on $\{v \in G\}$, for $\Lambda$ small enough,

$$c'\Lambda^{d/2} \leqslant \sigma([1-\Lambda, 1]) \leqslant C'\Lambda^{d/2} \,.$$

The proof is finished using Proposition 5.4.                                   ■

In section SM7, we prove some performance guarantees of the Jacobi polynomial iteration (1.3) under the assumption that the graph has spectral dimension $d$. As a corollary, we get performance results on two types of infinite graphs: the $d$-dimensional grid $\mathbb{Z}^d$ and supercritical percolation bonds in dimension $d$. This supports that the iteration (1.3) is robust to local perturbations of a graph.

**6. The Jacobi polynomial iteration with spectral gap.** In this section, we adapt the Jacobi polynomial iteration to the case where the spectral gap $\gamma$ of the gossip matrix $W$ is given. This allows us to obtain accelerated asymptotic rates of convergence, which compete with the state-of-the-art accelerated algorithms for gossip.

We assume that we are given the spectral dimension $d$ of the graph, which determines the density of eigenvalues near 1, and the spectral gap $\gamma = 1 - \lambda_2(W)$, the distance between the largest and the second largest eigenvalues. Given these parameters, we can approximate the spectral measure of $W$ with

$$\mathrm{d}\tilde{\sigma}(\lambda) = ((1-\gamma) - \lambda)^{d/2-1}\mathbf{1}_{\{\lambda \in (-1, 1-\gamma)\}}\mathrm{d}\lambda \,.$$

Following the recommendation of Proposition 4.2, this means that we should consider the polynomial iteration associated with the orthogonal polynomials w.r.t. $(1-\lambda)\mathrm{d}\tilde{\sigma}(\lambda) = (1-\lambda)((1-\gamma) - \lambda)^{d/2-1}\mathbf{1}_{\{\lambda \in (-1, 1-\gamma)\}}\mathrm{d}\lambda$. We do not know how to compute the recurrence formula for this measure; thus we used the orthogonal polynomials w.r.t. $((1-\gamma) - \lambda)\mathrm{d}\tilde{\sigma}(\lambda) = ((1-\gamma) - \lambda)^{d/2}\mathbf{1}_{\{\lambda \in (-1, 1-\gamma)\}}\mathrm{d}\lambda$, which is a rescaled version of a Jacobi measure. The corresponding polynomial method is called the *Jacobi polynomial iteration with spectral gap*.

A recursive formula for orthogonal polynomials w.r.t. $((1-\gamma) - \lambda)\mathrm{d}\tilde{\sigma}(\lambda)$ is derived in section SM6.3. Taking $\alpha = d/2$ and $\beta = 0$ in equations (SM6.3), and using the coefficients

$a_t^{(d)}, b_t^{(d)}, c_t^{(d)}$ defined in (5.2), we get the recursion

$$x^t = \frac{y^t}{\delta_t},$$

$$y^0 = \xi, \qquad \delta_0 = 1,$$

$$y^1 = a_0^{(d,\gamma)} W \xi + b_0^{(d,\gamma)} \xi, \qquad \delta_1 = a_0^{(d,\gamma)} + b_0^{(d,\gamma)},$$

(6.1) $\qquad y^{t+1} = a_t^{(d,\gamma)} W y^t + b_t^{(d,\gamma)} y^t - c_t^{(d,\gamma)} y^{t-1}, \qquad t \geqslant 1,$

$$\delta_{t+1} = \left( a_t^{(d,\gamma)} + b_t^{(d,\gamma)} \right) \delta_t - c_t^{(d,\gamma)} \delta_{t-1}, \qquad t \geqslant 1,$$

$$a_t^{(d,\gamma)} = a_t^{(d)} \left( 1 - \frac{\gamma}{2} \right)^{-1}, \qquad b_t^{(d,\gamma)} = b_t^{(d)} + \frac{\gamma}{2} \left( 1 - \frac{\gamma}{2} \right)^{-1} a_t^{(d)}, \qquad t \geqslant 0,$$

$$c_t^{(d,\gamma)} = c_t^{(d)}, \qquad t \geqslant 1.$$

**Theorem 6.1** (asymptotic rate of convergence). *Let $\gamma > 0$ be a lower bound on the spectral gap of the gossip matrix $W$, and let $d$ be any positive real. Let $\xi = (\xi_v^t)_{v \in V}$ be any family of initial observations and $x^t = (x_v^t)_{v \in V}$ be the sequence of iterates generated by the Jacobi polynomial iteration with spectral gap* (6.1). *Then*

$$\limsup_{t \to \infty} \| x^t - \bar{\xi} \mathbf{1} \|_2^{1/t} \leqslant \frac{1 - \gamma/2}{(1 + \sqrt{\gamma/2})^2}.$$

This shows that the Jacobi polynomial iteration with spectral gap enjoys linear convergence. The asymptotic rate of convergence is equivalent to $1 - \sqrt{2\gamma}$ as $\gamma \to 0$. This justifies that we obtain an accelerated asymptotic rate of convergence that compares with the state-of-the-art accelerated gossip methods (see Figure 4).

*Proof of Theorem* 6.1. In this section, we use the notation of section SM6.3. As $x^t = \pi_t^{(d/2,0,\gamma)}(W)\xi$, we have

(6.2) $\qquad \| x^t - \bar{\xi} \mathbf{1} \|_2^2 = \sum_{i=2}^n \langle \xi, u^i \rangle^2 \pi_t^{(d/2,0,\gamma)}(\lambda_i)^2 \leqslant \| \xi - \bar{\xi} \mathbf{1} \|_2^2 \left( \sup_{\lambda \in [-1, 1-\gamma]} |\pi_t^{(d/2,0,\gamma)}(\lambda)| \right)^2,$

where $\lambda_2, \ldots, \lambda_n$ are the eigenvalues of $W$ different from 1 that lie in $[-1, 1 - \gamma]$ by definition of $\gamma$, and $u^2, \ldots, u^n$ are the corresponding normalized eigenvectors.

$$\sup_{\lambda \in [-1, 1-\gamma]} |\pi_t^{(d/2,0,\gamma)}(\lambda)| \leqslant \frac{1}{|P_t^{(d/2,0,\gamma)}(1)|} \sup_{\lambda \in [-1, 1-\gamma]} |P_t^{(d/2,0,\gamma)}(\lambda)|$$

$$= \frac{1}{|\pi_t^{(d/2,0)}(\varphi_\gamma^{-1}(1))|} \sup_{\lambda \in \varphi_\gamma^{-1}([-1, 1-\gamma])} |\pi_t^{(d/2,0)}(\lambda)|$$

$$= \frac{1}{\left| \pi_t^{(d/2,0)} \left( \frac{1+\gamma/2}{1-\gamma/2} \right) \right|} \sup_{\lambda \in [-1, 1]} |\pi_t^{(d/2,0)}(\lambda)|$$

(6.3) $\qquad = \frac{1}{\left| P_t^{(d/2,0)} \left( \frac{1+\gamma/2}{1-\gamma/2} \right) \right|} \sup_{\lambda \in [-1, 1]} |P_t^{(d/2,0)}(\lambda)|,$

where $P_t^{(\alpha,\beta)}$ is the Jacobi polynomial; see section SM6.2. By Proposition SM2.6,

$$(6.4) \qquad \sup_{\lambda \in [-1,1]} |P_t^{(d/2,0)}(\lambda)| = \binom{t+d/2}{t} \underset{t\to\infty}{\sim} t^{d/2},$$

and by Proposition SM2.9 applied in $x = \frac{1+\gamma/2}{1-\gamma/2}$, there exists $c > 0$ such that

$$(6.5) \qquad P_t^{(d/2,0)}\left(\frac{1+\gamma/2}{1-\gamma/2}\right) \underset{t\to\infty}{\sim} ct^{-1/2}\left(\frac{(1+\sqrt{\gamma/2})^2}{1-\gamma/2}\right)^t.$$

Combining (6.3), (6.4), and (6.5), we get that there exists a constant $C$ such that

$$\sup_{\lambda \in [-1,1-\gamma]} |\pi_t^{(d/2,0,\gamma)}(\lambda)| \leqslant Ct^{(d+1)/2}\left(\frac{1-\gamma/2}{(1+\sqrt{\gamma/2})^2}\right)^t,$$

and we conclude using (6.2). ∎

Note that the asymptotic rate of convergence does not depend on $d$. However, the choice of $d$ may have an important effect during the nonasymptotic phase $t < 1/\sqrt{\gamma}$. In this phase, the spectral gap $\gamma$ can be neglected in the approximation of the spectral measure, and it is important that the densities of eigenvalues of $\sigma$ and $\tilde{\sigma}$ match near the upper edge of the spectrum. This is why one should choose $d$ as the spectral dimension of the graph.

In sections 5 and 6, we have used the polynomial point of view to build gossip algorithms suited to our priors on the graph structure (spectral dimension and spectral gap). In the supplementary material (section SM9), we reverse-engineer the message passing gossip iteration of [18] through the polynomial point of view. We show that this algorithm can be interpreted as an inner-product free polynomial iteration corresponding to a tree prior. This point of view allows us to derive convergence rates of the message passing gossip on regular graphs. This suggests that the polynomial point of view can be used more generally to analyze existing gossip algorithms.

**7. The parallel between the gossip methods and distributed Laplacian solvers.** There is a natural parallel between gossip methods and iterative methods that solve linear systems. Loosely speaking, simple gossip corresponds to gradient descent on the quadratic minimization problem associated to the linear system, shift-register gossip to Polyak's heavy-ball method, and the parameter-free polynomial iteration to the conjugate gradient algorithm (see [12] or [24] for references on these subjects). In this parallel, the fact that we can reach perfect gossip in $n$ steps (see Remark 4.3) translates into the finite convergence of the conjugate gradient algorithm in a number of iterations equal to the dimension of the ambient space. In the distributed resolution of linear systems, the problem that the recursion coefficients $a_t, b_t, c_t$ cannot be computed in a centralized manner has also appeared, and it has motivated the development of inner-product free iterations.

The Jacobi polynomial iterations presented above were motivated by the facts that (a) the parameter-free polynomial iteration is not feasible in the distributed setting of gossip, and (b) the gossip matrix $W$ exhibits a structure due to the low-dimensional manifold on which the agents live. Interestingly, the literature on multiagent systems deals with some minimization

problems with the same properties. Examples are given by the estimation of quantities on graphs from relative measurements, in which the agents $v \in V$ try to estimate some quantity $x_v, v \in V$, defined over the graph, from noisy relative measurements over the edges of the graph:

$$\xi_{v,w} = x_v - x_w + \eta_{v,w}, \qquad \{v, w\} \in E.$$

This problem has applications in network localization, where the $x_v$ are the positions of the agents and the $\xi_{v,w}$ come from measurements of the distances and directions between the neighbors. It also has similar applications in time synchronization of clocks over networks, where $x_v$ is the offset of the clock of node $v$, and to motion consensus, where $x_v$ is the speed of agent $v$. For an introduction to estimation on graphs from relative measurements and its applications, see [3] and references therein. Note that the quantities $x_v$ can only be determined up to a global constant from the measurements; either we seek the true solution up to a constant only, or we assume that some agents know their true value.

A natural approach to solving the problem is to determine estimates $y_v$ of $x_v$ that minimize

$$\tfrac{1}{2} \sum_{v,w} W_{v,w} \left( \xi_{v,w} - (y_v - y_w) \right)^2,$$

where $W_{v,w}$ are some weights on the edges of the graph. Indeed, this corresponds to finding the maximum likelihood estimator if the noise $\eta_{v,w}$ is i.i.d. Gaussian and $W_{v,w}$ is the inverse variance of $\eta_{v,w}$. The above minimization problem is a quadratic problem whose covariance matrix is the Laplacian $I - W$. It can be solved using gradient descent or spectral gap based accelerations like the heavy-ball method. However, the conjugate gradient algorithm cannot be applied here as it involves centralized computations. The Jacobi polynomial iterations developed in this paper can be adapted to this situation to develop accelerations exploiting the structure of the Laplacian $I - W$. Experimenting with how this performs in real-world situations is left for future work.

**8. Conclusion.** Gossip methods based on the spectral gap were designed to improve the slow convergence rate of simple gossip. However, these methods are paradoxically bad at averaging locally in the intermediate regime before consensus is reached. In this paper, we propose another acceleration of simple gossip based on (i) the polynomial-based point of view, which designs iterations that are efficient at all times, and (ii) the Jacobi approximation, which uses prior information on the spectral dimension of the graph, a more natural property than the spectral gap.

It would be interesting for future work to better understand the Jacobi polynomial iteration in the asynchronous setting, i.e., when a randomized gossip matrix is used, as this setting is closer to practical cases.

In general, this paper advocates for the use of the polynomial point of view to design a gossip algorithm, as it allows us to use different types of prior information about the graph (spectral gap, spectral dimension, tree-like structure, etc.) and gives tools to prove the convergence of the designed algorithms.

## REFERENCES

[1] S. Andres, M. T. Barlow, J.-D. Deuschel, and B. M. Hambly, *Invariance principle for the random conductance model*, Probab. Theory Related Fields, 156 (2013), pp. 535–580.

[2] M. Arioli and J. Scott, *Chebyshev acceleration of iterative refinement*, Numer. Algorithms, 66 (2014), pp. 591–608.

[3] P. Barooah and J. P. Hespanha, *Estimation from relative measurements: Electrical analogy and large graphs*, IEEE Trans. Signal Process., 56 (2008), pp. 2181–2193.

[4] R. Berthier, *Code for the Paper: Accelerated Gossip in Network of Given Dimension*, https://github.com/raphael-berthier/jacobi-polynomial-iterations, 2019.

[5] M. Biskup, *Recent progress on the random conductance model*, Probab. Surv., 8 (2011), pp. 294–373.

[6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.

[7] M. Cao, D. A. Spielman, and E. M. Yeh, *Accelerated gossip algorithms for distributed computation*, in Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computation, 2006, pp. 952–959.

[8] P. Diaconis and L. Saloff-Coste, *Moderate growth and random walk on finite groups*, Geom. Funct. Anal., 4 (1994), pp. 1–36.

[9] R. Diekmann, A. Frommer, and B. Monien, *Efficient schemes for nearest neighbor load balancing*, Parallel Comput., 25 (1999), pp. 789–812.

[10] A. D. Dimakis, A. D. Sarwate, and M. J. Wainwright, *Geographic gossip: Efficient averaging for sensor networks*, IEEE Trans. Signal Process., 56 (2008), pp. 1205–1216.

[11] B. Durhuus, *Hausdorff and spectral dimension of infinite random graphs*, Acta Phys. Polon., 40 (2009), pp. 3509–3532.

[12] B. Fischer, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, John Wiley & Sons, Chichester, UK, B. G. Teubner, Stuttgart, Germany, 1996.

[13] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Oxford University Press, New York, 2004.

[14] G. Grimmett, *What is percolation?*, in Percolation, Springer, Berlin, 1999, pp. 1–31.

[15] G. F. Lawler and V. Limic, *Random Walk: A Modern Introduction*, Cambridge Stud. Adv. Math. 123, Cambridge University Press, Cambridge, UK, 2010.

[16] J. Liu, B. D. O. Anderson, M. Cao, and A. S. Morse, *Analysis of accelerated gossip algorithms*, Automatica J. IFAC, 49 (2013), pp. 873–883.

[17] P. Mathieu and E. Remy, *Isoperimetry and heat kernel decay on percolation clusters*, Ann. Probab., 32 (2004), pp. 100–128.

[18] C. C. Moallemi and B. Van Roy, *Consensus propagation*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2005, pp. 899–906.

[19] B. Mohar and W. Woess, *A survey on spectra of infinite graphs*, Bull. London Math. Soc., 21 (1989), pp. 209–234.

[20] P. Müller and P. Stollmann, *Spectral asymptotics of the Laplacian on supercritical bond-percolation graphs*, J. Funct. Anal., 252 (2007), pp. 233–246.

[21] P. Nevai, *Géza Freud, orthogonal polynomials and Christoffel functions. A case study*, J. Approx. Theory, 48 (1986), pp. 3–167.

[22] P. G. Nevai, *Orthogonal Polynomials*, AMS, Providence, RI, 1979.

[23] M. Penrose, *Random Geometric Graphs*, Oxford Stud. Probab. 5, Oxford University Press, Oxford, UK, 2003.

[24] B. T. Polyak, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, New York, 1987.

[25] P. Rebeschini and S. Tatikonda, *Accelerated consensus via min-sum splitting*, in Proceedings of the 31st Conference on Advances in Neural Information Processing Systems, Long Beach, CA, 2017, pp. 1374–1384.

[26] S. Sardellitti, M. Giona, and S. Barbarossa, *Fast distributed average consensus algorithms based on advection-diffusion processes*, IEEE Trans. Signal Process., 58 (2010), pp. 826–842.

[27] K. SCAMAN, F. BACH, S. BUBECK, Y. T. LEE, AND L. MASSOULIÉ, *Optimal algorithms for smooth and strongly convex distributed optimization in networks*, in Proceedings of the 34th International Conference on Machine Learning, Vol. 70, Sydney, Australia, 2017, pp. 3027–3036.

[28] D. SHAH, *Gossip algorithms*, Found. Trends® Networking, 3 (2009), pp. 1–125.

[29] G. SZEGÖ, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ. 23, AMS, Providence, RI, 1939.

# SUPPLEMENTARY MATERIALS: ACCELERATED GOSSIP IN NETWORKS OF GIVEN DIMENSION USING JACOBI POLYNOMIAL ITERATIONS*

RAPHAËL BERTHIER†, FRANCIS BACH†, AND PIERRE GAILLARD†

The supplementary material is organized as follows.

## SM1. Details of the simulations of Section 3.
We recall that the code is available online [SM1].

*2D grid.* We run simulations on a $40 \times 40$ square lattice ($n = 1600$ vertices) endowed with the gossip matrix defined in (2.1) with $d_{\max} = 4$. The results are plotted in Figure 2A and a $20 \times 20$ grid is plotted in Figure 1A for visualization.

*3D grid.* We run simulations on a $12 \times 12 \times 12$ cubic lattice ($n = 1728$ vertices) endowed with the gossip matrix defined in (2.1) with $d_{\max} = 6$. The results are plotted in Figure 2B.

*2D percolation bond.* We build a 2D percolation bond by taking a $40 \times 40$ 2D grid, and keep each edge independently with probability $p = 0.6$. To avoid connectivity issues, we consider $G$ the largest connected component of the resulting graph, endowed with the gossip matrix defined in (2.1) with $d_{\max} = 4$. The results are plotted in Figure 2C and a $20 \times 20$ percolation bond is plotted in Figure 1B for visualization.

*3D percolation bond.* We build a 3D percolation bond by taking a $12 \times 12 \times 12$ 3D grid and keep each edge independently with probability $p = 0.4$. To avoid connectivity issues, we consider $G$ the largest connected component of the resulting graph, endowed with the gossip matrix defined in (2.1) with $d_{\max} = 6$. The results are plotted in Figure 2D.

*2D random geometric graph.* We build a 2D random geometric graph $G$ by sampling $n = 1600$ points uniformly in the unit square $[0, 1]^2$ and linking pairs closer than $1.5/\sqrt{n} = 0.0375$. To avoid connectivity issues, we consider $G$ the

largest connected component of the resulting graph. We build a gossip matrix $W$ on $G$ with the formulas: $W_{vw} = \max(\deg v, \deg w)^{-1}$ if $v \in \mathcal{N}(w)$ and $W_{vv} = 1 - \sum_{w \in \mathcal{N}(v)} \max(\deg v, \deg w)^{-1}$. The results are shown in Figure 2E.

*3D random geometric graph.* We build a 3D random geometric graph $G$ by sampling $n = 1728$ points in the unit cube $[0, 1]^3$ and linking pairs closer than $1.5/n^{1/3} = 0.125$. To avoid connectivity issues, we consider $G$ the largest connected component of the resulting graph. We build a gossip matrix $W$ on $G$ with the formulas: $W_{vw} = \max(\deg v, \deg w)^{-1}$ if $v \in \mathcal{N}(w)$ and $W_{vv} = 1 - \sum_{w \in \mathcal{N}(v)} W_{vw}$. The results are shown in Figure 2F.

**SM2. Toolbox from orthogonal polynomials.** In this section, we describe the tools from the theory of orthogonal polynomials that we use in this paper. The definition of the orthogonal polynomials $\pi_t$ w.r.t. some measure $\tau$ is given in Definition 4.1. We start by giving some general properties of orthogonal polynomials in Section SM2.1. We then describe two parametrized measures with respect to which orthogonal polynomials can be explicitly described: the Jacobi polynomials, in Section SM2.2, and the polynomials orthogonal to some measure of the form $(1 - \lambda^2)^{1/2}/\rho(\lambda)$, where $\rho$ is some polynomial, in Section SM2.3. We finally give in Section SM2.4 some asymptotic properties of the Jacobi polynomials as $t \to \infty$.

**SM2.1. General properties.**

PROPOSITION SM2.1 (from [SM8, Theorem 3.3.1]). *Let $\pi_t$ be a family of orthogonal polynomials w.r.t. some measure $\tau$ on some interval $[a, b]$. Then the zeros of $\pi_t$ are real, distinct and located in the interior of $[a, b]$.*

In Proposition 4.4, it is stated that the orthogonal polynomials satisfy a three-term recurrence relation. We write here the short proof as it is used in Section SM5.

*Proof of Proposition 4.4.* The polynomial $\lambda \pi_t(\lambda)$ of the variable $\lambda$ is of degree $t + 1$, thus it can be decomposed over the orthogonal basis $\pi_0(\lambda), \pi_1(\lambda), \ldots, \pi_{t+1}(\lambda)$:

$$\lambda \pi_t(\lambda) = \sum_{s=0}^{t+1} \frac{\langle \lambda \pi_t, \pi_s \rangle_\tau}{\langle \pi_s, \pi_s \rangle_\tau} \pi_s(\lambda) \,.$$

Note that $\langle \lambda \pi_t, \pi_s \rangle_\tau = \int \lambda \pi_t(\lambda) \pi_s(\lambda) \mathrm{d}\tau(\lambda) = \langle \pi_t, \lambda \pi_s \rangle_\tau = 0$ when $s \leqslant t - 2$ because in this case $\lambda \pi_s(\lambda) \in \mathbb{R}_{t-1}[X]$ and $\pi_t$ is orthogonal to $\mathbb{R}_{t-1}[X]$. Thus

$$\lambda \pi_t(\lambda) = \frac{\langle \lambda \pi_t, \pi_{t+1} \rangle_\tau}{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau} \pi_{t+1}(\lambda) + \frac{\langle \lambda \pi_t, \pi_t \rangle_\tau}{\langle \pi_t, \pi_t \rangle_\tau} \pi_t(\lambda) + \frac{\langle \lambda \pi_t, \pi_{t-1} \rangle_\tau}{\langle \pi_{t-1}, \pi_{t-1} \rangle_\tau} \pi_{t-1}(\lambda) \,,$$

with the convention $\pi_{-1} = 0$. Note that $\langle \lambda \pi_t, \pi_{t+1} \rangle_\tau$ is non-zero as otherwise it would imply that $\lambda \pi_t$ is a polynomial of degree smaller or equal to $t$, which is absurd. We get the recursion formula by denoting

$$(SM2.1) \quad a_t = \frac{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau}{\langle \lambda \pi_t, \pi_{t+1} \rangle_\tau} \,, \qquad b_t = -\frac{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau \langle \lambda \pi_t, \pi_t \rangle_\tau}{\langle \lambda \pi_t, \pi_{t+1} \rangle_\tau \langle \pi_t, \pi_t \rangle_\tau} \,,$$

$$c_t = \frac{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau \langle \lambda \pi_t, \pi_{t-1} \rangle_\tau}{\langle \lambda \pi_t, \pi_{t+1} \rangle_\tau \langle \pi_{t-1}, \pi_{t-1} \rangle_\tau} \,.$$

$\square$

**SM2.2. Jacobi polynomials.**

DEFINITION SM2.2 (from [SM8, Chapter IV]). *Let $\alpha, \beta > -1$. The Jacobi polynomials $P_t^{(\alpha,\beta)}$ are the orthogonal polynomials w.r.t. the Jacobi measure*

$$\sigma^{(\alpha,\beta)}(\mathrm{d}\lambda) = (1-\lambda)^\alpha (1+\lambda)^\beta \mathbf{1}_{\{\lambda \in (-1,1)\}}\mathrm{d}\lambda,$$

*normalized such that $P_t^{(\alpha,\beta)}(1) = \binom{t+\alpha}{t}$.*

EXAMPLE SM2.3 (from [SM8, Section 2.4]).

1. *The Chebyshev polynomials $T_t$ of the first kind are the orthogonal polynomials w.r.t. $\sigma^{(-1/2,-1/2)}(\mathrm{d}\lambda) = (1-\lambda^2)^{-1/2}$ and normalized such that $T_t(1) = 1$. They are, up to some rescaling, a family of Jacobi polynomials. They satisfy the trigonometric formula*

$$T_t(\cos\theta) = \cos(t\theta).$$

2. *The Chebyshev polynomials $U_t$ of the second kind are the orthogonal polynomials w.r.t. $\sigma^{(1/2,1/2)}(\mathrm{d}\lambda) = (1-\lambda^2)^{1/2}$ and normalized such that $U_t(1) = t+1$. They are, up to some rescaling, a family of Jacobi polynomials. They satisfy the trigonometric formula*

$$U_t(\cos\theta) = \frac{\sin(t+1)\theta}{\sin\theta}.$$

A remarkable property of the Jacobi polynomials is that their recurrence relation can be computed explicitly.

PROPOSITION SM2.4 (from [SM8, Section 4.4.5]). *Let $\alpha, \beta > -1$. The Jacobi polynomials $P_t^{\alpha,\beta}$ satisfy the three recurrence formula*

$$P_0^{(\alpha,\beta)}(\lambda) = 1, \quad P_1^{(\alpha,\beta)}(\lambda) = \frac{1}{2}(\alpha+\beta+2)\lambda + \frac{1}{2}(\alpha-\beta),$$

$$2(t+1)(t+1+\alpha+\beta)(2t+\alpha+\beta)P_{t+1}^{(\alpha,\beta)}(\lambda)$$
$$= (2t+\alpha+\beta+1)[(2t+\alpha+\beta+2)(2t+\alpha+\beta)\lambda + \alpha^2 - \beta^2]P_t^{(\alpha,\beta)}(\lambda)$$
$$- 2(t+\alpha)(t+\beta)(2t+\alpha+\beta+2)P_{t-1}^{(\alpha,\beta)}(\lambda).$$

EXAMPLE SM2.5. *The Chebyshev polynomials of the first and the second kind satisfy the same recurrence formula, but with different initializations:*

$$\begin{aligned} T_0(\lambda) &= 1, & T_1(\lambda) &= \lambda, & T_{t+1}(\lambda) &= 2\lambda T_t(\lambda) - T_{t-1}(\lambda), \\ U_0(\lambda) &= 1, & U_1(\lambda) &= 2\lambda, & U_{t+1}(\lambda) &= 2\lambda U_t(\lambda) - U_{t-1}(\lambda). \end{aligned}$$

PROPOSITION SM2.6 (from [SM8, Theorem 7.32.1]). *Let $\alpha, \beta \geqslant 1/2$. Then*

$$\max_{\lambda \in [-1,1]} \left| P_t^{(\alpha,\beta)}(\lambda) \right| = \binom{t+\max(\alpha,\beta)}{t}.$$

**SM2.3. Polynomials orthogonal w.r.t. $(1-\lambda^2)^{1/2}/\rho(\lambda)$, $\rho$ polynomial.** In this section, we present how one can compute the recurrence relation for some orthogonal polynomials w.r.t. a weight of the form $(1-\lambda^2)^{1/2}/\rho(\lambda)$, $\rho$ polynomial.

PROPOSITION SM2.7 (from [SM8, Theorem 1.2.1]). *Let $\rho$ be a real polynomial of degree $l$ which is non-negative for $\lambda \in [-1,1]$. Then there exists a polynomial $h$ of degree $l$ such that for all real $\theta$, $\rho(\cos\theta) = |h(e^{i\theta})|^2$.*

PROPOSITION SM2.8 (from [SM8, Theorem 2.6]).   *Let $\rho$ be a real polynomial of degree $l$ taking positive values on the interval $[-1, 1]$, and*

$$\tau(\mathrm{d}\lambda) = \frac{(1-\lambda^2)^{1/2}}{\rho(\lambda)}\mathrm{d}\lambda.$$

*Let $h$ be a polynomial of degree $l$ such that $\rho(\cos\theta) = |h(e^{i\theta})|^2$ (see Proposition SM2.7), and decompose $h(e^{i\theta}) = c(\theta) + is(\theta)$, $c(\theta)$ and $s(\theta)$ real. Then the polynomials*

$$\pi_t(\cos\theta) = c(\theta)U_t(\cos\theta) - \frac{s(\theta)}{\sin\theta}T_{t+1}(\cos\theta)$$

*are orthogonal w.r.t. $\tau$ when $l < 2(t+1)$.*

**SM2.4. Asymptotics for the Jacobi polynomials.** To prove the asymptotic performance guarantees of the polynomial iterations we build in this paper, we need the following asymptotic properties of the Jacobi polynomials.

PROPOSITION SM2.9 (from [SM8, Theorem 8.21.7]).   *Let $\alpha, \beta > -1$, and $\lambda > 1$ a real number. Then there exists a positive constant $c = c(\alpha, \beta, \lambda)$ such that*

$$P_t^{(\alpha,\beta)}(\lambda) \underset{t\to\infty}{\sim} \frac{c}{t^{1/2}}\left(\lambda + \sqrt{\lambda^2 - 1}\right)^t.$$

In the special case of the Chebyshev polynomials, we also have similar non-asymptotic bounds.

LEMMA SM2.10.   *For all $\lambda > 1$, for all $t \geqslant 0$,*

(SM2.2)
$$\frac{1}{2}\left(\lambda + \sqrt{\lambda^2 - 1}\right)^t \leqslant T_t(\lambda) \leqslant \left(\lambda + \sqrt{\lambda^2 - 1}\right)^t,$$

(SM2.3)
$$\left(\lambda + \sqrt{\lambda^2 - 1}\right)^t \leqslant U_t(\lambda) \leqslant (t+1)\left(\lambda + \sqrt{\lambda^2 - 1}\right)^t.$$

*Proof.* We start by deriving a classic expression for the Chebyshev polynomials. The identities

$$T_t(\cos\theta) = \cos(t\theta), \qquad\qquad U_t(\cos\theta) = \frac{\sin((t+1)\theta)}{\sin\theta},$$

can be interpreted as

$$T_t\left(\frac{z + z^{-1}}{2}\right) = \frac{z^t + z^{-t}}{2}, \quad U_t\left(\frac{z + z^{-1}}{2}\right) = \frac{z^{t+1} - z^{-(t+1)}}{z - z^{-1}}, \quad \text{for } z = e^{i\theta}.$$

The above equations are equalities of holomorphic functions on the unit circle, it implies that the identities must be true for all complex numbers $z \neq 0$; we use it here for real numbers $z$.

For $\lambda > 1$, write $\lambda = (z + z^{-1})/2$, $z > 1$. This is equivalent to $z = \lambda + \sqrt{\lambda^2 - 1}$. Then

$$T_t(\lambda) = \frac{z^t + z^{-t}}{2} = \frac{1 + z^{-2t}}{2}z^t = \frac{1 + z^{-2t}}{2}\left(\lambda + \sqrt{\lambda^2 - 1}\right)^t.$$

As $z > 1$,

$$\frac{1}{2} \leqslant \frac{1 + z^{-2t}}{2} \leqslant 1.$$

This proves the inequalities (SM2.2). Further,

$$U_t(\lambda) = \frac{z^{t+1} - z^{-(t+1)}}{z - z^{-1}} = \frac{1 - z^{-2t-2}}{1 - z^{-2}} z^t = \frac{1 - z^{-2t-2}}{1 - z^{-2}} \left( \lambda + \sqrt{\lambda^2 - 1} \right)^t.$$

As $z > 1$,

$$1 \leqslant \frac{1 - z^{-2t-2}}{1 - z^{-2}} \leqslant t + 1.$$

This proves the inequalities (SM2.3). □

PROPOSITION SM2.11 (from [SM8, Theorem 7.32.2]). *Let* $\alpha, \beta > -1$. *There exists two constants* $C_1, C_2 > 0$ *such that,*

$$\left| P_t^{(\alpha,\beta)}(\cos\theta) \right| \leqslant \begin{cases} C_1 \theta^{-\alpha - 1/2} t^{-1/2} & \text{if } 1/t \leqslant \theta \leqslant \pi/2, \\ C_2 t^\alpha & \text{if } 0 \leqslant \theta \leqslant 1/t. \end{cases}$$

**SM3. Some basic tools for the proofs.**

**SM3.1. Comparing integrals using a domination of the cumulative distribution function.**

LEMMA SM3.1. *Let* $\sigma, \tau$ *be two positive measures on some interval* $[a, b]$ *such that for all* $\lambda \in [a, b]$,

(SM3.1) $$\sigma([\lambda, b]) \leqslant \tau([\lambda, b]).$$

*Then for all continuous non-decreasing functions* $f : [a, b] \to \mathbb{R}_{\geqslant 0}$,

$$\int_{[a,b]} f(\lambda) \mathrm{d}\sigma(\lambda) \leqslant \int_{[a,b]} f(\lambda) \mathrm{d}\tau(\lambda).$$

*Proof.* For any $u \in \mathbb{R}_{\geqslant 0}$, denote $\lambda^*(u) = \min\{\lambda | u \leqslant f(\lambda)\}$.

$$\int_{[a,b]} f(\lambda) \mathrm{d}\sigma(\lambda) = \int_{[a,b]} \int_{\mathbb{R}_{\geqslant 0}} \mathbf{1}_{\{u \leqslant f(\lambda)\}} f(\lambda) \mathrm{d}u \mathrm{d}\sigma(\lambda)$$

$$= \int_{\mathbb{R}_{\geqslant 0}} \left( \int_{[a,b]} \mathbf{1}_{\{u \leqslant f(\lambda)\}} \mathrm{d}\sigma(\lambda) \right) \mathrm{d}u = \int_{\mathbb{R}_{\geqslant 0}} \sigma([\lambda^*(u), b]) \mathrm{d}u.$$

The proof is finished using (SM3.1) and similar equalities for $\tau$. □

.

**SM3.2. The gamma and beta function.** The gamma function $\Gamma$ and the beta function $B$ are defined as [SM5, Section 5.2, Section 5.12]

$$\Gamma(z) = \int_0^\infty e^{-u} u^{z-1} \mathrm{d}u, \quad z \geqslant 0;$$

$$B(a, b) = \int_0^1 s^{a-1}(1-s)^{b-1} \mathrm{d}s = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a, b > 0.$$

The asymptotic ratios of the gamma functions are given in [SM5, Eq. 5.11.12]: for $c, d \in \mathbb{R}$,

$$\frac{\Gamma(z+c)}{\Gamma(z+d)} \sim z^{c-d} \quad \text{as } z \to +\infty.$$

This gives the asymptotic of the beta function

$$(SM3.2) \qquad\qquad B(a,b) \sim \frac{\Gamma(b)}{a^b} \quad \text{as } a \to +\infty.$$

**SM4. Proof of Proposition 4.2.** Note first that as $d\tau(\lambda) = (1 - \lambda)d\sigma(\lambda)$ is a measure on $[-1, 1]$, if $\tilde{\pi}_0, \ldots, \tilde{\pi}_{T-1}$ is a sequence of orthogonal polynomials w.r.t. $\tau$, then the zeros of the polynomials $\tilde{\pi}_0, \ldots, \tilde{\pi}_{T-1}$ are located in the interior of $[-1, 1]$ (see Proposition SM2.1). In particular, $\tilde{\pi}_t(1) \neq 0$, $t < T$. Thus it is possible to build a family $\pi_0 = \tilde{\pi}_0/\tilde{\pi}_0(1), \ldots, \pi_{T-1} = \tilde{\pi}_{T-1}/\tilde{\pi}_{T-1}(1)$ of orthogonal polynomials normalized to take value 1 at 1, as it is done in Proposition 4.2.

The polynomial $\pi_t$ satisfies $\pi_t(1) = 1$ and $\deg \pi_t = t$. We now consider some polynomial $Q_t$ also satisfying $Q_t(1) = 1$ and $\deg Q_t = t$, and show that

$$\int \pi_t(\lambda)^2 d\sigma(\lambda) \leqslant \int Q_t(\lambda)^2 d\sigma(\lambda), \qquad \text{i.e.} \qquad \langle \pi_t, \pi_t \rangle_\sigma \leqslant \langle Q_t, Q_t \rangle_\sigma.$$

The polynomial $Q_t - \pi_t$ vanishes at 1 thus there exists a polynomial $R_{t-1}$ of degree at most $t - 1$ such that $Q_t(\lambda) = \pi_t(\lambda) + (1 - \lambda)R_{t-1}(\lambda)$. Then

$$\langle Q_t, Q_t \rangle_\sigma = \langle \pi_t, \pi_t \rangle_\sigma + 2\langle \pi_t, (1 - \lambda)R_{t-1} \rangle_\sigma + \langle (1 - \lambda)R_{t-1}, (1 - \lambda)R_{t-1} \rangle_\sigma.$$

Note that $\langle \pi_t, (1 - \lambda)R_{t-1} \rangle_\sigma = \langle \pi_t, R_{t-1} \rangle_\tau = 0$ because $\pi_t$ is orthogonal to all polynomials of degree smaller or equal to $t - 1$ w.r.t. $\langle ., . \rangle_\tau$. Moreover,

$$\langle (1 - \lambda)R_{t-1}, (1 - \lambda)R_{t-1} \rangle_\sigma = \int (1 - \lambda)^2 R_{t-1}(\lambda)^2 d\sigma(\lambda) \geqslant 0.$$

Thus $\langle Q_t, Q_t \rangle_\sigma \geqslant \langle \pi_t, \pi_t \rangle_\sigma$. This shows that $\pi_t$ is a minimizer.

We now show that the minimizer $\pi_t$ is unique. There is equality $\langle Q_t, Q_t \rangle_\sigma = \langle \pi_t, \pi_t \rangle_\sigma$ if and only if $\int (1 - \lambda)^2 R_{t-1}(\lambda)^2 d\sigma(\lambda) = 0$, i.e. $(1 - \lambda)R_{t-1}$ vanishes on $\text{Supp}\,\sigma$. But the cardinal of $\text{Supp}\,\sigma$ is at least $T$ while $(1 - \lambda)R_{t-1}$ is a polynomial of degree at most $t \leqslant T - 1$. Thus the equality case is reached if and only if $R_{t-1} = 0$, i.e. $Q_t = \pi_t$.

**SM5. The parameter-free polynomial iteration.** In this section, we give the details of the implementation of the parameter-free polynomial iteration in a centralized setting. We explicit the computation of the optimal coefficients $a_t$, $b_t$ and $c_t$. It is used in the simulation of Figure 4.

The parameter free polynomial iteration is Eq. (4.3), where the coefficients $a_t, b_t, c_t$, $t \geqslant 1$, are determined in Eq. (SM2.1). The results are repeated here for convenience:

$$x^0 = \xi, \qquad\qquad x^1 = a_0 W\xi + b_0\xi, \qquad\qquad x^{t+1} = a_t W x^t + b_t x^t - c_t x^{t-1},$$

$$a_t = \frac{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau}{\langle \lambda\pi_t, \pi_{t+1} \rangle_\tau}, \quad b_t = -\frac{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau \langle \lambda\pi_t, \pi_t \rangle_\tau}{\langle \lambda\pi_t, \pi_{t+1} \rangle_\tau \langle \pi_t, \pi_t \rangle_\tau}, \quad c_t = \frac{\langle \pi_{t+1}, \pi_{t+1} \rangle_\tau \langle \lambda\pi_t, \pi_{t-1} \rangle_\tau}{\langle \lambda\pi_t, \pi_{t+1} \rangle_\tau \langle \pi_{t-1}, \pi_{t-1} \rangle_\tau}.$$

where $\tau = (1 - \lambda)\sigma$, $\sigma$ is defined in (4.1). Note that the scalar products that appear in the formulas for $a_t, b_t, c_t$ can be computed from the iterates $x^t = \pi_t(W)\xi$, $t \geqslant 0$.

For instance,

$$\langle \lambda \pi_t, \pi_{t-1} \rangle_\tau = \int \lambda \pi_t(\lambda) \pi_{t-1}(\lambda)(1-\lambda) \mathrm{d}\sigma(\lambda)$$

$$= \sum_{i=1}^{n} \langle \xi, u^i \rangle^2 \lambda_i \pi_t(\lambda_i) \pi_{t-1}(\lambda_i)(1-\lambda_i)$$

$$= \langle W\pi_t(W)\xi, (I-W)\pi_{t-1}(W)\xi \rangle$$

$$= \langle Wx^t, x^{t-1} - Wx^{t-1} \rangle.$$

Note that the last line requires the computation of a scalar product $\langle ., . \rangle$ over $\mathbb{R}^V$, which means summing over $v \in V$. This is possible in simulations where we can centralize the information of the nodes $v \in V$. However in practical situation where the coordinates of $x^t$ are distributed among the nodes, such a computation requires many additional communication steps. This makes the parameter free polynomial iteration impractical.

The computation of the other scalar products give

$$b_t = -a_t \frac{\langle x^t - Wx^t, Wx^t \rangle}{\langle x^t, x^t - Wx^t \rangle}, \qquad c_t = a_t \frac{\langle x^t, x^{t-1} - Wx^{t-1} \rangle}{\langle x^{t-1}, x^{t-1} - Wx^{t-1} \rangle},$$

and as $a_t + b_t - c_t = 1$ (that follows from $\pi_t(1) = 1$ for all $t$), we get for $t \geqslant 1$,

$$\tilde{b}_t = -\frac{\langle x^t - Wx^t, Wx^t \rangle}{\langle x^t, x^t - Wx^t \rangle}, \qquad \tilde{c}_t = \frac{\langle x^t, x^{t-1} - Wx^{t-1} \rangle}{\langle x^{t-1}, x^{t-1} - Wx^{t-1} \rangle},$$

$$x^{t+1} = \frac{1}{1 + \tilde{b}_t - \tilde{c}_t} \left( Wx^t + \tilde{b}_t x^t - \tilde{c}_t x^{t-1} \right).$$

Similarly, one can compute that

$$x^1 = \frac{1}{1 + \tilde{b}^0} \left( W\xi + \tilde{b}^0 \xi \right), \qquad \tilde{b}^0 = -\frac{\langle \xi - W\xi, W\xi \rangle}{\langle \xi, \xi - W\xi \rangle},$$

which gives the initialization of the parameter-free polynomial iteration.

**SM6. Computation of the recursion coefficients of some orthogonal polynomials.**

**SM6.1. A rescaling lemma for orthogonal polynomials.** We start with a lemma giving the change in the recursion coefficients of orthogonal polynomials when the underlying measure undergoes an affine transformation. It is used in the next subsections.

LEMMA SM6.1. *Let $\sigma$ be a measure on $\mathbb{R}$, $\pi_0, \ldots, \pi_{T-1}$ a sequence of orthogonal polynomials w.r.t. $\sigma$ and*

(SM6.1) $$\pi_{t+1}(\lambda) = (a_t \lambda + b_t) \pi_t(\lambda) - c_t \pi_{t-1}(\lambda), \qquad t \geqslant 1,$$

*their recurrence formula (see Definition 4.1 and Theorem 4.4).*

*Let $\varphi : \lambda \mapsto \alpha \lambda + \beta$, $\alpha \neq 0$ be a linear function and $\tilde{\sigma}$ be the image measure of $\sigma$ by $\varphi$ (which means that for all measurable set $A$, $\tilde{\sigma}(A) = \sigma(\varphi^{-1}(A))$). Then a sequence of orthogonal polynomials w.r.t. $\tilde{\sigma}$ is given by the formula*

$$\tilde{\pi}_t(\tilde{\lambda}) := \pi_t \left( \varphi^{-1}(\tilde{\lambda}) \right) = \pi_t \left( \frac{\tilde{\lambda} - \beta}{\alpha} \right).$$

*These polynomials follow the recursion formula*

$$\tilde{\pi}_{t+1}(\tilde{\lambda}) = (\tilde{a}_t \tilde{\lambda} + \tilde{b}_t)\tilde{\pi}_t(\tilde{\lambda}) - \tilde{c}_t \tilde{\pi}_{t-1}(\tilde{\lambda})\,,$$

$$\tilde{a}_t = \frac{a_t}{\alpha}\,, \qquad \tilde{b}_t = b_t - \frac{a_t \beta}{\alpha}\,, \qquad \tilde{c}_t = c_t\,.$$

*Proof.* By change of variable,

$$\int \tilde{\pi}_t(\tilde{\lambda})\tilde{\pi}_s(\tilde{\lambda})\mathrm{d}\tilde{\sigma}(\tilde{\lambda}) = \int \pi_t\left(\varphi^{-1}(\tilde{\lambda})\right)\pi_s\left(\varphi^{-1}(\tilde{\lambda})\right)\mathrm{d}\tilde{\sigma}(\tilde{\lambda})$$

$$= \int \pi_t\left(\varphi^{-1}(\varphi(\lambda))\right)\pi_s\left(\varphi^{-1}(\varphi(\lambda))\right)\mathrm{d}\sigma(\lambda)$$

$$= \int \pi_t(\lambda)\pi_s(\lambda)\mathrm{d}\sigma(\lambda) = \mathbf{1}_{\{s=t\}}\,,$$

and $\deg \tilde{\pi}_t = t$ thus $\tilde{\pi}_0, \ldots, \tilde{\pi}_{T-1}$ are orthogonal polynomials w.r.t. $\tilde{\sigma}$. The recurrence relation for $\tilde{\pi}_t$ follows by evaluating the recurrence relation (SM6.1) for $\pi_t$ in $(\tilde{\lambda} - \beta)/\alpha$.                                                                                                   □

**SM6.2. Jacobi polynomials.** Let $\alpha, \beta > -1$. In this section, we derive, using the recurrence formula for the Jacobi polynomial $P_t^{(\alpha,\beta)}$ of Proposition SM2.4, a similar recurrence relation for the polynomials $\pi_t^{(\alpha,\beta)}$ orthogonal w.r.t. the Jacobi measure $\sigma^{(\alpha,\beta)}$, but normalized such that $\pi_t^{(\alpha,\beta)}(1) = 1$.

Substituting $P_t^{(\alpha,\beta)} = \binom{t+\alpha}{t}\pi_t^{(\alpha,\beta)}$ in the recurrence relation of Proposition SM2.4, we get

$$2(t+1)(t+1+\alpha+\beta)(2t+\alpha+\beta)\binom{t+1+\alpha}{t+1}\pi_{t+1}^{(\alpha,\beta)}(\lambda)$$

$$= (2t+\alpha+\beta+1)[(2t+\alpha+\beta+2)(2t+\alpha+\beta)\lambda + \alpha^2 - \beta^2]\binom{t+\alpha}{t}\pi_t^{(\alpha,\beta)}(\lambda)$$

$$- 2(t+\alpha)(t+\beta)(2t+\alpha+\beta+2)\binom{t-1+\alpha}{t-1}\pi_{t-1}^{(\alpha,\beta)}(\lambda)\,.$$

Using that $(t+1)\binom{t+1+\alpha}{t+1} = (t+1+\alpha)\binom{t+\alpha}{t}$ and $t\binom{t+\alpha}{t} = (t+\alpha)\binom{t-1+\alpha}{t-1}$, we can divide the above equation by $\binom{t+\alpha}{t}$. We get

$$2(t+1+\alpha+\beta)(2t+\alpha+\beta)(t+1+\alpha)\pi_{t+1}^{(\alpha,\beta)}(\lambda)$$

$$= (2t+\alpha+\beta+1)[(2t+\alpha+\beta+2)(2t+\alpha+\beta)\lambda + (\alpha+\beta)(\alpha-\beta)]\pi_t^{(\alpha,\beta)}(\lambda)$$

$$- 2t(t+\beta)(2t+\alpha+\beta+2)\pi_{t-1}^{(\alpha,\beta)}(\lambda)\,.$$

Summing up, we obtain the recursion formula

$$\pi_0(\lambda) = 1\,, \qquad \pi_1(\lambda) = a_0^{(\alpha,\beta)}\lambda + b_0^{(\alpha,\beta)}\,,$$

$$\pi_{t+1}^{(\alpha,\beta)}(\lambda) = \left(a_t^{(\alpha,\beta)}\lambda + b_t^{(\alpha,\beta)}\right)\pi_t^{(\alpha,\beta)}(\lambda) - c_t^{(\alpha,\beta)}\pi_{t-1}^{(\alpha,\beta)}(\lambda)\,,$$

with the recursion coefficients

$$a_0^{(\alpha,\beta)} = \frac{\alpha + \beta + 2}{2(1 + \alpha)}, \qquad b_0^{(\alpha,\beta)} = \frac{\alpha - \beta}{2(1 + \alpha)},$$

(SM6.2)
$$a_t^{(\alpha,\beta)} = \frac{(2t + \alpha + \beta + 1)(2t + \alpha + \beta + 2)}{2(t + 1 + \alpha + \beta)(t + 1 + \alpha)},$$

$$b_t^{(\alpha,\beta)} = \frac{(2t + \alpha + \beta + 1)(\alpha + \beta)(\alpha - \beta)}{2(t + 1 + \alpha + \beta)(t + 1 + \alpha)(2t + \alpha + \beta)},$$

$$c_t^{(\alpha,\beta)} = \frac{t(t + \beta)(2t + \alpha + \beta + 2)}{(t + 1 + \alpha + \beta)(2t + \alpha + \beta)(t + 1 + \alpha)}.$$

**SM6.3. Rescaled Jacobi polynomials.** Let $\alpha, \beta > -1$. In this section, we determine a recursion formula for the orthogonal polynomials $\pi_t^{(\alpha,\beta,\gamma)}$ w.r.t. the rescaled Jacobi measure

$$d\sigma^{(\alpha,\beta,\gamma)}(\lambda) = ((1 - \gamma) - \lambda)^\alpha (1 + \lambda)^\beta \mathbf{1}_{\{\lambda \in (-1, 1 - \gamma)\}} d\lambda,$$

The polynomials $\pi_t^{(\alpha,\beta,\gamma)}$ are normalized such that $\pi_t^{(\alpha,\beta,\gamma)}(1) = 1$.

Note that, up to a rescaling, $d\sigma^{(\alpha,\beta,\gamma)}$ is the image measure of the Jacobi measure $d\sigma^{(\alpha,\beta)}$ (defined in (SM2.2)) by the linear function $\varphi_\gamma(\lambda) = (1 - \gamma/2)\lambda - \gamma/2$. Thus Lemma SM6.1 gives a family of orthogonal polynomials $P_t^{(\alpha,\beta,\gamma)}$ w.r.t. $d\sigma^{(\alpha,\beta,\gamma)}$ and their recursion formula:

$$P_t^{(\alpha,\beta,\gamma)}(\tilde{\lambda}) = \pi_t^{(\alpha,\beta)}\left(\varphi_\gamma^{-1}(\tilde{\lambda})\right),$$

$$P_{t+1}^{(\alpha,\beta,\gamma)}(\tilde{\lambda}) = \left(a_t^{(\alpha,\beta,\gamma)} \tilde{\lambda} + b_t^{(\alpha,\beta,\gamma)}\right) P_t^{(\alpha,\beta,\gamma)}(\tilde{\lambda}) - c_t^{(\alpha,\beta,\gamma)} P_{t-1}^{(\alpha,\beta,\gamma)}(\tilde{\lambda}),$$

$$a_t^{(\alpha,\beta,\gamma)} = a_t^{(\alpha,\beta)}\left(1 - \frac{\gamma}{2}\right)^{-1}, \quad b_t^{(\alpha,\beta,\gamma)} = b_t^{(\alpha,\beta)} + \frac{\gamma}{2} a_t^{(\alpha,\beta)}\left(1 - \frac{\gamma}{2}\right)^{-1}, \quad c_t^{(\alpha,\beta,\gamma)} = c_t^{(\alpha,\beta)}.$$

However, the polynomials $P_t^{(\alpha,\beta,\gamma)}$ are not normalized such that $P_t^{(\alpha,\beta,\gamma)} = 1$. Indeed, $P_t^{(\alpha,\beta,\gamma)} = \pi_t^{(\alpha,\beta)}\left((1 - \gamma/2)^{-1}(1 + \gamma/2)\right)$. It is difficult to deduce the recurrence relation for $\pi_t^{(\alpha,\beta,\gamma)} = P_t^{(\alpha,\beta,\gamma)}/P_t^{(\alpha,\beta,\gamma)}(1)$ from the recurrence relation for $P_t^{(\alpha,\beta,\gamma)}$. One can circumvent this difficulty by using that the normalization $P_t^{(\alpha,\beta,\gamma)}(1)$ also follows the recurrence relation

$$P_{t+1}^{(\alpha,\beta,\gamma)}(1) = \left(a_t^{(\alpha,\beta,\gamma)} + b_t^{(\alpha,\beta,\gamma)}\right) P_t^{(\alpha,\beta,\gamma)}(1) - c_t^{(\alpha,\beta,\gamma)} P_{t-1}^{(\alpha,\beta,\gamma)}(1).$$

Summing things up, we get
(SM6.3)

$$\pi_t^{(\alpha,\beta,\gamma)}(\lambda) = \frac{P_t^{(\alpha,\beta,\gamma)}(\lambda)}{P_t^{(\alpha,\beta,\gamma)}(1)},$$

$$P_0^{(\alpha,\beta,\gamma)}(\lambda) = 1, \qquad P_0^{(\alpha,\beta,\gamma)}(1) = 1,$$

$$P_1^{(\alpha,\beta,\gamma)}(\lambda) = a_0^{(\alpha,\beta,\gamma)}\lambda + b_0^{(\alpha,\beta,\gamma)}, \qquad P_1^{(\alpha,\beta,\gamma)}(1) = a_0^{(\alpha,\beta,\gamma)} + b_0^{(\alpha,\beta,\gamma)},$$

$$P_{t+1}^{(\alpha,\beta,\gamma)}(\lambda) = \left(a_t^{(\alpha,\beta,\gamma)} \tilde{\lambda} + b_t^{(\alpha,\beta,\gamma)}\right) P_t^{(\alpha,\beta,\gamma)}(\lambda) - c_t^{(\alpha,\beta,\gamma)} P_{t-1}^{(\alpha,\beta,\gamma)}(\lambda), \qquad t \geqslant 1,$$

$$P_{t+1}^{(\alpha,\beta,\gamma)}(1) = \left(a_t^{(\alpha,\beta,\gamma)} + b_t^{(\alpha,\beta,\gamma)}\right) P_t^{(\alpha,\beta,\gamma)}(1) - c_t^{(\alpha,\beta,\gamma)} P_{t-1}^{(\alpha,\beta,\gamma)}(1), \qquad t \geqslant 1,$$

$$a_t^{(\alpha,\beta,\gamma)} = a_t^{(\alpha,\beta)}\left(1 - \frac{\gamma}{2}\right)^{-1}, \qquad b_t^{(\alpha,\beta,\gamma)} = b_t^{(\alpha,\beta)} + \frac{\gamma}{2} a_t^{(\alpha,\beta)}\left(1 - \frac{\gamma}{2}\right)^{-1}, \qquad t \geqslant 0,$$

$$c_t^{(\alpha,\beta,\gamma)} = c_t^{(\alpha,\beta)}, \qquad t \geqslant 1.$$

These equations give a practical way to compute the polynomials $\pi_t^{(\alpha,\beta,\gamma)}$ because all recursion coefficients can be computed explicitly using the formulas (SM6.2).

**SM6.4. Polynomials orthogonal to the modified Kesten-McKay measure.** In this section, we determine the recurrence formula for the orthogonal polynomials $\pi_t$ w.r.t. the modified Kesten-McKay measure

$$(1-\lambda)\sigma(\mathbb{T}_d)(\mathrm{d}\lambda) = \frac{d}{2\pi(1+\lambda)}\left(\frac{4(d-1)}{d^2} - \lambda^2\right)^{1/2}\mathbf{1}_{\left[-2\sqrt{d-1}/d, 2\sqrt{d-1}/d\right]}(\lambda)\mathrm{d}\lambda\,.$$

The polynomials $\pi_t$ are normalized such that $\pi_t(1) = 1$.

The measure $\mathrm{d}\sigma(\mathbb{T}_d)$ is, up to a rescaling factor, the image measure of

(SM6.4) $$\mathrm{d}\tilde{\sigma}(\lambda) = \frac{(1-\lambda^2)^{1/2}}{d+2\sqrt{d-1}\lambda}\mathbf{1}_{[-1,1]}(\lambda)\mathrm{d}\lambda$$

by the linear map $\varphi : \lambda \mapsto 2\sqrt{d-1}\lambda/d$. We thus compute a family of orthogonal polynomials w.r.t. $\tilde{\sigma}$ and then use Lemma SM6.1.

The orthogonal polynomials w.r.t. $\tilde{\sigma}$ are given by Proposition SM2.8. Following the cited theorem, we define $\rho(\lambda) = d + 2\sqrt{d-1}\lambda$ and

$$\rho(\cos\theta) = 2\sqrt{d-1}\cos\theta + d = |h(e^{i\theta})|^2\,,$$
$$h(e^{i\theta}) = \sqrt{d-1} + e^{i\theta} = \underbrace{\sqrt{d-1}+\cos\theta}_{:=c(\theta)} + i\underbrace{\sin\theta}_{:=s(\theta)}\,.$$

Then we have the following family $\tilde{p}_t$ of orthogonal polynomials w.r.t. $\tilde{\sigma}$.

$$\tilde{p}_t(\cos\theta) = c(\theta)U_t(\cos\theta) - \frac{s(\theta)}{\sin\theta}T_{t+1}(\cos\theta)\,,$$
$$\tilde{p}_t(\lambda) = (\sqrt{d-1}+\lambda)U_t(\lambda) - T_{t+1}(\lambda)\,,$$

where $T_t$ and $U_t$ denote the $t$-th Chebyshev polynomial of the first kind and the second kind respectively. As the Chebyshev polynomials $T_t$ and $U_t$ both satisfy the same recurrence relation

$$T_{t+1}(\lambda) = 2\lambda T_t(\lambda) - T_{t-1}(\lambda)\,,$$
$$U_{t+1}(\lambda) = 2\lambda U_t(\lambda) - U_{t-1}(\lambda)\,,\quad t \geqslant 1\,,$$

the same relation follows for $\tilde{p}_t$:

$$\tilde{p}_{t+1}(\lambda) = 2\lambda\tilde{p}_t(\lambda) - \tilde{p}_{t-1}(\lambda)\,,\quad t \geqslant 1\,,$$

with initial condition $\tilde{p}_0(\lambda) = \sqrt{d-1}$ and $\tilde{p}_1(\lambda) = 2\sqrt{d-1}\lambda + 1$.

Lemma SM6.1 gives the rescaled orthogonal polynomials $p_t(\lambda) = \tilde{p}_t\left(\varphi^{-1}(\lambda)\right)$ w.r.t. $\mathrm{d}\sigma(\mathbb{T}_d)$:

(SM6.5)
$$p_0(\lambda) = \sqrt{d-1}\,,\quad p_1(\lambda) = d\lambda + 1\,,\quad p_{t+1}(\lambda) = \frac{d}{\sqrt{d-1}}\lambda p_t(\lambda) - p_{t-1}(\lambda)\,,\quad t \geqslant 1\,.$$

As $\pi_t(\lambda) = p_t(\lambda)/p_t(1)$, it now remains to compute $p_t(1)$. The sequence $p_t(1)$, $t \geqslant 1$ satisfies a second-order recurrence relation with fixed coefficients, it thus can be solved explicitly

$$p_t(1) = \frac{1}{d-1}\left(d(d-1)^{(t+1)/2} - 2(d-1)^{(1-t)/2}\right)\,.$$

By substituting $p_t(\lambda) = p_t(1)\pi_t(\lambda)$ in (SM6.5), one obtains

$$\pi_0(\lambda) = 1, \qquad \pi_1(\lambda) = a_0\lambda + b_0, \qquad \pi_{t+1}(\lambda) = a_t\lambda\pi_t(\lambda) - c_t\pi_{t-1}(\lambda), \quad t \geqslant 1,$$

$$a_0 = \frac{d}{d+1}, \qquad b_0 = \frac{1}{d+1},$$

$$a_t = \frac{\frac{d}{d-1} - 2(d-1)^{-(t+1)}}{1 - \frac{2}{d}(d-1)^{-(t+1)}}, \qquad c_t = \frac{\frac{1}{d-1} - \frac{2}{d}(d-1)^{-t}}{1 - \frac{2}{d}(d-1)^{-(t+1)}}, \quad t \geqslant 1.$$

**SM7. Performance guarantees in graphs of spectral dimension $d$.** In this section, we seek to give theoretical support to the empirical observations of Section 3: Jacobi polynomial gossip improves on the non-asymptotic phase over existing methods. This is challenging because the analysis of gossip methods is simpler in the asymptotic regime. In our case, we use asymptotic properties of the Jacobi polynomials as $t \to \infty$.

In order to be able to run an asymptotic analysis without falling in the asymptotic phase of exponential convergence, we run our method on infinite graphs $G = (V, E)$. In infinite graphs, it is impossible for information to have reached every node in any finite time. In practice, the conclusions drawed on infinite graphs should be taken as approximations of the behavior on very large graphs.

Of course, it is impossible for any gossip method to estimate the average of the values in the infinite graphs: indeed, within time $t$ the node $v$ can only share information with nodes that are closer than $t$ (w.r.t. the shortest path distance in the graph). Even worse, the average of an infinite number of values is ill-defined. Thus additional assumptions on the observations $\xi_v$ are needed. Several choices could be possible here, to keep the discussion simple we assume that the observations $\xi_v$ are independent identically distributed (i.i.d.) samples from a probability law $\nu$. The agents then seek to estimate the statistical mean $\mu = \int_{\mathbb{R}} \xi d\nu(\xi)$ of $\nu$.

In practice, to build good estimates, the nodes should average their samples, thus it is natural to run gossip algorithms in this situation. An estimator performs well if it averages a lot of samples and averages them uniformly. Thus the mean square error (MSE) of the estimators measures the capacity of a gossip methods to average locally in the graph.

This statistical gossip framework was already present in [SM2] and is not only used for its technical advantages. It is also a reasonable modeling of gossip of signals with a statistical structure in large networks. For instance, in sensor networks, observations are measurements of the environment corrupted by noise. The purpose of the gossip algorithm is to average observations to get a better estimate of the ground truth. This statistical structure simplifies the underlying gossip problem: good estimates of the mean may not require using observations from nodes extremely far in the network.

Let us now sum up the setting. The network of agents is modeled by a (possibly infinite, locally finite) graph $G = (V, E)$, that we endow with a gossip matrix $W$. We consider a probability law $\nu$ on $\mathbb{R}$, and $\mu = \int_{\mathbb{R}} \xi \, d\nu(\xi)$ its statistical mean. Each agent $v \in V$ is given a sample from $\nu$:

$$\xi_v, v \in V \underset{\text{i.i.d.}}{\sim} \nu.$$

The following theorem gives the asymptotic MSE of the estimators built by the simple gossip method and the Jacobi polynomial iteration.

THEOREM SM7.1. *Fix a vertex $v$ and denote $d_s = d_s(G, W, v)$ the spectral dimension of the graph.*

1. *Let $x^t$ be the iterates of the simple gossip method* (2.2), *or the iterates of the shift-register gossip method* (2.3) *with some parameter $\omega \in [1, 2]$. Then*

(SM7.1)
$$\liminf_{t \to \infty} \frac{\ln \mathbb{E}[(x_v^t - \mu)^2]}{\ln t} \geqslant -\frac{d_s}{2}.$$

2. *Let $x^t$ be the iterates of the Jacobi polynomial iteration* (5.3) *with parameter $d = d_s$. Then*

(SM7.2)
$$\limsup_{t \to \infty} \frac{\ln \mathbb{E}[(x_v^t - \mu)^2]}{\ln t} \leqslant -d_s.$$

See Section SM8 for a proof. The above theorem shows that the asymptotic MSE of the Jacobi polynomial iteration can be upper bounded using the spectral dimension of the graph. The power decay of the MSE with the Jacobi polynomial iteration enjoys a better rate than with simple gossip and the shift-register iteration (regardless of the choice of $\omega$). Note that our bound (SM7.2) depends on $\tau$ only through the spectral dimension of $\tau$, implying that the Jacobi polynomial iteration is robust to the smaller-scale structure of the graph.

In some cases, this rate can be proved optimal using the Hausdorff dimension of the graph.

DEFINITION SM7.2 (Hausdorff dimension). *The Hausdorff dimension of the graph $G$ at vertex $v$ is, if it exists, the limit*

$$d_h = d_h(G, v) = \lim_{t \to \infty} \frac{\ln |B_t(v)|}{\ln t}.$$

*If $G$ is connected, then $d_h$ does not depend on the choice of $v$.*

PROPOSITION SM7.3. *Let $x^t = P_t(W)\xi$ be any polynomial gossip method on a graph $G$ with Hausdorff dimension $d_h$. Then*

(SM7.3)
$$\liminf_{t \to \infty} \frac{\ln \mathbb{E}[(x_v^t - \mu)^2]}{\ln t} \geqslant -d_h.$$

*Proof.* Note that the intuition lying behind the proposition is very simple: the unbiased estimator $x_v^t$ are linear combination of observations corresponding to vertices in the ball $B_v(t)$, thus it must have variance greater than $\operatorname{var} \nu / |B_v(t)| \approx \operatorname{var} \nu / t^{d_h}$.

A more rigorous argument goes as follows: using that $W$ is a gossip matrix, it is easy to show by induction that for all $s \geqslant 0$ and $v, w \in V$, if $(W^s)_{vw} > 0$, then there exists a path of length $s$ linking $v$ to $w$ in $G$. As $\deg P_t \leqslant t$, this implies that $P_t(W)e_v$ has at most $|B_v(t)|$ non-zero entries. Furthermore, the entries of $P_t(W)e_v$ sum to 1 because $W\mathbf{1} = \mathbf{1}$ and $P_t(1) = 1$. Thus, using the Cauchy-Schwarz inequality,

$$1 = \left( \sum_{w \in V} (P_t(W)e_v)_w \right)^2 = \left( \sum_{w \in V} (P_t(W)e_v)_w \mathbf{1}_{\{(P_t(W)e_v)_w > 0\}} \right)^2$$

$$\leqslant \|P_t(W)e_v\|_{\ell^2(V)}^2 \sum_{w \in V} \mathbf{1}_{\{(P_t(W)e_v)_w > 0\}} \leqslant \|P_t(W)e_v\|_{\ell^2(V)}^2 |B_v(t)|$$

$$\overset{\text{(Lemma SM8.1)}}{=} \mathbb{E}[(x_v^t - \mu)^2]|B_v(t)|.$$

Thus

$$\liminf_{t\to\infty} \frac{\ln \mathbb{E}[(x_v^t - \mu)^2]}{\ln t} \geqslant \liminf_{t\to\infty} -\frac{\ln |B_v(t)|}{\ln t} = -d_h \,.$$

Note that this lower bound it attained if $x^t$ is the local average of values:

$$x_v^t = \frac{1}{|B_t(v)|} \sum_{w \in B_t(v)} \xi_w \,.$$

Thus reaching this lower bound means that the polynomial gossip method averages locally. Theorem SM7.1 shows that it is the case with the Jacobi polynomial iteration if $d = d_s = d_h$.

COROLLARY SM7.4. *Assume that the spectral and the Hausdorff dimensions have the same value $d = d_h = d_s$. If $x^t$ are the iterates of the Jacobi polynomial iteration (5.3), we obtain the optimal asymptotic convergence rate*

$$\lim_{t\to\infty} \frac{\ln \mathbb{E}[(x_v^t - \mu)^2]}{\ln t} = -d_h \,.$$

**Application to the grid.** Proposition 5.2 states that the spectral dimension of $\mathbb{Z}^d$ is $d$, which coincides the Hausdorff dimension.

COROLLARY SM7.5. *Let $x^t$ be the iterates of the Jacobi polynomial iteration (5.3) on the grid $\mathbb{Z}^d$. Then we obtain the optimal asymptotic convergence rate*

$$\lim_{t\to\infty} \frac{\ln \mathbb{E}[(x_v^t - \mu)^2]}{\ln t} = -d \,.$$

Note that Theorem SM7.1 also gives that if $x^t$ are the iterates of the simple gossip method, then $\lim_{t\to\infty} \ln \mathbb{E}[(x_v^t - \mu)^2]/\ln t = -d/2$. (The theorem actually only gives the lower bound, but the proof technique, combined with the fact that the spectrum of $\mathbb{Z}^d$ is symmetric, actually gives the result.) This result could have been anticipated intuitively as follows. Under the simple gossip iteration, the information of the measurement $\xi_v$ diffuses following a simple random walk on the grid. According to the central limit theorem, at large $t$, the information is approximately distributed according to a Gaussian distribution of standard deviation $\sqrt{t}$, which is approximately supported by $\Theta(\sqrt{t}^d)$ nodes. This means that at time $t$, a node $v$ gets the information of $\Theta(t^{d/2})$ neighbors. As a consequence, the MSE $\mathbb{E}[(x_v^t - \mu)^2]$ is of the order of $t^{-d/2}$.

**Application to the percolation bonds.** Let $G$ be the random infinite cluster of a supercritical percolation in $\mathbb{Z}^d$ as defined in Proposition 5.5. The proposition gives that the spectral dimension of $G$ is a.s. $d$, which is also a lower bound for the Hausdorff dimension. But it is trivial that the Hausdorff dimension is smaller than $d$, thus the two coincide.

COROLLARY SM7.6. *Let $G$ be the random infinite cluster of a supercritical percolation in $\mathbb{Z}^d$, and $v \in \mathbb{Z}^d$. Let $x^t$ be the iterates of the Jacobi polynomial iteration (5.3). Then a.s. on the event $\{v \in G\}$,*

$$\lim_{t\to\infty} \frac{\ln \mathbb{E}_\xi[(x_v^t - \mu)^2]}{\ln t} = -d \,.$$

*Remark* SM7.7. The Jacobi polynomial iteration (5.3) is derived so that $x^t = \pi_t^{(\alpha,\beta)}(W)\xi$, where $\pi_t^{(\alpha,\beta)}$ are the orthogonal polynomials w.r.t. the Jacobi measure $\sigma^{(\alpha,\beta)}(d\lambda) = (1-\lambda)^\alpha(1+\lambda)^\beta d\lambda$ on $[-1,1]$, with $\alpha = d/2, \beta = 0$, $d$ is the spectral dimension. A curious reader could wonder what happens for other choices of $\alpha$ and $\beta$ (while keeping $d$ fixed). This question is investigated at length in Section SM8.5. The conclusion is that the natural choice $\alpha = d/2, \beta = 0$ is optimal (up to constant factors) but there are other choices that are optimal.

**SM8. Proof of Theorem SM7.1.** The proof is divided in four subsections. Section SM8.1 develops tools that we use both in the proof of the theorem. We then prove the theorem in Sections SM8.2, SM8.3 and SM8.4. Finally, in Section SM8.5, we discuss the choice of the parameters of the Jacobi polynomials in the Jacobi polynomial iteration. In all this section, we denote $\sigma = \sigma(G, W, v)$ the spectral measure of $G$.

**SM8.1. Preliminaries.** The first lemma relates the MSE of the estimator $x_v^t$ to the spectral measure.

LEMMA SM8.1. *Write $x^t = P_t(W)\xi$ using the polynomial gossip point of view. Then*

$$\mathbb{E}[(x_v^t - \mu)^2] = (\mathrm{var}\,\nu)\,\|P_t(W)e_v\|_{\ell^2(V)}^2 = (\mathrm{var}\,\nu)\int P_t(\lambda)^2 d\sigma(\lambda)\,.$$

*Proof.* As $P_t(1) = 1$, we have

$$\mathbb{E}[x^t] = \mathbb{E}[P_t(W)\xi] = P_t(W)\mathbb{E}[\xi] = P_t(W)\mu\mathbf{1} = P_t(1)\mu\mathbf{1} = \mu\mathbf{1}\,.$$

In words, the estimator $x_v^t$ is unbiased. Thus

$$\mathbb{E}[(x_v^t - \mu)^2] = \mathrm{var}\,x_v^t = \mathrm{var}\,\langle P_t(W)\xi, e_v\rangle_{\ell^2(V)}$$
$$= \mathrm{var}\,\langle \xi, P_t(W)e_v\rangle_{\ell^2(V)} = (\mathrm{var}\,\nu)\,\|P_t(W)e_v\|_{\ell^2(V)}^2\,,$$

using that $W$ is symmetric and that the $\xi_w, w \in V$ are i.i.d. random variables. Then

$$\mathbb{E}[(x_v^t - \mu)^2] = (\mathrm{var}\,\nu)\,\langle P_t(W)e_v, P_t(W)e_v\rangle_{\ell^2(V)} = (\mathrm{var}\,\nu)\,\langle e_v, P_t(W)^2 e_v\rangle_{\ell^2(V)}\,.$$

The proof is finished using the Definition 5.3 of the spectral measure. $\square$

In the statement of Theorem SM7.1, we have stated results in terms of the spectral dimension $d_s = 2\lim_{t\to\infty}\ln\sigma([1-\Lambda, 1])/\ln\Lambda$. In the proof here, we will be more precise. We show how the results of Theorem SM7.1 actually depend on different definitions of the dimension.

DEFINITION SM8.2. *Let $\tau$ be a probability measure on $[-1, 1]$. We define*
  1. *the right upper dimension $\overline{\dim}_\to\tau \in [0, \infty]$ of the measure $\tau$ as*

$$\overline{\dim}_\to\tau = 2\limsup_{\Lambda\to 0}\frac{\ln\tau([1-\Lambda, 1])}{\ln\Lambda}\,,$$

  2. *the right lower dimension $\underline{\dim}_\to\tau \in [0, \infty]$ of the measure $\tau$ as*

$$\underline{\dim}_\to\tau = 2\liminf_{\Lambda\to 0}\frac{\ln\tau([1-\Lambda, 1])}{\ln\Lambda}\,,$$

3. *the left upper dimension* $\overline{\dim}_{\leftarrow}\tau \in [0,\infty]$ *of the measure* $\tau$ *as*

$$\overline{\dim}_{\leftarrow}\tau = 2\limsup_{\Lambda\to 0}\frac{\ln\tau([-1,-1+\Lambda])}{\ln\Lambda}.$$

4. *the left lower dimension* $\underline{\dim}_{\leftarrow}\tau \in [0,\infty]$ *of the measure* $\tau$ *as*

$$\underline{\dim}_{\leftarrow}\tau = 2\liminf_{\Lambda\to 0}\frac{\ln\tau([-1,-1+\Lambda])}{\ln\Lambda}.$$

**SM8.2. Proof of Theorem SM7.1: simple gossip.** In the case of simple gossip, $P_t(\lambda) = \lambda^t$.

PROPOSITION SM8.3. *Let $\tau$ be a probability measure on $[-1,1]$. Then*

$$\liminf_{t\to\infty}\frac{\int\lambda^{2t}\mathrm{d}\tau(\lambda)}{\ln t} \geqslant -\frac{\min\left(\overline{\dim}_{\to}\tau, \overline{\dim}_{\leftarrow}\tau\right)}{2}$$

*Proof.* Let $d > \overline{\dim}_{\to}\tau$. As $\overline{\dim}_{\to}\tau = 2\limsup_{\Lambda\to 0}\ln\sigma([1-\Lambda,1])/\ln\Lambda$, there exists constants $c_1, c_2 > 0$ such that for all $\Lambda \in [0,2]$,

(SM8.1) $$\tau([1-\Lambda,1]) \geqslant c_1\Lambda^{d/2} = c_2\sigma^{(d/2-1,0)}([1-\Lambda,1])$$

where $\sigma^{(d/2-1,0)}(\mathrm{d}\lambda) = (1-\lambda)^{d/2-1}\mathrm{d}\lambda$. Then using jointly Lemma SM3.1 and Eq. (SM8.1),

$$\int\lambda^{2t}\mathrm{d}\tau(\lambda) \geqslant \int_{[0,1]}\lambda^{2t}\mathrm{d}\tau(\lambda) \geqslant c_1\int_0^1\lambda^{2t}(1-\lambda)^{d/2-1}\mathrm{d}\lambda = B(2t+1,d/2) \overset{(\text{SM3.2})}{\underset{t\to\infty}{\sim}} \frac{c_3}{t^{d/2}},$$

for some constant $c_3$. Thus

$$\liminf_{t\to\infty}\frac{\int\lambda^{2t}\mathrm{d}\tau(\lambda)}{\ln t} \geqslant -\frac{d}{2}.$$

This being true for all $d > \overline{\dim}_{\to}\tau$, this proves

$$\liminf_{t\to\infty}\frac{\int\lambda^{2t}\mathrm{d}\tau(\lambda)}{\ln t} \geqslant -\frac{\overline{\dim}_{\to}\tau}{2}.$$

The proof at the other edge of the spectrum is the same by symmetry. □

The proof of Theorem SM7.1 for simple gossip follows easily. Indeed, if $\tau = \sigma$ is the spectral measure of the graph, then $\overline{\dim}_{\to}\sigma = d_s$. Thus

$$\liminf_{t\to\infty}\frac{\ln\mathbb{E}[(x_v^t-\mu)^2]}{\ln t} \overset{(\text{Lemma SM8.1})}{=} \liminf_{t\to\infty}\frac{\ln\int\lambda^{2t}\mathrm{d}\sigma(\lambda)}{\ln t} \overset{(\text{Proposition SM8.3})}{\geqslant} -\frac{d_s}{2}.$$

**SM8.3. Proof of Theorem SM7.1: shift-register.** In the case of the shift-register gossip iteration, $P_t(\lambda)$ satisfies the second-order recurrence relation

(SM8.2) $$P_0(\lambda) = 1, \qquad P_1(\lambda) = \lambda, \qquad P_{t+1}(\lambda) = \omega\lambda P_t(\lambda) + (1-\omega)P_{t-1}(\lambda).$$

The case $\omega = 1$ corresponds to simple gossip: it has been treated above. We now assume $\omega \in (1,2]$.

PROPOSITION SM8.4. *Let $P_t$ be the polynomials defined in Eq.* (SM8.2) *with $\omega \in (1,2]$. Then*

$$P_t(\lambda) = (\omega - 1)^{t/2} \left[ \left( 2 - \frac{2}{\omega} \right) T_t \left( \frac{\omega}{2\sqrt{\omega - 1}} \lambda \right) + \left( \frac{2}{\omega} - 1 \right) U_t \left( \frac{\omega}{2\sqrt{\omega - 1}} \lambda \right) \right]$$

*where $T_t$ and $U_t$ are the Chebyshev polynomials of the first and second kind respectively (see Example SM2.3).*

*Proof.* Consider the rescaled version $Q_t$ of $P_t$ given by the formula

(SM8.3)
$$P_t(\lambda) = (\omega - 1)^{t/2} Q_t \left( \frac{\omega}{2\sqrt{\omega - 1}} \lambda \right).$$

If follows from Eq. (SM8.2) that

$$Q_0(\lambda) = 1, \qquad Q_1(\lambda) = \frac{2}{\omega} \lambda, \qquad Q_{t+1}(\lambda) = 2\lambda Q_t(\lambda) - Q_{t-1}(\lambda).$$

Thus the sequence $Q_t, t \geqslant 0$ satisfies the same recurrence relation as the Chebyshev polynomials, but with a different initialization. As a consequence, it must be a linear combination of the two sequences of Chebyshev polynomials: there exists $\mu, \nu \in \mathbb{R}$ such that for all $t$,

$$Q_t(\lambda) = \mu T_t(\lambda) + \nu U_t(\lambda)$$

The computation of the weights $\mu, \nu$ is straightforward from the initialization $Q_0, Q_1$. This proves the proposition. $\square$

PROPOSITION SM8.5. *Let $\omega \in (1,2]$. The polynomials $P_t, t \geqslant 0$ defined in Eq.* (SM8.2) *are the orthogonal polynomials w.r.t. the measure*

$$\tau(\mathrm{d}\lambda) = \frac{\left( \left( 2\sqrt{\omega - 1}/\omega \right)^2 - \lambda^2 \right)^{1/2}}{1 - \lambda^2} \mathrm{d}\lambda.$$

*Proof.* The orthogonal polynomials w.r.t. the measure

$$\tilde{\tau}(\mathrm{d}\lambda) = \frac{\left( 1 - \lambda^2 \right)^{1/2}}{\left( \omega/(2\sqrt{\omega - 1}) \right)^2 - \lambda^2} \mathrm{d}\lambda$$

are computed using Proposition SM2.8 with

$$\rho(\cos \theta) = \frac{\omega^2}{4(\omega - 1)} - \cos^2 \theta.$$

Simple computations give that $\rho(\cos \theta) = |h(e^{i\theta})|^2$ with

$$h(e^{i\theta}) = \left| \frac{\omega}{2\sqrt{\omega - 1}} \left[ \left( 2 - \frac{2}{\omega} \right) \frac{1 - e^{2i\theta}}{2} + \left( \frac{2}{\omega} - 1 \right) \right] \right|^2.$$

Proposition SM2.8 then gives that the polynomials $(2 - 2/\omega)T_t + (2/\omega - 1)U_t$ are orthogonal w.r.t. $\tilde{\tau}$. But these polynomials are the polynomials $Q_t$ defined in Eq. (SM8.3). We then use Lemma SM6.1 to prove that $P_t$ is orthogonal w.r.t. $\tau$. $\square$

LEMMA SM8.6. *Let $P_t$ be the polynomials defined in Eq. (SM8.2) with $\omega \in (1, 2]$ and $\tau$ a measure on $[-1, 1]$. Then*

$$\liminf_{t \to \infty} \frac{\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \geqslant -\frac{\min\left(\overline{\dim}_{\to}\tau, \overline{\dim}_{\leftarrow}\tau\right)}{2}$$

*Proof.*

$$\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda) \geqslant \int_{[2\sqrt{\omega-1}/\omega,1]} P_t(\lambda)^2 \mathrm{d}\tau(\lambda)$$

$$\overset{\text{(Proposition SM8.4)}}{\geqslant} c_1(\omega-1)^t \int_{[2\sqrt{\omega-1}/\omega,1]} T_t\left(\frac{\omega}{2\sqrt{\omega-1}}\lambda\right)^2 \mathrm{d}\tau(\lambda)$$

$$\overset{\text{(SM2.2)}}{\geqslant} c_2(\omega-1)^t \int_{[2\sqrt{\omega-1}/\omega,1]} \left(\frac{\omega}{2\sqrt{\omega-1}}\lambda + \sqrt{\frac{\omega^2}{4(\omega-1)}\lambda^2 - 1}\right)^{2t} \mathrm{d}\tau(\lambda).$$

where $c_i > 0$ is a constant independent of $t$. Let $d > \overline{\dim}_{\to}\tau$. As $\overline{\dim}_{\to}\tau = 2\limsup_{\Lambda \to 0} \ln \sigma([1-\Lambda, 1])/\ln \Lambda$, there exists constants $c_3, c_4 > 0$ such that for all $\Lambda \in [0, 2]$,

$$(\text{SM8.4}) \qquad \tau([1-\Lambda, 1]) \geqslant c_3 \Lambda^{d/2} = c_4 \sigma^{(d/2-1,0)}([1-\Lambda, 1])$$

where $\sigma^{(d/2-1,0)}(\mathrm{d}\lambda) = (1-\lambda)^{d/2-1}\mathrm{d}\lambda$. Then using jointly Lemma SM3.1 and Eq. (SM8.4),

$$\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda) \geqslant c_5(\omega-1)^t \int_{2\sqrt{\omega-1}/\omega}^{1} \left(\frac{\omega}{2\sqrt{\omega-1}}\lambda + \sqrt{\frac{\omega^2}{4(\omega-1)}\lambda^2 - 1}\right)^{2t} (1-\lambda)^{d/2-1}\mathrm{d}\lambda$$

$$\geqslant c_6(\omega-1)^t \int_0^{\cosh^{-1}(\omega/(2\sqrt{\omega-1}))} e^{2tu}\left(1 - \frac{2\sqrt{\omega-1}}{\omega}\cosh u\right)^{d/2-1} \sinh u \, \mathrm{d}u.$$

where in the last step we made the change of variable

$$\frac{\omega}{2\sqrt{\omega-1}}\lambda + \sqrt{\frac{\omega^2}{(4(\omega-1))}\lambda^2 - 1} = e^u, \qquad \text{i.e.} \qquad \lambda = \frac{2\sqrt{\omega-1}}{\omega}\cosh u.$$

Denote $u_{\max} = \cosh^{-1}(\omega/(2\sqrt{\omega-1}))$ to shorten notations. As $\cosh$ is a convex function, for $u \in [0, u_{\max}]$,

$$\cosh u_{\max} - \cosh u \leqslant \frac{\cosh u_{\max} - 1}{u_{\max}}(u_{\max} - u),$$

$$\Leftrightarrow 1 - \frac{2\sqrt{\omega-1}}{\omega}\cosh u \leqslant \left(1 - \frac{2\sqrt{\omega-1}}{\omega}\right)\left(1 - \frac{u}{u_{\max}}\right)$$

Moreover, choose some constant $u_{\min} \in (0, u_{\max})$ so that we can lower bound with a constant $c_7$: for all $u \in [u_{\min}, u_{\max}]$, $\sinh u \geqslant c_7$. This finally gives:

$$\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda) \geqslant c_8(\omega-1)^t \int_{u_{\min}}^{u_{\max}} e^{2tu}\left(1 - \frac{u}{u_{\max}}\right)^{d/2-1} \mathrm{d}u.$$

After the change of variable $w = 2t(u_{\max} - u)$, this gives

$$\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda) \geqslant c_8(\omega-1)^t \int_0^{2t(u_{\max}-u_{\min})} e^{2tu_{\max}} e^{-w}\left(\frac{w}{2tu_{\max}}\right)^{d/2-1} \frac{1}{2t} \mathrm{d}w.$$

Note that $e^{2tu_{\max}} = (\omega - 1)^{-t}$, thus there exists a constant $c_9 > 0$ such that

$$\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda) \geqslant c_9 \frac{1}{t^{d/2}} \int_0^{2t(u_{\max} - u_{\min})} e^{-w} w^{d/2-1} \, \mathrm{d}w \,.$$

This proves that

$$\liminf_{t \to \infty} \frac{\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \geqslant -\frac{d}{2} \,.$$

This being true for all $d > \overline{\dim}_\to \tau$, this proves

$$\liminf_{t \to \infty} \frac{\int P_t(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \geqslant -\frac{\overline{\dim}_\to \tau}{2} \,.$$

The proof at the other edge of the spectrum is the same by symmetry. □

**SM8.4. Proof of Theorem SM7.1: Jacobi polynomial iteration.** In this section, we use again the notation of Section SM6.2: in the case of the Jacobi polynomial iteration (5.3), we have $x^t = \pi_t^{(d_s/2,0)}(W)\xi$, where $\pi_t^{(\alpha,\beta)}$ is the rescaled Jacobi polynomial; $\pi_t^{(\alpha,\beta)} = P_t^{(\alpha,\beta)}/\binom{t+\alpha}{t}$ where $P_t^{(\alpha,\beta)}$ is the traditional Jacobi polynomial. Lemma SM8.1 suggests to study the quantity $\int \pi_t^{(d_s/2,0)}(\lambda)^2 \mathrm{d}\sigma(\lambda)$. However we study here the behavior of $\int \pi_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\sigma(\lambda)$ for any $(\alpha, \beta)$. This will be useful in Section SM8.5 to give a motivation for the choice $\alpha = d/2, \beta = 0$ complementary to the intuition developed in Section 5, and will allow us to discuss the performance of other choices.

PROPOSITION SM8.7. *Let $\tau$ be a probability measure on $[-1, 1]$ and $\alpha, \beta \geqslant -1/2$. Then*

$$\limsup_{t \to \infty} \frac{\ln \int \pi_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \leqslant -\min\left(2\alpha + 1, \underline{\dim}_\to \tau, 2(\alpha - \beta) + \underline{\dim}_\leftarrow \tau\right) .$$

Before proving this proposition, we use it to finish the proof of the theorem. If $\tau = \sigma$ is the spectral measure of the graph, then $\underline{\dim}_\to \sigma = d_s$. Thus taking $\alpha = d_s/2$, $\beta = 0$ in Proposition SM8.7, we get

$$\text{(SM8.5)} \quad \limsup_{t \to \infty} \frac{\ln \int \pi_t^{(d_s/2,0)}(\lambda)^2 \mathrm{d}\sigma(\lambda)}{\ln t} \leqslant -\min(d_s + 1, d_s, d_s + \underline{\dim}_\leftarrow \sigma) = -d_s \,,$$

as $\underline{\dim}_\leftarrow \sigma \geqslant 0$. One can conclude the proof using Lemma SM8.1.

We now turn to the the proof of Proposition SM8.7.

LEMMA SM8.8. *Let $\tau$ be a probability measure on $[-1, 1]$ and $\alpha \geqslant -1/2$, $\beta > -1$. Then*

$$\limsup_{t \to \infty} \frac{\ln \int_{[0,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \leqslant -\min(1, \underline{\dim}_\to \tau - 2\alpha) \,.$$

*Proof.* Let $d < \underline{\dim}_\to \tau$. As $\underline{\dim}_\to \tau = 2 \liminf_{\Lambda \to 0} \ln \tau([1 - \Lambda, 1])/\ln \Lambda$, there exists constants $C_1, C_2$ such that for all $\Lambda \in [0, 2]$,

$$\text{(SM8.6)} \qquad \tau([1 - \Lambda, 1]) \leqslant C_1 \Lambda^{d/2} = C_2 \sigma^{(d/2-1,0)}([1 - \Lambda, 1])$$

where $\sigma^{(d/2-1,0)}(\mathrm{d}\lambda) = (1-\lambda)^{d/2-1}\mathrm{d}\lambda$.

For the proof of this result, we use the asymptotic bound on the Jacobi polynomials given by Proposition SM2.11, thus we divide the integral

$$\int_{[0,1]} P_t^{(\alpha,\beta)}(\lambda)^2\mathrm{d}\tau(\lambda) = \int_{[\cos 1/t,1]} P_t^{(\alpha,\beta)}(\lambda)^2\mathrm{d}\tau(\lambda) + \int_{[0,\cos 1/t[} P_t^{(\alpha,\beta)}(\lambda)^2\mathrm{d}\tau(\lambda)\,,$$

and treat the two terms separately.

(a)

$$\int_{[\cos 1/t,1]} P_t^{(\alpha,\beta)}(\lambda)^2\mathrm{d}\tau(\lambda) \leqslant C_3 t^{2\alpha}\tau\left(\left[\cos\frac{1}{t},1\right]\right) \leqslant C_1 C_3 t^{2\alpha}\left(1-\cos\frac{1}{t}\right)^{d/2}$$

$$\leqslant C_4 t^{2\alpha-d}$$

for some constants $C_3, C_4$. Thus

$$\text{(SM8.7)} \qquad \limsup_{t\to\infty} \frac{\ln \int_{[\cos 1/t,1]} P_t^{(\alpha,\beta)}(\lambda)^2\mathrm{d}\tau(\lambda)}{\ln t} \leqslant 2\alpha - d\,.$$

(b)

$$\text{(SM8.8)}$$
$$\int_{[0,\cos 1/t[} P_t^{(\alpha,\beta)}(\lambda)^2\mathrm{d}\tau(\lambda) \leqslant C_5 t^{-1} \int_{[0,\cos(1/t)[} (\arccos\lambda)^{-2\alpha-1}\mathrm{d}\tau(\lambda)$$

We then use jointly Eq. (SM8.6) and Lemma SM3.1 with the function

$$f(\lambda) = (\arccos\lambda)^{-2\alpha-1}\mathbf{1}_{\{\lambda<\cos 1/t\}} + t^{2\alpha+1}\mathbf{1}_{\{\lambda\geqslant\cos 1/t\}}\,.$$

Note that $f$ is non-decreasing as $\alpha \geqslant 1/2$. We get

$$\int_{[0,\cos(1/t)[} (\arccos\lambda)^{-2\alpha-1}\mathrm{d}\tau(\lambda)$$

$$\leqslant C_2 \int_{[0,\cos(1/t)[} (\arccos\lambda)^{-2\alpha-1}(1-\lambda)^{d/2-1}\mathrm{d}\lambda$$

$$+ C_1 t^{2\alpha+1}\left(1-\cos\frac{1}{t}\right)^{d/2}$$

Now using the simple inequality $\arccos\lambda \geqslant \sqrt{2}\sqrt{1-\lambda}$, we get

$$\text{(SM8.9)} \qquad \int_{[0,\cos(1/t)[} (\arccos\lambda)^{-2\alpha-1}\mathrm{d}\tau(\lambda)$$

$$\leqslant C_5 \int_{[0,\cos(1/t)[} (1-\lambda)^{-\alpha+d/2-3/2}\mathrm{d}\lambda + C_6 t^{2\alpha+1-d}$$

for some constants $C_5, C_6$. Now if $\delta$ is a real number,

$$\text{(SM8.10)} \qquad \lim_{t\to\infty} \frac{\ln \int_0^{\cos 1/t} (1-\lambda)^\delta\mathrm{d}\lambda}{\ln t} = \max(0, -2\delta - 2)\,.$$

Indeed, if $\delta \neq -1$,

$$\int_0^{\cos 1/t} (1-\lambda)^\delta \mathrm{d}\lambda = \left[ -\frac{(1-\lambda)^{\delta+1}}{\delta+1} \right]_0^{\cos 1/t} = \frac{1}{\delta+1} \left[ 1 - \left(1 - \cos\frac{1}{t}\right)^{\delta+1} \right]$$

$$\underset{t\to\infty}{=} \frac{1}{\delta+1} \left[ 1 - t^{-2\delta-2} + o(t^{-2\delta-2}) \right] \sim C(\delta) t^{\max(0,-2\delta-2)} .$$

for some constant $C(\delta)$ depending on $\delta$. This proves the statement (SM8.10) for $\delta \neq -1$. The result for $\delta = -1$ follows easily by noting that both terms in (SM8.10) are decreasing in $\delta$.

Merging finally Eqs. (SM8.8), (SM8.9) and (SM8.10), we get
(SM8.11)

$$\limsup_{t\to\infty} \frac{\ln \int_{[0,\cos 1/t[} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \leqslant -1 + \max(0, 2\alpha + 1 - d)$$

$$= \max(-1, 2\alpha - d) = -\min(1, d - 2\alpha) .$$

Finally

$$\limsup_{t\to\infty} \frac{\ln \int_{[0,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t}$$

$$\leqslant \limsup_{t\to\infty} \frac{2 \max \left( \int_{[\cos 1/t,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda), \int_{[0,\cos 1/t[} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda) \right)}{\ln t}$$

$$\leqslant \max \left( \limsup_{t\to\infty} \frac{\ln \int_{[\cos 1/t,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t}, \limsup_{t\to\infty} \frac{\ln \int_{[0,\cos 1/t[} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \right)$$

$$\overset{\text{(SM8.7),(SM8.11)}}{\leqslant} \max(2\alpha - d, -\min(1, d - 2\alpha)) = -\min(1, d - 2\alpha) .$$

As this is true for all $d < \underline{\dim}_\rightarrow \tau$, the lemma is proved. $\qquad\square$

*Proof of Proposition SM8.7.* If we denote $\tilde{\tau}$ the symmetric measure of $\tau$ w.r.t. 0 (i.e. the image measure of $\tau$ by the map $\lambda \mapsto -\lambda$), we have

$$\int_{[-1,0]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda) = \int_{[0,1]} P_t^{(\alpha,\beta)}(-\lambda)^2 \mathrm{d}\tilde{\tau}(\lambda) = \int_{[0,1]} P_t^{(\beta,\alpha)}(\lambda)^2 \mathrm{d}\tilde{\tau}(\lambda)$$

Thus according to Lemma SM8.8,
(SM8.12) $\qquad\square$

$$\limsup_{t\to\infty} \frac{\ln \int_{[-1,0]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \leqslant -\min(1, \underline{\dim}_\rightarrow \tilde{\tau} - 2\beta) = -\min(1, \underline{\dim}_\leftarrow \tau - 2\beta) .$$

Finally, using that $\pi_t^{(\alpha,\beta)} = P_t^{(\alpha,\beta)}/\binom{t+\alpha}{t}$,

$$
\limsup_{t\to\infty} \frac{\ln \int_{[-1,1]} \pi_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \leqslant \limsup_{t\to\infty} \frac{\ln \int_{[-1,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} - 2\limsup_{t\to\infty} \frac{\ln \binom{t+\alpha}{t}}{\ln t}
$$

$$
\leqslant \limsup_{t\to\infty} \frac{\ln \left( 2\max \left( \int_{[-1,0]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda), \int_{[0,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda) \right) \right)}{\ln t} - 2\alpha
$$

$$
\leqslant \max \left( \limsup_{t\to\infty} \frac{\ln \int_{[-1,0]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t}, \limsup_{t\to\infty} \frac{\ln \int_{[0,1]} P_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\tau(\lambda)}{\ln t} \right) - 2\alpha
$$

$$
\overset{((\text{SM8.12}),\text{Lemma SM8.8})}{\leqslant} \max \left( -\min(1, \underline{\dim}_\leftarrow \tau - 2\beta), -\min(1, \underline{\dim}_\rightarrow \tau - 2\alpha) \right) - 2\alpha
$$

$$
\leqslant -\min(1, \underline{\dim}_\leftarrow \tau - 2\beta, \underline{\dim}_\rightarrow \tau - 2\alpha) - 2\alpha
$$

$$
= -\min(2\alpha + 1, 2(\alpha - \beta) + \underline{\dim}_\leftarrow \tau, \underline{\dim}_\rightarrow \tau).
$$

**SM8.5. Robustness of the Jacobi polynomial iteration and tuning of the parameters $\alpha$ and $\beta$.** In this section, we explore the generality of Proposition SM8.7, which goes beyond the proof of Theorem SM7.1, in order to derive some consequences for the robustness of the Jacobi polynomial iteration and for the tuning of the parameters $\alpha$ and $\beta$.

In the statistical framework of Section SM7, let $x^t = \pi_t^{(\alpha,\beta)}(W)\xi$ be the iterates of the Jacobi polynomial iteration on some graph with spectral measure $\tau = \tau(G, W, v)$, for some arbitrary choice of parameters $(\alpha, \beta)$. In particular, note that we do not impose the relation $\alpha = d/2$. Then Proposition SM8.7 gives

$$
\limsup_{t\to\infty} \frac{\ln \mathbb{E}\left[ (x_v^t - \mu)^2 \right]}{\ln t} \leqslant -\min\left( 2\alpha + 1, \underline{\dim}_\rightarrow \tau, 2(\alpha - \beta) + \underline{\dim}_\leftarrow \tau \right).
$$

Note that the rate of decay depends on the spectral measure $\tau$ only through the left and right (lower) dimensions of $\tau$. Again, this shows some robustness of the Jacobi polynomial iteration to finer details of the graph. Moreover, note that this inequality proves some rate of decay even when the parameter $\alpha$ of the Jacobi polynomial iteration is not half of the spectral dimension of the graph. This proves some robustness of the Jacobi polynomial iteration to a wrong tuning of the spectral dimension.

This last remark encourages to explore what choices of $\alpha$ and $\beta$ lead to optimal rates for a given graph. The Jacobi polynomial iteration introduced in Section 5.2 corresponds to the specific choice $\alpha = d_s/2, \beta = 0$, where $d_s$ is the spectral measure of the graph. Inspired by (SM8.5), we define optimality as follows

DEFINITION SM8.9. *Let $\alpha, \beta > -1, d_\leftarrow, d_\rightarrow \geqslant 0$. We say that $(\alpha, \beta)$ is optimal for $(d_\leftarrow, d_\rightarrow)$ if for any spectral measure $\sigma$ such that $\underline{\dim}_\rightarrow \sigma = d_\rightarrow$ and $\underline{\dim}_\leftarrow \sigma = d_\leftarrow$,*

$$
\limsup_{t\to\infty} \frac{\ln \int \pi_t^{(\alpha,\beta)}(\lambda)^2 \mathrm{d}\sigma(\lambda)}{\ln t} \leqslant -d_\rightarrow.
$$

The following theorem is an analogue of the optimality theorem SM7.1(2) in the general case.

THEOREM SM8.10. *Consider a graph $G$, a gossip matrix $W$ and a vertex $v$. Denote $\sigma = \sigma(G, W, v)$ the spectral measure of the graph. Let $\xi_v, v \in V$ be i.i.d. samples from a distribution of mean $\mu$.*

Let $\alpha, \beta > -1$ and define the polynomial iteration $x^t = \pi_t^{(\alpha,\beta)}(W)\xi$. If $(\alpha, \beta)$ is optimal for $(\underline{\dim}_{\leftarrow}\sigma, \underline{\dim}_{\rightarrow}\sigma)$, then

(SM8.13)
$$\limsup_{t \to \infty} \frac{\ln \mathbb{E}\left[(x_v^t - \mu)^2\right]}{\ln t} \leqslant -\underline{\dim}_{\rightarrow}\sigma \,.$$

In the section above, we prove that $(d_{\rightarrow}/2, 0)$ is optimal for $(d_{\leftarrow}, d_{\rightarrow})$ (for any $d_{\leftarrow}, d_{\rightarrow} \geqslant -1/2$). We now explore other choices. According to Proposition SM8.7, to prove that $(\alpha, \beta)$ is optimal for $(d_{\leftarrow}, d_{\rightarrow})$, it is sufficient to prove that

$$\min(2\alpha + 1, d_{\rightarrow}, 2(\alpha - \beta) + d_{\leftarrow}) = d_{\rightarrow} \qquad \Leftrightarrow \qquad \begin{cases} 2\alpha + 1 \geqslant d_{\rightarrow} \\ 2(\alpha - \beta) + d_{\leftarrow} \geqslant d_{\rightarrow} \end{cases}$$

(SM8.14)
$$\Leftrightarrow \qquad \begin{cases} \alpha \geqslant \frac{1}{2}(d_{\rightarrow} - 1) \\ \beta \leqslant \alpha + \frac{d_{\leftarrow} - d_{\rightarrow}}{2} \end{cases} \,.$$

This gives a wide range of optimal parameters. For instance, the parameter $\alpha$ can be chosen arbitrarily large. In Figures SM1B and SM2B, the shaded regions corresponds to region for $(\alpha, \beta)$ defined by (SM8.14) with $(d_{\leftarrow}, d_{\rightarrow}) = (2, 2)$.

Note however that we have only proved that (SM8.14) are *sufficient* conditions for the optimality Theorem SM8.10 to hold. To explore the tightness of our condition, we present in Figure SM1 the results of simulations on the 2D grid. The setting is the same as in Section 3 (see also Section SM1 for details). Note that for the 2D infinite grid, $d_{\leftarrow} = d_{\rightarrow} = 2$ (see Proposition 5.2 and the symmetry of the spectrum of $\mathbb{Z}^d$ that follows from [SM9, Eq.(7.4)]). The curves in Figure SM1A closest to the local averaging are those satisfying the condition (SM8.14), thus our condition seems tight.

Finally, note that the result (SM8.13) of Theorem SM8.10 gives the rate of the power decay of the MSE, but neglects constants and sub-polynomial factors. These factors depend on $(\alpha, \beta)$ and can be significant for extreme values of $(\alpha, \beta)$. For instance, in Figure SM2, we run simulations in the same setting as before, but for choices of parameters deeper in the optimality zone (SM8.14). The performance worses as $\alpha$ gets bigger. So contrarily to what is suggested by (SM8.14) and Theorem SM7.1, taking large values for $\alpha$ is a bad idea in practice. This can also be hinted at by the limit [SM5, Eq. (18.6.2)]

$$\lim_{\alpha \to \infty} \pi_t^{(\alpha,\beta)}(\lambda) = \left(\frac{1+\lambda}{2}\right)^t \,.$$

This means that, as $\alpha \to \infty$, the polynomial gossip $x^t = \pi_t^{(\alpha,\beta)}(W)\xi$ converges to the simple gossip $x^t = \tilde{W}^t \xi$ with the gossip matrix $\tilde{W} = (I + W)/2$. We know from Theorem SM7.1(1) that simple gossip is suboptimal.

Overall, theory and practice suggest that the choice $\alpha = \underline{\dim}_{\rightarrow}\sigma/2$, $\beta = 0$ that we make in Section 5.2 is relevant.

**SM9. Message passing seen as a polynomial gossip algorithm.** This section develops another application of the polynomial point-of-view on gossip algorithms. It is independent of the Jacobi polynomial iterations developed in Sections 5-6; we show that the message passing algorithm for gossip of [SM4] has a natural derivation as a polynomial gossip algorithm and uses this point of view to derive convergence rates.
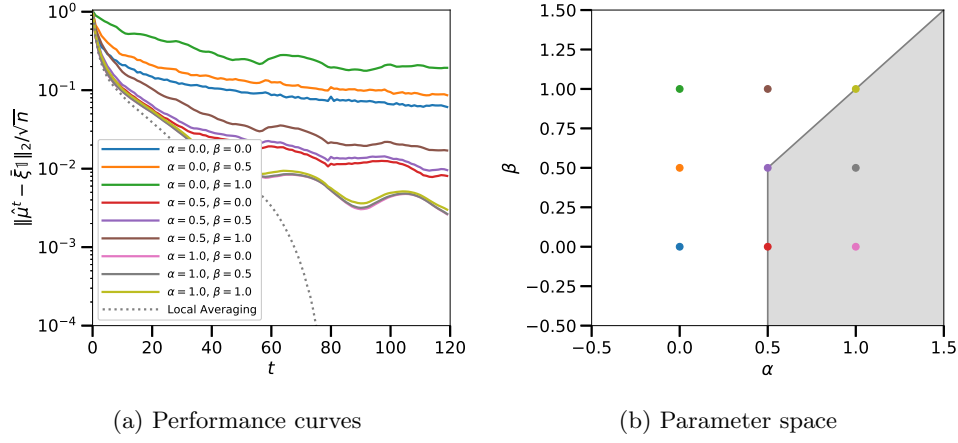
(a) Performance curves

(b) Parameter space

Fig. SM1: Simulations of polynomial iterations using Jacobi polynomials with different parameters $(\alpha, \beta)$: frontier tightness.
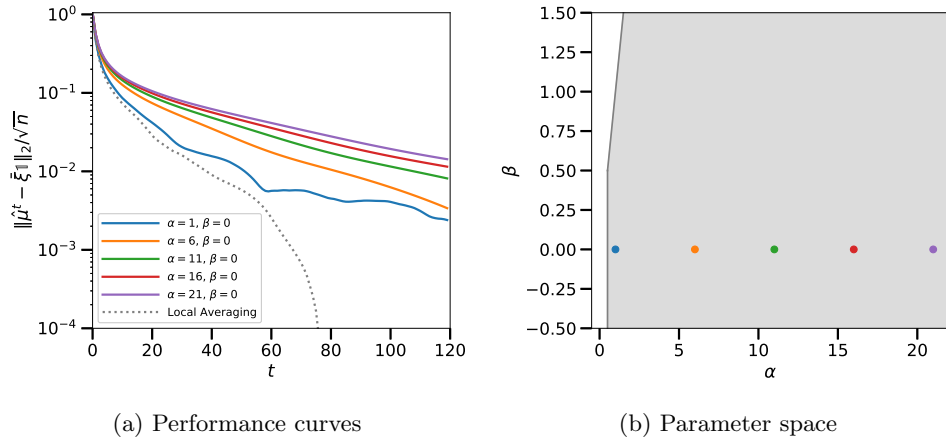


(a) Performance curves

(b) Parameter space

Fig. SM2: Simulations of polynomial iterations using Jacobi polynomials with different parameters $(\alpha, \beta)$: large $\alpha$ asymptotic.

The message passing algorithm of [SM4] (in its zero-temperature limit) defines quantities on the edges of the graph $G$ with the following recursion: for $v, w \in V$ linked by an edge in the graph $G$, it defines $K_{vw}^0 = 0$, $M_{vw}^0 = 0$, and

(SM9.1)

$$K_{vw}^{t+1} = 1 + \sum_{u \in \mathcal{N}(v),\, u \neq w} K_{uv}^t, \qquad M_{vw}^{t+1} = \frac{1}{K_{vw}^{t+1}} \left( \xi_v + \sum_{u \in \mathcal{N}(v),\, u \neq w} K_{uv}^t M_{uv}^t \right),$$

where $\mathcal{N}(v)$ denotes the set of neighbors of $v$. $K_{vw}$ and $M_{vw}$ are interpreted as

messages going from $v$ to $w$ in $G$: $M_{vw}^t$ corresponds to an average of observations gathered by $v$ and transmitted to $w$; $K_{vw}^t$ is the corresponding number of observations. We recommend [SM4, Section II.A] and Lemma SM9.2 for a detailed description of this intuition. At each time step $t$, the output of the algorithm is
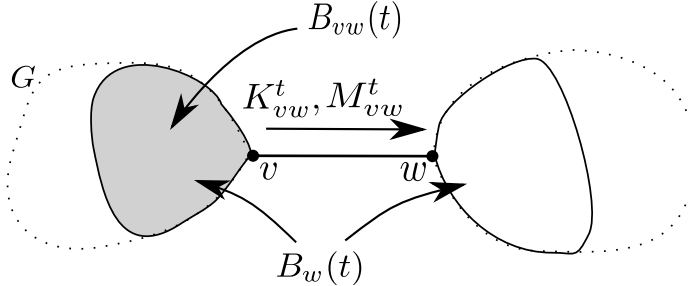
$$(SM9.2) \qquad x_v^t = \frac{\xi_v + \sum_{u \in \mathcal{N}(v)} K_{uv}^t M_{uv}^t}{1 + \sum_{u \in \mathcal{N}(v)} K_{uv}^t}.$$

This gossip methods performs exact local averaging on trees, as shown by the following proposition.

PROPOSITION SM9.1. *Assume that $G$ is a tree. Then for all $t \geqslant 1$, $v \in V$,*

$$x_v^t = \frac{1}{|B_v(t)|} \sum_{w \in B_v(t)} \xi_w .$$

*Proof.* Let $t \geqslant 0$ and $v, w \in V$ be two vertices linked by an edge in $G$. Define $B_{vw}(t)$ as the set of vertices $u$ in $B_w(t)$ such that all paths in the tree $G$ going from $u$ to $w$ pass though $v$.
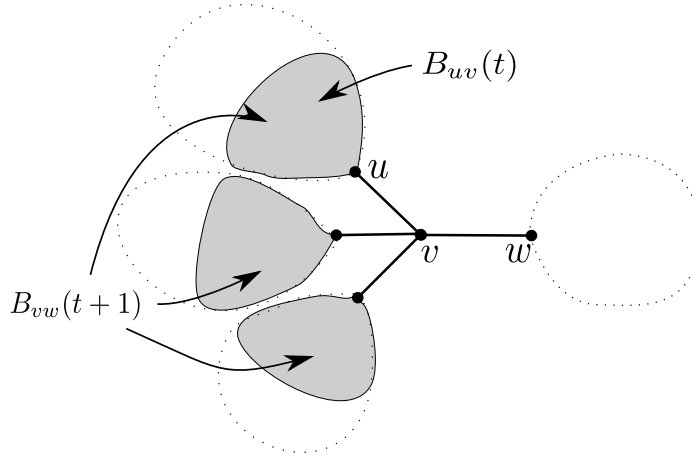


LEMMA SM9.2. *For all $t \geqslant 0$, for all $v, w \in V$ linked by an edge in $G$,*

$$K_{vw}^t = |B_{vw}(t)| , \qquad and \quad if \ t \geqslant 1, \quad M_{vw}^t = \frac{1}{|B_{vw}(t)|} \sum_{u \in B_{vw}(t)} \xi_u .$$

*Proof.* The proof goes by induction. The statement is trivial for $t = 0, 1$. For the induction, assume the result at time $t$ and note that

$$(SM9.3) \qquad B_{vw}(t+1) = \{v\} \cup \left( \bigcup_{u \in \mathcal{N}(v),\, u \neq w} B_{uv}(t) \right) ,$$

where all unions are disjoint. This essentially comes from the fact that $G$ has no loops.

Taking cardinal, we get that

$$|B_{vw}(t+1)| \overset{\text{(SM9.3)}}{=} 1+ \sum_{u\in\mathcal{N}(v),\, u\neq w} |B_{uv}(t)| \overset{\text{(induction)}}{=} 1+ \sum_{u\in\mathcal{N}(v),\, u\neq w} K_{uv}^t \overset{\text{(SM9.1)}}{=} K_{vw}^{t+1}.$$

This proves the induction for the first equality. The proof for the second equality is similar:

$$\frac{1}{|B_{vw}(t+1)|} \sum_{u\in B_{vw}(t+1)} \xi_u \overset{\text{(SM9.3)}}{=} \frac{1}{K_{vw}^{t+1}} \left( \xi_v + \sum_{u\in\mathcal{N}(v),\, u\neq w} \sum_{x\in B_{uv}(t)} \xi_x \right)$$

$$\overset{\text{(induction)}}{=} \frac{\xi_v + \sum_{u\in\mathcal{N}(v),\, u\neq w} |B_{uv}(t)| M_{uv}^t}{K_{vw}^{t+1}}$$

$$\overset{\text{(induction)}}{=} \frac{\xi_v + \sum_{u\in\mathcal{N}(v),\, u\neq w} K_{uv}^t M_{uv}^t}{K_{vw}^{t+1}} \overset{\text{(SM9.1)}}{=} M_{vw}^{t+1}. \quad \square$$

We now end the proof of Proposition SM9.1. As $B_v(t) = \{v\} \cup \left( \bigcup_{u\in\mathcal{N}(v)} B_{uv}(t) \right)$ with disjoint unions, using Lemma SM9.2, we get

$$\frac{1}{|B_v(t)|} \sum_{w\in B_v(t)} \xi_w = \frac{\xi_v + \sum_{u\in\mathcal{N}(v)} \sum_{w\in B_{uv}(t)} \xi_w}{1+ \sum_{u\in\mathcal{N}(v)} |B_{uv}(t)|} = \frac{\xi_v + \sum_{u\in\mathcal{N}(v)} K_{uv}^t M_{uv}^t}{1+ \sum_{u\in\mathcal{N}(v)} K_{uv}^t} \overset{\text{(SM9.2)}}{=} x_v^t.$$

Nothing prevents from running the message passing recursion (SM9.1)-(SM9.2) in a graph $G$ with loops. In the case of regular graphs, we are able to interpret the message passing algorithm as a polynomial gossip algorithm.

THEOREM SM9.3. *Assume $G$ is d-regular, meaning that each vertex has degree $d$, $d \geqslant 2$. Assume further that $W = A(G)/d$. Denote $\sigma(\mathbb{T}_d) = \sigma(\mathbb{T}_d, W, v)$ the spectral measure of the infinite d-regular tree at any vertex $v$ (see Definition 5.3). Then the output $x^t$ of the message passing algorithm (SM9.1)-(SM9.2) on $G$ can also be obtained as $x^t = \pi_t(W)\xi$ where $\pi_0, \pi_1, \dots$ are the orthogonal polynomials w.r.t. $(1-\lambda)\sigma(\mathbb{T}_d)(d\lambda)$.*

*Proof.* As noted by [SM6], the message passing iteration (SM9.1)-(SM9.2) indexed by the edges of the graph can be written as an iteration indexed by the vertices of the graph. We repeat here the elementary derivation of this statement in our particular case of $d$-regular graphs.

First, because $G$ is $d$-regular, it is an easy check from (SM9.1) that $K_{vw}^t$ does not depend on the edge $(v, w)$ (thus we denote it $K^t$) and it satisfies the recursion $K^0 = 0$, $K^{t+1} = 1 + (d-1)K^t$.

Let us now denote $S_v^t = \xi_v + \sum_{u \in \mathcal{N}(v)} K_{uv}^t M_{uv}^t$ and $L_t = 1 + dK^t$ so that $x_v^t = S_v^t / L_t$. We will now find recursions for $L_t$ and $S^t$:

$$L_{t+1} = 1 + dK^{t+1} \stackrel{\text{(SM9.1)}}{=} 1 + d(1 + (d-1)K^t) = 2 + (d-1)(1 + dK^t) = 2 + (d-1)L_t\,,$$

and

$$
\begin{aligned}
S_v^{t+1} = \xi_v + \sum_{u \in \mathcal{N}(v)} K^{t+1} M_{uv}^{t+1} &\stackrel{\text{(SM9.1)}}{=} \xi_v + \sum_{u \in \mathcal{N}(v)} \left( \xi_u + \sum_{w \in \mathcal{N}(u), w \neq v} K^t M_{wu}^t \right) \\
&= \xi_v + \sum_{u \in \mathcal{N}(v)} \left( S_u^t - K^t M_{vu}^t \right).
\end{aligned}
$$

As

$$
\begin{aligned}
\sum_{u \in \mathcal{N}(v)} K^t M_{vu}^t &\stackrel{\text{(SM9.1)}}{=} d\xi_v + \sum_{u \in \mathcal{N}(v)} \sum_{w \in \mathcal{N}(v), w \neq u} K^{t-1} M_{wv}^{t-1} \\
&= d\xi_v + (d-1) \sum_{w \in \mathcal{N}(v)} K^{t-1} M_{wv}^{t-1} = \xi_v + (d-1)S_v^{t-1}\,,
\end{aligned}
$$

we finally get

$$S^{t+1} = A(G)S^t - (d-1)S^{t-1}\,.$$

To sum up, we now have the simpler formulas for the message passing algorithm:

$$
\begin{aligned}
&L_{t+1} = 2 + (d-1)L_t\,, && L_0 = 1\,, \\
\text{(SM9.4)} \qquad &S^{t+1} = dWS^t - (d-1)S^{t-1}\,, && S^0 = \xi\,, \quad S^1 = \xi + dW\xi\,, \\
&x^t = S^t / L_t\,.
\end{aligned}
$$

In Section SM6.4, it is proved that $\pi_t(\lambda) = p_t(\lambda)/p_t(1)$ where $p_t$ satisfies the recursion formula

$$p_0(\lambda) = \sqrt{d-1}\,, \quad p_1(\lambda) = d\lambda + 1\,, \quad p_{t+1}(\lambda) = \frac{d}{\sqrt{d-1}}\lambda p_t(\lambda) - p_{t-1}(\lambda)\,, \quad t \geqslant 1\,.$$

Denote $q_t = (d-1)^{(t-1)/2} p_t$. It is an easy check that

$$q_0(\lambda) = 1\,, \quad q_1(\lambda) = d\lambda + 1\,, \quad q_{t+1}(\lambda) = d\lambda q_t(\lambda) - (d-1)q_{t-1}(\lambda)\,, \quad t \geqslant 1\,.$$

Using (SM9.4), one sees that for all $t$, $S^t = q_t(W)\xi$ and $L_t = q_t(1)$. Thus

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$$

$$x^t = \frac{S^t}{L_t} = \frac{q_t(W)\xi}{q_t(1)} = \frac{p_t(W)\xi}{p_t(1)} = \pi_t(W)\xi\,.$$

In words, the theorem above states that message passing corresponds to the best polynomial gossip algorithm when one *believes* the graph is a tree. This is not surprising as message passing algorithms are often derived by neglecting loops in a graph.

An easy follow-up of this theorem is that the iterates $x^t$ defined in (SM9.1)-(SM9.2) follow a second-order recursion (in $d$-regular graphs). Actually the spectral measure $\sigma(\mathbb{T}_d)$ of the infinite $d$-regular tree, also called the Kesten-McKay measure, can be computed explicitly (see [SM7, Section 2.2]),

$$\sigma(\mathbb{T}_d)(\mathrm{d}\lambda) = \frac{d}{2\pi(1-\lambda^2)}\left(\frac{4(d-1)}{d^2} - \lambda^2\right)^{1/2} \mathbf{1}_{\left[-2\sqrt{d-1}/d, 2\sqrt{d-1}/d\right]}(\lambda)\mathrm{d}\lambda.$$

The recurrence relation of the modified Kesten-McKay measure $(1-\lambda)\sigma(\mathbb{T}_d)(\mathrm{d}\lambda)$ is derived in Section SM6.4. It shows that

$$x^0 = \xi, \qquad x^1 = a_0 W\xi + b_0\xi, \qquad x^{t+1} = a_t Wx^t - c_t x^{t-1},$$
$$a_0 = \frac{d}{d+1}, \qquad b_0 = \frac{1}{d+1},$$
$$a_t = \frac{\frac{d}{d-1} - 2(d-1)^{-(t+1)}}{1 - \frac{2}{d}(d-1)^{-(t+1)}}, \qquad c_t = \frac{\frac{1}{d-1} - \frac{2}{d}(d-1)^{-t}}{1 - \frac{2}{d}(d-1)^{-(t+1)}}, \qquad t \geqslant 1.$$

Theorem SM9.3 gives a way to study the convergence of the message passing algorithms on $d$-regular graphs with loops. For instance, using asymptotic properties of the orthogonal polynomials w.r.t. $(1-\lambda)\sigma(\mathbb{T}_d)(\mathrm{d}\lambda)$, we obtain the convergence rate of the message passing algorithm as a function of the absolute spectral gap $\tilde{\gamma}$ of the matrix:

THEOREM SM9.4. *Assume $G$ is $d$-regular, meaning that each vertex has degree $d$, $d \geqslant 3$. Assume further that $W = A(G)/d$, and denote $\tilde{\gamma}$ its absolute spectral gap. Let $\xi = (\xi_v^t)_{v\in V}$ be any family of initial observations and $x^t = (x_v^t)_{v\in V}$ be the sequence of iterates generated by equations (SM9.1)-(SM9.2). Then*
1. *If $\tilde{\gamma} < 1 - 2\sqrt{d-1}/d$,*

$$\limsup_{t\to\infty} \|x^t - \bar{\xi}\mathbf{1}\|_2^{1/t} \leqslant \frac{(1-\tilde{\gamma}) + \sqrt{(1-\tilde{\gamma})^2 - 4(d-1)/d^2}}{1 + \sqrt{1 - 4(d-1)/d^2}}.$$

   *Moreover, the upper bound is reached if there exists an eigenvector $u$ corresponding to an eigenvalue of $W$ of magnitude $1 - \tilde{\gamma}$ such that $\langle u, \xi\rangle \neq 0$.*
2. *If $\tilde{\gamma} \geqslant 1 - 2\sqrt{d-1}/d$,*

$$\limsup_{t\to\infty} \|x^t - \bar{\xi}\mathbf{1}\|_2^{1/t} \leqslant \frac{2\sqrt{d-1}/d}{1 + \sqrt{1 - 4(d-1)/d^2}}.$$

A consequence of this theorem is that the rate of convergence of the message passing algorithm is $1 - c\tilde{\gamma} + o(\tilde{\gamma})$ as $\tilde{\gamma} \to 0$, for some constant $c$. This proves that message passing has a diffusive (or unaccelerated) behavior on graphs with a small spectral gap. Figure SM3 shows this diffusive convergence rate on the 2D grid.

*Proof of Theorem SM9.4.* Theorem SM9.3 states that $x^t = \pi_t(W)\xi$ where the $\pi_t$ are the orthogonal polynomials w.r.t. the modified Kesten-McKay measure $(1-\lambda)\sigma(\mathbb{T}_d)(\mathrm{d}\lambda)$. Then

$$(\text{SM9.5}) \quad \|x^t - \bar{\xi}\mathbf{1}\|_2^2 = \sum_{i=2}^{n}\langle \xi, u^i\rangle^2 \pi_t(\lambda_i)^2 \leqslant \|\xi - \bar{\xi}\mathbf{1}\|_2^2 \left(\sup_{\lambda\in[-(1-\tilde{\gamma}),(1-\tilde{\gamma})]} |\pi_t(\lambda)|\right)^2,$$

where $\lambda_2, \ldots, \lambda_n$ are the eigenvalues of $W$ different from 1, that lie in $[-(1-\tilde{\gamma}), (1-\tilde{\gamma})]$ by definition of the absolute spectral gap $\tilde{\gamma}$, and $u^2, \ldots, u^n$ are the corresponding normalized eigenvectors.

In Section SM6.4, we show that

$$\pi_t(\lambda) = \frac{p_t(\lambda)}{p_t(1)},$$

$$p_t(\lambda) = \tilde{p}_t(\varphi^{-1}(\lambda)), \qquad \varphi(\lambda) = \frac{2\sqrt{d-1}}{d}\lambda,$$

$$\tilde{p}_t(\lambda) = \left(\sqrt{d-1} + \lambda\right) U_t(\lambda) - T_{t+1}(\lambda),$$

where $T_t$ and $U_t$ are the Chebyshev polynomials of the first kind and of the second kind respectively. Denote $D = d/(2\sqrt{d-1})$. Then

(SM9.6) $$\sup_{\lambda \in [-(1-\tilde{\gamma}),(1-\tilde{\gamma})]} |\pi_t(\lambda)| = \frac{1}{|\tilde{p}_t(D)|} \sup_{\lambda \in [-(1-\tilde{\gamma})D,(1-\tilde{\gamma})D]} |\tilde{p}_t(\lambda)|.$$

If $\lambda \in [-1, 1]$, $|T_t(\lambda)| \leqslant 1$ and $|U_t(\lambda)| \leqslant t+1$. Thus

(SM9.7) $$\sup_{\lambda \in [-1,1]} |\tilde{p}_t(\lambda)| \leqslant \left(\sqrt{d-1}+1\right)(t+1) + 1.$$

We now discuss the different cases of the theorem.

**(1)** We assume $\tilde{\gamma} < 1 - 2\sqrt{d-1}/d$. As $\tilde{p}_t$ are orthogonal polynomials w.r.t. some measure on $[-1, 1]$, all zeros of $p_t$ are real, distinct and located in the interior of $[-1, 1]$ (see Proposition SM2.1). It follows that

(SM9.8)
$$\sup_{\lambda \in (1,(1-\tilde{\gamma})D]} |\tilde{p}_t(\lambda)| = |\tilde{p}_t\left((1-\tilde{\gamma})D\right)|, \qquad \sup_{\lambda \in [-(1-\tilde{\gamma})D,-1)} |\tilde{p}_t(\lambda)| = |\tilde{p}_t\left(-(1-\tilde{\gamma})D\right)|.$$

Merging Eqs. (SM9.6)-(SM9.8), we obtain

$$\sup_{\lambda \in [-(1-\tilde{\gamma}),(1-\tilde{\gamma})]} |\pi_t(\lambda)|$$

$$\leqslant \frac{1}{|\tilde{p}_t(D)|} \max\left(|\tilde{p}_t((1-\tilde{\gamma})D)|, |\tilde{p}_t(-(1-\tilde{\gamma})D)|, (\sqrt{d-1}+1)(t+1)+1\right).$$

LEMMA SM9.5.        1. *If $x > 1$, then there exists a constant $C(d, x) \neq 0$ such that*

(SM9.9) $$\tilde{p}_t(x) \underset{t \to \infty}{\sim} C(d, x)\left(x + \sqrt{x^2-1}\right)^t.$$

2. *If $x < -1$, then there exists a constant $C(d, x) \neq 0$ such that*

(SM9.10) $$\tilde{p}_t(x) \underset{t \to \infty}{\sim} C(d, x)\left(x - \sqrt{x^2-1}\right)^t.$$

*Proof.* In the proof of Lemma SM2.10, we developed the following formulas for the Chebyshev polynomials:

$$T_t\left(\frac{z + z^{-1}}{2}\right) = \frac{z^t + z^{-t}}{2}, \qquad U_t\left(\frac{z + z^{-1}}{2}\right) = \frac{z^{t+1} - z^{-(t+1)}}{z - z^{-1}}.$$

We write $x = (z + z^{-1})/2$ with $|z| > 1$. If $x > 1$, then $z = x + \sqrt{x^2 - 1}$, and if $x < -1$, then $z = x - \sqrt{x^2 - 1}$. Then

$$
\begin{aligned}
\tilde{p}_t(x) &= \left( \sqrt{d-1} + x \right) U_t(x) - T_{t+1}(x) \\
&= \left( \sqrt{d-1} + \frac{z + z^{-1}}{2} \right) \frac{z^{t+1} - z^{-(t+1)}}{z - z^{-1}} - \frac{z^{t+1} + z^{-(t+1)}}{2} \\
&\underset{t \to \infty}{\sim} \left[ \left( \sqrt{d-1} + \frac{z + z^{-1}}{2} \right) \frac{1}{z - z^{-1}} - \frac{1}{2} \right] z^{t+1} .
\end{aligned}
$$

The constant that appears is non-zero, thus the result is proved. $\qquad\square$

Using Lemma SM9.5, we get that there exists a constant $C(d)$ such that

$$
\sup_{\lambda \in [-(1-\tilde{\gamma}), (1-\tilde{\gamma})]} |\pi_t(\lambda)| \leqslant C(d) \left( \frac{(1-\tilde{\gamma})D + \sqrt{((1-\tilde{\gamma})D)^2 - 1}}{D + \sqrt{D^2 - 1}} \right)^t .
$$

Finally, using (SM9.5), this gives

$$
\|x^t - \bar{\xi}\mathbf{1}\|_2 \leqslant \|\xi - \bar{\xi}\mathbf{1}\|_2 C(d) \left( \frac{(1-\tilde{\gamma})D + \sqrt{((1-\tilde{\gamma})D)^2 - 1}}{D + \sqrt{D^2 - 1}} \right)^t .
$$

Dividing the numerator and the denominator of the fraction by $D$, we get the desired result.

We now turn to the second part of the statement. Let $u$ be an eigenvector of $W$ corresponding to an eigenvalue $\lambda$ of magnitude $1 - \tilde{\gamma}$ such that $\langle \xi, u \rangle \neq 0$. Then

$$
\|x^t - \bar{\xi}\mathbf{1}\|_2 \geqslant |\langle \xi, u \rangle| |\pi_t(\lambda)| = |\langle \xi, u \rangle| \frac{|\tilde{p}_t(\lambda)|}{|\tilde{p}_t(D)|} ,
$$

Using as before Lemma SM9.5, we get the desired lower bound.

*(2)* We now assume $\tilde{\gamma} \geqslant 1 - 2\sqrt{d-1}/d$. This means that $(1-\tilde{\gamma})D \leqslant 1$, and thus

$$
\sup_{\lambda \in [-(1-\tilde{\gamma})D, (1-\tilde{\gamma})D]} |\tilde{p}_t(\lambda)| \leqslant \sup_{\lambda \in [-1,1]} |\tilde{p}_t(\lambda)| \overset{(SM9.7)}{\leqslant} \left( \sqrt{d-1} + 1 \right) (t+1) + 1 .
$$

Combining with (SM9.5) and (SM9.6), we get

$$
\|x^t - \xi\mathbf{1}\|_2 \leqslant \|\xi - \bar{\xi}\mathbf{1}\|_2 \frac{1}{|\tilde{p}_t(D)|} \left( \left( \sqrt{d-1} + 1 \right) (t+1) + 1 \right) ,
$$

which gives the desired result using Lemma SM9.5. $\qquad\square$

However, the message passing algorithm can be competitive in situations with a large spectral gap. For instance, McKay's Theorem [SM7, Theorem 1.1] states that the spectral measure of a uniformly random $d$-regular graph on $n$ vertices converges to the spectral measure $\sigma(\mathbb{T}_d)$ of the $d$-regular tree (in law, for the weak-convergence topology). This suggests that the message passing algorithm is well-suited for uniformly sampled large regular graphs. We give simulations in Figure SM4 on uniformly sampled 3-regular graphs of size $n = 2000$. The results were averaged over 10 graphs. We observe that in this case, message passing matches closely the lower-bound. Note
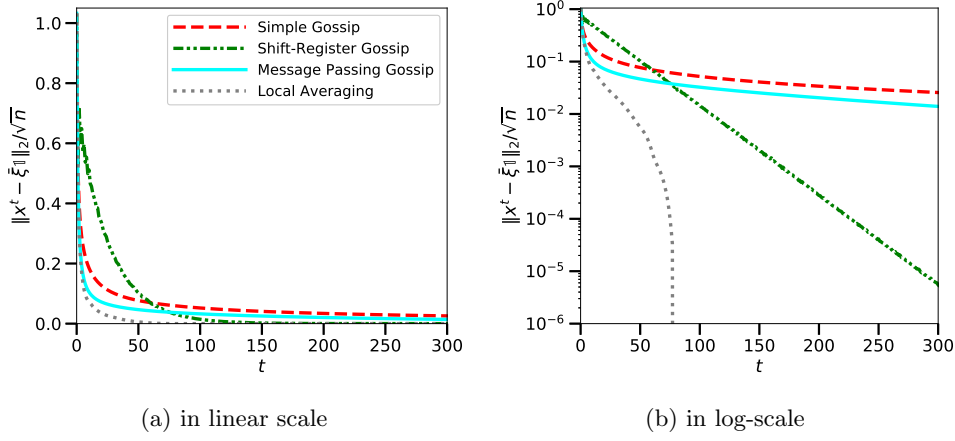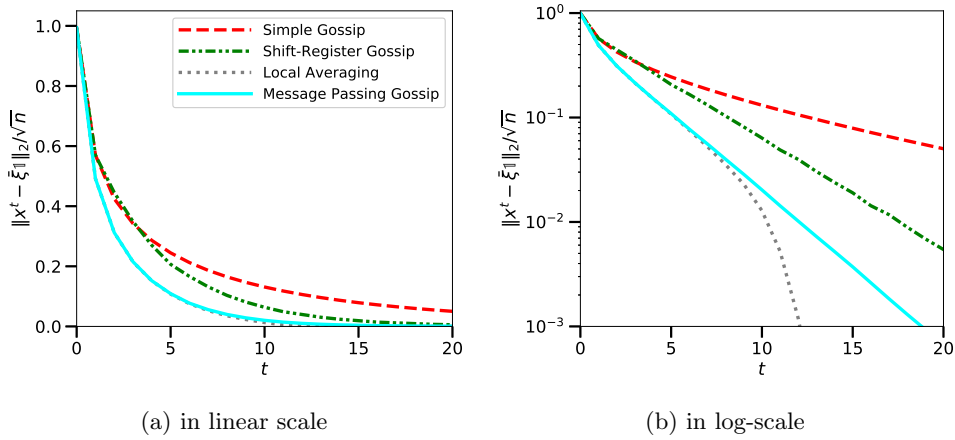
(a) in linear scale                              (b) in log-scale

Fig. SM3: Performance of different gossip algorithms running on the 2D grid.



(a) in linear scale                              (b) in log-scale

Fig. SM4: Performance of different gossip algorithms running a uniformly random 3-regular graph of size $n = 2000$.

that in this case, we do not have a diffusive rate of convergence because the spectral gap $\gamma$ does not converge to 0 as $n \to \infty$ (see [SM3] for a proof that $\gamma \to 1 - 2\sqrt{d-1}/d$ in probability).

## REFERENCES

[SM1]  R. BERTHIER, *Code for the paper: Accelerated gossip in network of given dimension.* https://github.com/raphael-berthier/jacobi-polynomial-iterations, 2019.

[SM2]  P. BRACA, S. MARANO, AND V. MATTA, *Enforcing consensus while monitoring the environment in wireless sensor networks*, IEEE Transactions on Signal Processing, 56 (2008),

pp. 3375–3380.

[SM3]  J. Friedman, *A proof of alon's second eigenvalue conjecture*, in Proceedings of the thirty-fifth annual ACM Symposium on Theory of computing, ACM, 2003, pp. 720–724.

[SM4]  C. C. Moallemi and B. V. Roy, *Consensus propagation*, in Advances on Neural Information Processing Systems, MIT Press, 2005, pp. 899–906.

[SM5]  F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST handbook of mathematical functions*, Cambridge University Press, 2010.

[SM6]  P. Rebeschini and S. Tatikonda, *Accelerated consensus via min-sum splitting*, in Advances on Neural Information Processing Systems, 2017.

[SM7]  S. Sodin, *Random matrices, nonbacktracking walks, and orthogonal polynomials*, Journal of Mathematical Physics, 48 (2007), p. 123503.

[SM8]  G. Szegö, *Orthogonal polynomials*, vol. 23, American Mathematical Soc., 1939.

[SM9]  W. Woess, *Random walks on infinite graphs and groups*, vol. 138, Cambridge University Press, 2000.